Cognitive Science and Technology

Amit Kumar
Gheorghita Ghinea
Suresh Merugu
Takako Hashimoto *Editors*

Proceedings of the International Conference on Cognitive and Intelligent Computing

ICCIC 2021, Volume 2

Springer

Cognitive Science and Technology

Series Editor

David M. W. Powers, Adelaide, SA, Australia

This series aims to publish work at the intersection of Computational Intelligence and Cognitive Science that is truly interdisciplinary and meets the standards and conventions of each of the component disciplines, whilst having the flexibility to explore new methodologies and paradigms. Artificial Intelligence was originally founded by Computer Scientists and Psychologists, and tends to have stagnated with a symbolic focus. Computational Intelligence broke away from AI to explore controversial metaphors ranging from neural models and fuzzy models, to evolutionary models and physical models, but tends to stay at the level of metaphor. Cognitive Science formed as the ability to model theories with Computers provided a unifying mechanism for the formalisation and testing of theories from linguistics, psychology and philosophy, but the disciplinary backgrounds of single discipline Cognitive Scientists tends to keep this mechanism at the level of a loose metaphor. User Centric Systems and Human Factors similarly should inform the development of physical or information systems, but too often remain in the focal domains of sociology and psychology, with the engineers and technologists lacking the human factors skills, and the social scientists lacking the technological skills. The key feature is that volumes must conform to the standards of both hard (Computing & Engineering) and social/health sciences (Linguistics, Psychology, Neurology, Philosophy, etc.). All volumes will be reviewed by experts with formal qualifications on both sides of this divide (and an understanding of and history of collaboration across the interdisciplinary nexus).

Indexed by SCOPUS

Amit Kumar · Gheorghita Ghinea ·
Suresh Merugu · Takako Hashimoto
Editors

Proceedings of the International Conference on Cognitive and Intelligent Computing

ICCIC 2021, Volume 2

Springer

Editors
Amit Kumar
BioAxis DNA Research Centre Limited
Hyderabad, India

Suresh Merugu
CMR College of Engineering
and Technology
Hyderabad, India

Gheorghita Ghinea
Department of Computer Science
Brunel University
Uxbridge, UK

Takako Hashimoto
Ichikawa, Chiba, Japan

ISSN 2195-3988 ISSN 2195-3996 (electronic)
Cognitive Science and Technology
ISBN 978-981-19-2357-9 ISBN 978-981-19-2358-6 (eBook)
https://doi.org/10.1007/978-981-19-2358-6

Contents

Teaching English as a Second Language: Improving Digital Literacy Skills

S. Rajeswari and E. Madhavi

Abstract English language teachers need to work and act to change efficaciously and call for current ways of teaching. Empowering teachers' professional knowledge helps them to understand contemporary educational practices and policies that are required for education. COVID-19 pandemic forcefully amended traditional learning environments to online teaching. The study investigates Information Communication Technology (ICT) as an alternative to traditional classrooms. The findings are significant that knowledge, interactions, and communities are pertinent and steered widely by innovations in portable computers and nominal price of information technology. Teaching methods may include classroom blogs, wikis, vlogs, glogster, podcasts, etc., transforming the personality of an individual connecting globally that are typically of the students' area of academics. The results suggest that understanding and acquiring English language become an active learning skill. It leads to progress, critical for a nation to build a skilled workforce, and also to help people improve their livelihoods as the country grows economically.

Keywords Contemporary practices · ICT · Teaching methods · Glogster · Podcasts

1 Introduction

Communication and digital media are developing rapidly. The world is changing increasingly and becoming complex as the global pool of data expands. Technology connects people and with resources all over the world. Present students must develop 'communication technology skills' [1] that are crucial. Students play a vital role for

S. Rajeswari
Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India

E. Madhavi (✉)
CMR College of Engineering and Technology, Hyderabad, Telangana, India
e-mail: madhavikoppula6@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_1

nation's prosperity. The 'creative competence' [2] of young leaners has become a key educational goal. Technology can promote creativity with a wide spread of digital communication software utilisation with devices, like computer, mobile phone using internet.

This paper evaluates using digital literacy for education competently. Interpreting of data helps to learn digital communication by assessing the reliability through research, creativity and communicate using suitable tools. To improve the skills, the learners can avail online assessments conducted by community colleges. Credentialing curricula of colleges as advantageous technology could cut costs and provide 'flexible learning environments' [3] for working people as well. These learning options are more challenging for the people without basic skills. Thus, digital skills play a vital role in various aspects of life, such as accessing information through social networks sites connecting with community.

2 Literature Review

Many researchers have described the concepts of literacy as 'social practice' [4] that acts as a 'functional system' [5]. Social interaction is key to broaden perspectives that allows for sharing and communicating ideas. Learning methods are constantly changing and as such new technologies are emerging. According to [6], literacy is defined as socially identified methods of generating meaningful content through encoded texts. Technology has changed meaningful ways of communication through the use of internet. It is significant that teachers need to understand the changes of new literacies and bring them into classrooms to engage their learners in the ways that best suits them. Digital technologies connect the learners with their prior knowledge which assists in conceptual development of science. It also capitalises on students' interest. According to [7], technology has become 'an indispensable tool' in today's education. Writing plays an important role that allows learners use their critical thinking skills. Weblog can be found to be a useful style for different learnings in classrooms to meet the needs of the learners. This dynamic approach provides better interaction of teachers and students expanding the walls of classrooms. This promotes 'peer learning' [8] and also understanding 'alternative viewpoints' [9]. Sykes et.al [10] opine wikis as 'collaborative tools' for content building and knowledge dissemination. Godwin-Jones [11] used podcasts as instructional material. This can be referred as an acceptable resource by teachers and students.

Digital literacy skills are required for more job opportunities and became a transversal competence that includes procedures, values, and attitudes. The significance of the digital skills is 'very rich' [12–14]. Enriching communicative competence, video conferencing is observed as a 'convenient tool' [15]. The strategies may vary in presentation and implementation from organisation to organisation. The constant change in competence, digital technology is continuously modernised to avoid and minimise risks of digital exclusion. Scholars generally conceptualised and investigated various digital tools. With limited research on digital platforms,

it is important to investigate the advanced teaching tools such as Zoom, Google Meet, Microsoft Teams, and WebEx. The paper uses both descriptive methodology to illustrate how teaching tools are used and real-time, online class experience with engineering students teaching English.

3 Findings

Reforms in Education System Leads to Progress

The raise in academic standards rises children's academic sights. When learners are regularly tested, teachers can assess where and how to improve the learners. When scores of the learners are known to their parents, parents are empowered to push for change. Making education accessible to all is the only way to harness this potential in an organised way by educating the masses.

Major reformation is needed at micro-level as well as macro-level. These reforms have to begin with an emphasis on the development of an individual. Information and Communication Technology (ICT) in education supports delivery of information and enhances learning. The present study has shown that ICT leads to an improved learning and better teaching practices. Recorded interactions took place in an engineering English class at a private college with B. Tech I year students, 20 females and 40 males that met weekly for 90 min via Google Classroom over 14 weeks. To support online classes with class time table, class news, assignments, discussion questions, and other class news, Google Classroom was exercised as a Learning Management System (LMS).

At the commencement of the class work, the learners were surveyed about their exposure to online classes and technology familiarity. About 83.33% of the students (58/60) had experienced with real-time, online classes (Fig. 1). The students' previous knowledge of Google Classroom or Zoom varied on 5 point Likert scale, with 2 being 'I don't know at all' and 10 being 'I know very well'. The rest with mixed results with the platform familiarity with 80% of respondents (48/60) scoring 3 or above with Google Meet or Zoom. Have you experienced real-time, online class before? (Google Meet/Zoom) 58 responses.

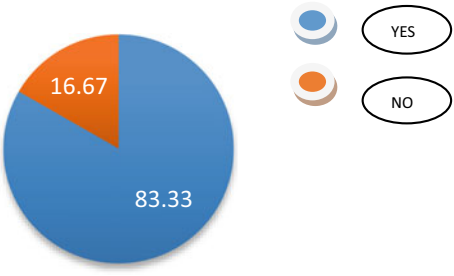


Fig. 1 Real-time, online class experience

The students completed four weeks course on Zoom with Google Classroom support to overcome the technological discomfort prior to qualitative study. The learners became comfortable with the instructional style and could successfully complete their individual presentations teacher and peer group evaluation.

4 Methods

4.1 Contemporary Practices

The digital transformation in all the levels of education system incorporated the buzz word 'Technology in Education'. The transformation should be built on principles of equality and openness that cater the needs of all the contemporary learners. New technologies like classroom blogs, wikis, interactive boards, podcasts, voice notes, and many more stimulate learners' engagement within the classroom. English language teachers can use innovative techniques like advertisements, English songs, movie clippings, sports commentaries, etc. can be used to engage students to become proficient at English language. This can be enhanced with a proper blend of e-learning tools and edification.

4.1.1 Classroom Blogs

Classroom blogs open extra opportunities for learners to educate themselves. These blogs encourage learners for collaborative work and provide opportunities for peer mentorship. Learners who hesitate to interact during classroom discussions could participate through classroom blogs to share information. Instructors can easily update information as it functions as a central source of data for students. The blog created by a teacher lists assignments, can link to important documents, and may use as 'online gallery' [16] of the learners works. The learners find classroom blogs more 'intuitive' [17] to use. They stimulate students' interest of the topic that develops their expertise in their domain. 'Learners interaction' [18] can be increased creating a community of learners. Such blogs help learners expose to diverse perspectives. Through this, students can understand 'alternative viewpoints' [9, 19]. The posts can be frequently updated and are available as searchable sources of all categories, archived in 'reverse chronological order' [20]. The blogs are often public which brings awareness of audience larger than teacher that yields to increased motivation, producing quality content, and be as 'editorial components' [21–23] of the posts.

4.1.2 Classroom Wikis

Teachers can establish an account for their students of respective classes. Based on the syllabus and topic under discussion in class, teachers can find an existing article, review the article with their learners, and encourage them to make appropriate edits for the improvement of the article. This type of activity enhances learners' critical thinking skills and language skills. The instructors can use wikis to post subject notes and propagate important information. They can also be used to make vocabulary lists. Thus, classroom wikis becomes an exceptional useful tool making students involve in curriculum.

4.1.3 Vlogs

YouTube became an educational tool that acts like a social media platform. It undergoes a process of institutionalisation turning into 'professional content creators' [24–26]. YouTube is like 'virtual village' [25] where user-generated content is shared online. Teachers may turn their classroom blogs into video blogs or vlogs. Video blogs can be organised around themes. Students' projects can be showcased by creating lessons available outside the classroom. Learners may opt the recording tools and edit the video files and upload them to their institution web servers. This leads to the development of valuable digital literacy skills. Vlogs can also be used to showcase students' musical talents. They help to promote students awareness on social topics.

4.1.4 Glogster

Glogging administers an excellent space for instructors to strengthen learners' creativity while organising and presenting their projects. Glogster is relatively a new tool for educators posting photographs and videos which encourages learners to participate in education activities. This tool allows instructors and learners to make virtual posters and share them with people. The posters may contain text, audio, video, images, and hyperlinks that enhance students' different technology skills. Such postings can address classroom policies and assignments.

Students of engineering English class were expected to give prepared formal presentations on any topic of their choice in Google Classroom. First week, the learners were shown sample formal presentation to provide format and presentation skills by the teacher handling the class. All the students were assigned to form groups and choose a topic for presentation. Each group was asked to submit an outline of the presentation for review. Second week, all the groups were asked to give formal presentations.

There were mixed results (Fig. 2) with preparation (excellent—30%, good—50%, average—15%, poor—5%), organisation (excellent—20%, good—40%, average—40%, poor—nil), content (excellent—20%, good—30%, average—30% poor—20),

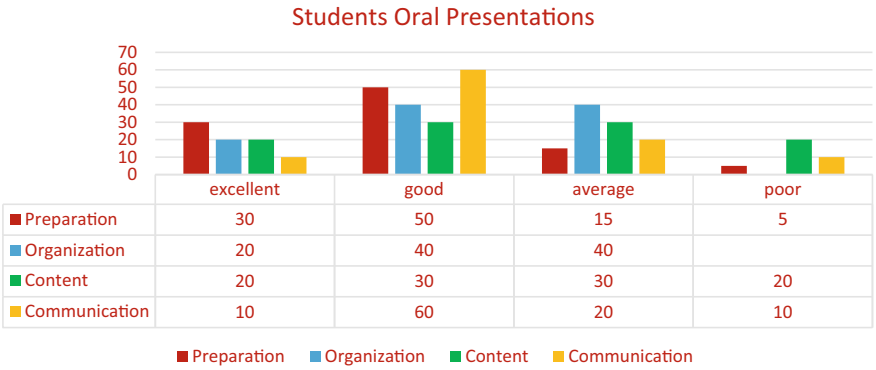


Fig. 2 Students oral presentations

and communication (excellent—10%, good—60%, average—20%, poor—10%) of the presentations.

The students became comfortable due to frequent online discussions and could successfully complete their presentations in groups, each group consisting of four students. The presentations were productive with the teacher's evaluation and peer group review uninterruptedly without any delay or technological guidance on time and thus mastering the mechanics of technology used in Goole Classroom. Students learning boosted tremendously improving their reading skills by availing authentic texts. Students could get awareness with 'good online texts' [27, 28]. Technology-integrated classroom teaching accommodates 'richer experience' [29] for learners. Various technological tools help students to conduct project works and make data manipulation feasible. Graphic displays show 'quantitative outcomes' [30]. Glogster acts as a useful publishing educational tool used in 'various English courses' [31]. The learners could develop critical understanding of different audio–video inputs which yields in decoding unfamiliar lexis.

4.1.5 Podcasts

Podcasts and glogster are a few of the digital tools that are used to draw attention of learners throughout the process of 'creating and sharing' [31]. Podcasting technology in educational frame works prompted expectations of considerable progress. Successful embedded podcasting 'beyond recording lectures' [32–34] offers potential benefits to student learning. Students could record their classroom activities, edit audio–video files, and upload them for digital distribution that helps them improve digital literacy. Adding visual information allows teachers to create resources that promote students' learning preferences. Using such technological tool benefits learning, enabling mobile learning. It empowers learners to 'manage their own time' [35] of learning that best suits them and also revise lectures to refine their understanding of complex content with clarity. Lecture videos and students presentations

Fig. 3 Feedback on audio–video/podcasts learning

were uploaded by the software in college portal for future reference by the students. Feedback was taken for the process and received mixed results (Fig. 3) with regular schedules—100%, authenticity—90%, consistency—80%, clear audio with good equipment—70% (equipment used by both teacher and students).

The process of capturing content-based audio–video learning allowed podcasts to be the most essential resource that reached an expectation position. Appropriate teaching–learning activities through podcasting beyond domain content gains momentum in education.

5 Results

The results of the new model education are more inclusive and collaborative, shifting the roles of teacher and students, slowly reshaping into digital world. The transformation begins to require greater responsibility of the learners charting their own path turning from passive student to an active learner.

Empowering to plan and implement their own learning with demonstrative creativity and construct knowledge is the key to success. Applying the existing knowledge, the learners could get new ideas and processes using technology. This yielded to use models and provided stimulations to investigate complex issues and identifying trends. A significant progress is observed with the spread of digital environments that developed cultural understanding and global awareness. Such environment could provide learners use digital media to work collaboratively and thus supporting individual learning and contributing to train others at a distance.

Critical thinking skills help learners plan and conduct their research that leads to informed decisions with appropriate use of digital tools and the resources. Thus, enhancing digital literacy contributes to strengthen knowledge through interactive digital texts.

6 Conclusion

Embedding technology into education became vital to students success. Rapid integrated technology into various aspects of life could prompt different responses that improved mobility and efficiency of interconnectedness. Research is supervised to better recognise the alternatives with respect to the subject of job market. The analysis by and large acknowledges that the demand for digital skills in almost all the occupations is expeditiously increasing. The need of digital skills became vital for carrying out many jobs. Candidates lacking basic digital literacy skills are disproportionately less educated. To bridge the gap, digital training is required especially to learners displaced from occupations. Education and training contributors need to encourage programmes that help the required to acquire digital literacy. The providers need to organise ICT training programmes that are most effective for teaching–learning digital skills. This leads to improve the ability to share knowledge across various platforms adapting new technological settings.

References:

1. The Global Economic Forum, 21st-century skills (2016). Retrieved from https://www.weforum.org/agenda/2016/03/21st-century-skills-future-jobs-students/
2. S. Vincent-Lancrin, C. Gonzalez-Sancho´, M. Bouckaert, F. de Luca, M. Fern´andez- Barrerra, G. Jacotin, Q. Vidal, in *Fostering Students' Creativity and Critical Thinking* (OECD , Paris, 2019)
3. J.E. Seaman, E. Julia, E. Allen, J. Seaman, *Grade Increase: Tracking Distance Education in the United States* (Babson Survey Research Group, Babson Park, MA, 2018)
4. J. Larson, J. Marsh, *Making Literacy Real* (Sage, Thousand Oaks, 2005)
5. J.P. Gee, What is literacy?, in *Literacy: A Critical Source Book*. ed. by E. Cushman, E.R. Kingten, B.M. Kroll, M. Rose (Bedford/St. Martins, Boston, 2001), pp. 525–544
6. C. Lankshear, M. Knobel, Researching new literacies: Web 2.0 practices and insider perspectives. E-Learning **4**(3), 224–240 (2007)
7. W. Richardson, *Blogs, Wikis, Podcasts, and Other Powerful Web Tools for Classrooms* (Corwin, Thousand Oaks, 2006)
8. K. Strampel, R. Oliver, Blogging for learning: Improving teaching strategies for implementing blogs in higher education, in Paper Presented at the Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, Honolulu, USA (2009). Retrieved from http://editlib.org/d/31921
9. A. Hemmi, S. Bayne, R. Land, The appropriation and repurposing of social technologies in higher education. J. Comput. Assist. Learn. **25**, 19–30 (2009). Retrieved from http://www.malts.ed.ac.uk/staff/sian/JCALpaper_final.pdf
10. J. Sykes, A. Oskoz, S.L. Thorne, Web 2.0, synthetic immersive environments, and mobile resources for language education. CALICO J. **25**(3), 528–546 (2008)
11. R. Godwin-Jones, Emerging technologies-skype and podcasting: disruptive technologies for language learning. Lang. Learn. Technol. **9**(3), 9–12 (2005)
12. European Commission, Europe's Digital Progress Report 2016, European Commission, Brussels. Available at: https://ec.europa.eu/digital-single-market/en/download-scoreboardreports, Accessed 12 Jan 2016
13. Eurofound, Digitisation of processes-literature review. Working Paper 17038 (Eurofound, Dublin, 2017)

14. Eurofound, Automation, digitisation and platforms: implications for work and employment. Working paper 18002 (Eurofound, Dublin, 2018)
15. R. Vurdien, Videoconferencing: Developing students' communicative competence. J. Foreign Lang. Educ. Technol. **4**(2), 269–298 (2019)
16. A. Sawmiller, Classroom blogging: What is the role in science learning? Clearing House J. Educ. Strat. Issues Ideas **83**(2), 44–48 (2010)
17. H.N. Kim, The phenomenon of blogs and theoretical model of blog use in Education contexts. Comput. Educ. **51**, 1342–1352 (2008). https://doi.org/10.1016/j.compedu.2007.12.005
18. B. Farmer, A. Yue, C. Brooks, Using blogging for higher order learning in large cohort university teaching: A case study. Aust. J. Educ. Technol. **24**(2), 123–136 (2008). Retrieved from http://www.ascilite.org.au/conferences/singapore07/procs/farmer.pdf
19. R. Philip, J. Nicholls, Group blogs: Documenting collaborative drama processes. Aust. J. Educ. Technol. **25**(5), 683–699 (2009). Retrieved from www.ascilite.org.au/ajet/ajet25/philip.pdf
20. Nardi, Bonnie A., Diane J. Schiano, and Michelle Gumbrecht. "Blogging as social activity, or, would you let 900 million people read your diary?." Proceedings of the 2004 ACM conference on Computer supported cooperative work (2004)
21. S. Carlson, Weblogs come to the classroom. Chron. High. Educ. **50**(14), A33 (2003)
22. S. Downes, Educational blogging. Educause, Sept/Oct, 14–26 (2004)
23. T. Martindale, D.A. Wiley, An introduction to teaching with weblogs. Draft Copy (2004)
24. J. Burgess, YouTube and the formalisation of amateur media, in *Amateur Media: Social, Cultural and Legal Perspectives*. ed. by D. Hunter et al. (Routledge, London, 2012), pp. 53–58
25. J. Kim, The institutionalization of YouTube: From user-generated content to professionally generated content. Media Cult. Soc. **34**(1), 53–67 (2012)
26. J. Morreale, From homemade to store bought: Annoying orange and the professionalization of YouTube. J. Consum. Cult. **14**(1), 113–128 (2014)
27. L.C. Larson, Digital readers: The next chapter in e-book reading and response. Read. Teach. **64**(1), 15–22 (2010)
28. L. Zawilinski, Hot blogging: A framework for blogging to promote higher-order thinking. Read. Teach. **62**(8), 650–661 (2009)
29. J. Ohler, Orchestrating the media collage. Challenging Whole Child, 161 (2009)
30. M. Roblyer, in *Integrating Educational Technology into Teaching*, 3rd edn. (Merrill Prentice Hall, Upper Saddle River, 2003)
31. D.B. Kent, Incorporating glogster in the University EFL curriculum. Arab World Engl. J. **1**(1), 130–170 (2010)
32. M. Sharples, J. Taylor, G. Vavoula, Towards a theory of mobile learning, in Proceedings of 4th mLearn Conference, ed. by J. Attewell, T. Brown, G.D. Bormida, Sharples, H.V.D. Merwe. mLearn, Cape Town (2005). http://www.mlearn.org.za/CD/papers/Sharples-%20Theory%20of%20Mobile.pdf
33. A. Herrington, J. Herrington, Authentic mobile learning in higher education, in Proceedings of the Australian Association for Research in Education International Educational Research Conference, ed. by P.L. Jeffery. University of Notre, Freemantle (2007). Dame. http://www.aare.edu.au/07pap/her07131.pdf
34. G. Salmon, M. Nie, Doubling the life of iPods, in *Podcasting for learning in universities*. ed. by G. Salmon, P. Edirisingha (McGraw Hill/Open University Press, Glasgow, 2008), pp. 1–11
35. M. Sharples, The design of personal mobile technologies for lifelong learning. Comput. Educ. **34**(3–4), 177–193 (2000)

ICT-Based Collaborative Learning Approach: Enhancing Students' Language Skills

Rukhiya Begum and Ravi Naga Dhana Lakshmi

Abstract The purpose of this study is to investigate the impact of ICT-based collaborative learning (CL) on the development of integrated skills and motivation in undergraduate students of English as a second language (ESL). To achieve this goal, forty students from St. Ann's Degree College in Telangana, India, were chosen on an experimental basis. The goal is to increase student collaboration and provide opportunities for them to practice using English correctly through collaborative learning techniques with the help of technology. This paper also discusses ICT as a tool in the CL approach, the importance of technology in ESL, and the challenges that teachers face when using ICT tools in the classroom. As a result, students contribute to the learning process and play an important role in successfully completing the task or project. Because these are the types of learners centric curriculum, and ICT tools are beneficial. Data is gathered using a Google forms-based questionnaire. After analyzing the data, the results revealed a significant improvement in the students' interactive skills through collaborative learning techniques in the classroom.

Keywords Collaborative learning · ESL · ICT tools · Google forms · Integrated skills · Learner-centric

R. Begum (✉)
Department of English, Koneru Lakshmaiah Deemed to be University, Guntur, Andhra Pradesh, India
e-mail: 786rukhiya.begum@gmail.com

Marri Laxman Reddy Institute of Technology and Management, Hyderabad, India

R. Naga Dhana Lakshmi
Department of English, St. Ann's College of Engineering and Technology, Chirala, Andhra Pradesh, India
e-mail: hs.drdhanalakshmi@sacet.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_2

1 Introduction

The modern method of teaching English has shifted from the "chalkboard model" to the "smartboard model." To create a communication environment in the ESL classroom, an ICT-based collaborative learning technique can be used. Anyone who wishes to benefit from modern education, research, science, trade, or technology understands how difficult it is to do so without a basic command of the English language and effective communication skills [1]. Collaborative learning is a teaching method in which groups of two or more students collaborate to solve problems, complete tasks, or learn new concepts. "Lev Vygotsky" recommends collaborative learning theory, in which he discusses how learners in the Zone of Proximal Development rely on one another to complete tasks that they would be unable to complete on their own. There is a diverse set of technological tools and resources for data transfer, storage, production, sharing, and exchange. Computers, the Internet (Websites, blogs, and emails), live broadcasting technologies (radio, television, and webcasting), recorded broadcasting technologies (podcasting, audio and video players, and storage devices), and telephones are just a few of the technological tools and resources available (According to UNESCO). "In Indian schools and other educational institutions, English is taught as a second language." Even though English is an international language, those who master and speak it fluently benefit academically, socially, and professionally [2]. Teachers of English face situations in which students are hesitant to speak due to inhibition, nervousness, fear of making mistakes, and a lack of self-confidence. Coming together is the first step, staying together is progress, and working together is success (Henry Ford). Collaborative learning (CL) is a method of learning and teaching in which groups of students collaborate to solve a problem or complete a task (Lev Vygotsky).

2 Literature Review

Information and communication technology (ICT) is a scientific, technological, and engineering discipline and management technique for handling information, as well as its application and connection to social, economic, and cultural issues [3]. Technology is used "to facilitate the acquisition of skills such as critical thinking, independent learning, communication and lifelong learning involving analysis, synthesis, evaluation, and organization of information" [4].

"Collaborative learning is better suited for elementary school children because they are not ready to work in groups because they lack the necessary social skills, whereas collaborative learning is better suited for college and university students because they already have the necessary social skills as well as a strong desire to achieve academic goals" [5]. Collaborative learning is a method of instruction in which students work in small heterogeneous groups to solve a problem, complete a

project, or achieve other instructional goals, with teachers serving as guides or facilitators [6]. "As an educational strategy, CL seeks to organize classroom activities into academic and social learning experiences." "It is a setting in which students work together in small groups to help themselves and others learn" [7]. In the teaching and learning process, computer-supported collaborative learning (CSCL) is an alternative method to a lecture-based paradigm. It is a popular learner-centered paradigm. CSCL results in increased learning achievement and motivation [8]. A learner-centered approach is a solution to a plethora of problems [9]. According to Shakibaei, Shahamat, and Namaziandost, "The role of teachers in using CL shifts from knowledge transmission to facilitators of student learning." This role entails facilitating, modeling, and coaching, and teachers who take on this role must maintain a safe, nonthreatening, and learner-centered environment [10]. When using this method, teachers frequently use collaborative learning activities such as group discussion, role-playing, and debate to help students practice some expressions learned in class after students each generated some ideas through individual learning [11]. Other three collaborative learning activities focused on integrating all four LSWR language skills. A learning together approach to second language acquisition creates a space for teachers and students to come together in their naturally separate groups to learn more about their roles in supporting their English language learning.

3 Need of ICT in the Teaching and Learning Process

We know that in this pandemic (COVID-19) situation, the scenario necessitates the use of ICT in educational settings. ICT benefits both traditional and remote education systems. It has the potential to change, accelerate, improve, and build abilities to excite and engage children, as well as provide economic sustainability for tomorrow's workforce and aid in the evolution of schools. It aids in creating an input-rich learning environment. It keeps students up to date on the most recent technological advancements.

4 Methodology

This is an experimental study. Peer activity was used to improve integrating skills. There were ten groups of forty students. There were ten groups of forty students- each group with four people. To solve a Mistry, the following Think–Pair–Share activity was used. This is a collaborative active learning strategy in which students work together to complete an instructor-assigned task. Students first think individually, then in pairs or groups, and finally as a class (share). The students watch video "secret for two." While carrying out a task, the following steps were taken.

4.1 Warm-Up

Explain the activity to the students who were working on the "secret for two" case. The activity began with the question "what do you mean by detective?" displayed on the screen by the projector. Students were instructed to solve the mystery by assisting one another and using every clue provided by the teacher. Students were told to work in groups to solve the mystery. Students read aloud the rules written on the screen. To complete the task successfully, the students were given clear instructions. As part of their teamwork, the students were assigned different roles.

4.2 Student Roles

1. Writer—only this person can write in the team; no one else can write.
2. Reader—only the chosen person can read the paper distributed to the entire team.
3. Presenter—only the chosen individual can present the answers to the entire team.
4. Monitor—only the chosen individual can keep track of the entire group's members.

For each activity, the students switched jobs. Nobody ever did the same job twice. Every student was very interested in learning about each other's strengths and weaknesses. Some students excelled at writing, while others excelled at explaining and solving mysteries.

4.3 Grouping Students

Three groups of students were formed. Each group was given the name of a famous detective and given instructions on what to do. Tags were created for student jobs such as writer, reader, and presenter. Before the students began working in groups, the task was explained to them.

4.4 Language Focused Task

Grammar and vocabulary were focused on after grouping. The students were told to follow the rules before passing the activity sheet. Other members, apart from the writer, can assist with grammar and vocabulary. The sheets were distributed to the students so that they could practice compound sentences.

4.5 Monitoring Groups

Students were being watched to see if they were all participating or not.

4.6 Collecting Responses Checking for Accuracy

Each group's responses were collected. Students created compound sentences, and two points were awarded for each correct answer. Each student actively participated in the task. Later, the presenter was asked to provide examples of compound and complex sentences. To include all of the students in the class, an audience poll was used.

4.7 Changing Roles

Roles were changed for the next activity.

4.8 Listening Activity Through Audio Aids

The teacher played the recorded lesson for the students to carefully listen to. Following that, each group was given a set of questions to answer in order to work as a team.

4.9 Collecting Responses and Checking for Comprehension

The instructor checked comprehension and meaning after collecting the responses, but not accuracy. Questions were posed to all of the groups, and the responses were collected and graded accordingly.

4.10 Changing Roles Again

Working on the paragraph: Students were asked to identify two pieces of evidence from the story that they believed would support the topic sentence. These could be excerpts from the story.

4.11 Solving the Mystery

Students were given questions to answer in a paragraph. They were also instructed to read and provide evidence based on the clue given to them.

4.12 Collecting Answers

The reader is instructed to read out the paragraph written by them when collecting answers. The emphasis was on well-formed paragraphs:

a. The topic sentence.
b. The topic sentence.
c. The concluding sentence.

4.13 Final Comprehension Check

The ending of the story was projected (video), and the students were asked to determine how many of them were successful in solving the mystery. This was perceived to be the part of collaborative learning in which students interacted with one another.

4.14 Wrap Up

After the activity was completed, the teacher awarded prizes to the winning teams.

5 Data Collection Procedures

Following the activity, data was gathered from students via questionnaires. Google forms were used to create the questions. This questionnaire is made up of ten closed-ended questions with multiple choice answers. They are closed-ended inquiries. To save time, the Google form was sent to their email addresses. The responses of the students are graphically analyzed.

6 Results

6.1 *Results for the Effectiveness of CL on Students' Integrating Language Skills Through Questionnaire Are as Follows*

The below graph says that there was a significant change in students participating actively in the computer-supported collaborative learning activity and improved students' learning motivation when compared with traditional teaching. In the students' questionnaire, the majority of the students, i.e., 40.5% prefer learning through collaborative learning which enhanced language skills. About 35.6% agree that collaborative learning is an effective way of instructional method. But 15.7% of students felt shy, inhibited, fear of making mistakes, and lacks of confidence. Whereas 8.2% was neutral (Figs. 1 and 2).

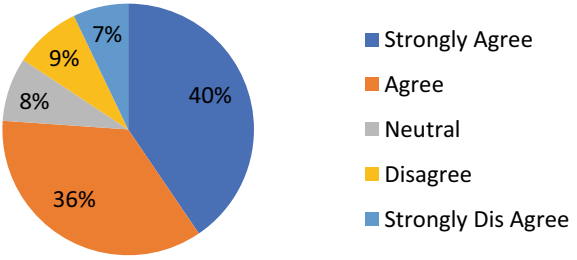


Fig. 1 Responses of the students in pie diagram in percentage

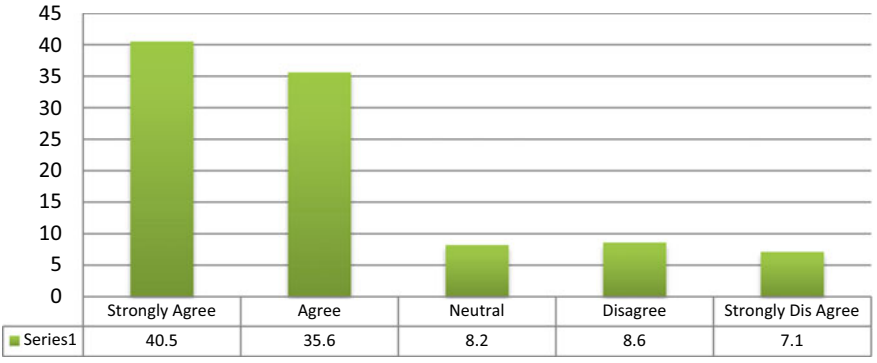


Fig. 2 Responses of the students in bar graph in percentage

7 Conclusion

Finally, the majority of students agree that collaborative learning improves their English language skills as well as their motivation to learn. Group or peer learning allows them to practice the language that helped them develop their self-esteem while also reducing inhibition. Group learning helps to deepen one's understanding of a topic by allowing them to interact with others.

Collaborative learning improves student interaction, confidence, and ability to learn. This activity can also be done via online platforms such as Zoom meetings, Google Meets, Microsoft Teams, and so on. These ICT tools allow students to learn at any time and in any location, while also providing a wealth of resources.

References

1. M. Pun, The use of multimedia technology in English language teaching: A global perspective. Crossing the border: Int. J. Interdisc. Stud. **1**(1), 29–38 (2013)
2. D. Efrizal, Improving students' speaking through communicative language teaching method at Mts Ja-alhaq, Sentot Ali Basa Islamic boarding school of Bengkulu, Indonesia. Int. J. Humanit. Soc. Sci. **2**(20), 127–134 (2012)
3. W. Haddad, S. Jurich, *ICT for education: Potential and potency. Technologies for education: Potential, parameters and prospects* (UNESCO and Academy for Educational Development, 2002), pp. 28–40
4. S. Haydar, Literacy drive and rehabilitation of the street children of Kolkata: An analytical study. Lulu.com. (2017)
5. A. Redes, Collaborative learning and teaching in practice. J. Plus Educ. **16**(2), 334–345 (2016)
6. R. Siltala, *Innovativity and cooperative learning in business life and teaching* (University of Turku, 2010), pp. 100–106
7. D.W. Johnson, R.T. Johnson, An educational psychology success story: Social interdependence theory and cooperative learning. Educ. Res. **38**(5), 365–379 (2009)
8. A.H. Kim, S. Vaughn, J.K. Klingner, A.L. Woodruff, C. Klein Reutebuch, K. Kouzekanani, Improving the reading comprehension of middle school students with disabilities through computer-assisted collaborative strategic reading. Remedial Spec. Educ. **27**(4), 235–249 (2006)
9. R. Mambwe, Social studies student teachers views on the implementation of learner-centred approach in Zambian Primary Schools. Int. J. Res. Innovation Soc. Sci. (IJRISS). **3**(3) (2019)
10. E. Namaziandost, L. Neisi, Kheryadi, M. Nasri, Enhancing oral proficiency through cooperative learning among intermediate EFL learners: English learning motivation in focus. Cogent Educ. **6**(1), 1683933 (2019)
11. C. Wang, T. Fang, Y. Gu, Learning performance and behavioral patterns of online collaborative learning: Impact of cognitive load and affordances of different multimedia. Comp. Educ. 143, 103683 (2020)

Modified Sierpinski Carpet Fractal Antenna for the Wireless Applications

Kuldip Kumar

Abstract The modified Sierpinski carpet fractal antenna is designed and fabricated to obtain multiband for wireless communication applications. The size of the patch antenna is of 30 mm × 30 mm. In terms of wavelength, the antenna size is 0.34 λ at the lowest frequency (where λ = wavelength at the resonant frequency). The obtained results of the proposed antenna fulfill existing need and the VSWR obtained is in acceptable range. Three measured resonant frequencies appeared at 3.4 GHz, 4.49 GHz and at 5.13 GHz for second iteration. The bandwidth obtained is 100 MHz, 50 MHz and 190 MHz, respectively, in these three bands. The gain obtained is 2 dBi, 2.4 dBi and 3.8 dBi at the respective three resonant frequencies. The measured results show that antenna is multiband in nature and is appropriate for wireless communication applications.

Keywords Fractal · Multiband antenna · Sierpinski carpet antenna · Scaling factor

1 Overview

With the progress of the wireless communication field, future system is likely to deliver multimedia and communication services based on high data rate. Numerous applications such as radar communication measurement system, imaging and hand-held wireless devices required integrated antenna of low cost, small size and low profile with appreciable broadband performance. Because the radiation method of employing different antennas for different frequency bands results in space limitation, multiband antenna [1–6] should be utilized to integrate more than one communication service in a wireless device.

Fractal geometry is a type of geometry that has a number of distinct characteristics [1]. Fractals are having space-filling curves, which means that electrically huge features can be competently packed into a tiny part, allowing the antenna to be smaller [7].

K. Kumar (✉)
ECE Department, Chandigarh University, Gharuan, Mohali, India
e-mail: kpahwa2002@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_3

Fractals are having the self-similar property in that some parts of it have the same shape as the general geometry, but only at a smaller scale [7, 8]. Fractals' self-similarity feature causes multiband behavior [9–11]. The self-similar features of fractal antenna allow them to operate on wide range of frequencies [2, 3].

The second iteration of a modified Sierpinski carpet shaped fractal antenna has been designed in this paper. When compared to a non-fractal square antenna of the same physical area, the modified Sierpinski carpet shaped fractal antenna shows an effective resonance at a lower frequency.

Fractal structure is generated by iterated function system (IFS). Fractal antenna structure is iterated by using scaling factor [9] which is expressed as

$$\xi = \frac{h_n}{h_n + 1} \tag{1}$$

where

ξ = scaling factor ratio
h = height of iterated antenna
n = iteration number.

The fractal antenna has a Sierpinski carpet shape in which two iterations have been performed. The experimental and simulated results show that antenna behavior is multiband in nature. With the increase in iteration order, resonant frequency gets lowered fulfilling the property of fractal antennas [12–14].

There are four sections in this study. The antenna design and structure are described in Sect. 2. Section 3 presents the simulated and measured results. The conclusion is found in Sect. 4.

2 Design of Antenna

The Sierpinski carpet fractal antenna has been modified. The IE3D software (MoM) is used to simulate the proposed antenna. The FR4 epoxy is used as a substrate. The substrate measures 1.575 mm thickness, and 4.3 is the substrate's dielectric constant (r). The square shape is cut out of the square patch antenna's center, indicating the first iteration, followed by the second iteration. The redesigned Sierpinski carpet shaped fractal antenna iterations are depicted in Fig. 1. The proposed antenna has a side length of 30 mm (without iteration). For the first iteration, a square with a 10 mm side length is carved out of the antenna's center. The fractal antenna contains four identical squares at its corners in the second iteration. 3.33 mm is the length of each square's side.

The length of the antenna is computed mathematically [8, 9] using Eq. (2).

$$L = \frac{c}{2 f_r \sqrt{\epsilon_{\text{eff}}}} - 2\Delta L \tag{2}$$

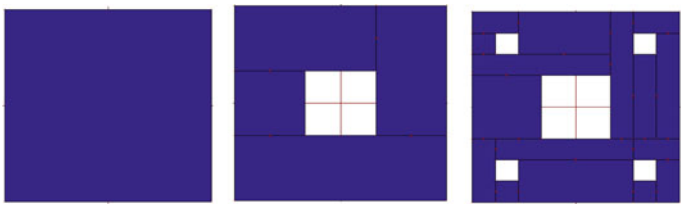


Fig. 1 Iterations of the proposed antenna

where

c = speed of light in vacuum
f_r = resonant frequency
ε_{eff} = effective permittivity and it is calculated using Eq. (3) [15, 16].

$$\in_{\text{eff}} = \frac{\in_r +1}{2} + \frac{\in_r -1}{2}\left(\frac{1}{\sqrt{1+12\frac{d}{W}}}\right) \tag{3}$$

where

ε_r = relative permittivity
d = thickness of substrate
W = width.

The width (W) is calculated by using Eq. (4)

$$W = \frac{c}{2f_r}\sqrt{\frac{2}{\in_r +1}} \tag{4}$$

Equation (5) is used to determine the patch's extended incremental length.

$$\Delta L = 0.412d\frac{(\in_{\text{eff}} +0.3)\left(\frac{W}{d}+0.264\right)}{(\in_{\text{eff}} -0.258)\left(\frac{W}{d}+0.8\right)} \tag{5}$$

The antenna is coaxially fed at the location $x = 6$ mm and $y = 8$ mm.
Figure 1 shows the iterations of the proposed antenna.
Figure 2 shows the snapshot of the fabricated antenna utilizing FR4 substrate. The coaxial probe SMA connector is used to feed the antenna.

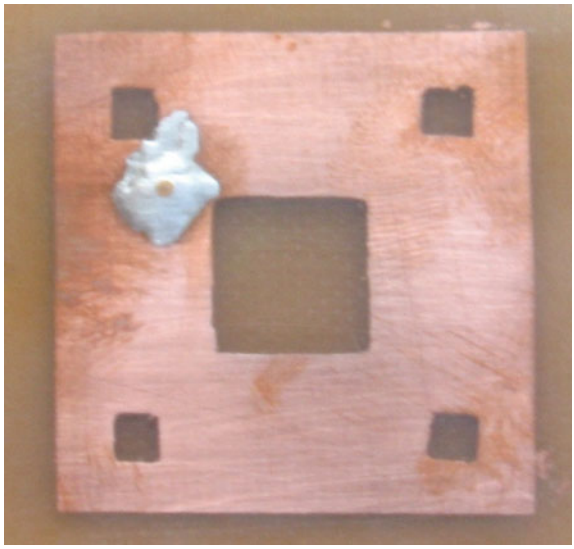


Fig. 2 Snapshot of the built antenna

3 Results and Discussion

3.1 Simulated and Experimental Results

The proposed antenna is simulated using IE3D electromagnetic simulator. The antenna configurations with different iterations in the previous section have been simulated. Figure 3 shows the measured return loss (S_{11}) characteristics of the modified Sierpinski carpet fractal antenna. The designed antenna resonates at three frequencies, i.e., having center frequency at 3.4 GHz, 4.49 GHz and 5.13 GHz at second iteration.

The resonance shifts downward as the iteration goes, and due to the space-filling technique, discrete multiband responses are recorded. With larger repetition orders, this result demonstrates the frequency dropping phenomenon.

Figure 4 depicts the experimental set-up using site analyzer to test the fabricated antenna.

Figure 5 depicts the simulated gain versus frequency plot of the designed antenna.

The gain versus frequency plot shows that the antenna has 2.0 dBi gain at 3.4 GHz whereas at 4.49 GHz it is 2.4 dBi and at 5.13 GHz, the gain is 3.8 dBi. The bandwidth obtained is 100 MHz, 50 MHz and 190 MHz at three respective resonant bands (Fig. 6; Table 1).

The value of VSWR lies from 1 to 2 at resonant frequencies. The VSWR is calculated by

$$\text{VSWR} = \frac{1 + S_{11}}{1 - S_{11}} \tag{6}$$

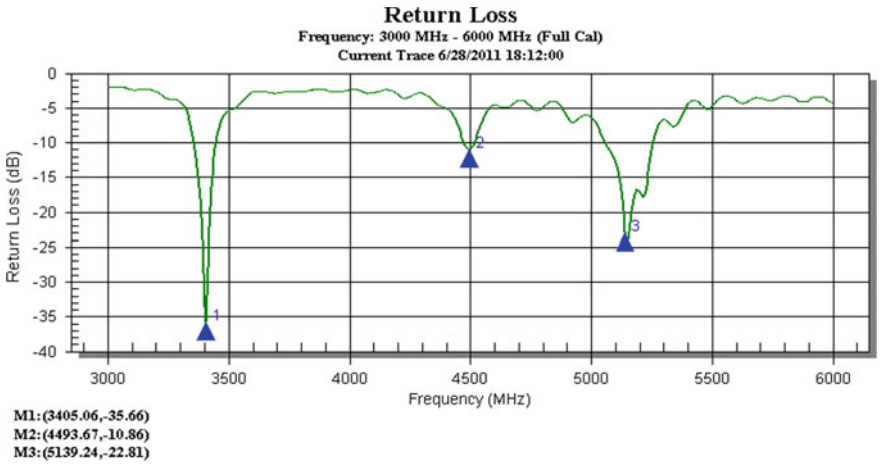


Fig. 3 Measured return loss of proposed antenna with second iteration

Fig. 4 Experimental set-up

4 Conclusion

The Sierpinski carpet fractal shape is used to design and fabricate a fractal antenna. The measured resonant frequencies of the fabricated antenna are 3.4 GHz, 4.49 GHz and 5.13 GHz with a gain of 2.0 dBi, 2.4 dBi and 3.8 dBi, respectively. The bandwidth obtained is 100 MHz, 50 MHz and 190 MHz at three respective resonant bands. Thus, the designed antenna is miniaturized and multiband with considerable bandwidth compatible for wireless communication applications.

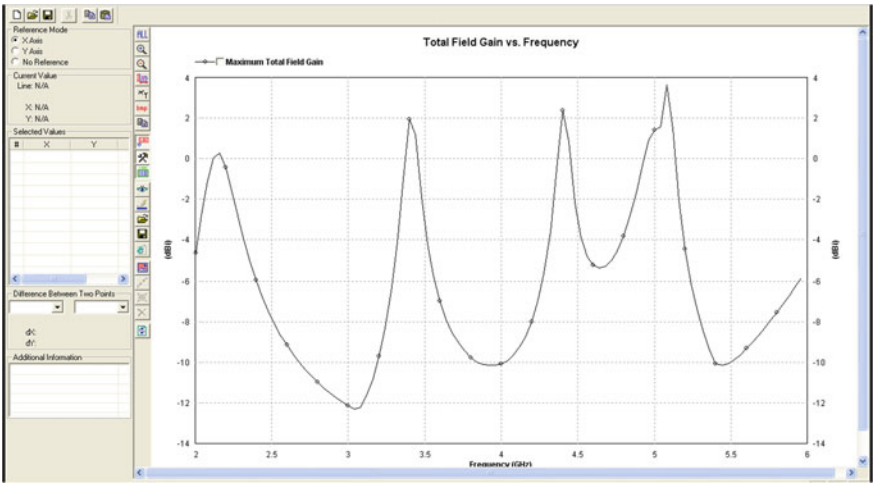


Fig. 5 Gain versus frequency plot

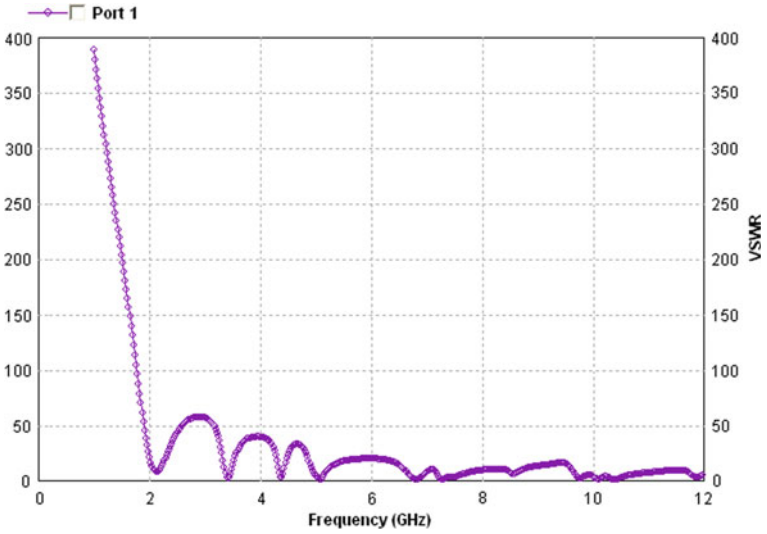


Fig. 6 VSWR characteristics of the proposed antenna

Table 1 Measured return loss of fabricated antenna with second iteration

Band	Resonant frequency (GHz)	Return loss (dB)
M1	3.40	−35.66
M2	4.49	−10.86
M3	5.13	−22.81

Acknowledgements Kuldip Kumar Pahwa is thankful to the management and head, ECE department of Chandigarh University, Mohali, India for their support.

References

1. A. Sabouni, S. Noghanian, M.S. Abrishamian, Analysis of microstrip fractal patch antenna by 3-D FDTD method, in *IEEE International Workshop on Antenna Technology* (2005)
2. C. Puente-Baliarda, J. Romeu, R. Pous, A. Cardama, Perturbation of the Sierpinski antenna to allocate operating bands. Electron. Lett. **32**, 2186–2188 (1996)
3. C. Puente, *On the Behaviour of the Sierpinski Multiband Fractal Antenna* (IEEE AP-46, 1998), pp. 517–524
4. C.T.P. Song, P.S. Hall, H. Ghafouri-Shiraz, Shorted fractal Sierpinski monopole antenna. IEEE Trans. Antennas Propag. **52**(10), 2564–2570 (2004)
5. D.C. Chang, CPW-fed circular fractal slot antenna design for dual-band applications. IEEE Trans. Antennas Propag. **56**(12) (2008)
6. J.P. Gianvittorrio, Y.R. Samii, Fractal antennas: a novel antenna miniaturization technique and applications. IEEE Antenna Propag. Mag. **44**(1) (2002)
7. L. Cao, S. Yan, H. Yang, Study and design of a modified fractal antenna for RFID applications, in *ISECS International Colloquium on Computing, Communication, Control and Management* (IEEE, 2009)
8. M. Comisso, Theoretical and numerical analysis of the resonant behaviour of the Minkowski fractal dipole antenna. IET Microw. Antenna Propag. **3**, 456–464 (2008)
9. M.N. Jahromi, Novel wideband planar fractal monopole antenna. IEEE Trans. Antennas Propag. **56**(12) (2008)
10. W. Wu, B.Z. Wang, X. Song, A pattern reconfigurable planner fractal antenna and its characteristic-mode analysis. IEEE Antennas Propag. Mag. **49**(3) (2007)
11. K.S. Aneesh, T.K. Sreeja, *A Modified Fractal Antenna for Multiband Applications* (ICCCCT, 2010), pp. 47–51
12. M.S. Maharana, G.P. Mishra, B.B. Mangaraj, Design and simulation of a Sierpinski carpet antenna for 5G commercial applications, in *IEEE WiSPNET Conference* (2017), pp. 1747–1750
13. M.R. Jena, B.B. Mangaraj, R. Pathak, A novel Sierpinski carpet fractal antenna with improved performances. Am. J. Electr. Electron. Eng. **2**(3), 62–66 (2014)
14. M. Gupta, V. Mathur, *Sierpinski Fractal Antenna for Internet of Things Applications* (ICEMS, 2016)
15. Y.B. Chaouche, M. Nedil, B. Hammache, M. Belazzoug, Design of modified Sierpinski Gasket fractal antenna for tri-band applications, in *2019 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting* (2019), pp. 889–890
16. T. Ali, B.K. Subhash, R.C. Biradar, A miniaturized decagonal Sierpinski UWB fractal antenna. Progr. Electromagn. Res. C **84**, 161–174 (2018)

COVID-19 Recovery Prediction Using Regression-Based Machine Learning Approaches

Gurman Kaur, Harsheen Kaur, and Harshpreet Singh

Abstract The number of people affected by the ongoing coronavirus has fluctuated rapidly and it has become strenuous to predict when will this pandemic end. To impede the spread of this virus, it is the need of the hour to maintain a social distance, wear masks and sanitize regularly. No doubt, the mortality rate has escalated, summing to a large percentage of population and destroying lives and economies with it. In India, mortality has climbed to as high as 3.2% as per the Indian Express, with recovery rate summing to 81.55% according to Times of India. Therefore, it has become prudent to determine and predict the effect and drawbacks of various factors such as testing, mortality rate and confirmed cases on the recovery rate. Due to progression and evolution in the discipline of machine learning, it has become practicable to get a middling figure of effects of these factors on death rate. Regression, one of the most broadly exhausted machine learning and statistics algorithm, is used to make predictions from data by learning the relationship between the features. In this article, regression algorithms are used to anticipate the same by using a cumulative data of all states in India. Study compares the ramification of the number of testing done and their impact on the recuperation of life due to the virus. Therefore, based on the research and computing, it was found that ridge regression gave the highest accuracy equivalent to 99.6%.

Keywords COVID-19 · Recovery · Testing · Regression · Machine learning

1 Introduction

As comprehended from history, mankind has been enduring through numerous epidemics and pandemics. One such pneumonia of unknown aetiology (novel coronavirus) that shook the entire population and economy of all the nations across the globe originated from Wuhan, Hubei Province of the People's Republic of China in December 2019. Initially named 2019 novel coronavirus, it was soon renamed

G. Kaur · H. Kaur · H. Singh (✉)
Guru Gobind Singh Indraprastha University, New Delhi, India
e-mail: harshpreets19@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_4

as SARS-CoV-2, due to previously identified—**severe acute respiratory syndrome coronavirus**. Eventually, this epidemic stirred its helm into a catastrophic and insidious global widespread disease. A person suffering from this virus could experience a disorder as customary as common cold to a malady such as extreme respiratory disease and death. According to worldometers.info, about 32,572,774 cases have been documented with an aggregate of recovered summing to 24,048,380 and deceased to 989,981 as of 25-September 20 [1]. Such sudden increase in the number of people infected by this virus led to vigorous trepidation and became a health emergency. All populace has united in this time of malaise and the federations working assiduously to enhance the medical facilities, laboratories and response procedures. COVID-19 has deleteriously spread into each state, town and district of India, taking with it millions of lives. India, the second country on the list of highest population in the world was struck with the coronavirus on January, 30, 20 with origin from China in a coastal state of Kerala [2]. Since then the dramatic increase in cases due to arrival of passengers through airways led the government to impose manifold lockdowns to hinder the spread of the virus. As a consequence of these lockdowns, GDP of India plummeted by 23.9% than primarily [3]. Therefore, in lieu of this loss, it has become imperative to find a solution to this mishap. It is judicious to determine the recovery based on the testing to brace up the nations to prepare accordingly.

Arora et al. [4] developed a prediction model using a recurrent neural network (RNN)-based long short-term memory (LSTM) variants such as deep LSTM, convolutional LSTM and bidirectional LSTM have predicted the positive reported cases in India on weekly and daily basis. They have reported their predictions of MAPE of less than 3% till May 14, 20. Rustam et al. [5] based on time forecasting techniques such as LR, LASSO, SVM and ES proposed to reckon the number of people to be infected by the novel virus. Three types of forecasts are made by the models namely number of new cases, number of deaths and number of recoveries for the coming 10 days. Correspondingly, Liu et al. [6] predicted the real-time COVID-19 activity in China and produced the 2-day forecast ahead of real time using the Augmented ARGONet. Vaishnav et al. [7] administered GMDH to analyze the influence of relaxation of lockdown in India and previse the number of active and confirmed cases during the same time period. They achieved an accuracy of 2.58% in GMDH model and 2.00% in forecasting. Various techniques for COVID-19 detection and diagnostics are being implemented throughout the world. One such effective technique is the analysis of computed tomography (CT) imagining. Barstugan et al. [8] comprehensively implemented and analyzed the use of CT images for early phase detection of COVID-19 by using machine learning methods. Detection is implemented on the CT images of the chest regions. According to their report, expert radiologists detected from the CT images that COVID-19 displays different behaviors from other viral pneumonias. By using four datasets formed by using different patch sizes from 150 CT images, different machine learning techniques such as GLCM, LDP and DWT were deployed for extraction of feature. Further, by classification and by cross validating the results up to tenfold during the implementation process they were able to attain a maximum accuracy of 99.68% on GLSZM. Kumar et al. [9] examined the different diagnostic methods for COVID-19 and explained the advantages and

disadvantages of implementing every technique thoroughly further examining the newer detection methods in the development phase and the feasibility of advanced platforms. According to their research, CT scans are highly sensitive and are used as an auxiliary diagnostic method for COVID-19. They concluded that more accurate conclusions can be obtained using a combination of CT scans and other sensitive diagnostic tests. High resolution CT of the chest is also proved as an essential tool for detection of SARS-CoV-2 at an early stage and to take rapid and necessary interventions. Liang et al. [10] reported the factors that impact the mortality rate across countries. Their study depicts that testing, government effectiveness rate and number of hospital beds are inversely proportional to the mortality rate. Nevertheless, the fatality is analogous to the proportion of population older than 65 and transport infrastructure quality score. For all nations studied by them, the mean mortality rate was approximately 3.70%. Yadav et al. [11] put forward the SVM method to scrutinize five different tasks: (I) Predicting the spread of coronavirus across regions. (II) Growth rates and mitigation type analyzation across countries. (III) Predicting how the epidemic will end. (IV) Analyzing the transmission rate of the virus. (V) Correlating the coronavirus and weather conditions.

The regression model offered in the study which predicts the recovery of people contaminated with the virus based on the dataset of 28 states and eight union territories of India. As per the research, it was found out that there are scarcely any papers based on recovery prediction of the novel coronavirus, expressed using testing. Therefore, several regression algorithms such as linear regression, support vector regression, ridge regression, decision tree regressor, random forest regressor, extra-tree regressor and elastic net are used for the same. After testing these multifold models on the concerned dataset, linear regressor and ridge regressor came out to be the highest in accuracies. As per the author's knowledge, research papers on recovery of COVID-19 in India with respect to the testing numbers have not been reported till now and thus, it is expected that the research proves to be beneficent and helpful.

The development of the paper is as follows: Sect. 2 gives the backdrop of profuse regression methodologies used. Section 3 gives a brief about the data used, preprocessing and visualization of the dataset. It contains heat maps to give correlation between columns, processed data pair plots and data pair plots after scaling. Section 4 gives a brief discussion of the methodology and the used libraries/packages, and the evaluation metrics of the model. Section 5 gives the final predictions and results, along with the fit line for the model. Finally, the study is concluded in Sect. 6 with a brief mention of future aspects of the study.

2 Methods Used

The algorithms implemented in this study are linear regression, support vector machine, ridge regression, elastic net, random forests, decision trees and artificial neural networks. This section would focus on giving a brief theoretical insight into

the actual regression algorithms used in this study—the idea behind them and their mathematical scopes.

2.1 Linear Regression (LR)

In linear models, there is a dependent variable which is calculated using other independent variables using the equation

$$a = cx + d \tag{1}$$

(In case of one independent variable) and where m is the slope of the line and c is the y intercept. The study is based on multiple independent variables together being responsible for the dependent variable, thus multivariate linear regression has been applied. The mathematical expression for this machine learning algorithm is as follows:

$$P_i = q + r_1 s_i^{(1)} + r_2 s_i^{(2)} + \cdots + r_n s_i^{(n)} \tag{2}$$

where

P_i = estimate of ith component (or label) of dependent variable y
n = number of independent variables,
s_i^j = ith component of the jth independent variable/feature (or feature, i.e., a known input)
q = intercept (or bias)
r_n = weight of the nth feature

2.2 Ridge Regression

Ridge regression (RR) yields a class of biased estimators indexed by a scalar non-negative parameter, from which the best estimator is determined, so as to determine a best choice for the ridge parameter [12].

If

$$\gamma = (\gamma_1, \gamma_p) \tag{3}$$

is a column vector of p regression coefficients, A is the $n \times p$ matrix of variants, standardized such that $A'A$ is in correlation form. $\beta = (\beta_1, \beta_n)'$ is a n-dimensional vector, ridge estimations of the regression coefficients are given by

$$\gamma_n = \left(A'A + kI\right)^{-1} A'\beta \tag{4}$$

where n controls the amount of shrinkage of the regression coefficients and I is the identity matrix [13]. This shrinkage parameter is said to be directly proportional to the model's robustness with collinearity. In this study, the shrinkage parameter α is given the value 0.001.

2.3 Elastic Net

Elastic net (ENET) shrinks the regression coefficients by combining L_1-norm penalty (lasso) and L_2-norm penalty (ridge) together [14]. After combining, the resulting ENET shrinks and selects, resulting in good predictive performance in simulated and real data [15].

Here, the model has been given a value of alpha (α) = 0.0001, with $l1$ penalty, by setting the $l1$_ratio parameter to 1. L_1 is a regularization parameter that selects variables according to the amount of penalization on the L_1-norm of the coefficients [16].

2.4 Support Vector Regression

Support vector regression (SVR), a supervised-learning approach that has been proven to be an effective tool in real-value function estimation, is characterized mainly by the use of kernels, margin control and the number of support vectors, and trains using a loss function, which equally penalizes high and low misestimates [17].

One of the main advantages of SVR is that its computational complexity does not depend on the input space dimensionality, and its structural risk minimization (SRM) principle which has excellent generalization capability, with high prediction accuracy, even superior to the empirical risk minimization (ERM) principle as adopted in neural networks [17, 18].

The mathematical explanation of this concept is out of the scope of this study, thus for proper mathematical concepts of support vectors and SVR, we refer you to [18]. In this study, authors have implemented this algorithm by using 'rbf' kernel, i.e., radial basis function kernel. RBF kernel function can be applied to any sample distribution through the choice of parameters and is being more and more used in the nonlinear mapping of SVMs. RBF kernel function expression is [19]

$$M\left(y_q, y_p\right) = \exp\left(-\beta \left\| y_q - y_p \right\|\right)^2 \tag{5}$$

2.5 Decision Trees

In decision tree (DT) modeling, an empirical tree represents a data segmentation that is created by applying a set of sequential rules, which can be used for prediction through the repetitive process of splitting [20]. While building a DT, a splitting criterion needs to be considered, which is like a function that defines how data should be split in order to maximize performance [21]. DT can be developed using many parameters such as maximum number of leaf nodes, maximum depth, minimum number of samples to be split, minimum weight fraction of the leaf, minimum impurity decrease, minimum impurity split and maximum number of features. In this study, DT has been trained to have maximum depth of 100, maximum number of leaf nodes as 25, and split the leaf nodes according to mean squared error criterion.

2.6 Random Forests

Random forest (RF) is an ensemble of unpruned regression trees which are created using training data's bootstrapping samples and random feature selection in tree induction, for prediction by aggregating (majority vote or averaging) the predictions of the ensemble [22]. RF is an ensemble of B trees $\{P_1(Y), P_B(Y)\}$, where $Y = \{y_1, y_x\}$ is a p-dimensional vector of descriptors. The ensemble produces B outputs

$$\left\{\hat{X}_1 = P_1(Y),\ \hat{X}_B = P_B(Y)\right\} \tag{6}$$

where $\hat{X}_B = 1$, B is the prediction for a value by the bth tree. Outputs of all trees are then aggregated to produce one final prediction, $\hat{Y}$, which is the average of the individual tree predictions [22]. The random behavior of a tree creates differences in individual trees predictions and therefore, the overall prediction of the forest is computed as the average of individual tree predictions, which is a multidimensional step function that predicts smooth functions as it focusses on aggregating a large number of different trees [23].

2.7 Extra Trees

The extra-trees (ET) algorithm works on building an ensemble of regression trees in such a way that it splits nodes by choosing cut-points fully at random and that it uses the whole learning sample (rather than a bootstrap replica as used by RFs and DT algorithm) to grow the trees [24]. This algorithm was considered better than previous ensemble methods because the latter one experienced modification in the splitting criterion algorithm while searching for optimal fit in tree growing [24].

2.8 Deep Learning Multilayer Perceptron

A multilayer perceptron (MLP) is defined by its architecture (number of layers, number of neurons per layer, etc.), the activation function used (transfer functions to compute the weighted sum of input and biases, which further decides if a neuron can be fired or not, and manipulates the presented data usually through gradient descent to produce an output for the neural network (NN), that contains the parameters in the data), the learning rate, specification of events—whether weight updating is carried out in 'real-time' or 'off-line' (at each epoch) [25, 26].

Authors have developed a relevant MLP neural network model, by constructing 12 dense input layers, eight dense hidden layers and one dense output layer, along with activation functions as—Sigmoid, ReLU (rectified linear unit) and linear, respectively. Detailed explanation for these mentioned transfer functions can be referred from [26].

3 Data and Methodology

Previous section gave a brief description about the type of regression algorithms to be used in this study. This section will further explain the proposed pipeline followed for the concerned study. The whole study pipeline from data analysis to machine learning is visualized in the form of a chart as shown in Fig. 1.

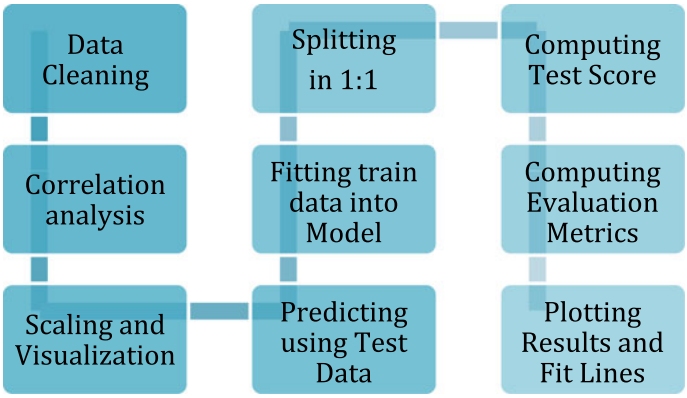


Fig. 1 Study pipeline from data preprocessing to result plotting

	State	Confirmed	Recovered	Deceased	Other	Tested
count	5585.000000	5.585000e+03	5.585000e+03	5585.000000	5585.000000	5.585000e+03
mean	17.857833	3.143109e+04	2.012032e+04	738.583706	6.935721	4.329278e+05
std	10.605885	1.610565e+05	1.093940e+05	3633.196205	44.649182	1.869239e+06
min	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.000000e+00
25%	9.000000	5.300000e+01	1.200000e+01	0.000000	0.000000	4.067000e+03
50%	18.000000	1.075000e+03	3.720000e+02	7.000000	0.000000	3.543000e+04
75%	27.000000	8.927000e+03	4.838000e+03	138.000000	0.000000	2.615870e+05
max	36.000000	2.719499e+06	1.992157e+06	52060.000000	507.000000	3.094126e+07

Fig. 2 Data description. The range of the values can be seen at extreme levels from minimum zero to maximum 30,941,260

3.1 Data Description

The dataset used in the study has been downloaded from the GitHub repository of covid19india.org which is COVID-19-India API [27], in the form of a csv file named 'states.csv', and imported into a data frame, for ease of the study. The initial data contained a total of 5585 entries, and 7 columns, and its brief description is given in Fig. 2.

3.2 Data Preprocessing and Visualization

Feature selection is a type of data engineering which focuses on selecting particular attributes/features from a set of numerous ones in the dataset, according to their contribution toward an efficient model. The original dataset mentioned above contained very few columns that did not have much relevance to the research aim. Hence, these columns were dropped directly before the preprocessing, namely—'Date', 'State', instead of applying feature selection algorithms. The NaN entries were too very limited and hence, treated directly by filling them with zero, since the presence of NaN entries prevents the model from developing a smooth fit. The cleaned data is then visualized in a correlation heat map to discover about the inter-column correlation measures, as shown in Fig. 3.

Correlation measure just gives the strength of linear relationship between the column attributes which can further be used to determine which type of model the concerned data might be fitted into.

After the successful correlation analysis, it becomes quite clear that the data, having a linear relationship, might fit well into any linear data. This further leads the study to visualize the data points into pair plots to confirm the above correlation. The data visualization can be seen from Fig. 4.

Regression models usually start overfitting when the predictor values have huge differences between them. This model overfitting leads to poor efficiency and

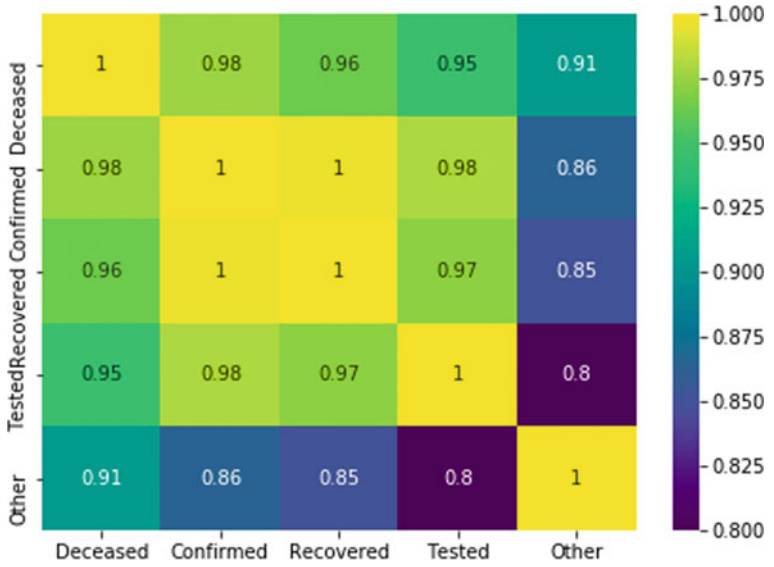


Fig. 3 Correlation heat map between the data attributes. This gives the correlation measure between the attributes which is further used to depict the extent of the linear relationship between them

increased mean squared error in the model predictions. To make the regression model stop from overfitting, it becomes important to scale the data between specified values. In this study, the data has been observed to have values as small as 0 and as large as 500,000. Thus, to avoid overfitting, it was required to scale the data between 0 and 1, using the min–max scaling algorithm, in which the data is linearly transformed between two values—a minimum value, and a maximum value, where in the study the minimum value is 0 and maximum value 1.

After successful data preprocessing and visualization, the data is split into two parts—train set and test set in the proportion of 1:1, i.e., 50% of the total data was distributed for training the respective models and the rest for testing the model fitting.

4 Methodology Used

The data analysis showcased in the previous sub-section clearly indicated that its study would be through supervised machine learning regression, since its data contained continuous values. Measures of correlation provide an initial impression of the extent of statistical dependence between variables, i.e., if the data variables are continuous, as is the case with the dataset used in this study, then a correlation coefficient must be calculated which can further be used as a measuring value of the relationship strength between them. The heat map for the same purpose has been plotted as shown in Fig. 1. In this study, the hypothesized dependent variable is

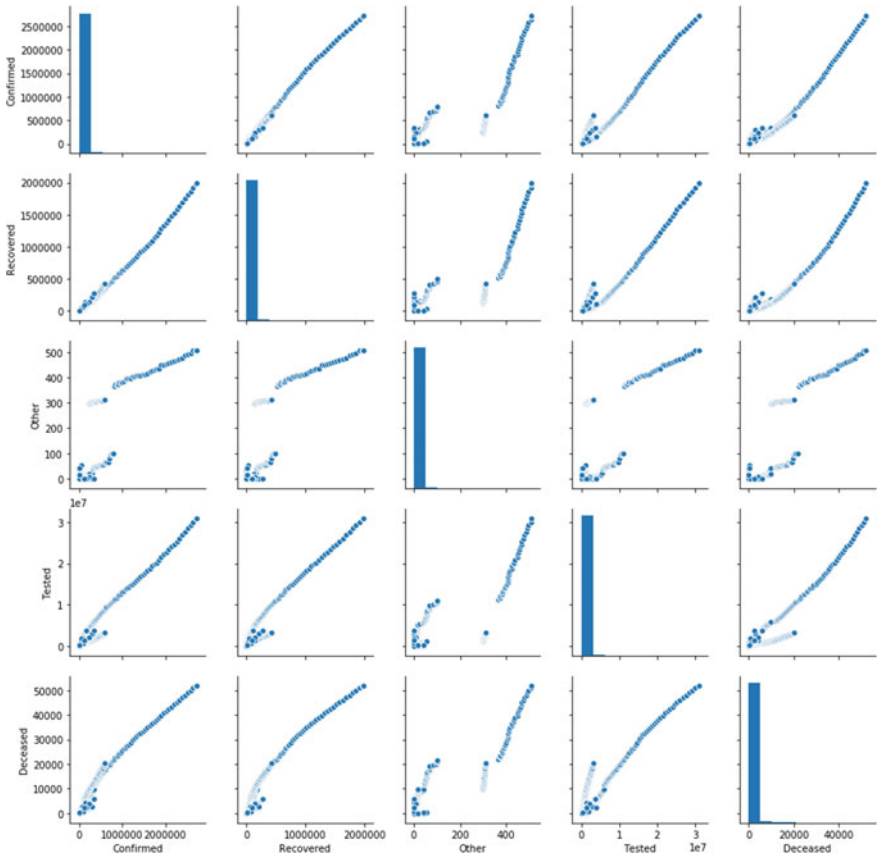


Fig. 4 Preprocessed data pair plot

the column 'recovered', and independent variables are the columns—'confirmed', 'tested', 'deceased' and 'other' for the study of the determination of recovered patients. It is to be noted that the implementation of the algorithms is done using available APIs such as—linear regression [28], support vector regression [28], ridge regression [28], elastic net [28], Lasso regression [28], random forests [28], decision trees [28], extra trees [28] and neural networks [29].

4.1 Performance Metrics

The task of a machine learning model is to learn the patterns and relations in a given set of data and train a model according to these relations so that any type of unseen or seen data can be used in future to generate relevant predictions. In order to verify, whether the model is behaving correctly as it is expected to, it is important to

measure the generated model predictions on a certain evaluation metrics to identify its accuracy in prediction and the minimum or maximum error generated by the model while predicting on the test set or any validated data. The metrics used in this study for measuring the performance are—root mean squared error (RMSE), mean absolute error (MAE), test accuracy, coefficient of determination (R^2) and cross-validation score (CV), and similarly, are used from the available sklearn API [28].

The coefficient of determination (R^2) is the criterion generally used in regression to test the adjustment of the model [30]. This coefficient is defined by

$$R^2 = 1 - \frac{\sum_i^{n-1} \left(y_i - \hat{y}_i\right)^2}{\sum_i^{n-1} (y_i - \overline{y})^2} \tag{7}$$

where y_i represent a data value from a dataset of length n, $\hat{y}_i$ its predicted/estimated value and $\overline{y}$ the mean of n data values.

Mean square error (MSE) corresponds to the prediction error per square (L_2–criterion method) and is defined by the following equation [30]

$$\text{MSE} = \frac{1}{p} \sum_j^{p-1} \left(a_j - \hat{a}_j\right)^2 \tag{8}$$

Lower the MSE value, more efficient is the model.

The root mean squared error (RMSE) criterion depends on the order of magnitude of the observed values and is defined by [30]

$$\text{RMSE} = \sqrt{(\text{MSE})} \tag{9}$$

The mean absolute error (MAE) criterion corresponds to standard L_1 and is similar to RMSE except that it is more robust due to its less sensitivity toward extreme values [30].

5 Predictions and Results

The study focuses mainly on performing regression for the prediction of the number of recovered people, when the data for the number of confirmed infected cases, deceased cases, and unknown cases and tested, is given. Among these, 'tested' is the only variable that can be controlled and varied, hence the target variable 'recovered' can be said to mainly depend on the number of people tested. The two variables have a positive correlation of 0.972 as can be seen in the showcased data analysis in previous sections. Thus, for the study different libraries available in the python were imported to implement different regressors. The statistical metrics computed

for the analysis of the performance of the models were all tabulated to give a clear comparative analysis of all the algorithms shown in Table 1.

It can be seen from the table that RR performed the best with a test score percentage of 99.61%, RMSE 0.323%, and R-square value of 99.628%. The relevant RR plots can be seen from Figs. 5 and 6. Figure 6 is developed using Yellowbrick library [31].

Table 1 Comparative analysis of the metrics computed for each algorithm

	Model	Test score	Mean squared error	Mean absolute error	R^2	RMSE
1	SVR	0.9457	0.01493	0.8651	95.3821	1.222
2	Ridge	**0.9961**	**0.0010**	**0.1268**	**99.628**	**0.323**
3	Decision tree regressor	0.9872	0.0035	0.159	98.7844	0.593
4	Random forest regressor	0.9900	0.0027	0.2615	98.9817	0.5227
5	Extra-trees regressor	0.9798	0.0055	0.4292	97.852	0.7442
6	Elastic net	0.9890	0.0030	0.2008	98.9108	0.5485
7	Linear regression	0.993	0.0010	0.1270	99.3571	0.324
8	ANN	NIL	0.0023	0.2681	99.1281	0.486

The data highlighted via the bold font signifies that the Ridge model generated the best results and the least amount of errors out of all the algorithms used

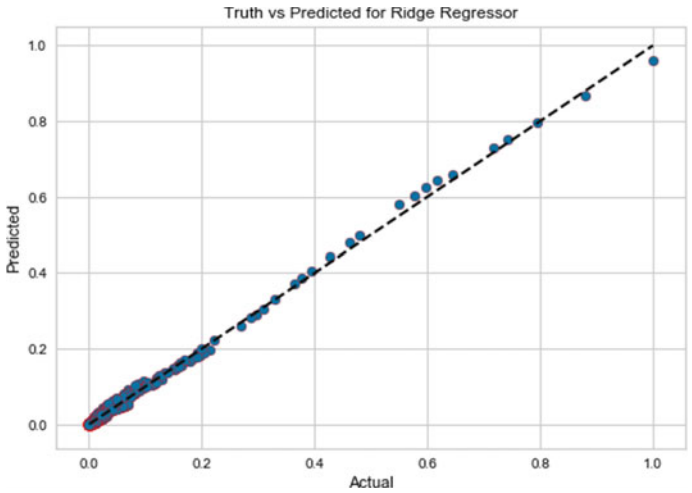


Fig. 5 Actual values are plotted on *x*-axis and predicted values on y-axis, along with a 45° line, using scikit-learn

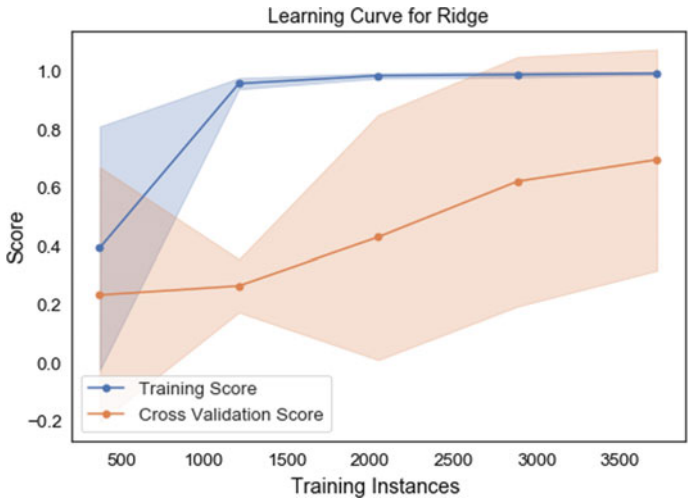


Fig. 6 Learning curve plotted for RR, where *x*-axis represents training instances and y-axis represents testing score. Here, the model experienced low CV score in the beginning and then successfully climbed more than 0.60

LR, being the next best model, performed with a test score percentage of 99.319%, MAE 0.127%, RMSE 0.324% and *R*-square value 99.35%. Among the mentioned regressors, SVR scored the least, in comparison with others with, test score 94.57%, MAE 0.865%, RMSE 1.222% and *R*-square of 95.38%. ANN being the next best had an explained variance of 99.128%, with model loss and validation loss plotted in Fig. 7.

6 Conclusion

With a sudden increase in the deployment of data mining and artificial intelligence in the field of medical and healthcare sciences, it is becoming convenient for scientists and scholars to optimize a solution for any health problem with the help of critical analysis. Similarly, since the COVID-19 pandemic has begun, artificial intelligence has paved many ways for epidemiological sciences to help study the virus characteristics and propose various ways in which it can be controlled. For the same purpose, many renowned organizations, both government and private, have granted open/protected access to the COVID-19 data for scientists and scholars to peruse. In this study, eight machine learning algorithms were deployed on the official India COVID-19 data, which contained 5585 rows and five columns of continuous values ranging from 0 to maximum 30,941,264. The chosen target was the number of recovered people, when the number of confirmed cases, tested, deceased and unknown cases was given. From the successful experimental study, although the average error

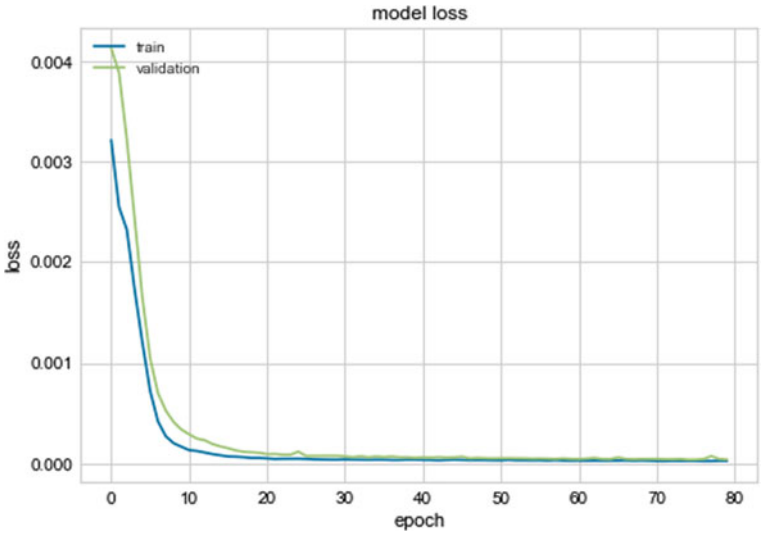


Fig. 7 Model loss per epoch in ANN

rate of all the regressors was very low due to scaled values, it was concluded that ridge regression performed the best among all the regressors with an accuracy of 99.6% and RMSE 0.323%. Various other regressors that showed similar promising results were LR, Lasso and neural networks. This study showcased a strong positive linear relation between recovery and testing, which can be easily justified in a way that with early detection of the virus it paves a way for the carrier to get an early treatment and care along with careful isolation from his/her surroundings, which can further help in combating critical case conditions that might prove to be fatal. While the study concluded with very promising results, it still is possible to pioneer more effective results with the addition of more dimensions such as type of testing used, isolation days, hospital facilities such as beds and medicines, number of tests done on a single patient, underlying health conditions and so on. Not only this, other statistical methods can also be applied to eliminate false numbers and decrease error rate, in the future studies.

References

1. Worldometer, https://www.worldometers.info/coronavirus/. Updated on 25 Sept, accessed on 25 Sept
2. V. Vara, https://www.pharmaceutical-technology.com/features/coronavirus-affected-countries-india-measures-impact-pharma-economy/. Updated on 7 Sept Accessed on 10 Sept
3. K. Subramanian, https://indianexpress.com/article/opinion/columns/gdp-contraction-indian-economy-6587087/. Updated on 8 Sept, accessed on 10 Sept
4. P. Arora, H. Kumar, B.K. Panigrahi, Prediction and analysis of COVID-19 positive cases using

deep learning models: a descriptive case study of India. Chaos Solitons Fractals 139. https://doi.org/10.1016/j.chaos.2020.110017
5. F. Rustam, A.A. Reshi, A. Mehmood, S. Ullah, B.W. On, W. Aslam, G.S. Choi, COVID-19 future forecasting using supervised machine learning models. IEEE Access **8**, 101489–101499 (2020). https://doi.org/10.1109/ACCESS.2020.2997311
6. D Liu, L. Clemente, C. Poirier, X. Ding, M. Chinazzi, J.T. Davis, A. Vespignani, M. Santillana, A machine learning methodology for real-time forecasting of the 2019–2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. ArXiv. http://www.ncbi.nlm.nih.gov/pubmed/32550248%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7280902
7. V. Vaishnav, J. Vajpai (2020) Assessment of impact of relaxation in lockdown and forecast of preparation for combating COVID-19 pandemic in India using group method of data handling. Chaos Solitons Fractals 140. https://doi.org/10.1016/j.chaos.2020.110191
8. M. Barstugan, U. Ozkaya, S. Ozturk, *Coronavirus (COVID-19) Classification Using CT Images by Machine Learning Methods*, vol 5, pp 1–10 (2020). http://arxiv.org/abs/2003.09424
9. R. Kumar, S. Nagpal, S. Kaushik, S. Mendiratta, COVID-19 diagnostic approaches: different roads to the same destination. VirusDisease **31**(2), 97–105 (2020). https://doi.org/10.1007/s13337-020-00599-7
10. L.L. Liang, C.H. Tseng, H.J. Ho, C.Y. Wu, Covid-19 mortality is negatively associated with test number and government effectiveness. Sci. Rep. **10**(1), 1–7 (2020). https://doi.org/10.1038/s41598-020-68862-x
11. M. Yadav, M. Perumal, M. Srinivas, Analysis on novel coronavirus (COVID-19) using machine learning methods. Chaos Solitons Fractals **139**, 110050 (2020). https://doi.org/10.1016/j.chaos.2020.110050
12. W. Vogt, Ridge regression. Dict. Stat. Methodol. (2015). https://doi.org/10.4135/9781412983907.n1697
13. E. Cule, M. De Iorio, Ridge regression in prediction problems: Automatic choice of the ridge parameter. Genet. Epidemiol. **37**, 704–714 (2013). https://doi.org/10.1002/gepi.21750
14. W. Liu, Q. Li, An efficient elastic net with regression coefficients method for variable selection of spectrum data. PLoS ONE **12**(2), e0171122 (2017). https://doi.org/10.1371/journal.pone.0171122
15. C.J.F.T. Braak, Regression by L_1 regularization of smart contrasts and sums (ROSCAS) beats PLS and elastic net in latent variable model. J. Chemom. **23**, 217–228 (2009). https://doi.org/10.1002/cem.1213
16. M.Y. Park, T. Hastie, L_1-regularization path algorithm for generalized linear models. J. R. Stat. Soc. Ser. B Stat. Methodol. **69**, 659–677 (2007). https://doi.org/10.1111/j.1467-9868.2007.00607.x
17. M. Awad, R. Khanna, Support vector regression, in *Efficient Learning Machines* (Apress, Berkeley, CA, 2015). https://doi.org/10.1007/978-1-4302-5990-9_4
18. C.H. Wu, J.M. Ho, D.T. Lee, Travel-time prediction with support vector regression. IEEE Trans. Intell. Transp. Syst. **5**, 276–281 (2004). https://doi.org/10.1109/TITS.2004.837813
19. S. Han, Q. Cao, M. Han, Parameter selection in SVM with RBF kernel function. World Autom. Congr. Proc. (2012)
20. G.K.F. Tso, K.K.W. Yau, Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. Energy **32**, 1761–1768 (2007). https://doi.org/10.1016/j.energy.2006.11.010
21. A. Mohi, U. Din, K. Syed, T. Rabani, Q. Rayees, Machine learning based approaches for detecting COVID-19 using clinical text data. Int. J. Inf. Technol. **12**, 731–739 (2020). https://doi.org/10.1007/s41870-020-00495-9
22. V. Svetnik, A. Liaw, C. Tong, J. Christopher Culberson, R.P. Sheridan, B.P. Feuston, Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. **43**, 1947–1958 (2003). https://doi.org/10.1021/ci034160g
23. U. Grömping, Variable importance assessment in regression: linear regression versus random forest. Am. Stat. **63**, 308–319 (2009). https://doi.org/10.1198/tast.2009.08199

24. P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees. Mach. Learn. **63**, 3–42 (2006). https://doi.org/10.1007/s10994-006-6226-1
25. F. Murtagh, Multilayer perceptrons for classification and regression. Neurocomputing **2**, 183–197 (1991). https://doi.org/10.1016/0925-2312(91)90023-5
26. C. Nwankpa, W. Ijomah, A. Gachagan, S. Marshall, *Activation Functions: Comparison of Trends in Practice and Research for Deep Learning* (2018), pp. 1–20. http://arxiv.org/abs/1811.03378
27. COVID-19 India Org Data Operations Group, covid19india org 2020 tracker, Accessed on 01 Aug 2020 from https://api.covid19india.org/. Published 2020
28. F. Pedregosa et al., Scikit-learn: machine learning in python. JMLR **12**, 2825–2830 (2011)
29. F. Chollet et al., in *Keras* (GitHub, 2015). Retrieved from https://github.com/fchollet/keras
30. B. Ait-Amir, P. Pougnet, A. El Hami, Meta-model development. Embed. Mechatron. Syst. **2**, 151–179 (2015). https://doi.org/10.1016/b978-1-78548-014-0.50006-2
31. B. Bengfort, R. Bilbro, Yellowbrick: visualizing the Scikit-learn model selection process. J. Open Source Softw. **4**, 1075 (2019). https://doi.org/10.21105/joss.01075

Regression to Forecast: An In-Play Outcome Prediction for One-Day Cricket Matches

R. Raja Subramanian, P. Vijaya Karthick, S. Dhanasekaran, R. Raja Sudharsan, S. Hariharasitaraman, S. Rajasekaran, and B. S. Murugan

Abstract The paper presents a technique for predicting the outcome of an ongoing cricket match. A robust prediction is obtained by taking into account the quality of players and form of the team in addition to the toss, host advantage, and day/night match facts. A novel approach is used to quantify the player quality and form of the team. At each stage, the explanatory power of the available parameters is analysed to include in the model, replacing few existing parameters, for effective prediction. The quantification of these parameters is subjected to dynamic logistic regression, as the parameters change as the match progresses. The model is evaluated against 187 ODI matches between 2015 and 2017, and a sound prediction accuracy is observed. The model is evaluated in five different scenarios of the match as follows: before match begins, after 30 overs in first innings, at the end of first innings, and after 30 and 40 overs in the second innings.

Keywords Dynamic logistic regression · Cricket forecasting · Match outcome prediction · Team strength calculation

R. Raja Subramanian (✉) · S. Dhanasekaran (✉) · S. Hariharasitaraman · B. S. Murugan
Kalasalingam Academy of Research and Education, Virudhunagar, India
e-mail: rajasubramanian.r@klu.ac.in

S. Dhanasekaran
e-mail: srividhans@gmail.com

P. Vijaya Karthick
Sir M. Visvesvaraya Institute of Technology, Bangalore, India
e-mail: vijaykarthik_is@sirmvit.edu

R. Raja Sudharsan
M Kumarasamy College of Engineering, Karur, India

S. Hariharasitaraman
VIT Bhopal University, Bhopal, India

S. Rajasekaran
PSN College of Engineering and Technology, Tirunelveli, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_5

1 Introduction

Cricket, being the second most popular sports in the world, has many fans and aspirants. Like soccer, cricket too has a huge betting market. Betting is prevalent in World Cup matches or matches between cricket rivals as follows: India versus Pakistan, Australia versus England in the Ashes, and among others. Betting requires the need for predicting the outcome of an ongoing match. Hence, we require an efficient algorithm to forecast the outcome of an ongoing match. The forecasting, besides being useful for betters, serves as a vital decision support source for management, coach, and captain to take crucial calls such as deciding whether to play aggressive or defensive, depending on the conditions in the match. Strategies of a captain have a major impact on the result of the match. Reference [1] analysed the 2011 and 2015 World Cup one-day international matches and found that the team batting second is more successful in their attempt to score more runs early in the innings by taking the risk of losing wickets. The decision to play aggressive or defensive requires knowledge about the players in the team and the strength of the opposition. This decision is dynamic, as the strength of each team varies throughout the match.

Several seminal research attempts in the field of cricket analytics concentrate on resetting targets in interrupted matches and ranking cricket teams and players. The pioneer in the research [2] proposed a technique to reset targets in an interrupted one-day cricket match. The technique is used commercially by International Cricket Council (ICC) by the name D/L method, till 2014. Duckworth and Lewis categorized the strength of each team as resources. The number of overs remaining and the wickets in hand together form the resource of the team. D/L method constructs a two-dimensional table of resources with each cell intersecting wickets lost and overs remaining. The resources are constant and same for all the teams, thus rendering the method unbiased. The targets are revised based on the resources. A modified version of the D/L method is proposed, which took into account the run scoring pattern for adjusting the scores [3]. As the match progresses, in addition to wickets and overs, runs required are also considered as a major parameter. DLS method, now popularly used in ICC, updating the D/L version is proposed by Stern [3]. DLS method mirrored D/L in all times, except in high-scoring matches. In such cases, DLS suggests a score more than that suggested by D/L. With the advent of the T20 matches, teams tend to score very high, unlike the matches in 1990s, during when D/L is proposed. Target adjustment and outcomes forecasting techniques use constant resource values for all the teams. When predicting the win probability, there is a need to determine the quality of the teams. The number of wickets in hand should be replaced by the quality of wickets in hand. Quality of wickets is calculated by determining the quality of players who have not lost their wickets. For instance, according to DLS, the percentage of resources remaining for the team with 10 overs left and 5 wickets being lost is 27.5. Being a constant value, it cannot effectively calibrate the strength of a team. The resource remaining should not be calculated based on the number of wickets lost, but by the quality of the wickets in hand. The quality of wickets in hand signifies the quality of batsmen yet to play. Many researches in forecasting cricket

match outcomes [3–6] have used the current state of the match with covariates such as wickets, runs, and overs to predict the outcome using learning techniques. To the best of our knowledge, techniques to use the quality of players to predict the match outcome are scarce.

2 Background and Related Works

Many researches and alternatives to D/L method evolved in the literature used dynamic programming model [7] to predict the outcome of a cricket match at different stages. Perston used the same model to revise targets in interrupted ODI matches. Perston [7] considered the run scoring rate and dismissal rate to carry out his experiments. When a nightwatchman enters the crease and goes back to pavilion with a duck, the model decreases the win probability of the batting side. It fails to consider the quality of the wicket lost. The model, though differs from D/L methods with respect to the construction of resource table, suffers from the same deficiency as the D/L method and the updated DLS method. Multinomial logistic regression [8, 9] is used by [4, 10, 11] to forecast the outcomes of a test cricket match. The methods consider ICC rating differences, effect of the pitch, and home field advantage as parameters to estimate. ICC ratings of a team describe the quality of the team but not the players composing the team. It is not the same team of players that plays every time. Hence, accounting the quality of the playing XI is required. For example, in Zimbabwe triangular series (Zimbabwe, India, and Sri Lanka) at Bulawayo, 2010 (see [12]), India lost to Zimbabwe in two consecutive matches. India with an ICC rating higher than Zimbabwe, and having won their previous series against South Africa, lost their matches against Zimbabwe, the team with a lower rank than either South Africa or India. One of the reasons is that India has not sent their usual players to Zimbabwe with MS Dhoni. Instead, the team composed of many young faces with Suresh Raina leading the side. Hence, team quality is determined by the playing XI.

We propose a metric as a replacement to the ICC rating to quantify the quality of players. Techniques [11] are proposed for rating cricket players considering the timely contribution of a player such as scoring more runs in second innings by a batsman and early wickets taken by the bowler. The technique did not consider the quality of runs scored based on the bowler against whom the batsman has scored and the quality of the batsman whose wicket is taken by the bowler. The quality of the opponent bowler and batsman determines the quality of the batsman and bowler, respectively. Mukherjee [13] proposed a technique to determine the quality of runs scored and quality of wickets taken. The method does not consider the nature of the pitch. A batsman hitting R runs in a batting pitch against a bowler X and another batsman who hits the same R runs against the same bowler X in a bowler friendly pitch will be treated as having the same quality. Hence, a method to determine the quality of a batsman based on pitch, quality of runs/wicket, strike rate, and form is required.

3 Proposed Metric to Quantify Player Quality

Cricket is a team sport where eleven players in each team compete. Like tennis, it is not possible to establish a direct one-on-one fight between players. The players play as a whole with few featuring as batsmen, who try to score runs for their team, and some as bowlers, who restrict the opponent team batsman from scoring runs and attempt to capture wickets. Some act as all-rounders, doing both batting and bowling. Ranking these players in their departments requires an unbiased way to estimate their performances. A batsman is evaluated not by the runs scored, but by the quality of those runs. The quality of a batsman (Q_P) is measured using Eq. 1:

$$Q_P = \tau_1 + \tau_2 + \tau_3 \tag{1}$$

$$\tau_1 = \alpha \frac{1}{n(X)} \sum_X \left(P_{\text{avgX}} / X_{\text{Cavg}} \right) \tag{2}$$

$$\tau_2 = \beta \frac{1}{n(G)} \sum_G \left(P_{rG} / X_{\text{ravg}} \right) \tag{3}$$

$$\tau_3 = \gamma \frac{1}{n(G)} \sum_G \left(P_{sG} / X_{\text{ravg}} \right) \tag{4}$$

where

Q_P is the quantitative measure of player (batsman) quality,
P_{avgX} is the average runs scored by P against bowler X,
X_{Cavg} is the career bowling average of the bowler,
P_{rG} is the average runs scored by P in ground G,
G_{ravg} is the average run rate in ground G,
P_{sG} is the strike rate of the player in ground G, and
$n(K)$ is the number of occurrences of K.

Equation 1 has three parts summed to obtain the player quality. First part (τ_1) checks the quality of runs. It is calculated by taking the ratio of batsman's average runs against the bowler and the bowler's career bowling average as shown in Eq. 2. Larger value of τ_1 refers to the fact that the batsman has scored runs against a quality bowler (bowler with lower career bowling average). Second part (τ_2) checks the versatility of the batsman across different pitches. In addition to scoring runs against a quality bowler, scoring runs in a bowler friendly pitch gains a credit. Hence, a batsman's average runs in a ground are normalized by the average runs per over in that ground to check the versatility of the player as depicted in Eq. 3. In limited over formats, taking a wicket can be compared to a couple of maiden overs. The ability of the batsman to adapt to the condition and strike the ball as soon as possible is represented in the third term (τ_3) (proficiency) of Eq. 1. The ratio of the strike rate of the batsman in a ground to the average runs per over in the ground is taken in Eq. 4. Each term has a summation term with limits depending on the number of bowlers,

the batsman faced and number of grounds, and the batsman played. The terms are then subjected to weighted averaging to obtain the batsman's quality. The values of the weights α, β, and γ are obtained by studying the significance of each of the parameters as follows: run quality, versatility, and proficiency using the logit model. Since all players strive to win the match, the outcome of the match is made as the dependent variable. The outcome is binary with two values (0—lost, 1—win). The draw and no-result outcomes are eliminated. The computed values of α, β, and γ are 0.573, 0.312, and 0.091, respectively. Table 1 depicts the top 10 unretired batsmen in ODI cricket. Players are ranked according to Q_P and then compared with ICC points.

$$Q_P = \tau_1' + \tau_2' \tag{5}$$

$$\tau_1' = \frac{1}{n(X)} \sum_X M\left(X_{\text{Cavg}} / P_{\text{avgX}}\right) \tag{6}$$

$$\tau_2' = \frac{1}{n(G)} \sum_G \left(G_{\text{ravg}} / P_{eG}\right) \tag{7}$$

where

Q_P is the quantitative measure of player (bowler) quality,
M is the number of times the batsman was dismissed by the bowler,
X_{Cavg} is the career batting average of the batsman,
P_{avgX} is the bowling average of P against X, and
P_{eG} is the economy of the bowler in ground G.

The bowler is judged not just by the wickets, but also by the quality of wicket or the quality of the batsman dismissed. This is represented as the ratio of batsman's career average to the bowler's bowling average against the batsman in Eq. 6. In addition to wickets, the ability of the bowler to save as many runs shows his strength. This is

Table 1 Ranking of the top 10 unretired batsmen in ODI cricket

Rank	Batsmen	Country	Q_P	Batting average	ICC points
1	Virat Kohli	India	7.128	58.10	876
2	AB de Villiers	South Africa	6.953	53.50	872
3	Rohit Sharma	India	6.423	44.55	816
4	David Warner	Australia	6.406	48.63	823
5	Joe Root	England	6.365	51.16	808
6	Quinton de Kock	South Africa	6.224	45.41	808
7	Kane Williamson	New Zealand	6.216	46.87	777
8	Babar Azam	Pakistan	6.050	55.11	813
9	Faf du Plessis	South Africa	5.907	43.79	773
10	Martin Guptill	New Zealand	5.833	43.43	764

Table 2 Ranking of the top 10 unretired bowlers in ODI cricket

Rank	Batsmen	Country	Q_P	Bowling average	ICC points
1	Kagiso Rabada	South Africa	8.357	27.57	679
2	Trent Boult	New Zealand	8.104	24.64	699
3	Mitchell Starc	Australia	7.068	20.96	658
4	Jusprit Bumrah	India	6.402	22.5	787
5	Chris Woakes	England	6.252	30.78	673
6	Josh Hazlewood	Australia	5.801	24.28	714
7	Imran Tahir	South Africa	5.420	24.81	777
8	Rashid Khan	Afghanistan	5.296	14.4	763
9	Hasan Ali	Pakistan	4.865	21.4	711
10	Yuzvendra Chahal	India	4.711	21.84	773

put forth by taking the economy of the bowler into account. Economy refers to the average number of runs, and the bowler concedes in an over. The ratio of the average runs per over in the ground to the bowler's economy rate in that ground is taken in Eq. 7. Each of the terms in Eqs. 6 and 7 has a summation term with limits based on the number of batsmen the bowler faced and number of grounds the bowler played, respectively. The terms are averaged and summed out to obtain the bowler's quality. Table 2 depicts the top 10 unretired bowlers in ODI cricket.

4 Form and Batting Resource of the Team

The proposed method to determine the form of a team in [3] is extended in our research. The method takes into account the results of the last five matches played by the team. A win in the most recent match gains more weight in determining the form. The weighted mean of the outcomes of the team in the last five ODIs determines the form of the team. The deficiency in this method is witnessed by Bangladesh's series defeat against England in its home, despite having a strong form of 0.879 as provided by [3], as a result of winning 4 out of 5 of the previous matches. This deficiency is due to the lack of exposure in the winning streaks of Bangladesh. Bangladesh's last five encounters are A, A, Z, Z, Z (A Afghanistan and Z Zimbabwe). These wins could not effectively provide Bangladesh's form. Hence, the weights are adjusted in the calculation of form of a team as in (8).

$$\text{form} = \frac{\sum_{t=1}^{5} w(t,\theta) y_t}{\sum_{t=1}^{5} w(t,\theta)} \tag{8}$$

where calculation of form of [3] in Eq. 8 takes the outcomes of the last t matches with weights, to provide more impact on the outcome of more recent matches. $y_t = 1$

(for win) or 0 (for losse) in match t. $w(t, \theta) = (1 - \theta)^{t-1}$ and $0 < \theta < 1$, and its value is estimated to be 0.2041.

$$y_t = \text{outcome} * (\alpha + ((R_k - R_x)/20)) \tag{9}$$

where outcome is the result of the match (1/0). R_k is the rank of team k whose form is to be determined. R_x is the rank of team x against which the team k has played. η is a constant and has a value of 0.50. The value of y_t has a minimum value of η if a team has won the match. The quality of the win is calculated by the later part of Eq. 9. It gives the difference between the ranking of the team under consideration and the opponent team, divided by a constant value 20. This value increases by 0.05 for a unit difference in ranking between the two teams. If a team with rank 9 defeats the team with rank 1, then the value y_t is calculated as shown in Eq. 10

$$y_t = 1 * \left(0.50 + \left(\frac{9-1}{20}\right)\right) = 0.90 \tag{10}$$

Hence, if teams with less ranking difference compete, then the value of y_t will be low and will increase by 0.05 for each ranking increase. This makes the value of y_t to be between 0.50 and 1 for a win outcome unlike Asif's method, where y_t is 1 for a win outcome. We omit wins against teams that are not within the top 12 rankings. The modified form computation gives a form value of 0.483 for Bangladesh taking their encounters A, A, SA, SA, SA (SA—South Africa) into consideration. The comparison between Asif's form calculation method and the proposed form calculation method is presented in Fig. 1. The form calculations are done for eight one-day international teams who played for the 2015 World Cup. It can be noted in Fig. 2 that the form of Australia and New Zealand is equal by Asif's form calculation. On the other hand, the form of Australia and New Zealand is 0.683 and 0.661, respectively, by the proposed model. Eventually, Australia with the highest form among all the other teams won the World Cup 2015.

In addition to the DLS resources available percentages, we also use the current batting resource the team has. This is obtained by summing up the batting quality of each not-out batsman. The relation between DLS resources and the proposed remaining batsmen quality percentages is depicted in Fig. 2. The data is collected from West Indies versus India ODI in 2009, where India lost 8 wickets in 25 overs, but still survived till 48.2 overs with Dhoni scoring 95 runs. Resource percentages of DLS take into account both the overs and the wickets remaining, while we consider only the wickets remaining and their quality. The resource remaining is still above 20 even after losing 8 wickets, for India, between 25 and 47 overs as shown in Fig. 2. This is because Dhoni was still batting in the pitch. DLS resource percentage does not take the quality of the remaining batsmen's wicket into consideration unlike our model. Hence even though wickets fell, it is possible to predict the outcome based on the batsmen yet to bat. Existing state-of-the-art models to predict match outcome fail if a batsman is retired hurt and comes later to bat again or if the batting

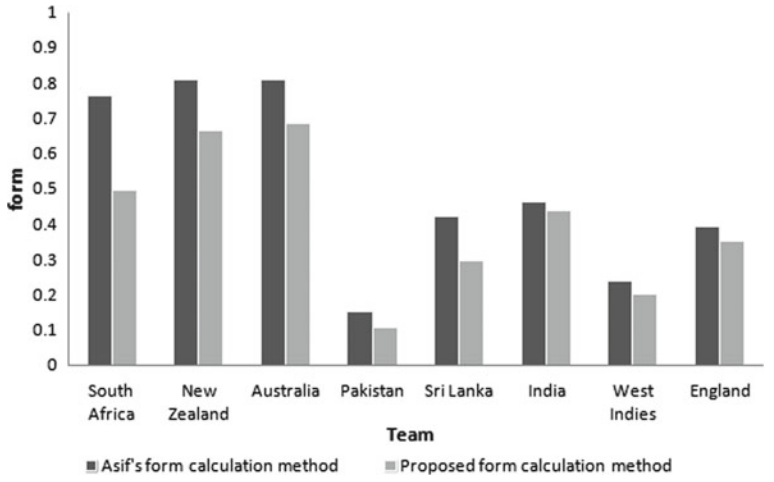


Fig. 1 Asif and McHale [3] form calculation procedure versus proposed form calculator

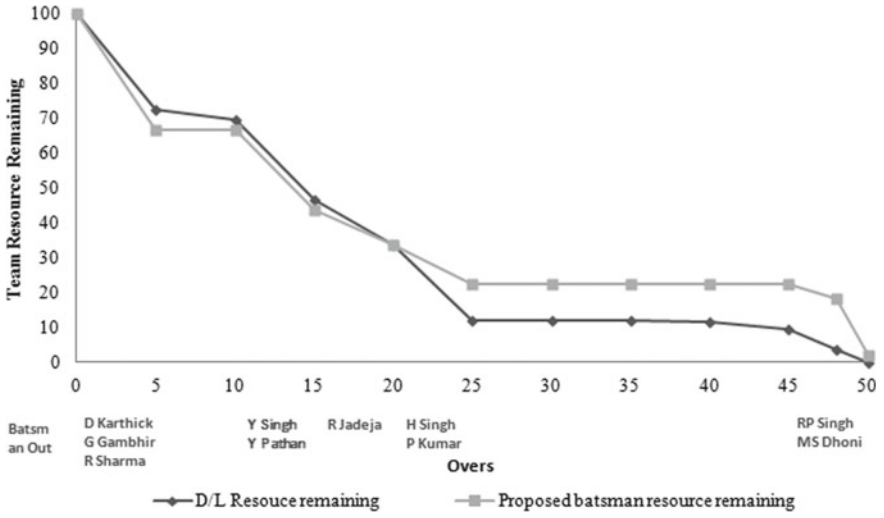


Fig. 2 Comparison between the DLS resource remaining and the proposed batsmen resource remaining percentages for team India batting against West Indies, 2009

order changes. Our model is flexible and robust in these cases by taking the batsmen resource remaining criterion into account.

5 Model to Forecast Win Probability

We use a dynamic logistic regression [14] model to predict the win probability of the chasing team. At different stages of the match, the parameter values vary, which makes the model dynamic. The outcome of the match is the dependent variable and dichotomous, taking value 0 for loss and 1 for win. Since a draw in an ODI match is a rare scenario, it is not included in the model. Predictor variables are represented as a matrix of independent variables X. The independent variables are form, team quality, toss, D/N, run rate, required run rate, batting resource remaining, and score quality. The logit function is given in Eq. 11

$$\text{logit}(p) = \beta_0 + \sum_i \beta_i X_i \tag{11}$$

Forecasting the win probability can be done in various stages of the match as follows: before start of the match, end of the first innings, and at various intervals during the first and second innings. Intervals constitute the completion of n overs in the innings. Before start of the match, form, team quality (t_Q), toss, host, and day/night (D/N) are the covariates available. Form has the ability to portray the outcome of the match with a higher confidence, as it provides the quality of wins the team had in its last five encounters. t_Q is another important parameter that estimates the strength of the team in terms of batting and bowling. A team's quality alone cannot bring victory to its team. It requires many other attributes such as coordination, wholistic team play, and leadership, for example, fielding of the team, left- and right-handed batting partnerships, left-arm bowler bowling for left-handed batsmen, and among others. These factors can be reflected by the form of the team, as the team has effectively employed to win. Host team has an advantage of having played more matches under the home conditions and pitches. Toss and D/N factors also influence the match. Typically, teams with a strong batting line-up consider bowling first, so that the target can be chased by their batsmen. Toss also has an influence in day and night matches. Dew makes it difficult for the bowlers to have control on the line, length, spin, and swing. Hence, captains prefer to bowl first.

Once the match has started, run rate (run_{curr}) is available. At the end of innings, score quality (s_Q) is available. (run_{curr}) is always a dominant factor in limited over formats. The explanatory power of run rate and other available parameters are depicted in Fig. 3a, where form and run rate in first innings contribute more towards the outcome. s_Q is the percentage increase in the score with respect to the average score in the pitch. The score quality determines the exploitation of the pitch by both the teams. It is possible to determine whether the bowling or the batting team dominates the first innings. In case of D/N match, if the team batting first has a smaller s_Q, then the match goes in favour of the opponent. This is due to the dew factor's influence on the bowling.

The effect of score quality in the prediction model is depicted in Fig. 3b. In addition to form, team quality and host contributing towards the outcome and score

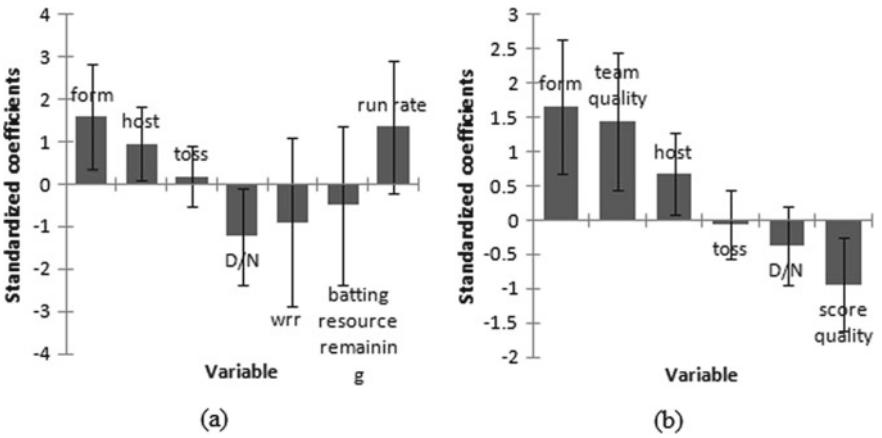


Fig. 3 Standardized coefficients of variables **a** after 30 overs in the first innings and **b** at the end of the first innings

quality have higher odds of predicting the outcome. Lower the quality of score of the opponent, higher the probability to win the match by the chasing team. Once the second innings begins, required run rate $\left(\text{run}_{\text{req}}\right)$, current team quality $\left(t_Q^{\text{curr}}\right)$, and wicket resource remaining (wrr) are the variables available. run_{req} is the variable that practically determines the outcome of the match. Decisions on changing style/order of batting/bowling are made, keeping in mind the run_{req}. t_Q^{curr} is the current team quality, obtained by subtracting quality of the players whose wicket is lost. wrr is the wicket resource remaining at a particular stage of the match (overs and wickets in hand) calculated through DLS method [15]. The relation between t_Q^{curr} and wrr is described in Sect. Sec5.

The standardized coefficients of the covariates are depicted in Fig. 4a and b. Figure 4a depicts the explanatory power of the variables after the completion of 30 overs in second innings where current batting quality of the team matters more in predicting the outcome. Figure 4b depicts the explanatory power of the variables at the end of 40 overs in the second innings. Form and current batting quality are dominant covariates at this stage, in determining the outcome. Intuitively, lower the required run rate, higher is the probability to win the match for the chasing team.

6 Model Predictions and Results

The model is constructed with 187 ODI matches played between 2015 and 2017. We predict the win probabilities at five stages of the match as follows: before start of the match, at the end of 30 overs in the first innings, end of first innings, and at the end of 30 overs and 40 overs in the second innings. The model predicts the probability of the chasing team to win the match. We created a method to quantify the form and

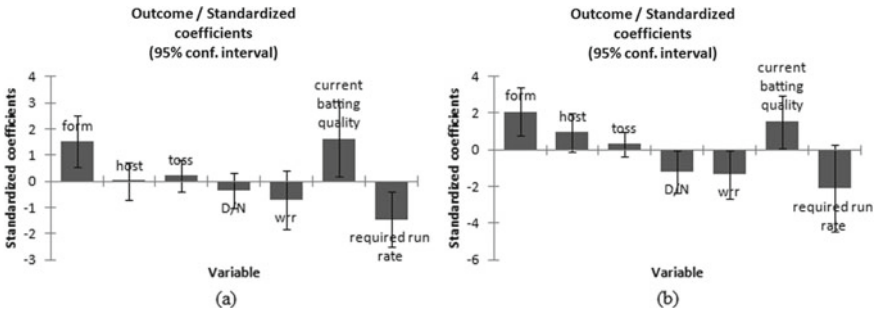


Fig. 4 Standardized coefficients of variables **a** after 30 overs in the second innings and **b** after 40 overs in the second innings

quality of the team. The quality of team is obtained by summing up the quality of the players in their domains (batting or bowling). In addition, covariates such as toss, host, D/N, run rate, and required run rate, whose values are directly obtained from the match, are used. DLS resource remaining, (wrr), is also used in predicting the outcome of the match. The robust method used to calculate the form and team quality makes them effectively predict the outcome of the match with high confidence.

Before start of the match, the accuracy of predicting the outcome of the match is 88.24% as stated in Table 3a. The form and team quality contribute more in explaining the outcome of the match. Once the match starts, with run rate of the team getting added to the input variables, we estimate the outcome of the match with an accuracy of 88.77% as given in Table 3b. Most of the matches tend to have a 120 plus score at the end of 30 overs and losing utmost 4 wickets. This imposes a restriction in predicting the win probability of the chasers. At the end of the first innings, we have the score quality of the team which batted first. Score quality plays a vital role in predicting the probability of the chasers to win. The model has correctly predicted 90.37% of the matches at the end of the first innings as depicted in Table 3c. Predictions become more accurate when parameters from second innings are used in the model. The prediction accuracy of the model is 92.51% and 95.19% at the end of 30 overs and 40 overs, respectively, in the second innings as given in Table 3d and e. The prediction becomes even more accurate as the match progresses.

7 Conclusion

In this paper, we present a dynamic logistic regression model to predict the outcome of ODI matches. The covariates used for the match are optimized at each stage of the match by the dynamic nature of the model. Matches are a complex interaction of individuals, teams, and playing conditions. It is not the same set of players that are going to represent for the team. Additionally, rankings need to be based upon prior competition levels, and a winning against poor teams does not equate to winning

Table 3 Predicted probabilities of the proposed model (a) before start of the match, (b) at the end of 30 overs in the first innings, (c) at the end of first innings, (d) at the end of 30 overs in the second innings, and (e) at the end of 40 overs in the second innings

From/to	Predicted outcomes		Total	%Correct	From/to	Predicted outcomes		Total	%Correct
	0	1				0	1		
(a)					*(c)*				
0	67	10	77	87.013	0	70	7	77	90.909
1	12	98	110	89.091	1	11	99	110	90.000
Total	79	108	187	88.235	Total	81	106	187	90.374
(b)					*(d)*				
0	71	6	77	92.208	0	72	5	77	93.506
1	15	95	110	86.364	1	9	101	110	91.818
Total	86	101	187	88.770	Total	81	106	187	92.513
(e)									
0	74	3	77	96.104					
1	6	104	110	94.545					
Total	80	107	187	95.187					

against closely ranked and competitive teams. Furthermore, environmental conditions are relevant as many teams can be accustomed to specific pitch types (pace, bounce, spin, carry, etc.) or weather (heat, cold, humidity, etc.). Batting teams conditioned on slow spinning wickets often struggle on fast bouncy wickets; alternatively, bowlers can struggle to replicate their best bowling with the different types of balls (Kookaburra, Duke, etc.). Our player quality measure is more robust as it is calculated based on the opposition's strength and the ability to play in different pitch types. The cumulative player quality categorized as the team quality is proposed on the assumption that the team includes sufficient percentage of batsmen, bowlers, and all-rounders. The two dynamic parameters are as follows: team quality and form which make the model robust in predicting match outcomes with a good accuracy. The proposed model exhibits a better prediction accuracy of 90.37% at the end of first innings which is greater than state-of-the-art form calculation models [3, 5]. We also plan to extend our work to build an unbiased model to predict the influence of weather in changing the state of the match.

References

1. P. O'Donoghue, Wicket loss and risk taking during the 2011 and 2015 Cricket World Cups. Int. J. Perform. Anal. Sports **16**, 80–95 (2017)
2. F.C. Duckworth, A.J. Lewis, A fair method for resetting targets in interrupted one-day cricket matches. J. Oper. Res. Soc. **49**, 220–227 (2004)

3. M. Asif, I. McHale, In-play forecasting of win probability in One-Day International cricket: a dynamic logistic regression model. J. Forecast. **32**, 34–43 (2016)
4. P.E. Allsopp, S.R. Clarke, Rating teams and analysing outcomes in one-day and test cricket. J. Roy. Stat. Soc. **167**, 657–667 (2004)
5. I. McHale, M. Asif, A modified Duckworth Lewis method for adjusting targets in interrupted limited overs cricket. Eur. J. Oper. Res. **225**, 353–362 (2013)
6. M. Bailey, S.R. Clarke, Predicting the match outcome in one day international cricket, while the game is in progress. J. Sports Sci. Med. **5**, 480–487 (2006)
7. I. Preston, J. Thompson, Rain rules for limited overs cricket and probabilities of victory. J. Roy. Stat. Soc. Ser. D **51**, 189–202 (2002)
8. S. Amara, R.R. Subramanian, Collaborating personalized recommender system and content-based recommender system using TextCorpus, in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 105–109. https://doi.org/10.1109/ICACCS48705.2020.9074360
9. R.R. Subramanian, N. Akshith, G.N. Murthy, M. Vikas, S. Amara, K. Balaji, A survey on sentiment analysis, in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021, pp. 70–75. https://doi.org/10.1109/Confluence51648.2021.9377136
10. P. Scarf, X. Shi, Modelling match outcomes and decision support for setting a final innings target in test cricket. J. Manag. Math. **16**, 161–178 (2005)
11. S. Akhtar, P. Scarf, Z. Rasool, Rating players in test match cricket. J. Oper. Res. Soc. **66**, 684–695 (2015)
12. www.espncricinfo/stats/statsguru. Accessed 12 Dec 2020
13. S. Mukherjee, Quantifying individual performance in Cricket A network analysis of batsmen and bowlers. J. Phys. A **393**, 624–637 (2014)
14. D. Tanouz, R.R. Subramanian, D. Eswar, G.V. P. Reddy, A.R. Kumar, C.V.N.M. Praneeth, Credit card fraud detection using machine learning, in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2021, pp. 967–972. https://doi.org/10.1109/ICICCS51141.2021.9432308
15. S. Stern, The Duckworth-Lewis-Stern method: extending the Duckworth-Lewis methodology to deal with modern scoring rates. J. Oper. Res. Soc. **67**, 1469–1480 (2016)

Analysis of OFDMA Over AWGN and Rician Channels

Manoj Kumar, Manish K. Patidar, and Narendra Singh

Abstract Wireless communications is a rapidly developing branch of the communication sector. The mobile communication is a branch that is used on broader canvas. But it has many challenges that must be studies, analyzed and compensated. A signal that is transmitted over a wireless channel has to face interference, fading, propagation path loss shadowing, etc. In wireless communication, it always requires a potential with the high-quality service. In this scenario, orthogonal frequency division multiplexing (OFDM) is best-suited communication technique. Orthogonal frequency division multiple access OFDMA is an effective technique for high bandwidth data communication; it converts the wideband signal into narrow band signals for communication. The transmission of each narrow band signals is done with orthogonal carrier. This paper presents simulation and implementation of multiuser orthogonal frequency division multiplexing (OFDM) using multiple modulation techniques. As well as, this paper presents the performance analysis of the implemented OFDMA over AWGN and Rician channels.

Keywords Orthogonal frequency division multiple access · AWGN channel · Rician channel · Signal-to-noise ratio · Symbol error rate · Bit error rate

1 Introduction

In a basic communication system, the information to be transmitted is modulated over an individual carrier frequency. In such a technique, symbol total occupies the available bandwidth. Such an implemented system can cause inter-symbol interference (ISI) at frequency selective channel. The grassroot idea of OFDMA is to separate the available spectrum into multiple orthogonal sub-channels. Hence in this technique each narrowband, Sub channel will experience mostly flat fading. At OFDMA, there may be a condition of overlapping sub-channels at frequency domain.

M. Kumar (✉) · M. K. Patidar · N. Singh
Department of Electronics and Communication Engineering, Jaypee University of Engineering and Technology, Raghogarh-Vijaypur, India
e-mail: manoj1985.111@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_6

In recent years, OFDMA technique has fetched an increased interest of researchers and developers. The European digital broadcast radio system is using this technique, also at asymmetric digital subscriber lines (ADSL), that is wired environment. OFDMA technique is used at digital subscriber lines (XDSL) to have high bit rate through twisted pair of wires.

Orthogonal frequency division multiplexing (OFDM), a specialized version of multicarrier modulation (MCM) with heavily spaced subcarriers as well as number of overlapping spectra, was patented at USA in 1970 s. At older frequency division multiple access system, steep band-pass filters were used. These filters are totally removed from the orthogonal frequency division multiple access system [1, 2]. Instead, OFDMA time domain waveforms are selected such that mutual orthogonality is confirmed considering that subcarrier spectra may overlap. It is understood that such a waveform can be generated using a fast Fourier transform (FFT) at the receiver and transmitter. For many years, the use of OFDMA technique was relatively limited. Though it was very smart and effective, it was not used that much due to its complexity, problems at implementation of real-time fast Fourier transform technique.

In earlier days, there were several issues regarding stability of transmitters and receivers due to oscillators in side it. Linearity required for RF power amplifiers was also an issue. After a year, many of the problems related to OFDMA were solved.

This paper will focus on orthogonal frequency division multiplexing access (OFDMA) research and simulation. Due to resistivity toward inter-symbol interference, OFDMA is most suitable for high-speed communication. At communication systems with increased information transfer speed, the required time for each transmission becomes very short. But at the same time, the delay time developed by multipath remains constant. Due to such conditions, ISI becomes a drawback of high data rate communication. OFDMA avoids this particular problem by transmitting several low-speed transmissions at a time [3, 4].

2 Proposed Technique

2.1 OFDM with AWGN Channel

In this proposed technique, we use two channels; first is AWGN, and second is Rician channel. Signal passes through this channel, and we observed the different performances. Here also we develop a simulation of OFDMA with AWGN, Rician channels which is shown in Figs. 1 and 2.

AWGN channel is basically used for communication as a noise model. It provides conditions to have effects of random processes occurring in environment.

The word additive refers that it might be added to any noise that is intrinsic to information system. The second important word in AWGN is white; white color refers to uniform emissions everywhere. Here, white refers to uniform emissions

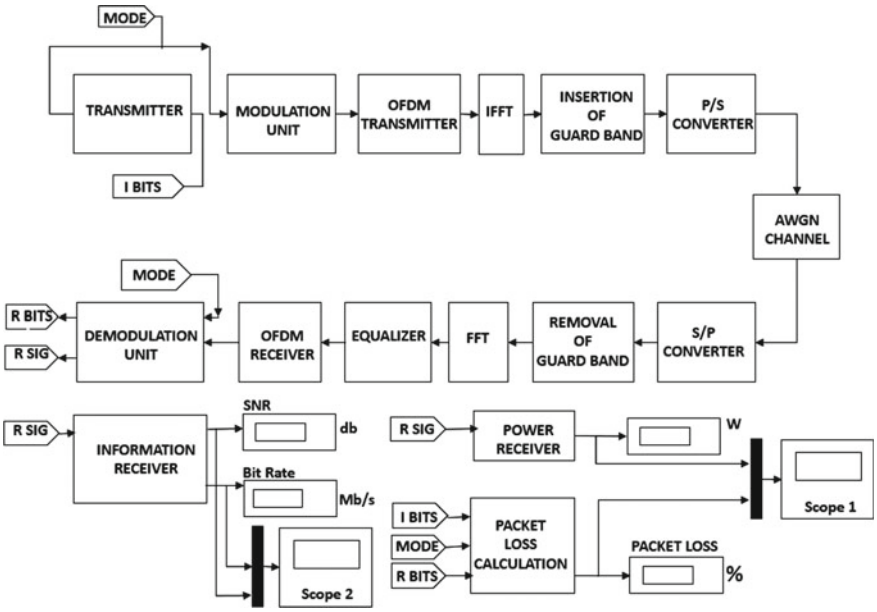


Fig. 1 Simulation diagram of OFDMA with AWGN channel

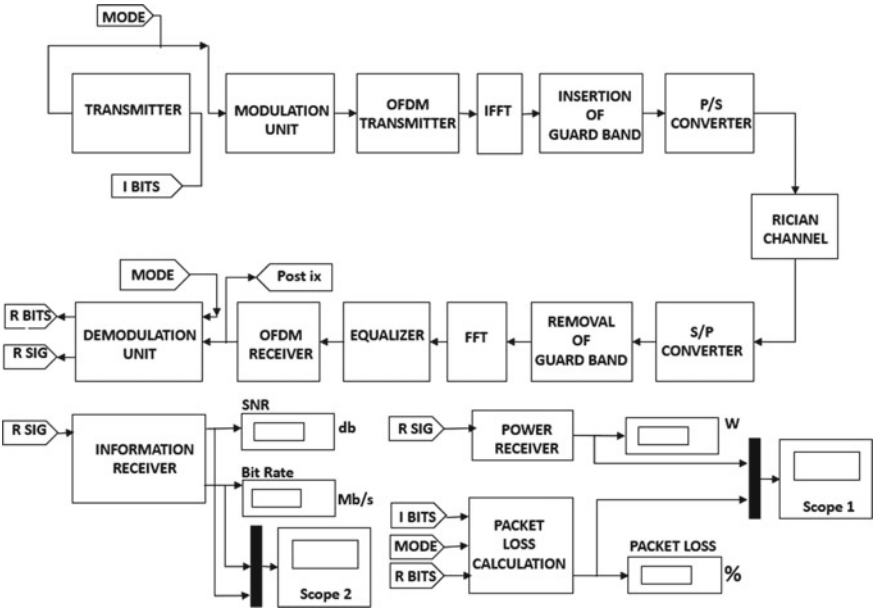


Fig. 2 Simulation diagram of OFDMA with Rician channel

throughout the frequency band for information system. The third important word Gaussian refers to normal distribution at time domain. Generally, AWGN channel is used for satellite communication. AWGN is basically used as a channel model in which the only improvement of communication is a linear addition of wideband or white noise with a constant spectral density and a Gaussian distribution of amplitude [5, 6]. This model is not used for fading frequency selectivity, nonlinear. It also provides simple mathematical model which is used for gaining.

This model is not that much good for terrestrial links due to problems such as multipath, interference, etc. The simulation diagram of OFDMA with AWGN channel is in Fig. 1.

2.2 OFDMA with Rician Channel

It is a stochastic model used for radio communication showing condition of partial cancelation of a radio signal in propagation. Radio propagation with multipath propagation allows signals to arrive at receiver end with several paths. In this situation, it may happen that one of the paths among multiple paths may remain shorter or lengthier one. In such a condition, there is a strong reflection of a particular signal as compared to other signals [7–9].

This situation leads to Rician fading effect. Rician channel represents Rician fading effect. Here, amplitude gain will be characterized by Rician distribution.

Signal received over a Rician multipath channel is mathematically represented as:

$$v(t) = C \cos w_c t + SN_{n=1} m \cos(w_c t + f_n) \tag{1}$$

where

C = amplitude of the line-of-sight component.
rn = the amplitude of the nth reflected wave.
f_n = the phase of the nth reflected wave.
$n = 1 \ldots N$ identify the reflected, scattered waves.

3 System Development

Thereafter, the proposed system is implemented for multiple users. To make it more effective and efficient, the proposed system is implemented with eight modulation and demodulation techniques.

The ultimate goal of this system is to provide multiple channels with well security and better transmission rate without any interference.

The proposed system is used for analysis of orthogonal frequency division multiplexing access over AWGN channel and Rician channel. For analysis, different

parameters such as symbol error rate, bit error rate, signal-to-noise ratio, packet loss, and received power are used.

3.1 Result Analysis

In this paper, first we take simulation of OFDMA with two different channels, and again by using MATLB, we find different parameters which are show in Figs. 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12. In design system, data was first encoded and then transmitted through the channel. The transmitted signal is distorted by noise which is additive white Gaussian noise. Here also we see Rician channel effect. The simulation is followed by m file. In this approach, the simulation is successfully done using different modulation techniques and the system is evaluated against multiple quality parameters such as symbol error rate, bit error rate, signal-to-noise ratio, packet loss and received power.

Figures 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 show the result of purposed system.

Figure 3 shows the real-time bitrate plot of AWGN and Rician channels. The above graph shows real-time bitrate of Rician channel is higher than AWGN.

Figure 4 shows the real-time SNR plot of AWGN and Rician channels. The above graph shows that real-time SNR of Rician channel is higher than AWGN.

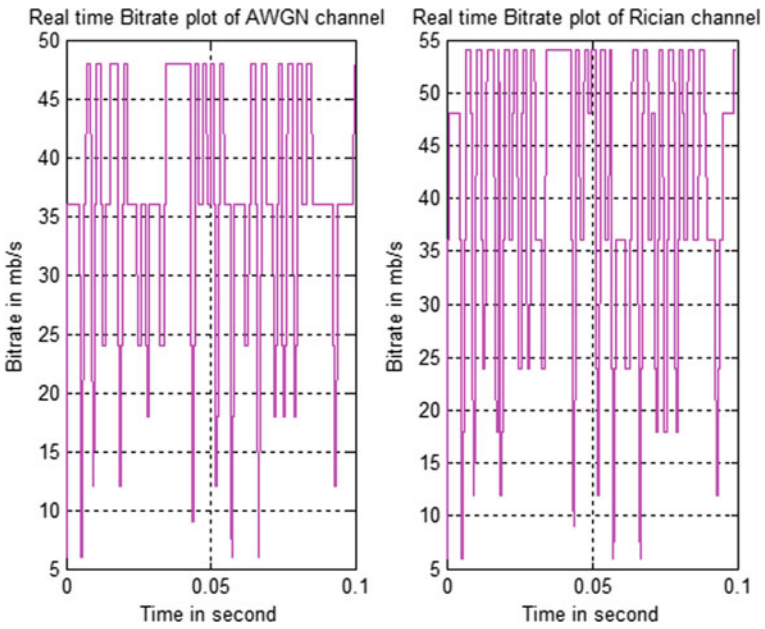


Fig. 3 Real-time bitrate plot of AWGN and Rician channels

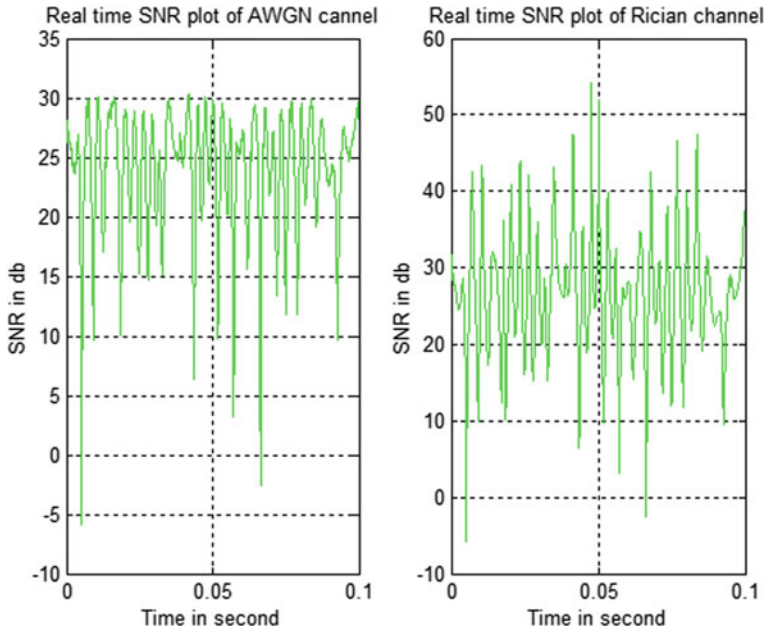


Fig. 4 Real-time SNR plot of AWGN and Rician channel

Figure 5 shows bit rate in mb/s of both channel; the above graph shows that the bit rate of Rician channel is higher than AWGN.

Figure 6 shows packet loss versus time for both channels. The above graph shows the packet loss in Rician channel is less than AWGN.

Figure 7 shows received power per second for both channels. The above graph shows the received power is maximum in Rician than AWGN channel.

Figure 8 shows signal-to-noise ratio versus bit error rate over AWGN and Rician channels. The graph indicates that Rician channel gives better performance than AWGN.

Figure 9 shows signal-to-noise ratio versus symbol error rate over AWGN and Rician channels. The graph indicates that Rician channel gives better performance than AWGN.

Figure 10 shows signal-to-noise ratio versus time in sec over AWGN and Rician channels. The graph indicates that the SNR of Rician channel is better than AWGN.

Figure 11 shows time in sec versus bit error rate over AWGN and Rician channels. The graph indicates that the BER of Rician channel is better than AWGN.

Figure 12 shows time in sec versus symbol error rate over AWGN and Rician channels. The graph indicates that the SER of Rician channel is better than AWGN.

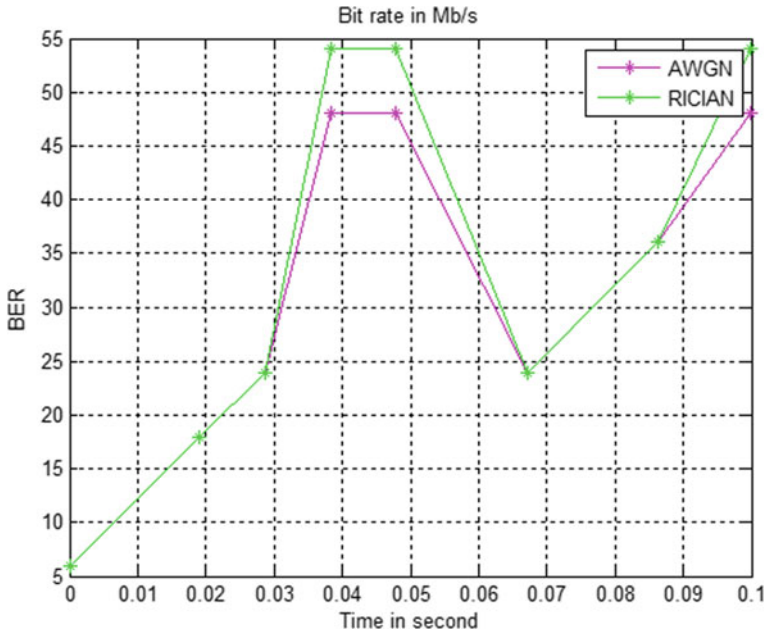


Fig. 5 Bit rate in Mb/s over AWGN and Rician channel

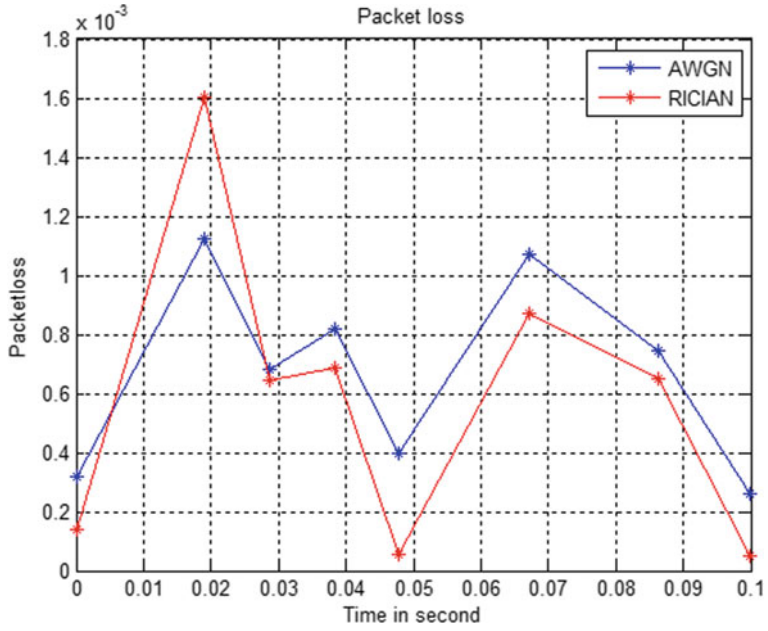


Fig. 6 Packet loss versus time (sec) over AWGN and Rician channels

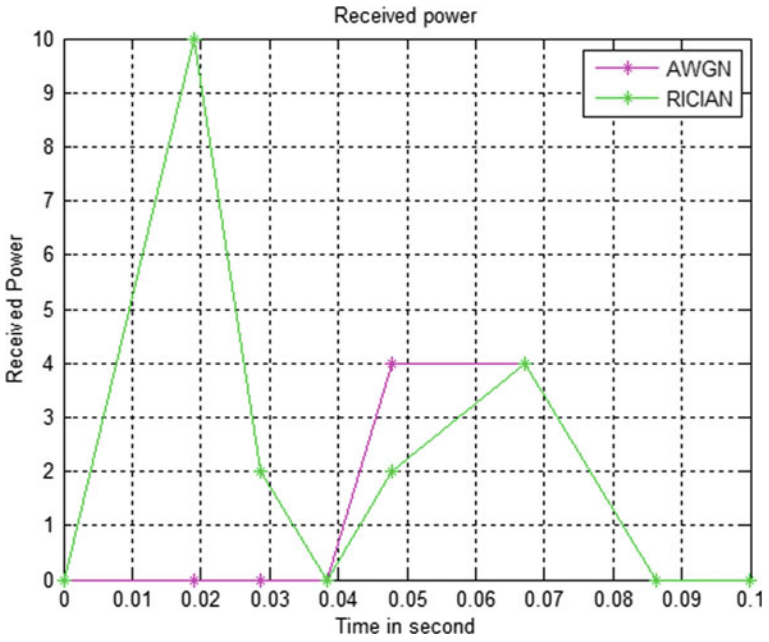


Fig. 7 Received power versus time (sec) over AWGN and Rician channels

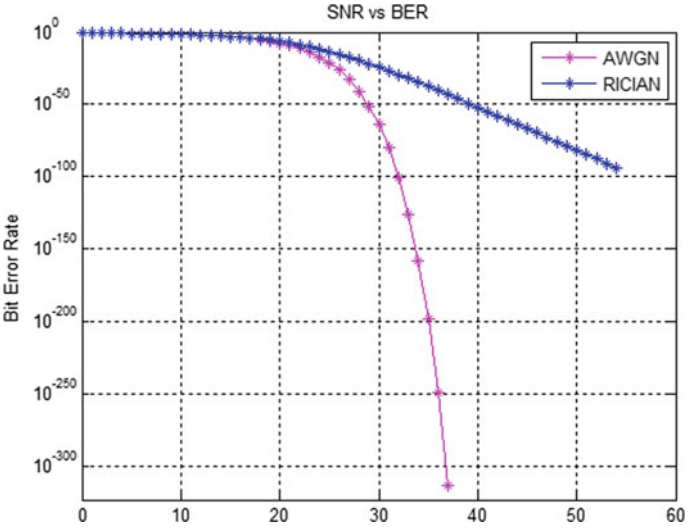


Fig. 8 SNR versus BER over AWGN and Rician channel

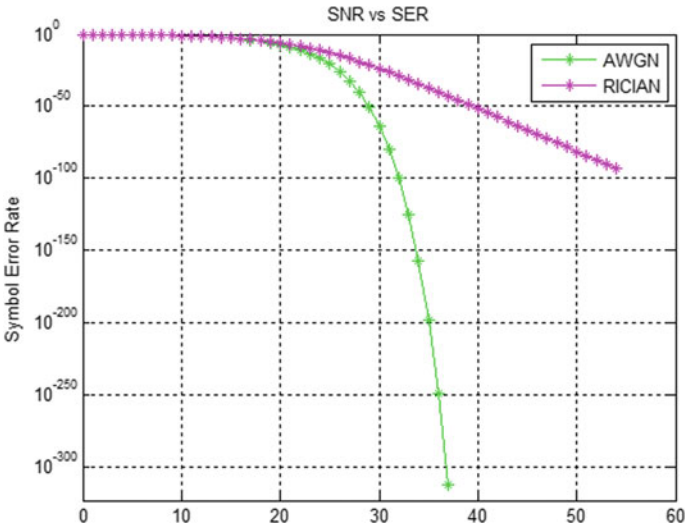


Fig. 9 SNR versus SER over AWGN and Rician channels

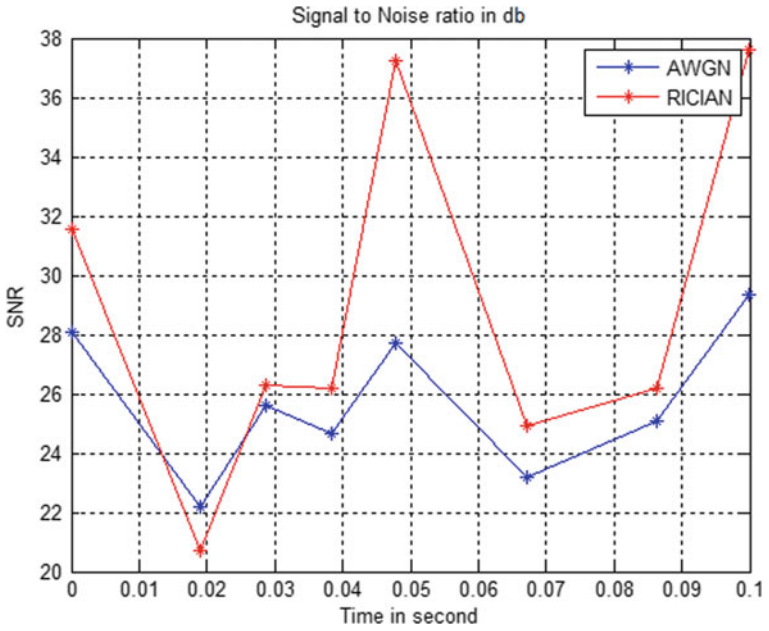


Fig. 10 SNR in db. versus time (sec) over AWGN and Rician channels

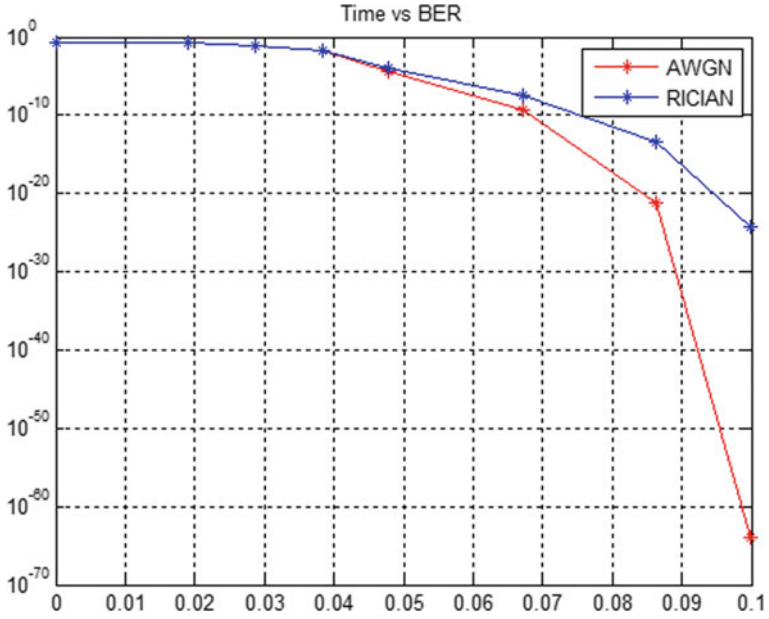


Fig. 11 Time (sec) versus BER over AWGN and Rician channel

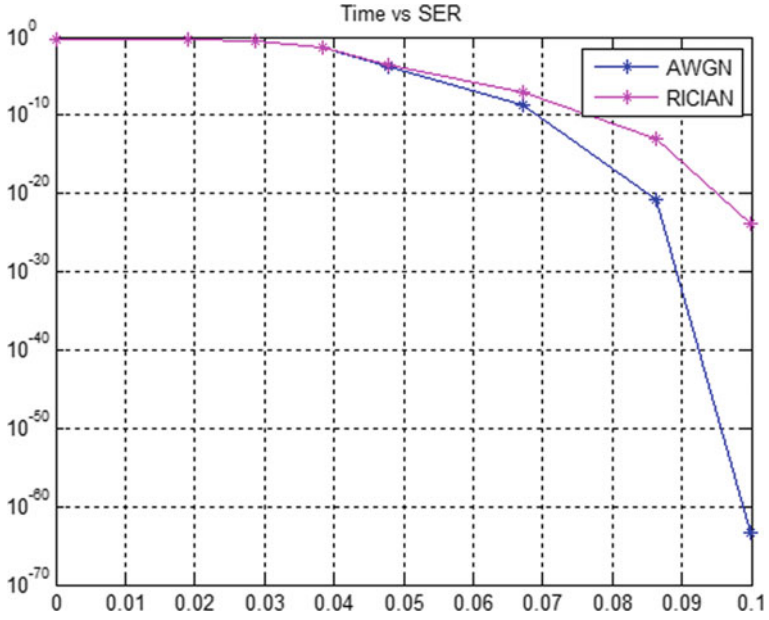


Fig. 12 Time (sec) versus SER over AWGN and Rician channels

4 Conclusion

The proposed system is testified and evaluated using MATLAB. For two different channels, AWGN and Rician channels, the system is evaluated against multiple quality parameters. Comparative analysis shows that Rician channel has an edge over AWGN channel in terms of symbol error rate, bit error rate, signal-to-noise ratio, packet loss, received power, etc. The proposed system is found effective in terms of less packet loss and better received power without inter-symbol interference in both channels.

References

1. D. Kivanc, G. Li, H. Liu, Computationally efficient bandwidth allocation and power control for OFDMA. IEEE Trans. Wireless Commun. **2**(6), 1150–1158 (2003)
2. M. Morelli, C.-C.J. Kuo, Man-on pun, synchronization techniques for orthogonal frequency division multiple access (OFDMA): a tutorial review. Proc. IEEE **95**(7), 1394–1427 (2007)
3. S. Pietrzyk, G.J.M. Janssen, Multiuser subcarrier allocation for QoS provision in the OFDMA systems, in *Proceedings IEEE 56th Vehicular Technology Conference* (Canada, 2002), pp. 1077–1084
4. H. Yin, S. Alamouti, OFDMA: broadband wireless access technology, in *IEEE Sarnoff Symposium* (New Jersey, USA, 2006), pp. 1–4
5. V. Dalakas, P.T. Mathiopoulos, F.D. Cecca, G. Gallinaro, A Comparative study between SC-FDMA and OFDMA schemes for satellite uplinks. IEEE Trans. Broadcast. **58**(3), 370–378 (2012)
6. A. Mráz, L. Pap, General performance analysis of M-PSK and M-QAM wireless communications applied to OFDMA interference, in *WTS'10: Proceedings of the 9th Conference on Wireless Telecommunications Symposium* (USA, 2010), pp. 18–24
7. P. Krishna, A.K. Tipparti, K.K. Rao, M-QAM BER and SER analysis of multipath fading channels in long term evolutions (LTE). Int. J. Sig. Process. Image Process. Pattern Recogn. **9**(1), 361–368 (2016)
8. I.B. Türk, B. Özbek, Radio resource management for user-relay assisted OFDMA-based wireless networks. Int. J. Electron. Commun. **70**(5), 643–651 (2016)
9. R. Khanai, S.S. Shahapur, D. Torse, Performance analysis of underwater acoustic communication using IDMA-OFDM-MIMO with Reed Solomon and Turbo Code. Int. J. Comput. Netw. Inf. Secur. (IJCNIS) **10**(12), 41–46 (2018)

A Review on Internet of Wearable Things for Pervasive E-Health Care: Energy Efficiency and Prospects

Partha Pratim Ray and Dinesh Dash

Abstract Internet of Wearable Things (IoWT) has emerged as a new paradigm of smart computing, especially for tiny, resource-constrained and Internet-enabled ecosystem. With advent of low-cost sensors and open market place such IoWT systems have become inevitably important for monitoring of numerous use cases, for example, health care, tracking, activity recognition and safety assurance. Huge prospects of IoWT simultaneously increase the energy alleviation problem, resulting into facilitation of unstable IoWT ecosystem. Unlimited consumption of energy due to the IoWT implementation can uplift the carbon foot-print. In this article, we present energy efficient approaches for minimizing power alleviation issue in IoWT. Firstly, we present background about the IoWT. We further discuss energy mitigating aspects to make IoWT ecosystem sustainable. Finally, we list a number of key issues and prescribe future direction to improve energy efficient scenario in IoWT.

Keywords Internet of Things · Wearables · Energy conservation · Health care · Smart applications

1 Introduction

Internet of Things (IoT) refers to the paradigm of exchanging information between the network of physical "things" based on the Internet infrastructure [1, 2]. In most of the cases, IoT comprises a set of embedded devices built from resource-constrained processing units, i.e., microcontrollers to process a specific task [3]. Such IoT devices

P. P. Ray (✉)
Department of Computer Applications, Sikkim University, Gangtok, Sikkim, India
e-mail: ppray@cus.ac.in

D. Dash
Department of Computer Science and Engineering, National Institute of Technology Patna, Patna, India
e-mail: dd@nitp.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_7

normally range from micro to macro scale to enable variety of applications in dynamic environment [4]. Sensors and actuators support IoT-based devices to gather information about a physical phenomenon and control the appropriate behavior in proper manner.

The "wearable form factor" in the IoT devices has helped to emerge a new dimension of tiny systems, called IoWT [5, 6]. The IoWT is useful to monitor some key use cases, for example, health care, localization, activity monitoring and support for safe environment maintenance. Due to the application specifics, such IoWT solutions need to capture continuous data from its periphery [7]. Thus, it results into the power alleviation problem [8].

It is expected that in coming days, the world shall observe billions of such IoWT implementations in numerous scenarios belonging to industry 4.0. Time has come to look into the minimization of power alleviation problem due to the excessive utilization of IoWT device pool [9, 10]. Stand-alone processing capabilities and dependency on the sensor-actuator notion have increased the burden of power mismanagement in current time. This concern is further elevated when IoWT is opted toward the life-saving and remote are monitoring use cases.

IoWT can certainly minimize the energy wastage while inculcating a range of strategies, for example, task-offloading, duty cycling, energy-aware routing, low-power embedded system design, use of adaptive transmission power control, compressive sensing and approximate computing [11, 12]. Though, selection of appropriate strategy is a difficult task, so it demands for a standardized policy realization.

This article presents a review on existing literature that advocate IoWT-based implications to mitigate the power alleviation problem. The main contributions of this article can be summarized as follows: (i) discussion on background of IoWT ecosystem, (ii) concise selection of appropriate application scenarios, (iii) identification of energy efficient strategies, (iv) key challenges thereto and (v) prescribing the future road map.

The article is organized as follows. Section 2 presents related works. Section 3 presents background of IoWT. Section 4 discusses energy efficient strategies for IoWT. Section 5 identifies key issues and present future directions.

2 Related Works

We perform a search on the public article repository including Scopus, Web of Science, IEEE Xplore, Google Scholar and Science direct to find related works. We used combination of following keywords while doing the search: "IoT AND Wearable AND Survey", "IoWT AND Survey", "IoT AND Wearable AND Energy AND Survey", and "IoT AND Wearable AND Energy AND Review". We searched articles during 2017–2020, with some filtering options. Based on initial screening, we find 52 articles. Later, we removed many articles based on following criteria, such as (i) not directly related to IoWT, (ii) no energy specific aware discussion, (iii) non full-text availability, (iv) redundant information and (v) non-survey or review articles. Finally,

we select 10 most related works for further analysis. Table 1 presents a comparative analysis of related works.

In [8], a systematic review on energy efficient IoWT design has been presented. Statistical analysis of existing IoWT solutions is investigated to identify energy-design trade-off. Performance evaluation perspectives are also inculcated to showcase IoWT-based market orientation.

A study reviews various IoT-based wearables for elderly care [13]. The main focus of this work is to check health status concerns that can be resolved using the IoT technology. A number of diseases and medical care including Parkinson's, Alzheimer's and ambient-assisted living (AAL) were investigated.

Dian et al. [14] present a survey to align wearables with the cellular IoT (CIoT) paradigm, i.e., connecting physical things to the distributed systems using existing cellular network technologies. The main focus of this work was to find wearables that can be used for vitals or non-vitals monitoring along with required sensors.

Investigation is carried out to check IoT-based telemedicine usage using smart devices [15]. The aim of this work is to prevent disease and promote healthy smart technology utilization among the masses.

Positive computing refers to the intelligent medical services that can improve well-being of a person in terms of depression and stress. Possibility of such a strategy is investigated in [16] to find research directions against the wearable-IoT users. This work deals with both mobile and wearable way outs to integrate with IoT-based frameworks. This study includes the behavior marker and device modality while performing the evaluation of such notion.

Baig et al. [17] have reviewed IoT-based wearables to assist the independent-elderly while providing fall and day-to-day activity detection.

IoT is further expanded to meet the smart device amalgamation for smart health care in [18]. The key focus of this research lies into the review of sensor subsystem design principles. Further, energy harvesting design is discussed along with on-node processing and wireless communication subsystem design.

IoT is capable to support medical devices. In [19], searches for hardware design considerations with respect to the existing context of medical IoT.

Fernández-Caramé et al. [20] present a review on the e-textiles that can be used to design IoWT devices. Such smart clothing frameworks are key to allow continuous monitoring of health and activity of the user.

Finally, Seneviratne et al. [21] present an in-depth discussion on various wearable devices in the context of research prototypes and products point of view. Wide range of wearables such as smart watch, smart eye-wear, smart garment, and e-patch systems are elaborated.

Novelty of this work: The novelty of this work lies into the discussion of IoWT-based implications for key application domains such as, health care, activity recognition, tracking and safety monitoring. We present discussions on identifying of the energy efficiency improving strategies and possible challenges that must be dealt with a priori installing into reality.

Table 1 Comparison of related works

Article	Year	IoWT	Energy	Contributions	Limitations
Qaim et al. [8]	2020	Yes	Partially Yes	Systematic review, energy efficient requirements of IoWT	No related work discussion, future direction missing
Stavropoulos et al. [13]	2020	Yes	No	Literature review, elderly care, sensors	No challenge, no future direction missing
Dian et al. [14]	2020	Yes	No	Survey, focus on applications, cellular IoT, challenges	No future direction
Albahri et al. [15]	2020	Partially Yes	No	Review, disease prevention, telemedicine	Unsuitable for energy-aware IoWT
Lee et al. [16]	2019	Partially Yes	No	Literature review, intelligent computing frameworks	No challenge, no future direction
Baig et al. [17]	2019	Partially Yes	No	Systematic review, independent living, sensors, monitoring	Inappropriate discussions, no future direction missing
Bali et al. [18]	2018	Partially Yes	No	Review, smart devices, subsystem design	No challenges, no future direction
Qureshi et al. [19]	2018	Yes	No	Review, medical IoT, hardware design	No challenge, no future direction
Fernández-Caramés et al. [20]	2018	Yes	No	Review, smart clothing, e-textile, medical, applications	No challenge, no future direction
Seneviratne et al. [21]	2017	Partially Yes	No	Survey, wearable devices, sensors, products	Not suitable for energy-aware IoWT
Proposed study	2020	Yes	Yes	Review, energy-aware IoWT KPIs, energy-aware challenges	Suitable for generic IoWT design

3 Background of IoWT

3.1 Requirements

Key Performance Metrics Performance of IoWT needs to be measured to improve overall service mitigation in a holistic manner. Six key performance metrics can be expressed as follows: (i) energy awareness: IoWT should be assessed against the energy-aware design considerations, (ii) societal impacts: applications of IoWT must cater societal factors such as equality, (iii) technology development: IoWT-centric technology development can be rapidly perceived, (iv) ecosystem openness: IoWT-based ecosystem most be able to include open source tools to improve faster business gain, (v) governance policies: strategies shall be imposed on the usage policy to enrich IoWT product distribution and (vi) community support: IoWT should enable community supported environment for seamless performance up-gradation.

IoT IoT is the core design module of IoWT. Gradual evolution from ad hoc network, wireless sensor network to mobile computing; IoWT has seen a lot change in the last decade, where personal area network (PAN) is further investigated and integrated with revolutionizing IoT paradigm. Body area network (BAN) and improved wireless BAN (WBAN) are two most closely notions of IoWT that have been fantastically dealt with [22]. TCP/IP protocol stack is further modified to work with IoWT. A range of smart on-board microcontrollers (e.g., ATMEL, TI, ARM, PIC, etc.) have changed the course of wearable extension in recent years. All sorts of microcontrollers with primary memory capacity starting from 4 KB to 8 GB are currently in use to deploy IoWT models. Moreover, cloud computing has become a part of existing IoT to enable IoWT to offload tasks. Machine learning techniques are also being investigated to form a miniaturized IoWT system. Further, several third-party programming interfaces are in practice that can help IoWT to emerge as lightweight solutions to several applications such as healthcare, rehabilitation, localization, tracking and activity recognition.

Sensors and Actuators Sensor refers to a electro-mechanical or chemical device that can convert physical phenomena to electrical signals for input to digital systems, i.e., microcontrollers. IoWT requires a range of sensors for example, accelerometer, temperature, pulse oximetry, near infrared, audio, gyroscope, inductive sensors, vibration, camera, ultrasound, capacitive, piezoelectric, health and force sensors [23]. Vitals and non-vitals can be easily measured by such sensors with help of IoWT that includes, orientation, direction, image, heart rate, respiration, blood pressure, blood oxygen, blood glucose, deflection and surface electromyography. Pneumatic pump, piezo buzzer, light emitting diodes and relay controlled valves are examples of some actuators that can be integrated with IoWT to control electrical signals in form of feedback physical notion.

Communication Technologies IoWT uses short, mid and large range communication technologies essentially in-built with the embedded design [24]. In short distance communication technologies, Bluetooth, Wi-Fi and visible light communication are

mostly used. For mid-range technologies, LoRA, Zigbee, Sigfox and low-power wide area network (LPWAN) are suitable for IoWT scenario. For long range, IoWT can depend on existing 4G-LTE, 5G, cellular IoT and satellite communication channels. Selection of any communication tool must be considered against energy utilization perspective as well as application requirement.

3.2 Application Scenarios

Smart Health care and Rehabilitation Most basic application of IoWT can be based on the smart healthcare facilitation. Such health service can be divided into two major parts, (i) vital monitoring and (ii) non-vital monitoring. IoWT is used in monitoring of following vitals such as, pulse rate, respiratory rate and body core temperature. Non-vitals, such as blood oxygen (SpO_2) and blood pressure monitoring, are also equally deployed using IoWT. Literature advocate that IoWT has been successfully tested and validated in following domain of health care: (i) smart walking, (ii) smart wheelchair, (iii) rise/fall detection and (iv) assistive nursing practices. Rehabilitation activities can also be monitored using IoWT. For example, orientation, force, distance, surface electromyography, deflections and acceleration.

Activity Recognition and Sports Athletes and regular end users can take help of activity recognition for better life style and physical fitness enrichment. Daily physical activities and out-of-hospital care can be augmented with IoWT. Especially, muscle movement during the yoga practice and some special sports, e.g., running, jumping and boxing. Various sorts of activities such as knee bending, arm movement, jumping around and waist positioning can be efficiently monitored by IoWT. Moreover, IoWT can be useful for all types of sports such as martial art, golf, weightlifting, basketball, rowing and hockey are tested as confirmation. Possible feature extraction from the raw input signals can be fed to IoWT with help of artificial intelligence strategies.

Tracking and Localization Tracking and localization of humans and animals can be seamlessly integrated with IoWT. For example, finding a bird's location in the air, trajectory of a wheel and pet movement can be assessed with IoWT. Video, audio and general signal patterns are classified to estimate the location. Sometimes, accurate localization is done with help of geographical position system (GPS). Location tracking of the dementia patients can help the care giver to find current status of the patient. Use of the received signal strength indicator (RSSI) and cellular IoT signals accurate positioning can be possible using IoWT. Further, disaster and search operations can be resolved using IoWT.

Safety Monitoring Safety monitoring is classified into three major types, (i) fall detection, (ii) fatigue detection and (iii) environment monitoring. Accidental fall causes sudden death or severe damage to a large number of elderly people each year. IoWT presents a good choice of fall detection applications to inform care givers. With

integration of statistical learning techniques, IoWT can act as assistive technology to the elderly people. A major portion of road accidents happens due to fatigue of drives. IoWT can facilitate a grate solution for fatigue monitoring of drives on the go. Environment monitoring can be assisted with IoWT, where targets stuck inside a coal mine, under sea, flood, mountain and forest fires can be safeguarded.

4 Energy Efficiency Improving Schemes

Efficient behavior of the IoWT paradigm depends on several crucial factors related to the energy usage policies. In this section, we discuss such important schemes that may be optimally used to improve energy efficiency of IoWT.

4.1 Task-Offloading

Currently, most of the IoWT end points use cloud computing services. Thus, task-offloading schemes can be used herein to improve the battery life of the device. Such process of task-offloading incurs tasks to get executed in the remote locations, i.e., cloud data centers. Mainly computation intensive jobs are offloaded to the cloud handlers for producing feedback results. It has a problem, i.e., unpredictable transmission delay within the network path. Instead of using cloud computing, recently emerged fog or edge computing can be used to minimize such delay. However, efficient task splitting and independent execution must be done in the vicinity of an IoWT end device.

4.2 Duty Cycling

IoWT devices come up with embedded short-range communication protocols such as Bluetooth and Wi-Fi that can establish connection with nearby peers for data sharing. Various types of on-board sensors also consume significant amount of energy that results in a faster “die out” scenario. Thus, duty cycling technique can be used with the IoWT ecosystem to conserve an optimum amount of energy. In this strategy, a number of modules of the IoWT device are instantly put into the sleep mode when not active. Use of artificial intelligence can be augmented with the medium access control (MAC) protocols to solve the wakeup and sleep cycle problem.

4.3 Energy-Aware Routing

Wearable-IoT devices frequently communicate with neighbor nodes as well as fog or cloud end points. Energy-aware routing schemes must be incorporated with the IoWT in order to enhance battery life of the device and point-of-care availability to the end users. While designing energy-aware routing, information related to the neighbor's energy status should be periodically updated. Such overhead needs to be resolved during the course of energy-aware routing protocol design.

4.4 Low-Power Embedded Design

Recent development of low-power digital equipment has brought an opportunity for the IoWT devices to be used ubiquitously. New type of application specific integrated circuit (ASIC) is being used for rapid prototype development, especially processing and memory units. Communication modules of any IoWT devices consume maximum energy. Despite using duty cycling algorithms, a large portion of time is devoted for channel sensing activities for collision avoidance [25]. Energy is more needed when IoWT needs to work in the dense populated nodes. Use of wakeup radio modules in wearables can minimize energy consumption, though trade-off between cost ad integration space complexity should be considered.

4.5 Low-Power Communications

Nowadays, most of the IoWT devices use wide range of communication tools that consists of Bluetooth, Wi-Fi, Zigbee and RF transceiver. It is true that choice of communication technology is related to the deployment aspects. However, it is important to consider that fact that an optimal selection of low-power communication protocol suite would certainly improve energy wastage. For example, LoRa, IEEE 802.11ah, narrow band-IoT (NB-IoT) and visible light communication (VLC) technologies can be used to reduce power consumption.

4.6 Adaptive Transmission Power Control

It is a fact that energy/bit transmission is 1000 times higher than energy/instruction. A transceiver can work in four different stages, namely send, receive, idle and sleep. Thus, selection of appropriate strategy toward utilization of adaptive transmission power control can help minimize overall energy usage. Artificial intelligence can be used to determine distance from the sender node to estimate transmission power

a priori. However, lightweight smart schemes must be developed to mitigate the learning experience with minimal bit error rate (BER).

4.7 Compressive Sensing

Compressive sensing refers to the process of signal acquisition and restructuring to achieve efficient behavior of energy consumption and performance evaluation. Normal the signal sparsity characteristic is used in achieving the optimal restructuring of signal with help of the Nyquist theorem. Compressive sensing is useful for the activity recognition application where fixed sensing intervals are considered.

4.8 IoWT Data Compression

IoWT devices produce an enormous amount of bit streams while in a continuous action. Many times, an IoWT application generates a large portion of redundant and un-correlated data which may not be of importance use. Thus, a data compression technique needs to be imposed over the wearable devices to efficiently handle the resultant data. It can reduce the size of data that is yet to be processed later in the processing chain.

4.9 Approximate Computing

Approximate computing refers to the production of "good enough" result out of a computation rather than accurate one. Some assistive healthcare applications can harness the significance of approximate computing approach to minimize power consumption. However, the key challenge is fixation of "enough threshold" value which is different in various use cases. Careful selection of approximation can reduce energy wastage and improve IoWT node life time thus reliability.

4.10 Lightweight Security

Standard security techniques are developed to run on top of full-fledged as well as targeted device pools. IoWT certainly requires modified security schemes with lightweight design approach. Investigation must be carried to seek for appropriate cipher and de-ciphering tools in the IoWT ecosystem. A significant trade-off between energy consumption and security must be well studied before use of security tools in the IoWT design frames.

5 Key Challenges and Future Directions

IoWT is a rapidly growing field of application that needs to cater evolutionary trends while putting the energy consumption factor on top of all. Numerous key challenges are identified in this section that must be well handled to gain sophisticated user experience of IoWT.

5.1 Challenges

Transmission Overheads IoWT devices need to transmit continuous data to remote cloud stations. Such huge data streaming incurs transmission overheads along with the IoWT-based services. Existing literature shows three main approaches to mitigate this issues such as, (i) embedded signal processing to avoid unnecessary data transmission by carrying out local processing, (ii) compressive sensing to avoid redundant data and (iii) embedded machine learning (ML) augmentation to perform complex task thus latency reduction.

Wireless Technology Inclusion of wireless technology in the IoWT should be carefully done to minimize energy consumption. A number of possible solutions can solve this issue that includes, (i) duty cycling implication to periodically power off communication modules, thus conserve energy, (ii) adaptive transceiver power control of transceiver based on distance and remote node, (iii) use of low-power wireless communication protocol, i.e., IPv6 low-power personal area network (6LoWPAN), (iv) multi-channel time division multiple access (TDMA) for allocation slots and (v) adaptive communication interval in IoWT environments.

Inefficient Routing IoWT-based routing algorithms need to be improved to minimize energy consumption. For example, (i) energy-aware neighborhood selection for optimal route search, (ii) use of multi-parameter cost function to select next hop on the route and (iii) data priority centric selective data routing.

Security and Privacy Security and privacy of IoWT are two pillars of attentions that must be carefully resolved while designing IoWT frameworks. Existing literature advocate following key requirements in this regard, such as (i) content agnostic privacy and encryption protocol elimination, (ii) lightweight cryptography protocols amalgamation, (iii) accelerated system-on-chip (SoC) utilization for execution of cryptographic primitives and (iv) trade-off mitigation between energy and security computation.

Processing and Storage Limitation Though, some of IoWT devices lack in processing and storage capabilities, new range of devices come with higher capacity. Thus, energy minimization can be easily harness with possible solutions as follows, (i) heterogeneous multi-core processor gateway design, (ii) task-offloading to reduce data computation on-board, (iii) inclusion of dew, edge and fog computing, (iv)

data aggregation, (v) data summarization, (vi) data compression tools, (vii) seamless resource sharing between IoWT devices.

Minimal Hardware Acceleration IoWT hardware is designed to perform specific tasks, for example, health data monitoring and activity recognition. In such applications, large data is produced which should be tackled with hardware acceleration tools. Besides, task-offloading and data compression schemes, identification of hardware acceleration facilities needs to be imposed.

Inefficient Energy Modules Energy is the umbrella term for IoWT devices. Thus, existing energy conversion modules must be well equipped with configurable data acquisition components. Further, transceiver coupling with near-zero wakeup radio can be added with the adaptive sampling strategies. Power down sensing modules can be used to improve energy efficiency in IoWT. For example, using a gyroscope sensor, instead of using two accelerometers could be used in activity recognition.

Battery Issues Energy harvesting schemes can be investigated to enhance battery capacity in IoWT nodes. However, inclusion of low-power and low-complexity SoC and approximate computing with IoWT uplifts energy conservation. Super capacitors, LM/M cylindrical lithium cells and VL rechargeable cells can be seen as alternatives. Efficient battery management units can be integrated with limited overhead in the IoWT devices to improve battery utilization.

Standardization and Regulation IoWT devices and application programming interface (API) combined suites are normally vertical siloed, i.e., non-monitored. There is no special governance neither standardization policies till date. Uniform policy is needed to minimize the social inertia for more utilization of IoWT solutions for pervasive use cases. Policy makers and international standardization bodies should come forward to devise a set of regulations to safeguard IoWT from possible human right and privacy violations.

5.2 Future Directions

Triboelectric Nanogenerator Triboelectric nanogenerator (TENG) refers to a physical phenomena where mechanical agitation can be transformed into the electric signal through triboelectricity and static-induction. TENG has gained a significant attention in recent past for development of low-power sustainable energy-aware systems such as IoWT. TENG-based energy harvesting can be seen as a promising technology for IoWT-centric energy conservation. A range of designs can be sought for embedding TENG with IoWT that includes vertical separation mode, lateral sliding and atomic electrode mode. Various forms of mechanical energy such as vibration, human body motion and force are suitable for IoWT integration. Though, existing TENGs face choice of materials and optimal surface design for enough power generation, the issues in the reliable design constraints with IoWT must be investigated.

Unified Architecture IoWT follows unspecified and non-standardized architectural provisions for deployment in several use case scenarios. Such practice appears to be suitable for the sustainable start-up growth and product-based business formulation. Thus, a unified architecture should be paved to augment energy-aware protocols along with autonomous framework management. Renewable energy harvesting platform can be envisioned with such unified architecture to improve energy conscious IoWT applications.

On-Demand Service IoWT solutions are deployed in a variety of highly dynamic situations, for example, location tracking of a dementia patient. Due to the dynamic nature of these applications, IoWT must cope up with the on-demand service mitigation approach that would facilitate the user's need when stuck in an emergency case. Instead of continuous polling, IoWT wakeup radios with help ML-based techniques can provide support to the end users and care givers at remote location.

Next-Generation Computing 5G is at the verge of deployment in various corner of the world by 2023. It shall provide an opportunity for the IoWT business to grow with minimum energy consumption metrics. It is expected that development toward the 6G might follow with immediate effect of 5G implications. This would result into integration of IoWT with low-energy aware quality of service (QoS) metrics.

6 Conclusion

In this article, we discuss how IoWT can accommodate certain energy efficient strategies to improve the issues with excessive energy consumption. We observe that special care should be imposed while designing IoWT systems to minimize energy utilization. It is expected that existing IoWT shall be aligned with new energy mitigating policy both in industry and social level. Further, investigations should be carried out to develop holistic architecture for lightweight service provisioning.

References

1. F. Wu, T. Wu, M.R. Yuce, Design and implementation of a wearable sensor network system for IoT-connected safety and health applications, in *IEEE 5th World Forum on Internet of Things (WF-IoT)* (Limerick, Ireland, 2019), pp. 87–90. https://doi.org/10.1109/WF-IoT.2019.8767280
2. Z. Zhou, H. Yu, H. Shi, Human activity recognition based on improved Bayesian convolution network to analyze health care data using wearable IoT device. IEEE Access **8**, 86411–86418 (2020). https://doi.org/10.1109/ACCESS.2020.2992584
3. P. Sundaravadivel, A. Fitzgerald, P. Indic, i-SAD: an edge-intelligent IoT-based wearable for substance abuse detection, in *IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS)* (Rourkela, India, 2019), pp. 117–122. https://doi.org/10.1109/iSES47678.2019.00035

4. M.T. Sqalli, D. Al-Thani, Towards an integrated health and wellness program using human-IoT interaction, in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)* (Doha, Qatar, 2020), pp. 283–290. https://doi.org/10.1109/ICIoT48696.2020.9089538
5. P. Sundaravadivel, C. Tumwesigye, S.P. Mohanty, E. Kougianos, iMED-tour: an IoT-based privacy-assured framework for medical services in smart tourism, in *2020 IEEE International Conference on Consumer Electronics (ICCE)* (Las Vegas, NV, USA, 2020), pp. 1–5. https://doi.org/10.1109/ICCE46568.2020.9043085
6. S. Bao, T.N. Gia, W. Chen, T. Westerlund, Wearable health monitoring system using flexible materials electrodes, in *IEEE 6th World Forum on Internet of Things (WF-IoT)* (New Orleans, LA, USA, 2020), pp. 1–2. https://doi.org/10.1109/WF-IoT48130.2020.9221282
7. B. Mukhopadhyay, O. Sharma, S. Kar, IoT based wearable knitted fabric respiratory monitoring system. IEEE Sensors, 1–4. https://doi.org/10.1109/ICSENS.2018.8589540
8. W.B. Qaim et al., Towards energy efficiency in the Internet of wearable things: a systematic review. IEEE Access **8**, 175412–175435 (2020). https://doi.org/10.1109/ACCESS.2020.3025270
9. T. Wu, F. Wu, C. Qiu, J.-M. Redout, M.R. Yuce, A rigid-flex wearable health monitoring sensor patch for IoT-connected healthcare applications. IEEE Int. Things J. **7**(8), 6932–6945 (2020). https://doi.org/10.1109/JIOT.2020.2977164
10. K. Miyamoto, K. Hashimoto, M. Kasaoka, M. Kakumu, Wearable sensors corresponding to various applications in medical, healthcare field, in *27th International Workshop on Active-Matrix Flatpanel Displays and Devices (AM-FPD)* (Kyoto, Japan, 2020), pp. 115–118. https://doi.org/10.23919/AM-FPD49417.2020.9224474
11. M. Haghi et al., A flexible and pervasive IoT-based healthcare platform for physiological and environmental parameters monitoring. IEEE Int. Things J. **7**(6), 5628–5647 (2020). https://doi.org/10.1109/JIOT.2020.2980432
12. D. Chloros, D. Ringas, Fluspot: seasonal FLU tracking app exploiting wearable IoT device for symptoms monitoring, in *5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)* (Corfu, Greece, 2020), pp. 1–7. https://doi.org/10.1109/SEEDA-CECNSM49515.2020.9221843
13. S.T. Stavropoulos, A. Papastergiou, L. Mpaltadoros, S. Nikolopoulos, I. Kompatsiaris, IoT wearable sensors and devices in elderly care: a literature review. Sensors **20**(10), 2826 (2020)
14. F. John Dian, R. Vahidnia, A. Rahmati, Wearables and the Internet of things (IoT), applications, opportunities, and challenges: a survey. IEEE Access **8**, 69200–69211 (2020). https://doi.org/10.1109/ACCESS.2020.2986329
15. A.S. Albahri, J.K. Alwan, Z.K. Taha, S.F. Ismail, R.A. Hamid, A.A. Zaidan, O.S. Albahri, B.B. Zaidan, A.H. Alamoodi, M.A. Alsalem, IoT-based telemedicine for disease prevention and health promotion: state-of-the-art. J. Netw. Comput. Appl. **173**, 102873 (2020). https://doi.org/10.1016/j.jnca.2020.102873
16. U. Lee, K. UHan, H. Cho, K.M. Chung, H. Hong, S.J. Lee, Y. Noh, S. Park, J.M. Carroll, Intelligent positive computing with mobile, wearable, and IoT devices: literature review and research directions. Ad Hoc Netw. **83**, 8–24 (2019). https://doi.org/10.1016/j.adhoc.2018.08.021
17. M.M. Baig, S. MAfifi, H. Gholamhosseini, F. Mirza, A systematic review of wearable sensors and IoT-based monitoring applications for older adults–a focus on ageing population and independent living. J. Med. Syst. **43**(8), 233 (2019)
18. H. Baali, H. Djelouat, A. Amira, F. Bensaali, Empowering technology enabled care using IoT and smart devices: a review. IEEE Sens. J. **18**(5), 1790–1809 (2018). https://doi.org/10.1109/JSEN.2017.2786301
19. F. Qureshi, S. Krishnan, Wearable hardware design for the internet of medical things (IoMT). Sensors **18**(11), 3812 (2018)
20. T.M. Fernndez-Carams, P. Fraga-Lamas, Towards the Internet of smart clothing: a review on IoT wearables and garments for creating intelligent connected e-textiles. Electronics **7**(12), 405 (2018)

21. S. Seneviratne et al., A survey of wearable devices and challenges. IEEE Commun. Surv. Tutor. **19**(4), 2573–2620 (2017). https://doi.org/10.1109/COMST.2017.2731979
22. K. Miyamoto, K. Hashimoto, M. Kasaoka, M. Kakumu, Wearable sensors corresponding to various applications in medical, healthcare field, in *27th International Workshop on Active-Matrix Flatpanel Displays and Devices (AM-FPD)* (Kyoto, Japan, 2020), pp. 115–118. https://doi.org/10.23919/AM-FPD49417.2020.9224474
23. O.S. Albahri et al., Fault-tolerant mHealth framework in the context of IoT-based real-time wearable health data sensors. IEEE Access **7**, 50052–50080 (2019). https://doi.org/10.1109/ACCESS.2019.2910411
24. A.R. Maria, P. Sever, S. George, MIoT applications for wearable technologies used for health monitoring, in *2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (Iasi, Romania, 2018), pp. 1–4. https://doi.org/10.1109/ECAI.2018.8679069
25. T. Wu, F. Wu, C. Qiu, J.-M. Redout, M.R. Yuce, A rigid-flex wearable health monitoring sensor patch for IoT-connected healthcare applications. IEEE Int. Things J. **7**(8), 6932–6945 (2020). https://doi.org/10.1109/JIOT.2020.2977164

An Analysis on Tesla's Stock Price Forecasting Using ARIMA Model

Aishwariya Subakkar, S. Graceline Jasmine, L. Jani Anbarasi, J. Ganesh, and C. M. Yuktha Sri

Abstract Predicting stock price in the short term is always a demanding task over the decades. Every common thought of a human being is to devote money to stock markets which results in tremendous output. Strong research time was invested in stock market prediction as it is an important phenomenon of the economy. This has also sparked the development of improved models for stock price analysts. The autoregressive integrated moving average (ARIMA) is a type of moving average model that is used to for building a stock price forecasting model. In this paper, ARIMA models have been used to predict the stock price of daily trading stock prices published in the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE). ARIMA model is analyzed for time series prediction, and the results obtained using this model show strong accuracy for short term and daily stock prediction, and this engages with other model for predicting stock price.

Keywords Analysis · Finance · Forecasting · Time series · Stock prediction · ARIMA

A. Subakkar · S. Graceline Jasmine (✉) · L. Jani Anbarasi · J. Ganesh · C. M. Yuktha Sri
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India
e-mail: graceline.jasmine@vit.ac.in

A. Subakkar
e-mail: aishwariya.subakkar2018@vitstudent.ac.in

L. Jani Anbarasi
e-mail: janianbarasi.l@vit.ac.in

J. Ganesh
e-mail: ganesh.j2018@vitstudent.ac.in

C. M. Yuktha Sri
e-mail: cm.yukthasri2018@vitstudent.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_8

1 Introduction

A stock market is a place where investors buy and sell-out securities of public companies which hold shares, bonds, and other monetary apparatus. Machine learning (ML) plays a key role to analyze the trend of the stock market. In stock market, forecasting the basic objective has been in aiding to buy the stock, sell stock, or hold stock decision [1]. Today many researchers are fascinated regarding the machine learning techniques which provides descent accuracy rates of prediction. ML used various parameters to foresee the stock scores and contributed numerous prototypes. From the stock market emergence, there are various researchers with variety of methods and movements to study stock price prediction. From an economic view, investors maximum use the sequential fundamental analysis, and technical analysis to forecast stock markets. The above stated methods are theoretical, and they don't acceptingly reflect the correlation among the data. Idrees et al. [2] give a clear-cut understanding of the a ARIMA which helps in analysis the forecasting methodology for future values of a time series data and also helps in predicting stock market values. Adebiyi et al. [3] used New York stock exchange (NYSE) and Nigeria stock exchange's (NSE) future stock value to forecast and ARIMA model in order to forecast the accurate values. To provide the best model, they worked on BIC comparing and computing, adjusted R Square, and S.E. of regression to find perfect match of the fitting model. Akpinar and Yumusak [4] predicted the natural gas of a household value in turkey by using ARIMA (p, d, q) Box—Jenkins approach. They worked on joined model over sequential traditional model and showed that Mean Absolute Prediction Error (MAPE) was decreased to 2.2% while comparing with sequential traditional non-merged model. Due to the large data availability, MAPE is used. This research talks about [5] estimation model and prediction of newspaper sales. The number of sales of newspaper from January 2016 to March 2017 is taken as historical data. This approach tends to get more precise data and can be reported from an object. Box—Jenkins method helped in forecasting the behavior of data-points.

This paper has been organized as follows. Section 2 focuses on the related work on methodologies performed to predict stock market using TELSA stock data. Also, the time series forecasting model for precise prediction stock with respect to the observed values and future forecasting is also detailed. Section 3 explains the results and discussion part which deals with the forecasting. It also explains the comparison of fit performance metric for various statistical models. Section 4 details the conclusion and future scope of this research.

2 Related Work

The stock prediction of Tesla includes acquiring data source from yahoo finance, prediction based on various statistical models such as a time series (ARIMA). This section details the ARIMA model with its detailed mathematical flow.

2.1 ARIMA Model

ARIMA models provide a technique to forecast time series. The most widely used approaches are exponential smoothing and ARIMA models to time series forecasting. This scheme also provides supporting procedure to the problem. The exponential smoothing technique is placed on the trend and seasonality of description in the data, and on the other hand, ARIMA model aims to express the autocorrelations from the given data. Before getting to know ARIMA model approach, the concept of stationarity and differencing time series has to be analyzed. Stationary time series main property is they do not depend on the time. Trends, or with seasonality in time series, are not stationary. The trend and seasonality will cause an effect of the time series value at different times. However, a white noise series must look much the same at any point in time, and the series is stationary no matter when you observe it. Continuous cyclic behavior (but with seasonality or no trend) is stationary. The stated behavior with time series can be confusing. This is due the unknown fixed length of the cycles, so before we analyze the series, we cannot be sure where the high point and low point of the cycles lie on. In general, long-term stationary time series don't have expected patterns. Time plots will show the series to be horizontal (although some continuous cyclic actions are possible), with continuous variance.

In Fig. 1a, the Tesla Stock Value was non-stationary in (Tesla Stock Price for 200 consecutive days), but the everyday change in stock values were stationary in Fig. 1b, (Daily Tesla Stock Price for 200 consecutive days). This helps in understanding how to make a non-stationary time series to stationary time series in order to calculate the differences among continuous observations. This is known as differencing. Variance of a time series can be done with the help of transformations of logarithms to stabilize. Differencing can stabilize the mean of a time series by eliminating changes in the level of a time series, and therefore, working on to eliminate the trend and seasonality Box—Jenkins method or ARIMA method is taken for analyzing a short-term prediction. The proposed model uses the above approach for daily trading in stock market. The same result won't fit the long-term modeling. ARIMA technique may be explained as the combo of two autoregressive (AR) model that is integrated with the moving average (MA) model, denoting the notation of ARIMA (p, d, q). d is the value of differencing, q is the degree of MA process, P is the degree of process of AR, autoregressive model with the order of AR (p) model of ARIMA (p, 0, 0) is stated as follows Eq. (1):

$$X_t = \theta_0 + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \cdots + \theta_p X_{t-p} + e_t \tag{1}$$

where: t = stationary time series, θ_0 = constant, θ_p = parameter of autoregressive model (AR), e_t = residual time (t). Moving average model with the order of the MA (q) or ARIMA (0, 0, q) is stated as follows:

$$X_t = \phi_0 - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_q X_{t-q} + e_t \tag{2}$$

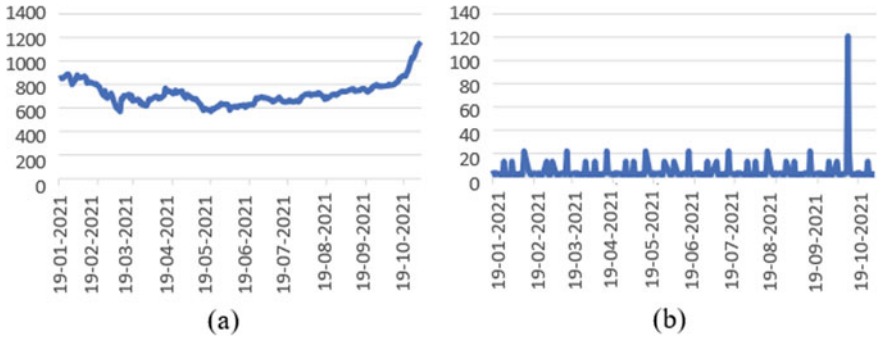


Fig. 1 **a** Tesla stock value for consecutive 200 days and **b** daily Tesla stock value for consecutive 200 days

where: X_t = stationary time series, ϕ_0 = constant, ϕ_q = coefficient of the model which shows the moving average weights, e_t = residual tense used.

Each of essential components of the autoregressive integrated moving average (ARIMA) approach is explained in sequential manner: The "AR" in ARIMA stands for autoregression, and this indicates the approach is analyzed on a dependent relationship among current data and previous data. The letter "I" represents "integrated," which signifies that the data is static and if data is non-static, then the integrated part is to convert data into static. Time series data that has been made "static" by removing observations from previous data, and it is referred to as "stationary data". The "MA" stands for moving average model, and this means that forecasting outcome is linearly dependent on previous data. Each of the AR, I, and MA components is embedded as a parameter in the model. The following is an example of ARIMA parameter notation and explanation.

2.2 *Performance Metrics*

In order to make sure the results derived from ARIMA is specifically accurate with reduced levels of error, the metrics like *R* Squared, Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are taken into account and are utilized to analyze the accuracy of ARIMA model. Various metrics used to measure the performance of ARIMA shown in Eqs. (3), (4), (5), (6) and (7).

$$R\text{-Squared} = \frac{\sum\left(\beta_i - \hat{\beta}_i\right)^2}{\sum\left(\beta_i - \overline{\beta}\right)^2} \tag{3}$$

$$\text{MSE} = \sum_{t=T+1}^{T+h} \left(\hat{\beta}_t - \beta_t\right)^2 / h \tag{4}$$

$$\text{RMSE} = \sqrt{\sum_{t=T+1}^{T+h} \left(\hat{\beta}_t - \beta_t\right)^2 / h} \tag{5}$$

$$\text{MAE} = \sum_{t=T+1}^{T+h} \left|\hat{\beta}_t - \beta_t\right| / h \tag{6}$$

$$\text{MAPE} = 100 \sum_{t=T+1}^{T+h} \left|\frac{\hat{\beta}_t - \beta_t}{\beta_t}\right| / h \tag{7}$$

R-Squared technique is a statistical measure of regression provides an understanding of how much variation of a dependent variable is defined by the independent variable(s). *R* Squared are stated as percentages from 0 to 100, and in this approach, value ranges from 0 to 1. (MSE) is expressed as Mean or Average of the square of difference between real values and estimated values. RMSE maintains the estimation value. This technique provides an accurate phenomenon and reduces large errors. The capacity to differentiate various time series is restricted with this approach. MAE verifies error magnitude for an accurate set of prediction. It also describes how close the actual outcome value while comparing with the forecast value. The direction of the forecasts is not considered in this metric. Moreover, the precisions of continuous variables are determined in this metric. MAPE provides differentiation of different time series data values without describing the error percentage. This metric is unique and this analyzed the parameters are huge [6, 7].

3 Experimental Results and Discussion

In this section, experiments for forecasting the trend of stock market value of Tesla (from July 10, 2015 to July 10, 2020) are conducted to analyze the effectiveness of our proposed model. The first four tasks included the high value, low value, close value, and adjacent close value from years to days is computed to check the prediction performance.

3.1 ARIMA Model for Tesla Stock Price

Tesla Company launched its Initial Public Offering (IPO) in June 29, 2010. NASDAQ Trading, Tesla provided thirteen million shares at a price value of seventeen dollar per share. This enhanced a total value of $226 million. At one point of time, Tesla

Stock Value had a smooth price growth, and it gave way to a dependable pattern of volatility that preceded a huge drop. Until recent records, it could be argued Wall Street at New York found out that Tesla was not a tech company rather Tesla was an automobile company, and this changed its expectations about development for the stock value.

Autocorrelation lag plot of Tesla adjusted close value closely follows a line as shown in Fig. 2a. Tesla adjusted close value series was split into 80% of training data information and 20% of test information data. The forecasting was preformed to over the look at dataset. We have used training dataset to suit the model and generates a prediction for every day within the taken dataset of Tesla Stock value from (July 10, 2015 to July 10, 2020). A forecast was performed and dependence on evaluation data values in previous time steps were considered. ARIMA approach was re-created once every new evaluation value is received to get the training forecast. In all observations, AR manually tracked and hold on in an exceedingly list known as history with training dataset information and to that new observations AR added to every iteration.

From Fig. 2, we analyze that the data is not random but linear and has non-stationary points. Different accuracy metrics are detailed in Table 1.

Grid explore for best (p, d, q) setting was performed on training dataset. *R* Squares, MSE, RMSE, MAE, and MAPE scores were calculated from prediction on the look at dataset and therefore the actual worth in at dataset. The ARIMA model obtained using Tesla stock value provides an appreciable result. Figure 2b gives a clear explanation of predicted stock value over actual stock value of Tesla stock price. This model gives a best prediction accuracy and to be comparative fast with other alternatives, in terms of training, fitting time, and complexity.

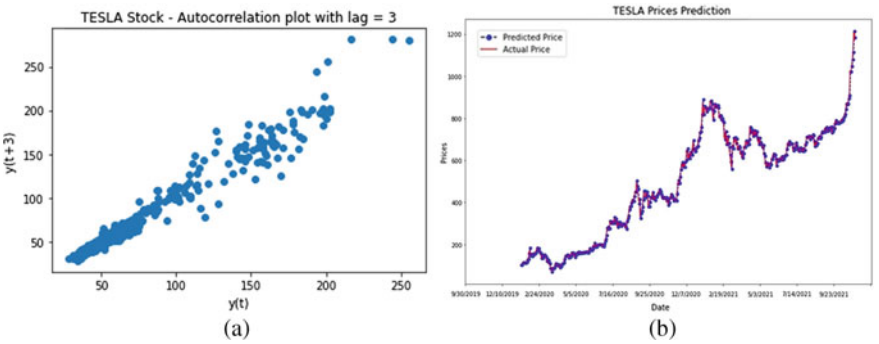


Fig. 2 **a** Autocorrelation plot of Tesla stock value. **b** Tesla stock value prediction

Table 1 Different accuracy metrics value

Criteria	*R* squared	MSE	RMSE	MAE	MAPE
Values	0.9933	441.4330	21.0103	14.2046	101.03700

4 Conclusion

This paper delves into the process of developing an ARIMA model to forecast stock prices. The findings obtained with the best ARIMA model demonstrate that ARIMA models have a promising future in predicting stock costs. ARIMA models will be able to compete fairly with rising statement strategies based on the results obtained. The model's effectiveness is evaluated using the *R*-Squared value, RMSE, MSE, MAE, and MAPE for the considered dataset, and the time series models predict the cases like the observed value than the other models that are compared. The proposed model generated accurate forecasting achieving *R*-Squared value as 99% for ARIMA. Hence, we have performed the future forecasting of the new cases in the upcoming month using the time series model ARIMA. For future works, it is intended (i) Deep Neural Network approaches, (ii) to perform data augmentation to deal with small data samples, and (iii) to tune hyper-parameters using multi-objective optimization for forecasting models compared.

References

1. M.M.R. Majumder, M.I. Hossain, in *[IEEE 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)—Cox's Bazar, Bangladesh (2019.2.7–2019.2.9)] 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)—Limitation of ARIMA in Extremely Collapsed Market: A proposed Method* (2019), pp. 1–5. https://doi.org/10.1109/ECACE.2019.8679216
2. S.M. Idrees, S. Mohammad, M.A. Alam, P. Agarwal, A prediction approach for stock market volatility based on time series data. IEEE Access **7**, 17287–17298 (2019)
3. A.A. Adebiyi, A.O. Adewumi, C.K. Ayo, Stock price prediction using the ARIMA model, in *2014, UKSimAMSS 16th International Conference on Computer Modelling and Simulation*
4. M. Akpinar, N. Yumusak, Forecasting household natural gas consumption with ARIMA model: a case study of removing cycle, in *2013, 7th International Conference on Application of Information and Communication Technologies*. https://doi.org/10.1109/ICAICT.2013.6722753
5. G. Incesu, B. Asikgil, M. Tez, Sales forecasting system for newspaper distribution companies in Turkey, Pak. J. Stat. Oper. Res. **8**(3), 685–699 (2012)
6. G.A. Tularam, T. Saeed, Oil-price forecasting based on various univariate time-series models. Am. J. Oper. Res. **6**, 226–235 (2016).G.E.P. Box, G.M. Jenkins, in *Time Series Analysis: Forecasting and Control*, revised edn. (Holden-Day, USA, 1976)
7. R. Nallathamby, C.R. Rene Robin, D.H. Miriam, Optimizing appointment scheduling for out patients and income analysis for hospitals using big data predictive analytics. J. Ambient Intell. Hum. Comput. **12**, 5783–5795 (2021). https://doi.org/10.1007/s12652-020-02118-4

Multi-regressions Based-Traditional Machine Learning Approach for Energy Forecasting System

S. Balaji and S. Karthik

Abstract Energy monitoring and predicting play a vital role in setting up an industry. The cost spent by the industry for its energy consumption in producing its products acquires an important position in evaluating the profit percentage of the industry. For many decades, the concept of energy prediction is practiced and effective approaches are analyzed based on various parameters that are to be considered. Recently, various machine learning approach provides feasible solution in energy prediction with higher levels of accuracy. In this paper, we discuss about various regression-based energy prediction models for energy consumed by the machineries in the industry level.

Keywords Energy prediction · Regression models · Machine learning algorithm

1 Introduction

Currently, these energy utilization issues are important in the world. In developed countries, the rapid rise in energy consumption is observed. Energy consumption by the machineries in an industry plays a vital role in setting up an industry. So, when energy consumption is not considered as key factor, it may even lead to a great loss for the industry. In this research work, a proposal is made to predict energy requirements of an industry using pre-fetched data from sensors connected to the various machineries and implementing an algorithm through machine learning techniques. This proposal also provides a method for efficient management of energy requirements at the time of demand in energy consumption using various IOT modules.

S. Balaji (✉) · S. Karthik
Department of ECE, College of Engineering and Technology, SRM Institute of Science and Technology, Vadapalani Campus, No. 1, Jawaharlal Nehru Road, Vadapalani, Tamil Nadu, India
e-mail: balajis3@srmist.edu.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_9

2 Literature Survey

Luis M. Candanedo in "Data-driven prediction models of appliance energy usage in a low-energy home," data filtering was used to eliminate non-predictive parameters. Four mathematical models have been used by him for training, (a) multiple linear regression, (b) radial kernel-based supporting vector machine, (c) random forest, (d) gradient boosting machines (GBM), and Luis M. Candanedo found (GBM) as best model with 97% of variance (R^2) in training set and 57% in testing set when compared with other models [1]. ML is commonly used to characterize an algorithm for the machine that learns from existing data. These algorithms usually use a large amount of data and fairly few input features for the process of learning [2]. Blackouts in the past are crucial for the forecast level and future procurement in statistical approaches. There are several references to the Graf-Haubrich method [3]. For each input parameter, a density function is generated using this approach. The density functions are twisted jointly assuming that the parameters are probabilistically unrelated [4–7] suggest a similar approach. Further different implementation constraints for deep learning systems in terms of hardware complexity, power consumption and hardware optimization are discussed in the work [8].

3 Proposed Methodology

This system seems suitable in industries and also in home and made easy to analyze from the Internet. To overcome the problems faced previously, this system gives a clear analysis of the actual and the predicted values over a period of time that is, mostly during the summer seasons, the consumption of power will be very high which is comparatively less in winter seasons, thus we can predict the power consumption that will be required in the winter season approximately so that if there is any form of current looting occurs can be detected easily from the excess value when compared with the predicted value. This system is made even more user friendly that we can monitor the power consumption according to the time period that we allot to monitor the level of consumption and as well as the data will be stored in the Internet that can be used to analyze the rise or fall of the usage of power. If there is a great gap between the power consumption's actual and anticipated value will be plotted as in the form of graph; it shows a error rate from which the variation between the predicted and the real value will be indicated. Thus, this system predicts the overall consumption of energy for a long and short term period of time and stores as data in the Internet from which the analysis can be done. Here, a various regression techniques are applied to the datasets obtained from the IOT models and hardware components which consist of various sensor and wireless units. The block diagram is shown in Fig. 1.

Linear Regression Model

A linear regression model is best suited for predicting the future dataset with the available dataset. This method contributes a good level of accuracy in predicting

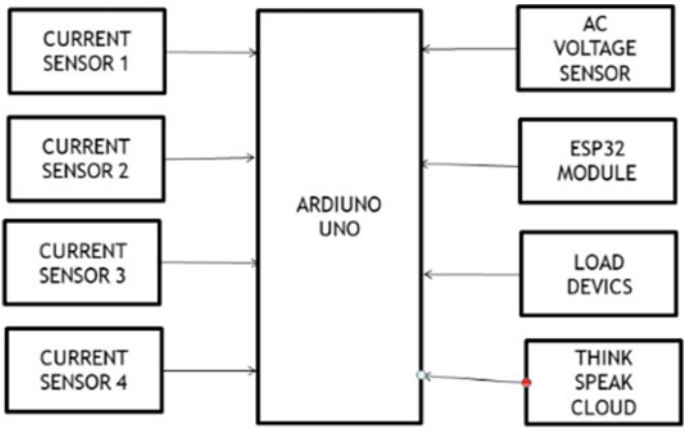


Fig. 1 Block diagram

Energy consumed by the industry. Linear regression is a linear model, a model which assumes a linear relationship between the variables of input (x) and the single output (y). More precisely, they can be determined from a linear input variables (x) combination. The method is referred to as simple linear regression when there is a single input variable (x). In multiple linear regression, the input is multiple variable. Linear regression model output is given in Table 1. Figure 2 shows Prediction versus Actual plot.

$$Y = a + bX, \tag{1}$$

where Y is the output which is dependent variable, X is the input and it is independent variable, b is the slope of the line and a is the y-intercept.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r + \varepsilon. \tag{2}$$

Table 1 Result comparison of performance metrics of four regression models

Performance metrics	Linear regression model	Stepwise linear regression model	Robust linear regression model	Interactions linear regression model
RMSE	46.797	47.557	19.979	56.17
R-squared	0.19	0.16	0.85	– 0.17
MSE	2189.9	2261.6	399.18	3155.1
MAE	15.947	16.01	7.4399	21.395
Prediction speed (obs/s)	560	610	430	470
Training time (s)	0.59744	3.8912	7.7368	0.80768

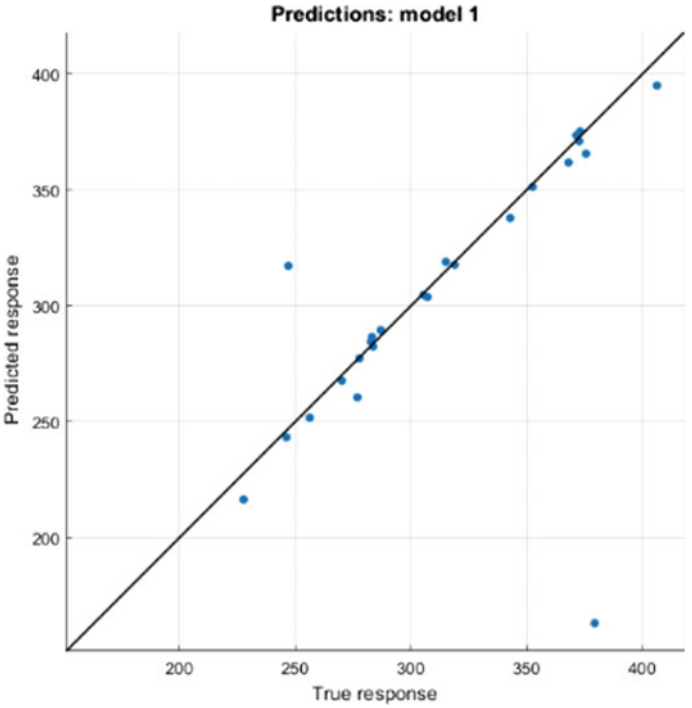


Fig. 2 Prediction versus actual plot

Stepwise Linear Regression

A way of performing regressing several variables is done in stepwise linear regression model, at the same time as simultaneously eliminating those that are not crucial. Thereby stepwise regression effectively does a set of multiple regressions, each time removing the weakest associated component. The key approaches are forward selection, that includes beginning the model with no variables in it, checking each variable addition using a specified criterion-based model fit, inserting the variable, which then included provides the majority of statistically important improvement in fitting, and then the process is repeated if none of the options enhance the model statistically significantly. Figure 3 shows the flowchart of stepwise linear regression. Here, Table 1 gives step-by-step linear model outcomes.

$$b_{j.\text{std}} = b_j \left(\frac{s_{x_j}}{s_y} \right) \tag{3}$$

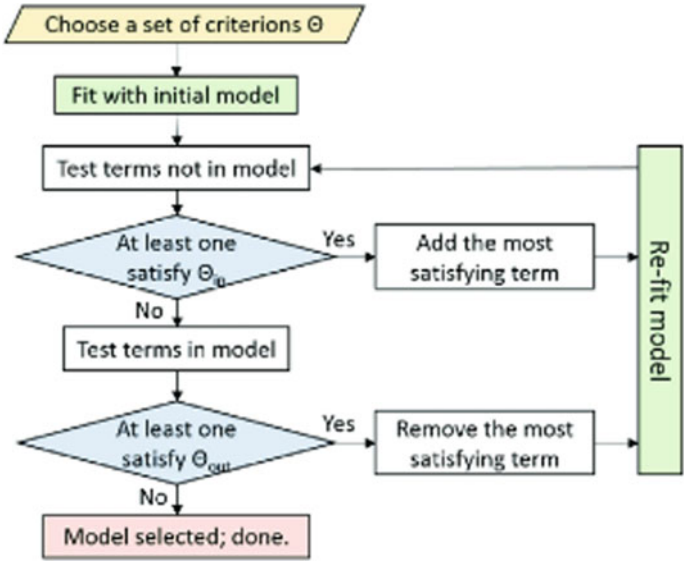


Fig. 3 Flowchart

There are some other models existing, which were grouped under supervised learning methods like artificial neural network or *k*-nearest neighbor and classification algorithms, which can be used to estimate capacity for balancing (ANN) [7].

Robust Linear Regression

For least square regression, the substitute is robust regression, here the influential observations or outliers pollute the data, and influential observations can be identified using this. When least squares regression is in use, for certain situation, we can also use robust regression there. When fitting least squares of regression, we may find some data points or outliers which is at higher leverage. We determined those points of data are not errors due to entry of data, or they from a population other than most of our data. We have decided that these data points are not mistakes in the data entry nor are they from a population other than most of our results. And we have no good reason to rule them out of the study. One of the first-class approaches is robust regression, because it is a negotiation among fully removing those points in the study and treating them all in OLS regression equally and also incorporating all the data points.

Measuring the observations differently relies in how well the examinations have been handled, and this is the principle behind robust regression [8, 9]. To describe precisely, it is a sort of regression of the reweighted and weighted least squares. When the effect of an independent variable on a dependent variable changes, depending

on the values of one or more independent variables, an interaction effect exists. Totaling terms of interaction to a model of regression will significantly enhance knowledge of the connections between the model's variables and allow further testing of hypotheses.

4 Result

Here, the value of the power can be calculated from the current and voltage sensors and the data can be stored and monitored in cloud, thus the analysis of the predicted and the actual value will be plotted in a graph which can be used to find the error rate and the overall analysis can be done from the graph, all the data will be transmitted from the Arduino to the ESP32 from which it can pushed to the cloud for analysis. After using different regression models, we now compare them to arrive the best regression models based on dissimilar performance-based metrics such as root mean square error (RMSE), MSE and *R*-squared and MAE.

5 Conclusion

In this work, a methodology of energy prediction is done by implementing traditional machine learning algorithm based on various regression models, and their results are compared and analyzed with various parameters such as root mean square error (RMSE), MSE and *R*-squared and MAE. The goal is to predict the approximate amount of energy required to be consumed according to the seasons in an industry or home from the actual consumption of energy over a length of time for the purpose of lowering the excess consumption when there is more requirement of power consumption so that if there is any current looting that can also be predicted from the energy forecast rate and this can be analyzed through datasets obtained from the Internet of Things connected to the system. From the result comparison Table 1, we can conclude that robust linear regression, Model 3 is best regression model for energy prediction since its RMSE value is very less and *R*-squared value is higher when compared with other three models discussed here.

References

1. L.M. Candanedo, V. Feldheim, D. Deramaix, Data driven prediction models of energy use of appliances in a low-energy house. Energy Build. (2017)
2. S. Seyedzadeh, F.P. Rahimian, I. Glesk, M. Roper, Machine learning for estimation of building energy consumption and performance, a review, in *Visualization in Engineering* (2018)

3. Consentec, H.-J. Haubrich, Gutachten zur Höhe des Regelenergiebedarfs, in *VgbPowertech* (2008) [Online]. Available: http://www.consentec.de/wpcontent/uploads/2011/12/Gutachten-zur-Höhe-des-Regelenergiebedarfes.pdf. Accessed: 25 Aug 2016
4. C. Maurer, S. Krahl, H. Weber, Dimensioning of secondary and tertiary control reserve by probabilistic methods. Eur. Trans. Electr. Power **19**(4), 544–552 (2009)
5. C. Breuer, C. Engelhardt, A. Moser, Expectation-based reserve capacity dimensioning in power systems with an increasing intermittent feed-in, in *International Conference on the European Energy Market, EEM* (2013)
6. D. Jost, M. Speckmann, F. Sandau, R. Schwinn, A new method for day-ahead sizing of control reserve in Germany under a 100% renewable energy sources scenario. Electr. Power Syst. Res. **119**, 485–491 (2015)
7. A. Ohsenbrügge, S. Lehnhoff, Dynamic dimensioning of balancing power design of balancing power, in *23rd International Conference on Electricity Distribution Paper 0309* (2015), pp. 15–18
8. X. Wang, X. Mao, H. Khodaei, A multi-objective home energy management system based on the internet of things and optimization algorithms. J. Build. Eng. **33**, 101603 (2021)
9. Y. Wen, Z. Wang, Y. Zhang, J. Liu, B. Cao, J. Chen, Energy and cost-aware scheduling with batch processing for instance-intensive IoT workflows in clouds. Futur. Gener. Comput. Syst. **101**, 39–50 (2019)

Comparative Study of Various Machine Learning and Deep Learning Techniques for Energy Prediction and Consumption Using IoT Modules

S. Balaji and S. Karthik

Abstract Due to the rise in worldwide energy consumption, prediction of energy in the buildings which has recently become a critical subject of cost effectiveness and energy conservation (Almalaq and Zhang in IEEE Access 7:1520–1531, 2018 [1]). The insecurity of energy supply poses a threat to the systems' energy efficiency and long-term viability. To plan the use of the captured energy, well-defined power management systems are required. Because of the complexities of the demand patterns that are available, predicting future electricity demand is a difficult task (Kang et al. in Appl Sci MDPI J 10:7241, 2020 [2]). Power management modules are used by certain intelligent IoT node networks to identify the frequency and voltage of each component of their hardware in actual environments. As per the predicted energy harvest. End-users rely on businesses that supply commercialized electrical energy in steady mode and safe electricity. As a result, developing effective models for prediction will be a critical stage to plan electronic power system operations (Kang et al. in Appl Sci MDPI J 10:7241, 2020 [2]). Because of its relevance in decreasing energy waste, the energy management system for building (BEMS) have become a topic of hot nowadays. However, due to issues such as low forecast accuracy. Energy usage forecasting is one of the application of BEMS' which is remained as static (Shapi et al. in Dev Built Environ J 5:100037, 2021 [3]). Serious problem about lowering the energy usage in building requires using the effective prediction mechanism for correctly estimating energy in the future consumption (Almalaq and Zhang in IEEE Access 7:1520–1531, 2018 [1]). Here in the research article, we have performed a detailed comparative study of various machine learning and deep learning techniques for energy prediction and consumption using IoT modules and their performance with real-time data sets and historical-based data sets.

Keywords Harvesting of energy · Fusion based multi-algorithm · Internet of the Things · Evolutionary computation consumption of energy · Machine learning genetic algorithms · Recurrent neural networks

S. Balaji (✉) · S. Karthik
Department of ECE, College of Engineering and Technology, SRM Institute of Science and Technology, Vadapalani Campus, No. 1, Jawaharlal Nehru Road, Vadapalani, Tamil Nadu, India
e-mail: balajis3@srmist.edu.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_10

1 Introduction

Energy efficiency and user comfort have become increasingly important in recent years as a result of excessive electrical energy waste in residential buildings [4]. Forecasting energy demand is essential for planning power generation upgrades and preparing for normal operations [2]. In an intelligent power management system, predicting electric energy consumption (EECP) is a critical and difficult problem. The EECP plays an important role in developing a national energy development policy [5]. The basis of proper power management is accurate harvested energy forecast of energy harvesting Internet of Things (IoT) nodes, which should be low overhead [6]. As a result, several academics have been working on developing reliable energy forecast methods for energy harvesting wireless sensors and IoT nodes [6]. Weather conditions, geographical location, structural features, occupancy, and other factors all influence energy consumption, making it impossible to estimate a building's energy usage exactly. Because of the increased energy demand induced due to massive increase in world's population [4]. Energy usage projection is a challenge for energy producing firms.

2 Related Works

Shapi [3] have attempted to solve the challenges in this study by creating a prediction energy usage model in machine learning platform which is based on Microsoft Azure cloud. The SVM approach yielded the best result, as it was the best method for two tenants, Tenant A1 has RMSE values of 4.7506789 and Tenant 2 has RMSE values of 3.5898263. Additionally, the SVM result for Tenant 1 and Tenant 2 reveals a lower mean absolute error of 12.09 and 43.97, respectively, despite the fact that a lower RMSE result for these two tenants was observed with *K*NN. When consumption average was determined from demand, SVM forecasted demand with greater accuracy, achieving a lower MAPE than the other approaches for all renters. The conceptual framework using this algorithm has taken 13–18 h for training, whereas the *K*NN method, which performed lesser than SVM, have taken only 40 s for training. The advantage of cloud-based predictive model building is that it is not reliant on the performance of the hardware it runs on. The time it takes to run the algorithm is a restriction of this study. Therefore, to implement SVM algorithm, a very powerful platform or computer should be employed.

Xie et al. [6] have suggested fusion framework based new multi-algorithm, combines the findings of numerous forecasting algorithms which improves accuracy. Solar radiation predictor is a fusion type with three algorithms being implemented. Real-time implementation with data of solar radiation demonstrates that it reduces the level of percentage forecast error around 10–26% for varying forecasting intervals. Its difficulty shows that it is lower enough for it to be execute in real time on embedded systems. Here, they have presented a new (HEP) mechanism for IoT

nodes based on multi-algorithm fusion. Its a framework that combines the predictions of multiple algorithms into a single final result. Their approach of prediction for Wireless sensor network (WSN) nodes is more accurate than standard prediction methods while maintaining a cheap cost. The combination of algorithms focused on several factors improves effectiveness and statistical correctness. On lightweight systems, it can run on the board. Furthermore, it is unaffected by the number and kind of core algorithms. This framework is both migrateable and scalable. The weighted average of all the forecast outcomes yields the fusion result. (1) Algorithm of the exponential-based weighted moving average (EWMA) model, (2) Model with a weather-conditioned moving average (WCMA), (3) solar energy forecasting algorithm (SEPAD) is an additive decomposition based solar energy forecasting algorithm. MAF also provides a more accurate estimate. Its accuracy beats any of the fundamental algorithms because to the use of multiple models fusion. Furthermore, the underlying algorithms are scalable and replaceable. This technology has also been used to construct a solar energy harvesting prediction solution. Experiments with this forecaster it reveal that the suggested approach can enhance accurateness around 10–26% on variety of situations. Subsequently, this will save occurrence of loss of power and decrease the battery limit prerequisite or stretch the forecast span. It has demonstrated force efficient and pertinent to constant expectation on the IoT hubs.

Almalaq et al. [1], this research offers a novel hybrid forecasting model it collaborates long- short-term memory with genetic algorithm. Data sets from public building are taken for implementing prediction. Here, the output reveals that the proposed system preforms excellently well. One of these attempts was with deep neural network topologies and DL algorithms, which yielded a promising prediction result. By changing the number of hidden neurons and window size prediction as well as reviewing two hidden layers, the suggested methodology integrates the GA with the LSTM method. The system used for forecasting was implemented on two varieties of public databases of commercial and residential structures. The suggested model outperformed traditional prediction approaches such as ARIMA, decision tree, *K*NN, GA-ANN, MLP, and LSTM when compared. For the residential building example study, the best percentage of decrease in comparison to the normal LSTM is 17.319%, and for the commercial building case study, it was 10.669%. This research provides a creative mixture prediction strategy based on the developmental deep learning strategy, which combines hereditary calculation with extended momentary memory while streamlining target work with time window slacks and the institution's hidden neurons. By extending the window size forecast and number of covered up neurons, as well as multiple secret layers, the suggested strategy joins LSTM with GA technology. Two public datasets of private and company structures were used to test the expectation framework. Over the investigated regular forecast techniques like decision tree, ARIMA, GA-ANN, *K*NN, LSTM, and MLP, the proposed model introduced preferable exhibition. For the private structure contextual analysis, the best degree of reduction in examination with the traditional LSTM is 17.319%, and for the commercial building contextual analysis, it is 10.669%.

Kang et al. [2] this research focuses into the effectiveness of fundamental models of deep learning for forecasting power related with electrical, like power consumption, supply ability, and facility ability. Convolutional based neural networks, recurrent based neural networks, as well as a mixture model that incorporates both RNN and CNN model are among the deep learning models that they have developed. To test these models, they have used real-time data from the Korea Power Exchange. The day by day accounts of office limit, supply limit, and energy utilization are included in this data. For the three features forecasting, the CNN based model greatly beats rest of the models, according to the observed measurements. RNN, CNN, and a hybrid model that incorporates both RNN and CNN model were all investigated as deep learning architectures. The CNN-based model greatly excels than hybrid models RNN and in the experiments. We also compared the CNN model's performance to that of SVM and ANN models. For the previous 7 days and the future 1 day, CNN's best R^2 value is 0.992.

Balaji and Karthik [7], in this paper a *K*NN based algorithm was implemented for Electrical energy theft identification and LSTM algorithm was implemented for energy prediction, and it was concluded that LSTM performed well in predicting energy with higher accuracy and *K*NN for successfully identifying abnormality in Energy consumption due to energy theft. NodeMCU module, current sensor and Voltage sensors were being used in measuring the energy consumed by various loads in a given power Socket. From the real-time dataset obtained from the sensors, it is processed and the prediction datas are obtained by implementing LSTM algorithm using deep learning environment tools. The performance metrics includes mean absolute percentage Error, mean absolute error, 3. R^2 all the values of error rate was comparatively low which ensures higher accuracy.

del Real et al. [8], here the application of Deep learning algorithms to estimate demand in energy is investigated in this research. For this purpose, a mixed design is proposed by the authors that combines a "convolutional neural network" with an "artificial neural network" aiming at a goal in maximizing the benefits of both structures. In particular, the proposed approach's overall MAPE metric yields some error of about 1.4934%. This is somewhat ok than the 1.494% obtained with the RTE forecast data. They have compared the proposed solution. As stated in output segment, this recommended method outperformed other prevailing approaches. The MAPE of "linear regression, regression tree, and support vector regression "(linear) techniques were all above 10%, while the MAPE of support vector regression (polinomic) was 9.2218%, and the MAPE of ARIMA and ANN was slightly less than 3%. Because the proposed structure produced a MAPE of 1.4934%, it leads to decided that the ANN + CNN technique outstrips the previous models.

Dennis van der Meer, et al. [9], here the architecture of a system related to energy management (EMS) proficient of anticipating photo voltaic power generation and optimising the flow of power in-between the photo voltaic system, battery electric vehicles, and the grid at work is presented in this study. The goal is aimed to reduce charging costs while reducing grid energy demand by increasing PV self-consumption and, as a result, increasing BEV sustainability. An "autoregressive integrated moving average model" to estimate Photo Voltaic power output and a "mixed

integer linear programming "method to efficiently allocate power to reduce charging cost make up the developed EMS. The findings suggest that the proposed EMS can greatly cut charging costs while enhancing PV self-consumption and lowering energy usage. EMS greatly decreased total cost by minimizing interchange of energy between the grid and growing self-usage, all while upholding consumer comfort and meeting demand in energy. More specifically, the overall cost was decreased by 118.44%, while profit climbed by 427.45% in the instance of one charging point. The EMS looked excellent in terms of prediction accuracy, with an R^2 of 0.986 and an RMSE of 0.395 kW. Byrne et al. [10], there is a tremendous growth in the implementation of energy based on large-scale storage devices in recent years.

Le et al. [5] this research presents the EECP-CBL model, which uses a mix in "Bi-directional Long Short-Term Memory" and "Convolutional Neural Networks" to evaluate the usage of electric energy. The first module with wo CNNs in this system retrieves significant data's of numerous variable quantity of data set related to private house electric power grid consumption (IHEPC). The values collected through Bi-LSTM based module is transmitted to the final module, which is made up of two fully connected layers and will estimate future electric energy usage. The experiments were carried out to evaluate the suggested model's prediction performance to that of existing models for the IHEPC dataset with numerous modifications. On numerous versions of the IHEPC dataset, the practical output shows that the EECP-CBL framework works well when compared with other models in many aspects of performance metrics for electric energy prediction in, short term, real time, medium term, and finally even long-term time span.

3 Conclusion

From this above proven results of various research publication that were discussed here, we could ensure the importance of energy consumption and prediction in various aspects. For energy prediction many algorithm's where implemented which is compared in Table 1. In many instance, we could see a very promising results are obtained when energy prediction implemented using deep learning tools. When analyzing the best algorithm for energy prediction, we could conclude that there are many advantages and disadvantages in different algorithm which have been experimented and each algorithm has its own features in predicting accurate results. herefore, we conclude that deep learning algorithm works best for different situation with increased complexity.

Table 1 Detailed comparison of various energy prediction approach

References	Algorithm proposed	Performance metrics	Advantage and findings
[3]	1. SVM 2. ANN 3. *K*-NN	RMSE, NRMSE, and MAPE	1. SVM Higher accuracy *K*NN does with 40 s 2. SVM takes longer training 13–18 h to train
[6]	1. EWMA 2. WCMA 3. (SEPAD) 4. MAF	Power losses	1. MAF also gives a more accurate estimation 2. MAF has low Power loss 33.1
[1]	1. Combining genetic algorithm with long short-term memory	RMSE	1. GA with LSTM better performance than the compared conventional prediction methods
[2]	1. CNN 2. RNN 3. Hybrid model is combined one	R^2, MAE	1. CNN outperforms
[4]	1. ML algorithm	–	1. ML algorithm performs well but the process is complex
[7]	1. *K*NN algorithm 2. LSTM	R^2, MAE	1. *K*NN outperforms for abnormality detection LSTM outperforms for Energy Prediction
[8]	1. Mixed architecture consisting of CNN with ANN 2. RTE	MAE-, MAPE, MBE, MBPE	1. Mixed architecture outperforms RTE
[9]	1. ARIMA model 2. A mixed-integer linear programming (MILP) framework	R^2, RMSE	1. Total cost was reduced by 118.44%, whereas profit was increased by 427.45% 2. Accuracy of the EMS performed to satisfactory level, achieving an R2, of 0.986 and RMSE of 0.395 kW

References

1. A. Almalaq, J.J. Zhang, Evolutionary deep learning-based energy consumption prediction for buildings. IEEE Access **7**, 1520–1531 (2018)
2. T. Kang, D.Y. Lim, H. Tayara, K.T. Chong, Forecasting of power demands using deep learning. Appl. Sci. MDPI J. **10**, 7241 (2020)
3. M.K.M. Shapi, N.A.R., Lilik J. Awalin, Energy consumption prediction by using machine learning for smart building: case study in Malaysia. Dev. Built Environ. J. **5**, 100037 (2021)
4. A.S. Shah, H. Nasir, M. Fayaz, A. Lajis, A. Shah, Review on energy consumption optimization techniques in IoT based smart building environments. Inf. MDPI J. **10**, 108 (2019)

5. T. Le, M.T. Vo, B. Vo, E. Hwang, S. Rho, S. W. Baik, Improving electric energy consumption prediction using CNN and Bi-LSTM. Appl Sci, MDPI J, **9**(20), 4237 (2019)
6. X. Li, N. Xie, Multi-algorithm fusion framework for energy prediction of energy harvesting IoT node and implementation. Chin. J. Electron. **29**(5) (2020)
7. S. Balaji, S. Karthik, Energy prediction system using internet of things, in *6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, Conference Paper*. IEEE
8. A.J. del Real, F. Dorado, J. Durán, Energy demand forecasting using deep learning applications for the French grid. Energies **13**, 2242 (2020)
9. D. van der Meer, G.R.C. Mouli, G.M.-E. Mouli, L.R. Elizondo, P. Bauer, Energy management system with PV power forecast to optimally charge EVs at the workplace. IEEE Trans. Ind. Inf. 2634624 (2016)
10. R.H. Byrne, T.A. Nguyen, D.A. Copp, B.R. Chalamala, I. Gyuk, Energy management and optimization methods for grid energy storage systems. Battery energy storage and management systems. IEEE Access **6**, 13231–13260 (2017)

Emulating SoCs for Accelerating Pre-Si Validation

S. Karthik, K. Priyadarsini, E. Poovannan, and S. Balaji

Abstract Verification cycle is exceeding the design cycle these days, due to drastic increase in the complexity of the design of SOCs. Hence, we need to look after new methods that will accelerate the verification cycle and also improve the quality of test. The main objective of our paper is to demonstrate one such verification methodology that gives out better quality of test than the conventional validation methods and also validates the design in much accelerated pace. The important goals which were set before starting the project were early RTL Checkout with real devices using the SV test content. This would also eventually improve the quality of test and reduction in the time taken in Pre-Si validation for at least some of the IP's in a SoC. The aim of this paper is to resourcefully carry out simulation on numerous industrial benchmark circuits on a zybo board containing Xilinx Zynq System on Chip with an emphasis on attaining lesser simulation time. Zybo board a heterogeneous platform (PS-PL) has been utilized in this work for the implementation purpose.

Keywords SoC · System Verilog · Emulation · Profiling · ZynQ

1 Introduction

As design teams move from the HW IP development phase into the full SoC Integration verification phase, efficiency of validation becomes a critical issue. For example, let's consider a SoC that is specifically targeted at a video application. During the SoC full chip verification, the verification team will need to verify that the SoC can

S. Karthik · E. Poovannan (✉) · S. Balaji
Department of ECE, College of Engineering and Technology SRM Institute of Science and Technology, Vadapalani Campus, Chennai, India
e-mail: pe1314@srmist.edu.in

S. Karthik
e-mail: karthiks1@srmist.edu.in

K. Priyadarsini
Department of Data Science and Business System, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Chennai, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_11

properly handle a full frame of video data—ideally in a matter of minutes versus waiting for days of simulation to complete. Hardware-assisted speedup in test bench execution becomes compelling under these circumstances. However, acceleration becomes even more compelling when it is accomplished without sacrificing other important aspects and techniques of a comprehensive functional verification flow, such as constrained random, assertion-based verification techniques, and coverage driven. According to several industry reports, there is a need for research to develop new methodologies to be used in pre-silicon validation [1–3] in order to cope with the complexity of designs, as well as to reduce the time needed to declare product release qualification (PRD). This project is an effort in this direction. This work is a research project which can help reduce the turn-around time in system validation.

2 HW/SW Co-Design

The HW/SW co-design feature in zynq-based boards enables you to prototype any design. The design to be verified is partitioned between the ARM processor and the FPGA fabric of Zynq SoC. In broad means, we use the ARM processor to implement slow-rate functions and you use the FPGA fabric to implement faster functions. When the algorithm is profiled, we can find the functions which runs slower and which one faster. The main goal of hardware/software design requires implementing some part of application in software and other in hardware. So, by use of this approach, we can solve the difficulties faced in logic simulation which take more time for larger designs.

3 Experimental Platform

To increase the performance, speeding up an application [1] is performed by allocating/distributing the load among the processors in a PS-PL platform. The Zybo board is an all-inclusive kit has used for this paper work. The board consists of all the necessary intercommunication interfaces, on-chip RAMS, and auxiliary interfaces to implement an extensive range of applications. The main component on this board is All programmable system on chip (ZynqSoC) [4–7] which contains computational resources supported out by ARM based programmable section and FPGA based programmable logic. They support complex applications that comprise real-time video, DIP, code acceleration, motor control, real-time OS, development of Android/LINUX/Ubuntu, embedded processing, and prototyping.

4 Profiling

Profiling [8–10] is a method used in analyzing a program to help in optimizing software. Profiling tool in XILINX computes execution time for all function present in the C code. It permits you to find in an application, the part of the code that you wrote ingests its time and the corresponding functions which invoked or called the other functions when that gets executed. Performance profilers are software development tools intended to help you analyses the performance of your applications and develop badly performing sections (with respect to time) of your application. They offer measurements of how lengthy a routine takes to execute, how frequently it is called, where it is called from, and how much of total time at some spot is spent executing that routine. The profile output has many parameters. The parameter 'Name' gives the name of the function and the location where it resides. The next one is 'Samples' which tell number of times the timer interrupt has detected the function execution in the program. The third parameter is 'CALL' which provides about number of times calls are made to the function and 'Time/Call' give time usage and '%Time' talk about percentage of time spent in the function.

In this work, the function which consumes extra time will be mapped on to programmable logic section or in lay man words the function will be mapped in the hardware rather than in the PS section.

5 Implementation Flow

The flowchart (see Fig. 1) shows the complete methodology that was followed. First Zynq hardware is created by Xilinx Vivado tool, necessary IPs are called and integrated with the hardware. Next step is to export the created hardware to Xilinx SDK (Software Design Kit). Xilinx SDK is a tool for accelerating an application and developing embedded solutions. In that tool, the board support packages needed are created. Then after creation of BSPs, the application is profiled to find out the candidate to be exported to hardware. Once it is found, custom IP of the function which will be mapped to PL section is created. Profiling results can be found in gmon.out file. The synthesizable function is found and highlighted in the table which is con-verted as IP core.

Table 1 displays the results after profiling Hamming code and pie chart representation of profiling.

6 Simulation Time/Speed up for Benchmark Design

Once the benchmark designs were profiled the synthesizable function which consumed more time were identified and exported to PL section, i.e., hardware part of

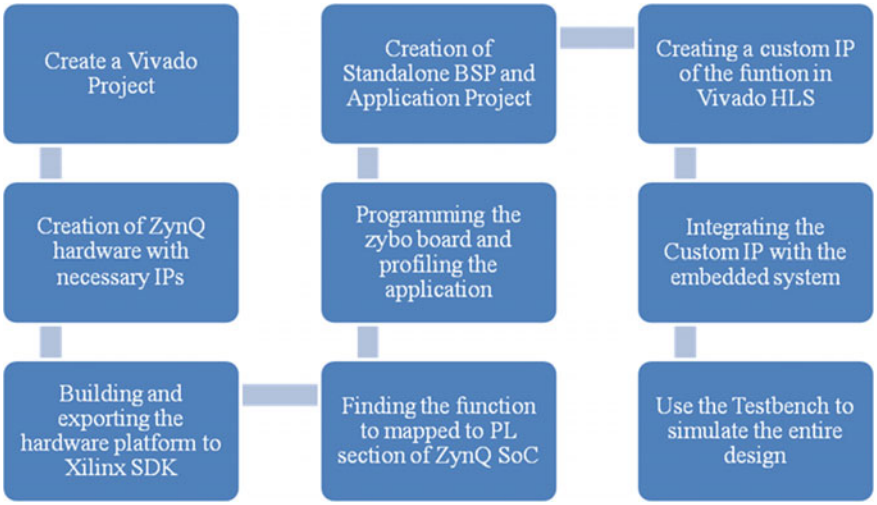


Fig. 1 Implementation flow

Table 1 Profiling parameters on Hamming code

Name	Samples	Calls	Time/call (us)	% Time
	2357			100
xiips_masterend	1227	75	23.5	23.35
H1.c	120	66	35.32	32.44
outtbyte.c	22	20	22.51	10.24
xinttx_l.c	10	3	16.35	5.21
xii_cache.c	3	4	13.46	1.1
xbaasic_type.c	2	5	1.3	1.09

Zynq. With the use of test bench simulation was carried out on Vivado High level synthesis tool. Total simulation time taken of various designs was compared with results obtained in existing methodologies and speedup factor was calculated (Table 2).

7 Results and Discussions

In this paper, an attempt to improve the current verification methods in terms of speed was made. As an example, simulation of standard hamming code took 12,420 ns and using the hardware/software co-design approach and profiling we could achieve speed up factor 1.7227. Through the results obtained, one can obviously observe that we succeeded in that attempt. A novel method of verification that can accelerate the

Table 2 Speed up factors of various bench mark designs

Application	Simulation time taken in normal simulation	Simulation time taken using Hw/Sw Co design approach	Speed up factor
Hamming code	12,420	7211	1.7227
Arithmetic and logic unit	8762	4235	2.069
Advanced encryption standard	23,441	12,301	1.904
Peripheral component interconnect	34,451	24,221	1.4212

speed of verification and also improve the quality of verification was devised. The beauty of this implementation setup was that it can be used for pre-si verifications. Hence, a solution to ever increasing complexity in the verification methodologies is found. This emulation technique aims at accelerated speed, iterative test runs, verification, and interaction with the real-world entities, even before the silicon is up. The speed up factor for the benchmark circuit obviously show the success of our methodology.

References

1. S. Karthik, K. Priyadarsini, V. Jeanshilpa, VLSI systems for simultaneous in logic simulation, in *International Conference on Recent Trends in Electrical, Control and Communication (RTECC)* (2018), pp. 23–27. https://doi.org/10.1109/RTECC.2018.8625665
2. W.K. Lam, *Hardware Design Verification: Simulation and Formal Method-Based Approaches*, 1st edn. (Prentice Hall PTR, USA, 2008)
3. S. Karthik, S. Saravana Kumar, A review on parallel logic simulation. Int. J. Pure Appl. Math. **114**(12), 191–199 (2017)
4. M.J. Sarmah, R.S. Pandey, System monitoring using the Zynq-7000 AP SoC processing system with the XADC AXI Interface. XAPP1182 (v1.0) November 18 (2013)
5. N. Kai, R. Nishinohara, H. Koide, A SIMD parallelization method for an application for LSI logic simulation, in *41st International Conference on Parallel Processing Workshops* (2012), pp. 375–381. https://doi.org/10.1109/ICPPW.2012.54
6. C. Tropper, Parallel discrete-event simulation applications. J. Parallel. Distrib. Comput. **62**(3), 327–335 (2002)
7. M. Rieger, Application specific integrated circuits (ASICs), in *The Electronic Design Automation Handbook*, ed. by D. Jansen (Springer, Boston, MA, 2003). https://doi.org/10.1007/978-0-387-73543-6_16
8. T.B. Ahmad, M. Ciesielski, Parallel multi-core verilog HDL simulation using domain partitioning, in *IEEE Computer Society Annual Symposium on VLSI* (2014), pp. 619–624. https://doi.org/10.1109/ISVLSI.2014.47
9. T.B. Ahmad, M. Ciesielski, An approach to multi-core functional gate-level simulation minimizing synchronization and communication overheads, in *Microprocessor Test and Verification Conference* (2013)

10. Z. Yuxuan, W. Tingcun, K. Yaowen, F. Xiaoya, Z. Meng, Z. Lili, Logic simulation acceleration based on GPU, in *Proceedings of the 18th International Conference Mixed Design of Integrated Circuits and Systems—MIXDES* (2011), pp. 608–613

Optimization of Power Flow in DC Microgrid Connected to Electric Vehicle Charging Station

Sumant Sarmokadam and Ribu Mathew

Abstract The load on the power grid due to electric charging station is very difficult to predict. The management of load on charging station during peak load condition is very important task to maintain power and voltage stability of the system. The charging infrastructure plays important role for electric vehicles (EVs) promotion. The effective way to manage excess power demand of EVs on charging station during peak load condition is through hybrid power system. This paper discusses the scheme to optimize power flow for DC microgrid connected to charging station. The proposed hybrid system is simulated in MATLAB® Simulink. In simulation, the power flow management for different load conditions on charging station is discussed and voltage across the DC grid is maintained within permissible limit. The simulation result shows that DC microgrid voltage variation is less than ± 0.5% during different load conditions, which addresses the issues related to system voltage stability.

Keywords Charging station · DC microgrid · Energy management · Electric vehicle

1 Introduction

Several countries are trying to encourage the improvement in transportation electrification, which address the number of serious issues such as green gas emission, increasing cost of fossil fuels [1]. The EVs whose powertrain consist of electric motor with energy storage system (ESS) is considered to be one of the effective methods to minimize greenhouse gas emissions and oil dependence.

S. Sarmokadam (✉) · R. Mathew
School of Electrical and Electronics Engineering, VIT Bhopal University, Bhopal, India
e-mail: sumant.sarmokadam2019@vitbhopal.ac.in

S. Sarmokadam
St. Vincent Pallotti College of Engineering Technology, Nagpur, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_12

The increase in adaptability of EVs can have substantial effect on the power distribution system. The use of photovoltaic (PV) system to feed EV is one of the effective ways of charging. The combination of pollution-free EVs with minimum carbon PV power generation can address issues of greenhouse gases produced due to internal combustion engine vehicle (ICEV). Two methods of EVs charging using PV system—first one is PV grid and second one is standalone—are widely discussed in [2]. The intermittent and irregular power output from PV system make it difficult to manage the power flow in grid. In some of the research work, PV generation into grid is considered to be partial. The basic problem for PV systems is to synchronize with intermittent energy production to manage dynamic power demand of the existing system. To address such issues, the ESS along with intermittent power generation sources play a very effective role [3]. The large number of EVs causes adverse effects on distribution network, which causes large voltage dip and overloading of components connected to the system, mainly during charging of large numbers of EVs at a same instant—normally during peak load period. In [4], the improvement in penetrations is possible in the domestic networks by monitoring the charging rate of individual electric vehicles connected to charging poles, with small advancement or changes in present residential network infrastructure. The ESS helps the system to manage peak demand of the system, which causes reduction in the system running cost. The renewable energy sources (RES) along with coordinated operation of ESS help the grid to maintain balance between supply and demand. The hybrid system plays important role in managing the peak demand on the system. The major problem with renewable energy sources is its unpredictable nature. Hence, the ESS is becoming important part of hybrid system [5–12].

From the studies, most of the researcher focuses on EV charging station with PV system, ESS and AC grid for energy management. The ESS system used acts as a load during light load condition and sometimes during the peak load condition as well. Differing from the earlier research, new energy management scheme is proposed, in which ESS is used as the input source in both the situations, i.e., light load and peak load condition. This paper presents energy management scheme for grid-connected EV charging station which minimize the load demand and power fluctuation on AC grid. The main aim of control scheme is to reduce EVs charging burden on the existing distribution network. To address this issue, control scheme for effective utilization of RES along with ESS is proposed. The control scheme takes constant power from AC grid and manages peak demand of charging station through PV system and two energy storage systems (ESS1 and ESS2)—ESS1 operates in charging discharging mode and ESS2 acts as a DC source which supplies power to DC microgrid during heavy load condition. AC grid supplies constant power to proposed system only during emergency. So the proposed charging system acts as a constant load on the AC only during emergency condition. It helps the system to address the issues related to power management and power system stability.

2 Energy Management for EV Charging

Normally, EVs draw large amount of energy from charging station during charging, since it acts as a high-power appliance in the power system network. If operation of such high-power appliances is uncoordinated, then it introduces the new power demands, which requires emergency supply system to maintain reliability and stability of power grid. The proper supervision of charging of EVs without affecting the current power system network is one of the important areas of research to increase EV promotion. Following are some possible solutions developed by researchers to reduce influence of EV charging on the public grids:

1. Linear programing (LP) is mostly applied to optimize the EV charging. It increases the aggregator's earnings and decreases EV charging cost. The EV charging using a hierarchical control method for multiple aggregators was recommended to reduce the cost of energy and load on the system during peak period [6].
2. During quick charging of EVs, the charging infrastructure acts like peak load in distribution grid, which effects power system voltage stability. In [7], ESS control scheme is used to stabilize the voltage so that it should not cross the maximum limit or drop below the minimum limit.
3. The impact on charging load is more serious if EV charging process is continuous when grid operates in peak load condition. The swapping of battery during peak load is one of the effective tools to reduce impact on utility grid during peak load conditions [8].
4. The distributed energy generation system normally uses on-site power to supply peak load, which helps to improve the behavior of utility grid during the period of peak load without changing grid capacity. In on-site sources, the PV is one of the most preferable choices. Because of intermittent properties of PV, an ESS is used to balance the PV sources to obtain constant power output [9, 10].
5. In [11], grid to vehicle (G2V) and vehicle to grid (V2G) control strategy are discussed to support grids with high penetration of renewable energy.

3 System Description

The proposed system considered for energy management consists of DC microgrid, PV array, AC grid, two energy storage systems (ESS1 and ESS2) and EVs charging station as shown in Fig. 1. The AC grid connected through AC–DC converter which supplies constant power to the DC microgrid during emergency period only. The power taken from the AC grid is constant. The change in power demand due to load variation is managed through PV system, ESS1, EES2 and AC grid.

The DC grid operates in isolation, which reduces burden on main grid and helps to maintain system stability. The MPPT is used to obtain maximum power (P_{EV}) from PV array. The EES1 operates in both charging and discharging mode, it charges

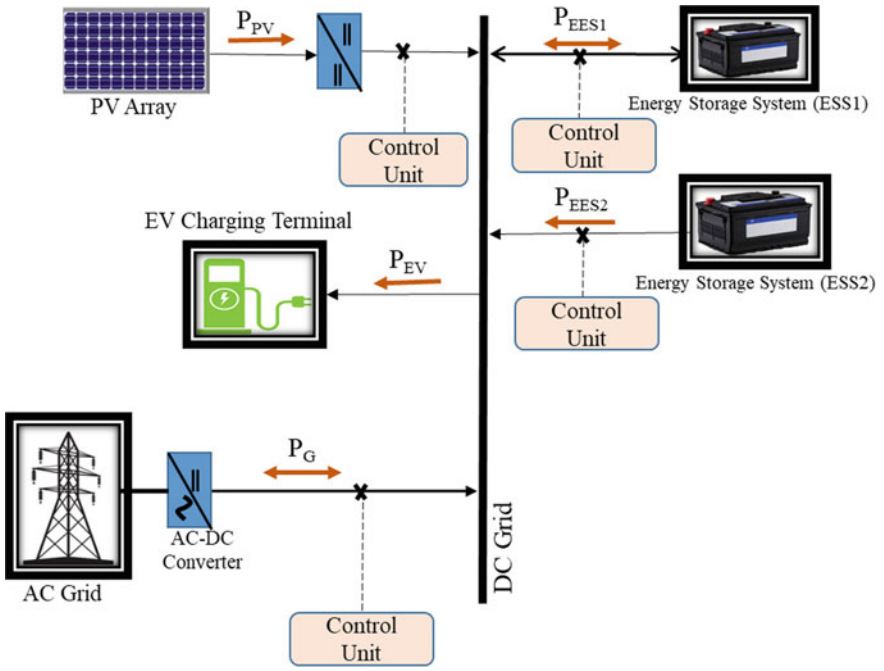


Fig. 1 Proposed DC microgrid system

during light load condition and discharges during peak load condition. The ESS2 is assumed to be fully charged and it acts as a backup energy source during peak load condition.

The EES2 supplies power to the DC microgrid when load on the system exceeds the PV and ESS1 rated capacity. The parameters considered for simulation are given in Table 1.

For balanced operation of system, generation and demand should be same, i.e.,

$$P_{\mathrm{PV}} + P_G + P_{\mathrm{ESS2}} = P_{\mathrm{EV}} + P_{\mathrm{ESS1}} \tag{1}$$

Table 1 System parameters for simulation

Parameter	Values
Bus voltage	24 V
PV power	213 W
Charging terminal power	120 W
ESS1 rating	10 Ah
ESS2 rating	10 Ah

where P_{PV} is PV output power, P_G is grid power, P_{ESS2} is ESS2 output power. The grid supplies constant power to the charging station only during emergency condition and assists to maintain stability and voltage profile.

4 Result and Discussion

The simulation is carried out for the proposed control scheme in MATLAB Simulink. The different load conditions under study are given in Table 2. The ESS1 is charged through PV array when load on the charging station is low and discharged when load on the system is moderate and high. ESS2 acts as DC source when load on the system is high along with PV and ESS1.

The proposed control scheme is verified for different load variation on the charging station. The simulation results are discussed with different power scenario. The power demand on the charging station is changed purposefully to generate different scenarios. The line losses are not considered during simulation.

Figure 2 shows the power output of PV array. The power output of PV array is 213 W assuming constant for all load conditions given in Table 2. During light load condition, PV array supplies power to charge ESS1 as well as to charging station. In this case, no power is taken from the grid, also ESS2 maintains the initial charge.

Table 2 Different load conditions on DC microgrid

Load on charging station	Status of PV system	Mode of ESS1	Mode of ESS2	Status of AC grid
Light	On	Charge	Off	Off
Moderate	On	Charge	Off	Off
High	On	Discharge	Discharge	Off

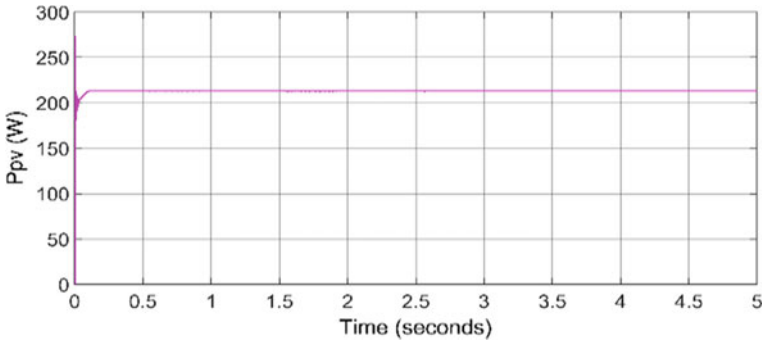


Fig. 2 PV output power

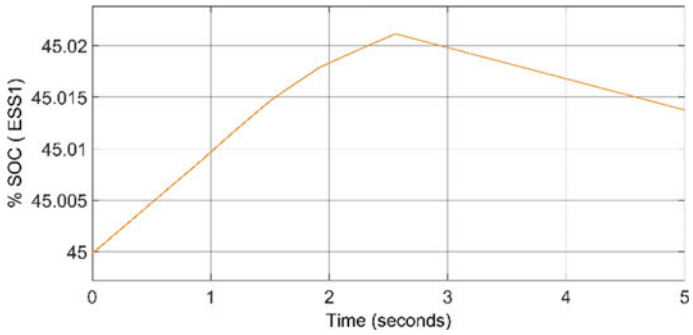


Fig. 3 % SoC of ESS1

The *% SoC* of ESS1 is shown in Fig. 3. When the load on charging station is light and moderate, PV output power is sufficient to manage ESS1 charging and EVs load on charging station. The load on the station is light up to $t = 0.9$ s., moderate from i = 0.9 s. to $t = 2.6$ s. and high after $t = 2.6$ s. At moderate load, ESS1 charges at lower rate. When the load on the system is high and PV output power is deficient to supply load on charging station, then additional power requirement is taken care by ESS1. Hence, at $t = 2.6$ s, ESS1 operates in discharging mode.

Figure 4 shows the load power variation at charging station. Initially, load on the charging station is 120 W which is managed by PV array. The load on the charging station varies at moderate rate from $t = 0.9$ s. to $t = 1.9$ s., during this period, PV array along with ESS1 manages the load on the system. When the load on the system is high, i.e., 240 W, then ESS2 along with PV array and ESS1 starts suppling power to the load. The grid acts as a constant input source for DC microgrid only during emergency condition. In this simulation, emergency condition is not considered. Load on charging station varies from 120 to 240 W. The average power on the charging station is 170 W.

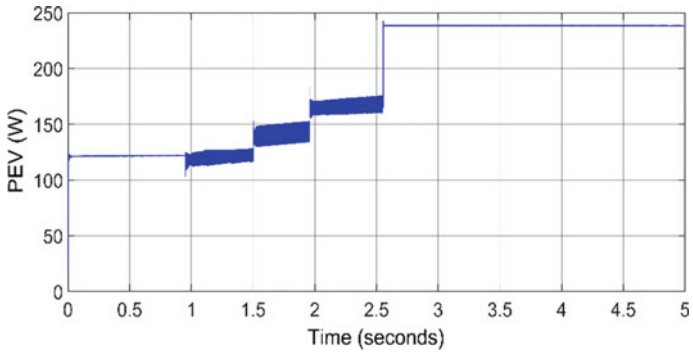


Fig. 4 Load on charging station (P_{EV})

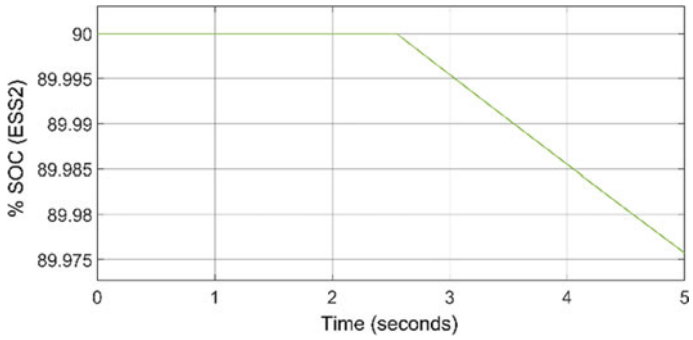


Fig. 5 % SoC ESS2

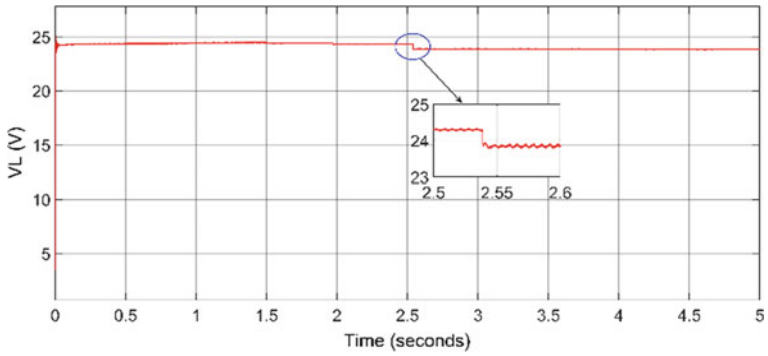


Fig. 6 DC microgrid voltage

The *% SoC* of EES2 is shown in Fig. 5. The ESS2 acts as a DC input source for microgrid. When the load on charging station is high, then ESS2 operates in discharging mode and supplies required power to the load. At $t = 2.6$ s, load on system is high and ESS2 starts discharging.

The voltage variation at DC grid is shown in Fig. 6. The operating voltage of DC grid is 24 V. When load changes, the voltage across DC microgrid almost remains constant due to effective optimum utilization of available energy sources.

From the waveform, it is clear that the mean value of voltage at DC microgrid during complete simulation period is 24.11 V and is almost equal to rated operating voltage of DC microgrid.

5 Conclusion

This paper presents control scheme to optimize power flow in DC microgrid connected to EVs charging station. The DC microgrid consists of renewable energy

sources, energy storage systems, AC grid and EV charging station. The proposed DC microgrid is simulated for different load conditions on charging station. The control scheme optimizes the power flow between PV array and energy storage system in such a way that DC microgrid draws power from AC grid only during emergency load condition. The impact of minimum load demand as well as load fluctuations is reduce on AC grid that helps to maintain system stability. Control scheme also helps to maintain constant voltage at DC microgrid. The percentage voltage fluctuation at DC microgrid is less than $\pm$ 0.5%.

References

1. K. Qian, C. Zhou, M. Allan, Y. Yuan, Modelling of load demand due to EV battery charging in distribution systems, IEEE Trans. Power Syst. **26**(2), (2011)
2. A.R. Bhatti, Z. Salam, K.P. Yee, A critical review of electric vehicle charging using solar photovoltaic. Int. J. Energy Res. **40**, 439–461, (2016)
3. Y. Riffonneau, S. Bacha, F. Barruel, S. Ploix, Optimal power flow management for grid connected PV systems with batteries. IEEE Trans. Sustain. Energy **2**(3), (2011)
4. P. Richardson, D. Flynn, A. Keane, Optimal charging of electric vehicles in low-voltage distribution systems. IEEE Trans. Power Syst. **27**(1), (2012)
5. M.O. Badawy, Y. Sozer, Power flow management of a grid-tied PV-battery powered fast electric vehicle charging station, in *Proceeding IEEE Energy Conversion Congress and Exposition* (2015), pp. 4959–4966
6. L. Yao, W.H. Lim, T. Shih Tsai, A real-time charging scheme for demand response in electric vehicle parking station. IEEE Trans. Smart Grid **8**(1), (2017)
7. D.-J. Kim, K.-S. Ryu, B. Kim, Optimal operation strategy of ESS for EV charging infrastructure for voltage stabilization in a secondary feeder of a distribution system. Energies **13**, 179 (2020)
8. D. Wang, F. Locment, M. Sechilariu, Modelling, simulation, and management strategy of an electric vehicle charging station based on a DC microgrid. Appl. Sci. **10**, 2053 (2020).https://doi.org/10.3390/app10062053
9. H.N.T. Nguyen, C. Zhang, J. Zhang, Dynamic demand control of electric vehicles to support power grid with high penetration level of renewable energy. IEEE Trans. Transp. Electrification **2**(1), (2016)
10. M.O. Badawy, Y. Sozer, Power flow management of a grid tied PV-battery system for electric vehicles charging. IEEE Trans. Ind. Appl. **53**(2), (2017)
11. H.N.T. Nguyen, C. Zhang, Md.A. Mahmud, Optimal coordination of G2V and V2G to support power grids with high penetration of renewable energy. IEEE Trans. Transp. Electrification **1**(2), (2015)
12. R.R. Deshmukh, M.S. Ballal, An energy management scheme for grid connected EVs charging stations (2018). 978-1-5386-2462-3/18©2018. IEEE

Static Hand Gesture Prediction Using Inception V3

S. B. Anusha, G. H. Samyama Gunjal, and N. S. Manjushree

Abstract Sign language is a language for communication that uses the structure of hand, facial expressions, and body language to communicate, and this is a means of communication for deaf and dumb through the use of the hands and other parts of the body. American Sign Language (ASL), which consists of static hand gestures, is one of the widely used sign languages around the world. There arises a communication barrier, when hearing people who do not understand the sign language wants to communicate with deaf and dumb people, which is a major problem in the present society. Hence, this work is focused on developing the static hand gesture recognition system, which acts as a bridge of communication between hearing people and deaf and dumb people. In the proposed work, deep learning and artificial neural network concept are used, since it provides a wide range of algorithms that can recognize objects by learning the features of the object. The proposed work is to predict the static hand gestures of ASL alphabets using Inception v3 model, which uses transfer learning to train the model.

Keywords First American Sign Language (ASL) · Inception v3 · Transfer learning

S. B. Anusha (✉) · G. H. Samyama Gunjal · N. S. Manjushree
Department of CSE, UVCE, Bangalore University, Bengaluru, India
e-mail: anushasb98@gmail.com

G. H. Samyama Gunjal
e-mail: samyamagh@uvce.ac.in

N. S. Manjushree
e-mail: manju123ns@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_13

1 Introduction

Sign language is a type of language which is mostly used by deaf and dumb community, and the language uses hand structures, facial expressions, and body language for communication. People who have hearing problems frequently use sign language instead of spoken languages to convey their message. Deaf people can use sign languages more easily than spoken languages. The communication barrier arises when hearing people who do not know the sign language want to communicate with the deaf and dumb people.

Signs and gestures are considered as the most natural and common way to communicate between people through body movements. Although signs and gestures are classified as non-verbal communication, they can effectively convey the communication messages among hard to hear and hearing-impaired people. The sign language is the most common way to express the words or vocabularies for deaf and dumb people.

American Sign Language (ASL) is one of the most commonly used sign languages around the world. American Sign Language provides the deaf community a way to interact inside the community itself as well as to the exterior world. ASL is largely used by deaf and dumb people, who cannot hear and speak. The fundamentals of ASL are the ASL alphabet so that using the alphabet, the people can communicate letter by letter, by signing of names, place names, age, numbers, dates, years, and words.

The proposed work is focused on developing the static hand gesture recognition system by using ASL dataset, which consists images of hand gestures of alphabets from A to Z at various conditions like different brightness level, different background, different zooming level, and at various positions of hand gesture like top right, top left, middle, bottom right, bottom left corners of image. This static hand gesture recognition system acts as a bridge of communication between hearing people and deaf and dumb people.

Deep learning and artificial neural network concepts are used in proposed work, since it has a wide range of applications in the field of image processing. Artificial intelligence which consists of multiple branches where deep learning is one among them, which highlights the concept of multi-layered perceptron learning. Deep learning and artificial neural network concepts have a wide range of algorithms in the field of image processing, with the help of these algorithms, it is possible to build a system that can recognize objects by learning the features of the images.

The proposed work is to predict the static hand gestures of ASL alphabets using Inception v3 model, which uses transfer learning to train the model. Inception v3 is convolution neural network architecture from inception family, which is one of the popular models for image processing and classification, since it has a deeper network with auxiliary classifier as regularizer which improves the convergence of very deep neural networks. Compared to the other Inception models such as v1 and

v2, Inception v3 will give higher accuracy and also computationally less expensive. The Inception v3 model has symmetrical and asymmetrical building blocks. The model is consisting of convolutions, activation layers, and fully connected layers.

The goal of the proposed work is to train an algorithm that allows to classify images of various hand gestures and signs such as thumbs up and finger count. In the proposed system, the Inception v3 model which helps to classify the images by training dataset, the dataset has different classes of static hand gestures, using deep learning Inception v3 network, and the model gives better accuracy on training data, using the different approaches of deep learning like data augmentation and transfer learning.

2 Literature Review

Gesture recognition is one of the active research techniques, and there are various approaches proposed on gesture recognition on different sign languages.

Aly et al. [1], the convolutional neural network architecture is used for feature learning of American Sign Language (ASL) alphabet images, and the depth images are captured using low-cost Microsoft Kinect sensors. The local features are extracted using the unsupervised principal component analysis network (PCANet), and for the classification process, the support vector machine (SVM) is used.

Dima and Ahmed [2] have proposed the YOLOv5 model as a solution to find the letters and numbers of each provided gesture. The YOLOv5 model provides excellent accuracy as it is lightweight and fast. The model uses a benchmark dataset (MU HandImages ASL) to train and test the model.

Adithya and Rajesh [3], the deep learning convolution neural network (CNN) architecture is used for the gesture recognition of the hand, which is the primary component of sign language vocabulary. The two publicly available datasets are used and compared to know which dataset achieved the better recognition accuracies. The computational burden is reduced with the classical approaches that eliminate the need for manual detection and segment.

Das et al. [4] have used a skilled artificial intelligence tool, convolutional neural network (CNN) for the recognition of native American Sign Language (ASL) by using ASL images. The ASL dataset of 1815 images which are 26 English alphabets was used to train and validate the model.

Hasan et al. [5] have proposed a deep convolutional neural network (DCNN) model for recognition of sign language alphabets. To train the model, MINST ASL dataset is used, which are available in Kaggle.

Avola et al. [6], the recursive neural network (RNN) is used for the recognition of sign language. The leap motion controller (LMC) sensor is used to extract the features. The features and angles formed by the human finger bones are formed

using RNN. The massive range of gestures described through the ASL which defines the effectiveness of the selected angles.

In [7], the convolutional neural networks method is used for the ASL alphabet recognition with inference fusion and multiview augmentation, and the more perspective views of the original data are generated by the augmented approach that makes the training more powerful and decreases the overfitting of the model.

Chung et al. [8], the multiple hand gestures are recognized using deep learning convolutional neural network method, and the images which are used to train are resized into 100×120 for region of interest (ROI). The two CNN architectures are used for the study, they are AlexNet and VGGNet.

Lee et al. [9], the long short-term memory recurrent neural network (LTMRNN) with K-nearest neighbor (KNN) method is used for recognize the hand gesture. The K-nearest neighbor method is one of the classification method that classifies the images.

Nagi et al. [10], the deep learning neural network (NN) including the convolutions and max pooling technique of supervised feature learning is used for classification and feature learning for the dataset images, and the images are captured using colored gloves. The noisy edges are eliminates by color segmentation and smoothed by morphological image processing.

Wang et al. [11] used deep convolutional neural network to create a hand gesture recognition application. The depth images are trained, and the images are captured using Microsoft Kinect depth camera.

Pigou et al. [12] use the convolutional neural network (CNN) model used to train Italian gestures. The model is able to generalize about users and the environment. The system uses Microsoft Kinect to recognize the gesture.

De Smedt et al. [13], the hand gesture recognition system is developed for 3D dynamic hand poses. The dynamic hand pose information is then trained using the convolutional neural network. The hand skeletal data is extracted effectively with the use of hand kinematic descriptors from the number of gestures which are there in datasets.

Dhall et al. [14], the hand gesture recognition is done using multi-layer convolutional neural network. It is an efficient model to use on ImageNet dataset with adequate image preprocessing.

Hossain et al. [15], the Bengali sign language hand images are recognized using convolutional neural network. The hand bone features are extracted and classified into the standard communication alphabet.

Szegedy et al. [16], the classification and detection using standard ImageNet dataset. The concept of inception model is based on convolutional neural network architecture. The 22 layer deep neural network is called as GoogleNet, the quality of which is assessed in the context to classification and detection.

Mitra and Acharya [17] show the survey for gesture recognition methods. The survey compares the different models which used for gesture recognition. The application involves in hidden Markov models, and the efficiency is calculated by varying particle filters, condensation of layers, skin color, optical flow, and connection of the model.

3 Static Hand Gesture Recognition System

Deep learning model for static hand gesture prediction is implemented using neural network architecture, using deep learning models, it is able to pull the required and useful information from the input images Sajetha and Gunjal [18], Kamalapurkar and GH [19]. Deep learning automates the process of extracting features from metadata by preprocessing it through series of various layers VM et al. [20]. The proposed system utilizes the Inception v3 model, which uses the transfer learning method to train and test the ASL dataset images.

Inception v3 is a convolutional neural network from inception family that has deep network of 48 layers, through which image analysis and object detection are done. Inception v3 is the third version of Google's inception convolution neural network, which is pre-trained on ImageNet database. This pre-trained network is trained with ASL dataset, and the trained model predicts the hand gestures of alphabets.

3.1 ASL Data

American Sign Language (ASL) is the natural language which as the same linguistic property as spoken language. It is widely used by deaf and dumb community for their communication purpose.

The ASL data required for the proposed static hand gesture system is taken from Kaggle, which provides repositories of various dataset. ASL data consists hand gesture images of 29 classes, which are divided into 26 alphabets of sign language and 3 additional signs, they are SPACE, DELETE, and NOTHING. The total number of images is 87,000 which are of 200 $\times$ 200 pixels. Figure 1 shows some images of ASL data. The ASL data is classified into different classes based on their texture of finger, position, and structure of the figures held.

3.2 Implementation

Inception v3 is the supervised learning model introduced by Google, which is mainly used for image processing using the transfer learning technique. The data preprocessing step of the Inception v3 model will reshape and rescale the dataset images before training the datasets. Inception v3 is flexible of preprocessing which standardize the image by cropping and resizing.

Inception v3 model is pre-trained with standard ImageNet data, through which the model learns how to extract the particular feature from the provided ASL alphabet images, transfer learning method along with data augmentation technique is used to create the deep learning Inception v3 model for ASL dataset.

Fig. 1 Images in dataset

Transfer learning is the technique which gain the knowledge while solving the problem and apply it for related problem, which helps to attain the greater accuracy without the need of extensive training and computational power. Data augmentation is the technique which uses to train the model for various real-world scenarios; in the proposed work, the dataset images are augmented by shift zooming out up to 120% and brightness shift up to 20% darker lighting condition.

The total number of ASL images used are 87,000 among which 78,300 images are used for training, and 8700 images are used for validation, in order to make the model suitable for the experiment out of 311 layers, the first 248 layers are locked, the last 2 inception blocks are left for training the model, fully connected layers at the top of the inception network are removed, then the own set of fully connected layers are created and added after the inception network. Inception network used in the proposed system consists of 29 softmax units and 1024 ReLU units for the prediction of 29 classes.

Firstly, the model is trained with predefined ImageNet dataset, then the model is able to retrain the final layer, so that it retain the knowledge that inception model had learned during the training of ImageNet data and then apply to ASL dataset to train and test the network. The architecture for the proposed Inception v3 model is shown in the Fig. 2 that includes convolutions, different blocks, max pooling, concat, activation layer, and fully connected layer.

Convolution determines the pixel value for the hand gesture feature of images which is calculated using Eq. 1. The 1×1, 3×3, 5×5 convolution filters used in the proposed Inception v3, where 1×1 convolution filter in the Inception v3 model will reduce the dimensions of data passing through the network, and it learn pattern across depth of the images and enable the network to learn more, and the 3×3 and

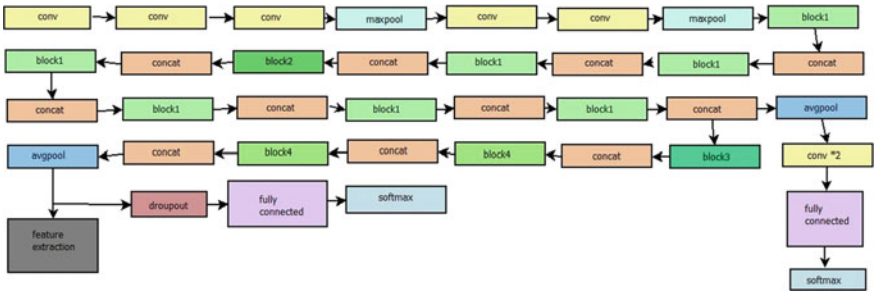


Fig. 2 Block diagram of Inception v3

5×5 convolution filter in the network allow to learn spatial patterns at different scales.

$$z^t = h^{t-1} * W^t \tag{1}$$

where

W is learnable parameter

z^t is pre-activation output of layer l

h^t is activation of layer l

$*$ is the discrete convolution operator

Figures 3, 4, 5, and 6 show the detailed view of each block in Inception v3 model, which includes convolutions, average pooling, max pooling, and concat. Average pooling calculates the average value for patches and down samples the feature map, whereas max pooling selects the maximum element from the region. Average pooling and max pooling are calculated using Eq. 2 and 3, respectively.

$$f_k^{(m)} = \max_{x \in X_k} x \tag{2}$$

$$f_k^{(a)} = \frac{1}{|X_K|} \sum_{x \epsilon X_k} x \tag{3}$$

where

X_k represents N dimensional feature vector

$f_k^{(m)}$ represents max pooling

$f_k^{(a)}$ represents average pooling.

Activation layer decides whether the neuron is activated or not by using activation functions such as ReLU and softmax. The rectified linear activation function (ReLU) is a piecewise linear activation function that will produce the output directly as

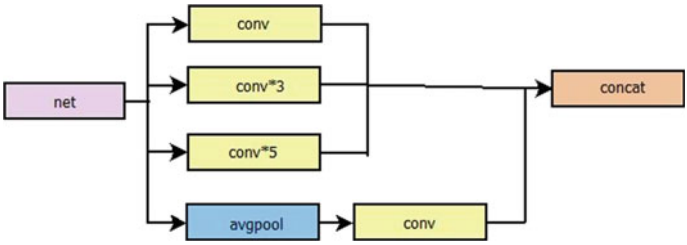


Fig. 3 Details of block 1

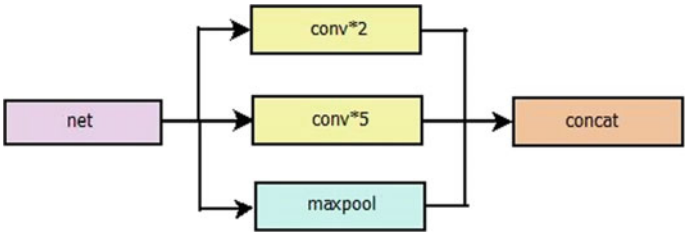


Fig.4 Details of block 2

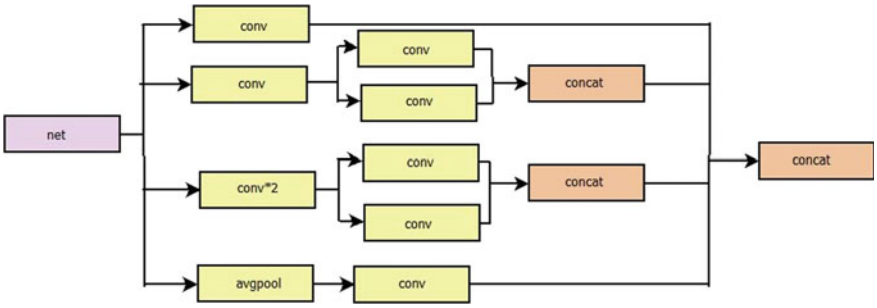


Fig. 5 Details of block 3

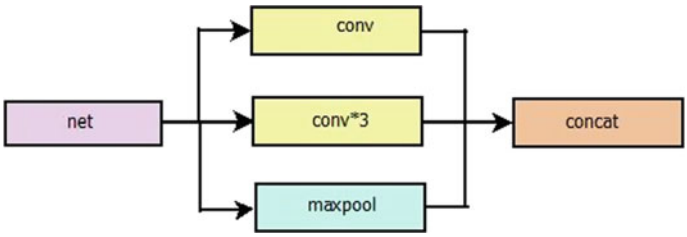


Fig. 6 Details of block 4

positive if the input is positive; otherwise, output will be zero so that the features of the image can be highlighted. Softmax is a activation function presents just before the output layer, which assigns decimal probabilities to each of the 29 classes in a proposed model, these decimal probabilities are added up to 1.0. ReLU and softmax are calculated using Eq. 4 and 5, respectively.

$$ReLU = \max(0, z_i) \tag{4}$$

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \tag{5}$$

where

σ indicates the softmax activation function

z is the input vector

z_i *and* z_j are the standard input and output vectors.

Fully connected layers in a neural networks are those layers where all the inputs from one layer are connected to each activation unit of the next layer. The equation for calculating fully connected layer is provided in Eq. 6.

$$z^t = W^t h^{t-1} \tag{5}$$

where

z^t is pre-activation output of layer l

h^t is activation of layer l

W is learnable parameter

4 Result

The American Sign Language (ASL) dataset is consisting of total 87,000 images in which 78,300 images are trained using the Inception v3 model, and remaining 8700 images are used to validate the model. The training accuracy for the proposed model is 98.87%, the training loss is 0.11, validation accuracy is 95.75%, validation loss is 0.19, and test accuracy is 96.43%. The results of the proposed system are shown in Table 1. The proposed Inception v3 model is trained for total 25 epochs. For training, categorical cross entropy was used to measure the loss along with stochastic gradient descent optimizer with learning rate of 0.0001 and momentum of 0.9 to optimize the model.

Figure 7 shows the plotted result of training and validation accuracy for American Sign Language dataset. The training and validation accuracy are gradually increasing

Table 1 Training and validation results

Metric	Values
Training accuracy	0.9887 (~98.87%)
Training loss	0.1100
Validation accuracy	0.9575 (~95.75%)
Validation loss	0.1926

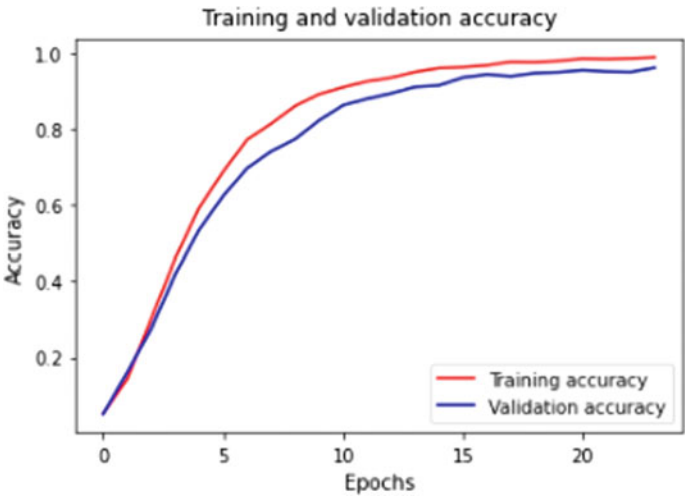


Fig. 7 Training and validation accuracy for ASL dataset

with the epochs. Training accuracy provides accuracy of the Inception v3 model while training the model with the classified training ASL dataset. Validate accuracy provides the accuracy of the trained Inception v3 model with validation dataset.

Figure 8 shows the plotted result of training and validation loss for American Sign Language dataset. The training and validation loss are gradually decreasing with the epochs. Training loss provides how the training ASL data is fitting to the Inception v3 model in each epochs during training, while the validation loss provides how well the trained Inception v3 model is fitted to the validation ASL dataset.

Figure 9 shows the testing result of the Inception v3 model, where the images shows their respective actual and predicted classes of ASL images. The testing accuracy of the proposed Inception v3 model is 96.43%.

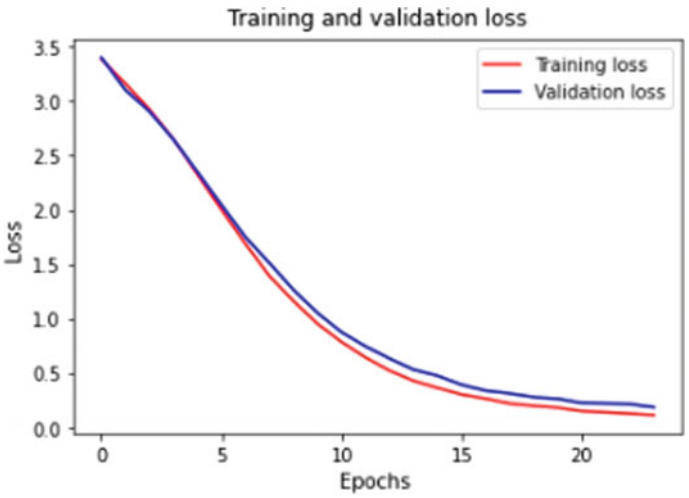


Fig. 8 Training and validation loss for ASL dataset

The Table 2 shows the comparison result of the existing hand gesture recognition system such as PCANet, CNN, and RNN with the proposed Inception v3 model, and it shows that Inception v3 model for static hand gesture recognition system gives better accuracy compare to existing systems.

5 Conclusion

Static hand gesture prediction using Inception v3 model is focused on predicting the hand gestures of American Sign Language (ASL) alphabets using the Inception v3 network. The model uses transfer learning technique and data augmentation technology to look at the effectiveness of Inception v3 model on TensorFlow platform in the prediction of American Sign Language alphabets. Data augmentation method is to expand the present dataset and convey to perform the experiment. Inception v3 is able to classify and detect the American Sign Language alphabet images. In the proposed system, it able to get the training accuracy of 98.87% and test accuracy of 96.43%. The comparative result indicates that the proposed model scored the best accuracy among other models.

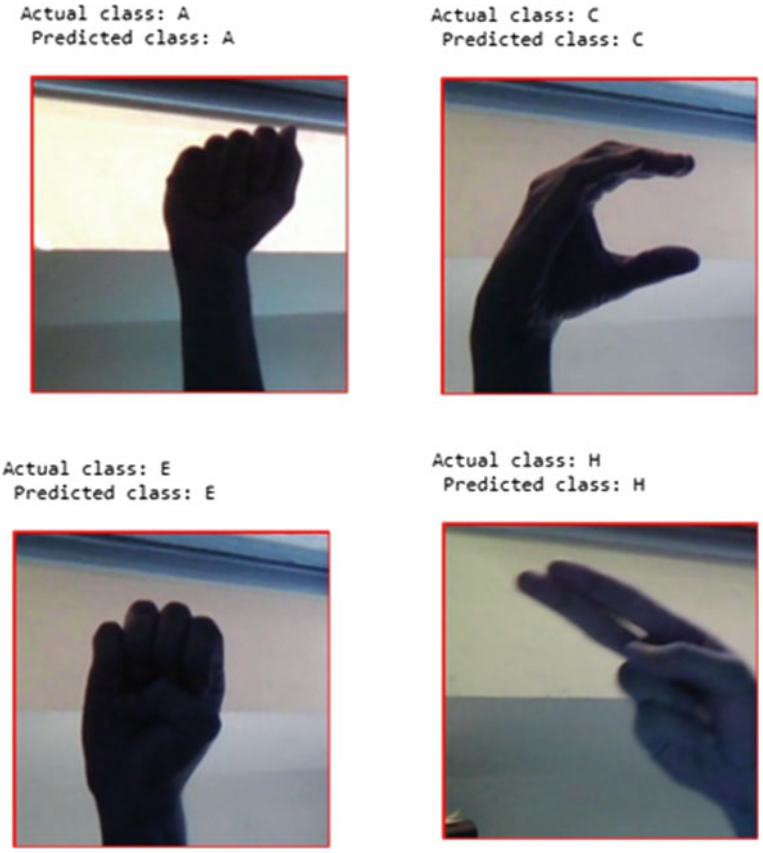


Fig. 9 ASL alphabet predicted result

Table 2 Comparing the proposed result with existing results

Method	Results (%)
PCANet Aly et al. [1]	88.7
CNN Adithya and Rajesh [3]	94.7
RNN Avola et al. [6]	91.43
Proposed system	98.87

References

1. W. Aly, S. Aly, S. Almotairi, User-independent american sign language alphabet recognition based on depth image and pcanet features. IEEE Access **7**, 123138–123150 (2019). https://doi.org/10.1109/ACCESS.2019.2938829
2. T.F. Dima, M.E. Ahmed, Using yolov5 algorithm to detect and recognize american sign language (2021), pp. 603–607. https://doi.org/10.1109/ICIT52682.2021.9491672

3. V. Adithya, R. Rajesh, A deep convolutional neural network approach for static hand gesture recognition. Procedia Comput. Sci. **171**, 2353–2361 (2020). https://www.sciencedirect.com/science/article/pii/S1877050920312473, https://doi.org/10.1016/j.procs.2020.04.255. *Third International Conference on Computing and Network Communications (CoCoNet'19)*
4. P. Das, T. Ahmed, M.F. Ali, Static hand gesture recognition for American sign language using deep convolutional neural network (2020), pp. 1762–1765. https://doi.org/10.1109/TENSYMP50017.2020.9230772
5. M.M. Hasan, A.Y. Srizon, A. Sayeed, M.A.M. Hasan, Classification of sign language characters by applying a deep convolutional neural network (2020), pp. 434–438. https://doi.org/10.1109/ICAICT51780.2020.9333456
6. D. Avola, M. Bernardi, L. Cinque, G.L. Foresti, C. Massaroni, Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures. IEEE Trans. Multimedia **21**, 234–245 (2019). https://doi.org/10.1109/TMM.2018.2856094
7. W. Tao, M.C. Leu, Z. Yin, American sign language alphabet recognition using convolutional neural networks with multiview augmentation and inference fusion. Eng. Appl. Artifi. Intell. **76**, 202–213 (2018). https://www.sciencedirect.com/science/article/pii/S0952197618301921, https://doi.org/10.1016/j.engappai.2018.09.006
8. H.Y. Chung, Y.L. Chung, W.F. Tsai, An efficient hand gesture recognition system based on deep CNN (2019), pp. 853–858. https://doi.org/10.1109/ICIT.2019.8755038
9. C. Lee, K.K. Ng, C.H. Chen, H. Lau, S. Chung, T. Tsoi, American sign language recognition and training method with recurrent neural network. Expert Syst. Appl. **167**, 114403 (2021). https://www.sciencedirect.com/science/article/pii/S0957417420310745, https://doi.org/10.1016/j.eswa.2020.114403
10. J. Nagi, F. Ducatelle, G.A. Di Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, L.M. Gambardella, Max-pooling convolutional neural networks for vision-based hand gesture recognition, in *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (2011). https://doi.org/10.1109/ICSIPA.2011.6144164
11. C. Wang, Z. Liu, S.C. Chan, Superpixel-based hand gesture recognition with kinect depth camera. IEEE Trans. Multimedia **17**, 29–39 (2015). https://doi.org/10.1109/TMM.2014.2374357
12. L. Pigou, S. Dieleman, P.J. Kindermans, B. Schrauwen, Sign language recognition using convolutional neural networks, in *European conference on computer vision* (2015), pp. 572–578
13. Q. De Smedt, H. Wannous, J.P. Vandeborre, Heterogeneous hand gesture recognition using 3d dynamic skeletal data. Comput. Vis. Image Underst. **181**, 60–72 (2019). https://www.sciencedirect.com/science/article/pii/S1077314219300153, https://doi.org/10.1016/j.cviu.2019.01.008
14. I. Dhall, S. Vashisth, G. Aggarwal, Automated hand gesture recognition using a deep convolutional neural network model (2020), pp. 811–816. https://doi.org/10.1109/Confluence47617.2020.9057853
15. S. Hossain, D. Sarma, T. Mittra, M.N. Alam, I. Saha, F.T. Johora, Bengali hand sign gestures recognition using convolutional neural network (2020), pp. 636–641. https://doi.org/10.1109/ICIRCA48905.2020.9183357
16. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions (2015)
17. S. Mitra, T. Acharya, Gesture recognition: a survey. IEEE Trans. Syst. Man, and Cybern. Part C (Appl. Rev.) **37**, 311–324 (2007).https://doi.org/10.1109/TSMCC.2007.893280
18. T. Sajetha, G.S. Gunjal, An approach to face recognition using feed forward neural network. Int. J. Comput. Appl. Technol. Res. (IJCATR) **6**, 172–212 (2017)
19. S. Kamalapurkar, S.G. GH, Online portal for prediction of heart disease using machine learning ensemble method (prhd-ml) (2020), pp. 1–6. https://doi.org/10.1109/B-HTC50970.2020.9297918
20. M. Vm, S.G. Gh, S. Kamalapurkar, Air pollution prediction using machine learning supervised learning approach. Int. J. Sci. Technol. Res. **9**, 118–123 (2020)

Evaluation on Radiation Exposure in Human Eye Tissue Model with IFA

S. Jemima Priyadarshini, J. Immanuel Nargunathan, S. Anitha Christy, and A. Kanimozhi

Abstract The eye is a vital organ in every human in day-to-day life. The human eyes are subjected to exposure to radiation from devices such as mobile phones and some other diagnostic devices. The radiation emitted from these devices can be calculated with a constraint specific absorption rate (SAR). Though these devices are non-ionizing radiation, prolonged exposure can even lead to damage to eye tissue and result in a loss of eyesight. There is a requirement for validation of these emitting radiation on human vital tissues. This paper intends to set up a simulation environment to study these radiation absorptions on human eye tissues. The human eye tissue models of cornea, sclera, vitreous humor, lens cortex, and lens nuclei are designed as tissue models with permittivity and conductivity properties. The prominent communication emitting inverted F antenna (IFA) is placed at an equidistant, and the behavior of these eyes tissues toward absorption is studied. The SAR is computed using Altair FEKO at 900 MHz operating frequency. This paper also investigates the effect of tissues on the performance of IFA with an analysis of gain and reflection coefficient values.

Keywords Human · Eye tissues · Specific absorption rate · GSM · IFA · Dielectric properties

S. Jemima Priyadarshini (✉)
Immanuel Hospital, Chennai, India
e-mail: mailjemi@gmail.com

J. I. Nargunathan
Sri Balaji Medical College and Hospital, Chennai, India

S. A. Christy
Govt. Stanley Medical College, Chennai, India

A. Kanimozhi
VHS-M.A. Chidambaram College of Nursing, Chennai, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_14

1 Introduction

Radiofrequency can cause adverse alterations due to non-ionizing radiation can induce tumors in various parts of humans. Several studies indicate the human brain is more susceptible to absorptions due to its proximity with its device [1]. The communication devices have been revolutionized with facetime calling and for selfie pictures taken in the upfront face. In both these cases, the devices are placed proximal to the eyes, and human eyes are inevitably under radiation exposure. There is conclusive research finding stating that radiofrequency causes stress in rat's eyes affecting ocular vision. The extended study also designates radiations from radiofrequency devices to have the potential to damage eye tissues [2]. The SAR estimates these radiations that are absorbed by the exposed eye tissues in terms of average or as peak values [3]. The IFA is suggested for its compatibility in fabrication and its widespread use [4].This paper focuses on eye tissues such as the cornea, sclera, vitreous humor, lens sorted, and lens nuclei that are designed as tissue models with dielectric equivalent properties [5].

1.1 Specific Absorption Rate

The SAR is a calculable value for the valuation of energy absorption of a bare exposed surface. When the exposed human surface is laid open to electromagnetic waves, there is a manifestation of absorption due to its field. The depiction of SAR is with associations of power absorption and exposed tissue properties. It is stated in units of watts per kilogram (W/Kg) [6]. The SAR is measured by the unveiled near field E (V/m) by the radiating device, and the general equation is given by the following Eq. (1).

$$SAR = \frac{\sigma |E|}{\rho} (\text{W/Kg}) \tag{1}$$

In the above equation, the E indicates the field near the radiating device (V/m), σ and ρ represent the conductivity (S/m) and mass density (Kg/m^3) of the tissue.

1.2 Inverted F Antenna

Inverted F antenna was initially designed and projected for missile-based communication, but a modified version is used for mobile phone communication. It is a quarter-wave monopole antenna bent with a conductive rod placed on the ground plane as the main arm. The length of the main arm is shorted with an offset point for feed as an alternative to the base feed [7].

Table 1 IFA design values

IFA parameters	Values (mm)
The length and width of planar arm (L1 and W1)	23.8 and 2.5
The width and length of the folded arm (W2 and L2)	3 and 5.3
The width and length of the feed arm (W3 and L3)	0.85 and 5.3
The offset of feed from the top	6.97

2 Tissue Modeling and Experimental Setup

2.1 Modeling Background

There is a stringent restriction for any real-time experiments conducted on human samples with exposure studies. This examination is done on numerical human eye models and their supporting calculations with method of moments (MoM) [8]. MoM method is more ideal for simple boundary conditions and computation efficiency.

2.2 Geometry and Dimensions of IFA

IFA limits backward radiation so contributes to reduced SAR values. The following IFA design values of the planar, folded, and feed arm are designed in CADFEKO. The dimensions are listed in Table 1.

2.3 Dielectric Properties of Eye Tissues

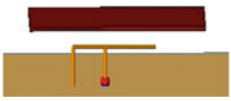
The eye tissue model with a length of 200 mm and width of 100 mm with a height of 0.3 mm is designed, and the dielectric values are applied to the eye model. The tissue dielectric properties are obtained from the biomedical data simulated over a 900 MHz frequency [9]. The conductivity σ, relative permittivity $\varepsilon_{r,}$ and mass density ρ of the tissue over the 900 MHz frequency are applied to the eye tissue structure. The designed tissue equivalent model with dielectric properties is enumerated in Table 2.

Simulation setup. The conduction of SAR simulation is experimented with within an electromagnetic setup using the computational MoM technique [10]. The tissue equivalent is kept within a distance of 0.1 mm from the simulated IFA for SAR assessment. The designed tissue medium is placed above IFA and is set for EM interaction as shown in Fig. 1. The IFA is positioned at free-space condition, and the exposure area field is created for SAR computation**.**

Table 2 Dielectric properties of eye tissues

Tissue	ε_r	σ
Cornea	55.23	1.39
Sclera	55.27	1.16
Vitreous humor	68.90	1.63
Lens cortex	46.57	0.79
Lens nuclei	35.84	0.48

Fig. 1 IFA with eye tissue



3 Results and Validation

3.1 Operation Characteristics

S_{11} or reflection coefficient (return loss) is a ratio of given input power to the output radiated power for a particular frequency represented in dB value. If the value tends to be zero dB, it suggests no power as output radiation is obtained from the IFA antenna. For any antenna simulation S parameters are the basic tester value to check whether the antenna is effective in terms of radiating for the operation range over a particular frequency. In this experiment, the S_{11} is verified for all the eye tissues [11]. The return loss is computed for eye tissues, and S_{11} values are shown in Table 3. The S_{11} characteristics of the cornea are shown in Fig. 2.

The return loss S_{11} values of IFA for all the five eye tissues are obtained and results are shown in Table 3. The values in the table specify that for all eye tissues, the IFA operates optimally, and it can generate an acceptable field characteristic for SAR computation [12].

Table 3 Reflection coefficient

Tissue	S_{11}
Cornea	− 27.2
Sclera	− 26.3
Vitreous humor	− 24.3
Lens cortex	− 36.6
Lens nuclei	− 24.8

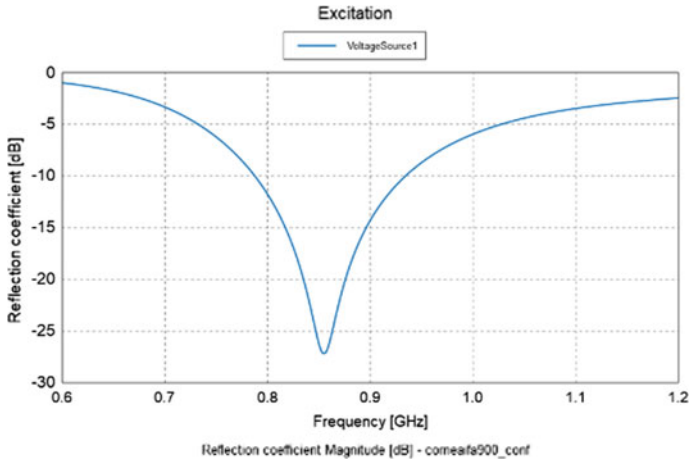


Fig. 2 Return loss of cornea (dB)

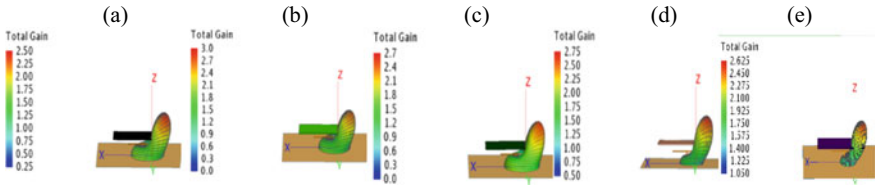


Fig. 3 3D Far field characteristics of **a** cornea, **b** sclera, **c** vitreous humor, **d** lens cortex, **e** lens nuclei

3.2 Far field in 3D

The field analysis of eye tissues is examined with derived gain values. The IFA exhibits radiation up to a particular distance. The maximum distance where its field is revealed is shown graphically as a far field radiation pattern. The far field pattern of eye tissues with total gain in dB is given in Fig. 3.

3.3 Gain and Near Field

The antenna gain measures the capacity of the antenna or device to translate the given input power to the electromagnetic waves for free-space propagation. The gain of IFA with each dielectric eye tissue is measured, and tissue relations with IFA are studied. The general gain of IFA is 3 dB, and the obtained values are compared in Fig. 4. The sclera tissue helps in achieving a maximum gain of 3 dB, whereas in the case of the cornea, the gain of IFA is lowered to a 2.5 dB value. The scattering

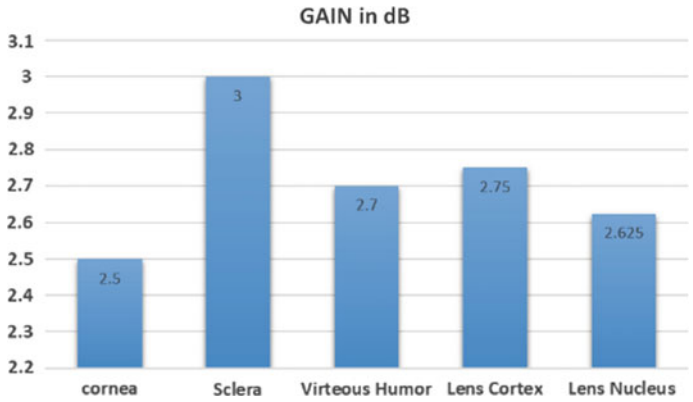


Fig. 4 Gain of the eye tissues (dB)

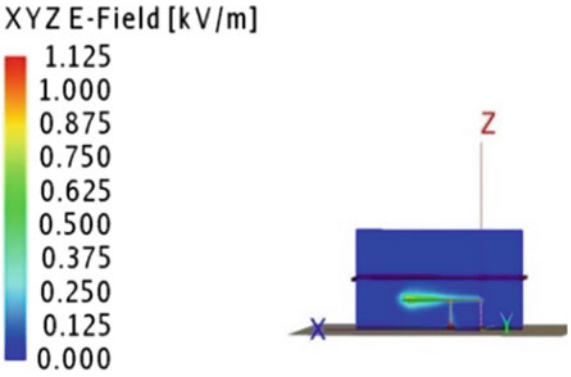


Fig. 5 Near field distribution

around and near the IFA contributes to near field [13]. The estimated near field value is a maximum of 1.125 kV/m and shown in Fig. 5.

3.4 SAR Comparison

The SAR 1 g is estimated from the SAR maximum value that is averaged and acquired until a mass of 1 g is achieved. Similarly, SAR in point to point is the volume which is calculated for an average of 10 g. A cubic mass is favorable for gram computations [14]. In this experiment, linear eye tissue equivalents are used for comparison, and hence, peak SAR value is desirable. The tissues and their behavior are analyzed in this SAR computation displayed in Fig. 6.

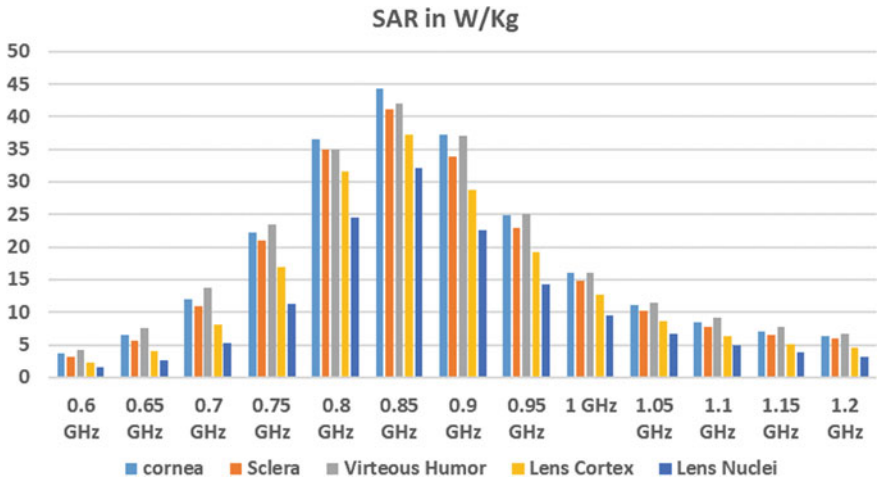


Fig. 6 SAR comparison of eye tissue equivalents (mW/Kg)

Each tissue has different medium characteristics contributing to different SAR values, and cornea medium tends to absorb more radiation 44mW/Kg followed by vitreous humor of 42 mw/Kg. The lens nuclei show resistance to absorption with a minimum SAR value of 32 mW/Kg. The sclera matter absorbs more than the lens cortex but still less than cornea or vitreous humor.

4 Conclusion

This paper aims to research and study SAR impact on five different human eye tissues for IFA at 900 MHz frequency. This work further settles that the eye tissues when positioned with IFA can modify the gain and S_{11} characteristics of IFA. There are individual characteristics of tissues detected, and the difference in dielectric medium properties is applied to the investigational study. The homogenous linear eye model is designed for evaluation using the MoM method in FEKO. The reflection coefficient or return loss of – 24 dB to − 36dB is achieved. The gain is wide-ranging between 2.5 dB and 3dB for the eye tissues. There is a maximum SAR value obtained near 850–900 MHz band with SAR peak values of 44 mW/Kg for cornea and a minimum value of 32 W/Kg for lens nuclei. The future work can be extended to larger cubic mass equivalence for 1 g and 10 g comparison with existing reduction methods [15]. The analysis can also be extended for the wide range of human tissues for detailed investigation in the area of biomedical research.

References

1. J. Zhang, A. Sumich, G. Wang, Acute effects of radiofrequency electromagnetic field emitted by mobile phone on brain function. Bioelectromagnetics **38**(5), 329–338 (2017)
2. P.S. Jemima, D.J. Hemanth, Investigation and reduction methods of specific absorption rate for biomedical applications: a survey. Int. J RF Microw. Comput. Aided Eng. **28**, e21211 (2018)
3. E.D. Eker, B. Arslan, The effect of exposure to 1800 MHz radiofrequency radiation on epidermal growth factor, caspase-3, Hsp27, and p38MAPK gene expressions in the rat eye. Bratisl Lek Listy. **119**(9), 588–592 (2018)
4. A.Y. Simba, T. Hikage, S. Watanabe, T. Nojima, Specific absorption rates of anatomically realistic human models exposed to radiofrequency electromagnetic fields from mobile phones used in elevators. IEEE Trans. Microwave. Theory. Tech. **7**(5), 1250–1259 (2009)
5. O.S. Dautov, E.M. Egypt, Application of FEKO program to the analysis of SAR on human head modeling at 900 and 1800 MHz from a Handset Antenna, in *IEEE International Symposium on Electromagnetic Compatibility and Electromagnetic Ecology* (2005), pp. 274–277
6. IEEE Standard-1528 IEEE Recommended Practice for Determining the Peak Spatial-Average Specific Absorption Rate (SAR) in the Human Head from Wireless Communications Devices. Meas. Tech. (2003)
7. R. King, C. Harrison, D. Denton, Transmission-line missile antennas. IRE Trans. Antennas Propag. **8**(1), 88–90 (1960)
8. S. Caorsi, G.L. Gragnani, M. Pastorino, Use of redundant testing functions in moment-method solutions for block models. IEEE Trans. Microw. Theory Tech. **41**(2), 305–310 (1993)
9. C. Gabriel, The dielectric properties of biological materials, in *Radiofrequency Radiation Standards*, ed. by B.J. Klauenberg, M. Grandolfo, D.N. Erwin. NATO ASI Series 274 (1995)
10. https://www.altair.com/feko
11. M.A. Jensen, Y. Rahmat-Samii, EM interaction of handset antennas and a human in personal communications, in *Proceedings*, vol. 83 (IEEE, 2009), pp.7–17
12. https://www.antenna-theory.com/
13. R. Dolecek, V. Schejbal, Estimation of Antenna gain. IEEE Antennas Propag. Mag. **5**(11), 124–125 (2009)
14. H. Kawai, K. Ito, Simple evaluation method of estimating local average SAR. IEEE Trans. Micro. Theory Tech. **52**(8), 2021–2029
15. J.P. Stephen, D.J. Hemanth, An investigation on specific absorption rate reduction materials with human tissue cube for biomedical applications. Int. J. RF Microw. Comput. Aided Eng. **29**, 21960–21972 (2019)

A Novel Approach for Sentiment Analysis and Opinion Mining on Social Media Tweets

Mulatu Gebeyaw Astarkie, Bhoomeshwar Bala, G. J. Bharat Kumar, Swapna Gangone, and Yagnam Nagesh

Abstract Social media is a powerful tool for people to share their thoughts and feelings. People post their personal feelings and thoughts on any topic, person, policy or product for marketing or social attraction. This text information can be broken down into facts and opinions. Reflect people's feelings about government policies, products, national and international personalities, the constitution, incidents and events. Sentiment analysis is an area of study that analyzes people's opinions, ratings, attitudes and emotions in general from the written language. The texts of the reviews are edited to give an accurate description of the author's opinion on the subject. This paper proposes an analysis of the feelings of the writers about the subject and comments of the users of social networks in the context of Ethiopia with Facebook as a platform. In this paper, natural language processing (NLP) and machine learning (ML) based on a novel sentiment analysis and opinion mining framework used to conduct sentiment analysis and opinion mining of posts and user comments on social media via a Facebook application.

Keywords Sentiment analysis · Opinion mining · Classification · Emotions

1 Introduction

Social media is any digital platform that enables users to quickly create content and share it with the public. Facebook, Twitter, LinkedIn, Instagram, Snapchat, YouTube are some of the most popular social media platforms. It is a powerful tool for people to share their thoughts and feelings. It is a huge and growing source of texts ranging from

M. G. Astarkie · B. Bala (✉) · Y. Nagesh
Department of Information Technology, Debre Tabor University, Debre Tabor, Ethiopia
e-mail: drbhoomi08@gmail.com; drbhoom08@dtu.edu.et

Y. Nagesh
e-mail: nageshyagnam1@dtu.edu.et

G. J. Bharat Kumar · S. Gangone
Department of Computer Science, Mettu University, Mettu, Ethiopia

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_15

everyday observations to engaging discussions. People post their personal feelings and thoughts on any topic, person, policy or product for marketing or social taste. Such posts often receive hundreds or thousands of comments and it becomes difficult for a reader to read all of them in order to gauge public opinion. Sometimes you just want to know the opinion, behavior, trend or general thought being discussed there, or to determine whether those opinions are positive or negative For example, in the case of product marketing, the company wants to assess the success of an advertising campaign or the introduction of a new product or which products or services are popular and what people like or dislike about the special features of a product. Individuals or organizations want to hear the public's opinion on any government policy, law or situation in the country. In such situations, automated sentiment analysis and polling can help a lot.

Sentiment analysis is the machine learning process of analyzing text (social media data, news articles, emails, etc.) to assess the polarity of opinions (positive to negative mood) and emotion, topic, tone, etc. The spread of social networks makes it possible to share opinions about different aspects of life, and the millions of messages appear on the web. A basic goal is to classify the text as a positive or negative emotion. We believe that by their very nature, it is possible to more accurately classify emotions in Facebook messages. Facebook messages are more concise than ratings and easier to classify. The ability to include more characters allows for better writing and more accurate rendering of emotions on your part. Sentiment analysis is an area of study that analyzes people's opinions, ratings, attitudes and emotions in general from the written language. The texts of the reviews are processed to provide an accurate description of the author's opinion on the subject. Sentiment analysis is one of the most active research areas in natural language processing and social media mining as well as text/multimedia data mining. The growing importance of sentiment analysis coincides with the popularity of social media platforms such as Facebook, Twitter and Instagram. It is the most powerful social network in the world. Facebook provides a unique opportunity to share and share your ideas, feelings and opinions with the public, organizations and government in ways that your website and blog cannot even reach.

Therefore, a novel sentiment analysis and opinion mining framework are proposed in this research paper, which uses different natural language processing (NLP) and machine learning (ML) techniques to adjust the sentiment of the user's opinion about different entities measure up. The proposed framework is used to conduct sentiment analysis and the extraction of opinions from user posts and comments on social networks through a Facebook application.

2 Literature Review

An extensive literature search was carried out to analyze the gap in sentiment analysis approaches used by various authors. Some of the peer-reviewed research is presented in this section as follows:

Kumar et al. [1] The sentiment analysis of electronics tweets using the big data framework looked at the various sales tweets that were used to examine consumer sentiments about electronics. The experimental results of the proposed work will be helpful for various companies in making business decisions and further improving product sales. In today's scenario, people produce millions of tweets every year. But dealing with those huge unstructured tweets is not possible through the traditional platform. Hence, the big data framework like Hadoop and Spark is used to ingest this type of big data.

Mathapati et al. [2] "Sentiment Analysis and Opinion Mining of Social Networks: A Review" discussed the need for automated analysis techniques to extract sentiments and opinions submitted in user comments. The words provide a detailed analysis of the customer ratings. The tagged data required to train a classifier is expensive, and therefore, the domain adaptation technique is used to overcome it. With this technique, the individual classifier is designed to classify homogeneous and heterogeneous entries from different domains focuses on the collection of existing methods of social media sentiment analysis and opinion research techniques.

Rastogiet al. [3] "A Sentiment Analysis Approach to Facebook User Recommendations" discussed the system of offering new friends who have similar interests but have different opinions. The motivation for this work is that users have similar interests but have different opinions about it. In this thesis, a user recommendation technique based on a novel weighting function is proposed, which not only takes into account the interests of the users, but also their feelings. This feature allows building a better recommendation process than other content-based approaches. First results of a comparative analysis show the advantages of the novel approach compared to some modern user recommendation systems.

Gupta et al. [4] "Sentiment Analysis of Twitter and Facebook Data with MapReduce" discussed the fun Twitter and Facebook data source for opinion extraction or sentiment analysis, and this large amount of data can be used to find women's feelings about a particular topic or product. A system is proposed that involves collecting data from the social network using the Twitter and Facebook APIs. Big data challenges are then answered with Hadoop through the map reduction framework, in which the entire data is mapped and reduced to smaller data for easier handling and finally includes the analysis of the collected data and the presentation of the results through graphics.

Isah et al. [5] "Social Media Analysis for Product Safety Using Text Mining and Sentiment Analysis" discussed user-generated content from social media platforms that can provide early information about allergies, adverse events and counterfeit products. This article represents ongoing work through contributions that include: developing a framework for collecting and analyzing the opinions and experiences of drug and cosmetic product users through machine learning, text mining and analysis of emotions; applying the branding analysis framework projected on Facebook comments and Twitter data and describing how to develop a product safety lexicon and training data to model a machine learning classifier to predict customer sentiment.

Gursoy et al. [6] "Social Media Mining and Sentiment Analysis for Brand Management" discussed that corporations want to benefit more from big data studies.

Although it affects the different dynamics of the company in different areas, social media services have become very important for the marketing and CRM departments of companies. In this way, communication with customers is always recognized and the use of big data in these areas is seen as one of the most important steps for companies to become a great brand. 3 large companies in the manufacturing, technology and food industries in Turkey were examined. The data was obtained with the help of API and web crawler.

Nigam et al. [7] "Investigating online shopping sites through Facebook pages with Graph API and FQL query" discusses posting our details and adding and rejecting people as friends and creating the so-called virtual human social network, ideas, feelings, status, pictures, etc. This document introduces the manual implementation of how to find out the full status of the Facebook account from social networks to get the total number of people on our network, our friends, their records by gender, their status, their identifications, shared photos, such as to see me in contrast to the different reactions of people in our network to our posts, everything can be calculated as a summary with the help of the analysis software R with an additional Facebook R package.

Zamaniet al. [8] "Sentiment Analysis: Determining People's Emotions on Facebook" discussed to identify opinion extraction and sentiment analysis devices to extract both English and Malay words on Facebook. This effort begins with converting unstructured information into meaningful lexicons after Facebook content has been extracted and stored in a database. After physical identification in sentiment analysis, emotions are divided into happy (positive), unhappy (negative) and unemotional. As a case study, a problem in the inspection result is published and the results of the students' answers are resolved. This study is critical to enabling stakeholders such as managers and entrepreneurs to follow disputes to improve their facilities.

Patil and Thakare [9] "Analyzing the Variations in Public Sentiment on Twitter and Facebook" discussed sharing views, ideas, expressions, feelings and opinions on social networking sites such as Twitter and Facebook. In this work, analyzing the variations in public opinion over an explicit period of time toward an explicit goal on both Twitter and Facebook, this type of analysis is useful in several areas in order to draw appropriate conclusions and degrade public opinion.

Salloum et al. [10] "A Survey on Text Mining on Social Media: Facebook and Twitter Insights" discussed a common practice of not writing a sentence with proper grammar and spelling on social media sites, causing various types of uncertainties such as lexical, syntactic and semantic and conditional leads. With this type of insecure data, it is inflexible to find out the actual order of the data. This study aims to describe how social media studies have used text analysis and text mining methods to categorize key topics in information focused on studying text mining studies related to Facebook and Twitter; the two most important social networks in the world.

3 Proposed System

This paper proposes learning sentiment-specific embeddings, dubbed sentiment embeddings, to sentiment analysis. For more powerful continuous word representations, we use the sentiment of text to preserve the effectiveness of word contexts. As illustrated in the Fig. 1 the embedding space captures both sentiment and context evidence. This allows us to identify the closest neighbors that are semantically similar. We also prefer to have the same sentiment polarity so it can distinguish between good and bad. Sentiment embeddings are learned from tweets. We use positive and negative emotions to create pseudo-sentiment labels for sentences with manual annotations. Urban dictionary provides lexical-level sentiment supervision. This is based on a limited list of sentiment seeds that have been manually annotated. We propose to learn sentiment embeddings that encode the sentiment of text in continuous word representation. Sentiment embeddings are learned from tweets that contain positive and negative emotions. We do not need any annotations. Three sentiment analysis tasks are used to verify the effectiveness and efficiency of sentiment embeddings. Empirical experiments show that sentiment embeddings perform better than context-based embeddings when compared to several benchmark datasets.

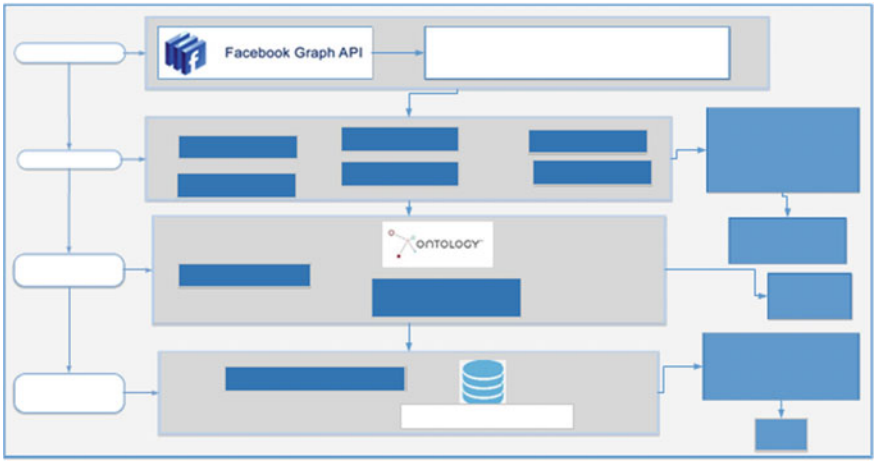


Fig. 1 Proposed system architecture

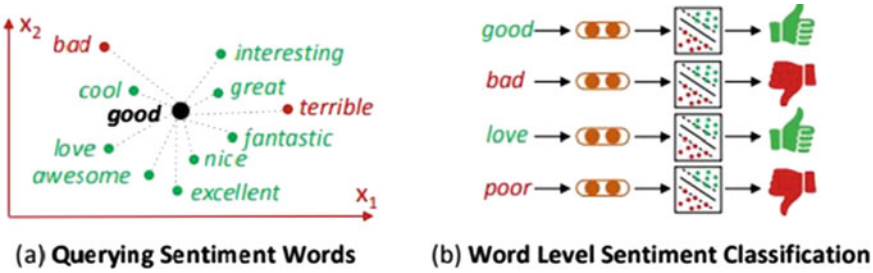


Fig. 2 Design of word-level sentiment analysis of querying sentiment words and word-level sentiment classification

4 Results and Discussions

4.1 Word-Level Sentiment Analysis

We examine whether sentiment embeddings can be used to find similarities among the section's sentiment words. As illustrated in the Fig. 2 two settings are used to test word-level sentiment analysis: querying neighboring sentiment words within embedding space and word-level classification.

4.2 Querying Sentiment Words

Better sentiment embedding will map positive words into close vectors and distinguish between positive and negative words. In the vector space for sentiment embedding, words that are near a positive word such as "good" should be dominant by words like "cool," "awesome" and "great," while words like "bad," "terrible" or "nasty" should surround a negative word. To test whether sentiment embeddings can be used to find similarities between sentiment words, we first query the neighboring sentiment words within the existing sentiment dictionary. We first need to find the N_w nearest words in the sentiment dictionary when we have a sentiment word. The similarity between two words (e.g., The cosine (or the ratio) of their word embeddings is used to measure how close two words are. Then, we calculate the percentage of neighboring words with the same sentiment polarity as our target sentiment word. This percentage of polarity consistency corresponds with an accuracy evaluation measurement, which is as follows.

$$\text{Accuracy} = \frac{\sum_{i=1}^{\#\text{Lex}} \sum_{j=1}^{Nw} \delta\omega\left(\omega_i, c_{ij}\right)}{\#\text{Lex}\ X\ N\omega}$$

where #Lex denotes the number of words in a sentiment lexicon, W_i is the ith term in the dictionary. *Cij* is the closest word to w_i in the lexicon. DW (w_i, *Cij*) is an indicator function with a value of 1 for sentiment polarity. 0 is for the reverse case. Higher accuracy means that sentiment embedding is better at capturing similarities between sentiment words. To illustrate the experimental protocol further, let us take N_w = 10 and "good" as an example. Fig. 3 illustrates this. First, we need to find the ten closest words within the embedding space. These include "cool," "love," "awesome" and "bad." There are eight words, except for "terrible" and "bad" that share the same sentiment polarity as "good" among these ten neighbors. The accuracy of the word "good," therefore, is 80%. Three benchmark sentiment lexicons were used in our experiments: BL-Lexicon (MPQA) and NRC-Lexicon (NRC-Lexicon). Table 1 gives the statistics for each lexicon. To evaluate the effectiveness of sentiment embeddings, we set N_w at 10 and 30 separately. To learn word embeddings, we compare the following embedding algorithms.

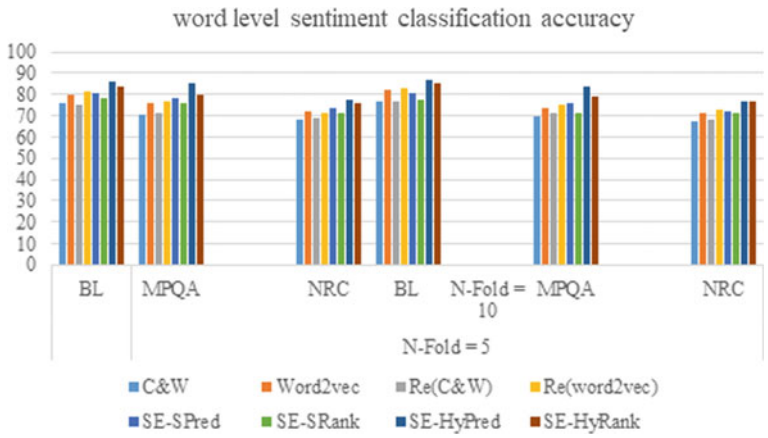


Fig. 3 Word level sentiment classification accuracy

Table 1 Word-level sentiment classification accuracy

Embedding	*N*-fold = 5			*N*-fold = 10		
	BL	MPQA	NRC	BL	MPQA	NRC
C&W	76.2	70.5	68.1	76.6	69.8	67.5
Word2vec	80.1	75.7	71.7	81.8	73.4	71.2
Re(C&W)	75.0	71.3	68.6	76.6	71.4	68.2
Re(word2vec)	81.5	76.9	71.6	83.1	75.3	73.1
SE-SPred	80.8	78.2	73.3	80.4	76.0	72.4
SE-SRank	78.6	76.2	71.3	77.8	71.1	71.2
SE-HyPred	86.0	85.3	77.5	87.0	83.4	76.9
SE-HyRank	83.9	80.1	75.7	85.6	79.2	77.0

C&W Model: The C&W model represents an algorithm for learning word embeddings. It is also the context ranking model. The Word2Vec model is used to extract the notion of relatedness across words or products such as semantic relatedness, synonym detection, concept categorization, selectional preferences, and analogy. Mikolov and colleagues developed CBOW (and SkipGram) to learn continuous word vectors. They then embedded these algorithms in the widely used toolkit "word2vec." CBOW is used in our experiments. It is analogous to our context prediction model. * embed—Lipson and Labutov learn task-specific embeddings using a sentiment-specific corpus and a background embedding. Re (C&W and Re (Word2vec) stands for different word embeddings in background embedding.

Word-level sentiment classification to further examine the effectiveness of sentiment embeddings when capturing similarity between sentiment words, we conduct word-level sentiment classification. To determine whether a sentiment word has a positive or negative meaning, we perform binary polarity classification using three sentiment lexicons (BL), MPQA and NRC. Continuous word embeddings are used to classify a word. N-fold cross-validation is performed for each dataset. Each dataset includes $(N - 1)$ parts that are used as training data and one as test data. As the final evaluation metric, we train the N number of supervised classifiers and average their classification accuracy. Table 1 gives experimental results for $N = 5, N = 10$.

4.3 Sentence-Level Sentiment Classification

Emotional incorporations are used to classify feelings in a supervised learning environment. We use embeddings of feelings instead of manual functions to create the characteristic of a sentence. Sentiment classifiers are constructed from sentences that have been manually annotated with sentiment polarity. Use a semantic composition-based approach. It is a simple idea to create sentence-level features using sentence sentiment inlays. This principle of compositionality says that the meanings of longer expressions (z simple and efficient way to learn compositionality in vector-based semantics [66]). Each grouping layer grouping p uses the operations of embedding and the vector of p to order the order of the columns in each look-up table Z(s) is the concatenation results of various grouping functions. This means that there are no additional parameters in a composition component. Then we can use the constructed sentence representation to build an SVM classification engine and compare it to other modern functions for a fair comparison. You can test whether sentence insertion functions can be built into existing function sets for an additional increase in performance. Other semantic composition models such as the convolution neural network or the recursive neural network can also incorporate feelings.

The emotion-embedding space can make it possible to separate words with similar contexts and opposite mood polarity labels, such as "good" or "bad". To encode sentiment information and context-level information at the same time, we use multiple neural networks to create consistent word embeddings. Three sentiment analysis tasks were used to check the effectiveness of three sentiment embeds. We show that

sentiment embeddings can be used to find similarities between sentiment words in word-level sentiment analysis. Sentiment embeds can be used to predict the sentiment of sentences by capturing discriminatory characteristics. Emotional embedding helps to compare the similarities between words when working on a linguistic level, for example, when building an emotional lexicon. The most powerful models are hybrids, which capture both moods and information from the context.

5 Conclusion

This paper introduces a novel sentiment analysis and sentiment mining framework that uses various natural language processing (NLP) techniques to preprocess textual data from social media users and machine learning (ML), as well as techniques the user's opinion The proposed framework will be used to carry out sentiment analysis and the extraction of opinions from the publications and comments of users on social networks through an application of Facebook: in the future, the proposed framework may be extended to include data in other Ethiopian languages on social media, and we will expand our system to handle multimedia data including video, audio and images.

References

1. G.N. Alemneh, A. Rauber, S. Atnafu, Negation handling for Amharic sentiment classification, in *The Fourth Widening Natural Language Processing Workshop, USA* (2020)
2. A.P. Jain, V.D. Katkar, Sentiment classification for under-resourced language using Word2Vec neural network: amharic language social media text, in *International Conference on Computer Science and Information Technology, Switzerland* (2020)
3. Y.Y. Aklilu, *Exploring Neural Word Embeddings for Amharic Language*, M.Sc. Thesis, Near East University, Turkey (2019)
4. M. Mukhtar, *A Hybrid Sentiment Classification for Amharic Book Reviews*, M.Sc. Thesis, St. Mary University, Addis Ababa (2019)
5. Y. Mekonnen, *Dictionary and Rule Based (Hybrid) Approach of Amharic Sentiment Analysis for KANA TV*, M.Sc. Thesis, Bahirdar university, Bahirdar, Ethiopia (2018)
6. G. Abreham, *Opinion Mining From Amharic Entertainment Texts*, M.Sc. Thesis, Addis Ababa University, Addis Ababa (2014)
7. T. Tilahun, *Linguistic Localization of Opinion Mining from Amharic Blogs*, M.Sc. Thesis, Addis Ababa University, Addis Ababa (2014)
8. A. Alemu, *Sentiment Mining for Discovering Hidden Knowledge in Opinionated Amharic Text Review*, M.Sc. Thesis, University of Gondar, Gondar (2013)
9. M. Belete, *Sentiment Analysis for Amharic opinionated text*, M.Sc. Thesis, Ababa university, Addis Ababa, Ethiopia (2013)
10. G. Selama, *Sentiment Mining Model for Opinionated Amharic Texts*, M.Sc. Thesis, Addis Ababa University, Addis Ababa, Ethiopia (2010)

Cloud Control Cold Storage System

Rajalakshmi Alavanthan, P. Sivakumar, and P. Arokiya Prasad

Abstract The grain storage plays an important role in the economies of both developed and developing countries. But, because of lack of knowledge and awareness on the impact of humidity and temperature on food items; several times, food protection is not well enough preserved. Therefore, automatic grain monitoring can help us increase the grain storage operating levels, reduce grain losses, and decrease labour intensity. In this work, an Internet of Things (IoT)-based cloud is utilized to gather the necessary data, namely, temperature, humidity hot load and weather forecasting report of the specific locations is derived from social open source. It also reduces manual contribution for operation and maintenance and improves the smartness of the cold storage system. Real-time IoT-based temperature and humidity values are obtained through the DHT-11 sensor along with the integration of ESP-8266 Node MCU module. The same has been simulated through suitable software and experimentally evaluated the effectiveness of the IoT-based monitoring system; the same has been compared with the traditional monitoring system.

Keywords Cold storage · Cloud control · IoT · Food security

1 Introduction

India, with a population of over 1.27 billion, is very critical in ensuring food security. Grain production has been gradually growing due to advanced production technologies but inadequate infrastructure facilities and unfavourable environmental conditions result in food grain storage losses. Before the next harvest, the food grains must

R. Alavanthan (✉) · P. Sivakumar · P. Arokiya Prasad
Rajalakshmi Engineering College, St. Joseph College of Engineering, Chennai, India
e-mail: arajalakshmi@gmail.com

P. Sivakumar
e-mail: sivakumar.p@rajalakshmi.edu.in

P. Arokiya Prasad
e-mail: arokiyaprasad@stjoseph.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_16

be processed, because they are the main economy of the country. Insects and mould can infect grains. In Indian agriculture, the main food security problem is inadequate infrastructure and lacks efficient supply chain management. India is one of the largest producers of food grains. Agricultural products contribute about 13.8 per cent to the country's Gross Domestic Product. Indeed, India is one of the countries with significant food shortages. The paper addresses two supply chain management models designed to streamline India's supply chain [1]. First is the ITC E-Choupal supply chain model. With E-Choupal, farmers are constantly informed regarding (worldwide) market price, crop demand, etc. This is done with the aid of Kiosk. The information is gathered from various sources such as Department of Agriculture, Government Colleges, Department of Indian Meteorology, Input Firms and retailers. The weather knowledge allows the farmers to take proper care of the crops.

E-Choupal is effectively used in more than 38,000 villages across various states in India through 6500 kiosks. Another supply chain model is the dependency fresh supply chain system that obtains knowledge from farmers and city processing centres. Collection centres function as intermediates to collect food grains and meet public demand via retail outlets. The centres provide the food grains with a price band that lets the farmers get full profitability. The food grains must be covered against the natural degrading factors such as humidity and temperature. The storage area infrastructure plays a crucial role in quality preservation. The two listed models have better infrastructure for maintaining the quality of the food products. Groundnuts are used for oil processing, and are often eaten directly [2]. Research was conducted to avoid groundnut spoilage. The groundnuts if stored for the long term are infected by the pest such as bruchid beetle. Usually used to preserve food, aluminium phosphate is detrimental to health as it has significant toxic effects on the lungs, heart and blood vessels. Thus it took advantage of alternative storage solutions. Many storage bags such as the lino bag, gunny bag and fertilizer bag were used for grain preservation. The study concluded that by suppressing the C. Serratus insect pest the super grain bag showed better results in the winter season [3].

Food grains are vulnerable to rodent attacks and deterioration due to other environmental factors if stored for long periods under unsuitable conditions. The explanation for rodent attacks is lack of hygienic locations to store the food. Apart from rodent attacks variations in temperature, humidity and concentration of gas lead to food grain spoilage. Models such as Partial Least Square (PLS) and Takagi–Sugeno (TS) are available for monitoring grain quality and predicting how long the food grains will last [3]. The dynamic PLS system with decoupling and reduction of dimensions describes the multi-input multi-output system as a combination of several single input single output system. The nonlinear grain growth process can be effectively managed using the TS fuzzy modelling system. A framework was planned to incorporate both PLS and TS that were designated as the TS-PLS model. The model has predictive error and is efficient as compared to conventional models such as PLS and TS models [4]. A grain storage monitoring and control system was designed in real time [5]. The machine uses the sensor for the humidity and temperature to calculate the food grain consistency. The sensors sense the temperature and humidity and send the data to the administrator via ethernet or wireless devices. Then the administrator takes the

appropriate steps according to the information obtained. The device gives the grain humidity and temperature which starts to decompose when it reaches the threshold value. The system offers efficient reliability and the grain status. Any control system built should be cost efficient and capable of producing accurate results [6].

Mechanical ventilators and mechanical vapour compression grain chillers allow the grains to be stored and cooled down. These mechanical devices consume significant quantities of energy, and are thus not economical. So, solar-powered adsorption chillers were designed to overcome these drawbacks [7].

The solar-powered adsorption chillers are composed of subsystems including solar water heating unit, silica gel water adsorption chillers, a cooling frame, and a fan coil unit. The solar water heating unit is used to supply hot water to run the adsorption chillers. There are two identical adsorption systems in the silica gel adsorption chiller, and a second stage evaporator with methanol working fluid [8, 9]. Using the cooling tower, the condensers and the absorbers cool. The fan coil device is used to refrigerate the vapour. The performance of the adsorption chillers is measured using the performance coefficient and cooling power. The fresh air from the solar-powered adsorption chillers is used for ventilation in the test grain bins. The developed solar-powered adsorption chillers are a good refrigeration system for storing grain at low temperatures [10, 11]. Four solar-powered adsorption chillers were mounted in China for storage of grain at low temperatures until August 2011.A carbon dioxide sensor was developed using polyaniline boronic acid [12] to detect the spoilage present in stored grains. Primary model of the IoT-based PV power control implementation as described in [13] is further modified to attain the proposed method. PV source power sharing for cold storage system control approaches in [14] however IOT control is not attained. Phase power balance with economic consideration is processed [15, 16] however weather forecasting point is not concentrated. The proposed cloud data integrated real-time system along with weather forecasting and economic power balance-based control implemented in cold storage systems.

2 Implementation of Cloud Control Cold Storage System

The integrated system proposed includes four major food items, namely rice, wheat, rava and maida. These food products are selected taking into account the two realities, rice and wheat being the main staple food products and hence the latter two on the premise that they are easily infested. Mass and temperature transfer also happens when there is a substantial difference between the temperatures of the inside and outside bin doors. For example, if the temperature inside the bin is normal, and outside the door bin temperature. And, the moisture content that should be maintained changes with the time of storage. The values above regarding long periods of storage are listed. Temperature is a key parameter which affects the ability to retain moisture in the air. If the temperature is high then the capacity to retain moisture in the air increases, resulting in increased levels of moisture content. Therefore, a rational choice is made for temperature stage. Increasing the rate of food infestation often play

a role in lighting conditions. Hence by implementing the proposed system, the food grains are therefore controlled and shielded from harm by keeping the environmental conditions within the approved limits.

Figure 1 shows the block diagram of the implementation of the proposed cold storage system which is completely controlled by cloud service. The system has a think-speak cloud in which all the necessary cold storage parameters, namely temperature, humidity, and hot loads are fed to it through Arduino Duo and Node MCU (ESP 8266). The parameters which are sensed is compared with the set values and if any parameters exceed its limit value, then the corresponding control actions is taken. Here, the control strategies have cooler for temperature rise, spray water system for humidity and ventilator system for hot loads. According to the cold storage system parameters, the corresponding control strategies were activated.

The system has a think-speak cloud in which all the necessary cold storage parameters, namely temperature, humidity, and hot loads are fed to it through Arduino Duo and Node MCU (ESP 8266) (Fig. 2).

The parameters which are sensed is compared with the set values and if any parameters exceed its limit value, then the corresponding control actions is taken. Here, the control strategies have cooler for temperature rise, spray water system for humidity and ventilator system for hot loads. According to the cold storage system parameters, the corresponding control strategies were activated (Fig. 3).

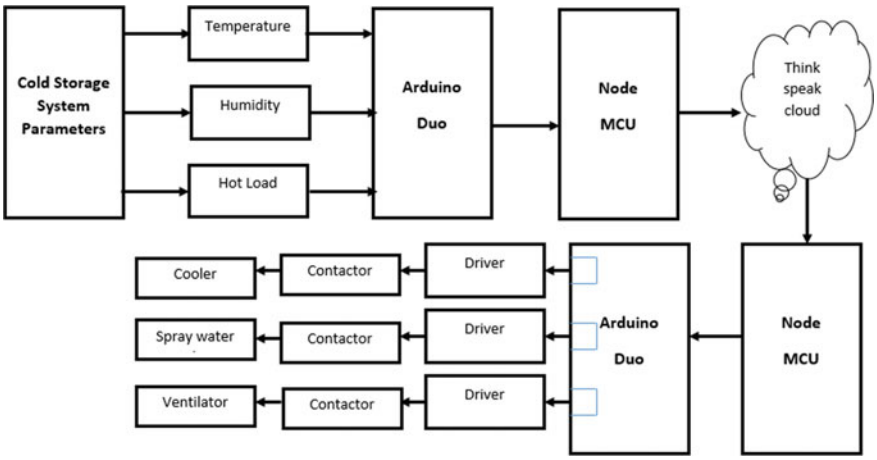


Fig. 1 Block diagram of implementation of the proposed cold storage system using Node MCU

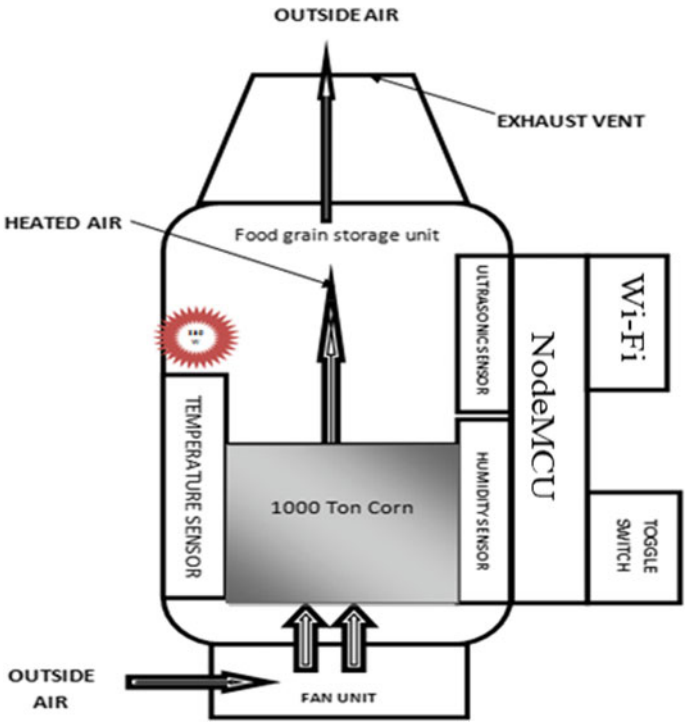


Fig. 2 Proposed grain monitoring system architecture

Fig. 3 Monitoring and controlling system

3 Simulation

Figure 4 shows the flowchart of control and monitoring of the proposed system using Node MCU. Node MCU is connected to the virtual terminal. Here, virtual terminal may be a replacement of the Wi-Fi module. The sensor value is continuously.

monitoring by the Node MCU. The sensor value are often within the limit of a particular range means no function to be activated. Throughout this initiative, the work delineate aims to vogue and incorporate the storehouse environmental monitoring framework powered by NODEMCU and Wi-Fi that can be used in large granaries. Using Node MCU the planning for monitoring and controlling of grain storage units is completed successfully and even the resulting images obtained during

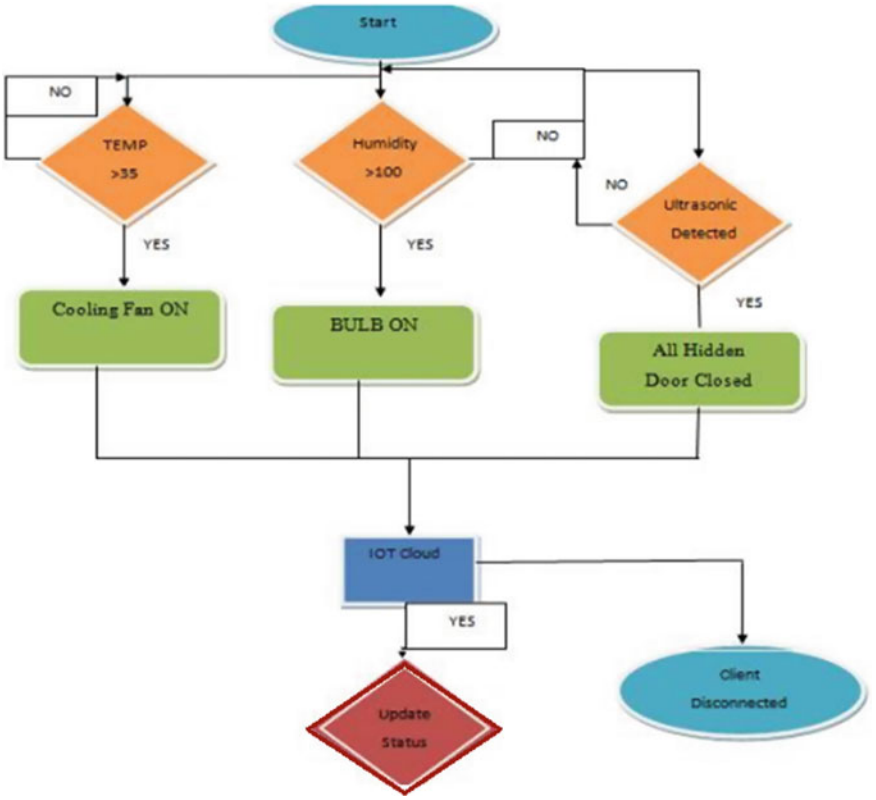


Fig. 4 Flowchart of Monitoring and controlling of cold storage Using Node MCU

simulation using PROTEUS simulator are provided as screenshots. The simulation result obtained during normal running condition is shown in Fig. 5.

4 Hardware Implementation

The grain condition data, to illustrate, temperature, and mugginess information is collected and put away by multi-sensor. On the off chance that the dampness and temperature sensor esteem is an increment, the further activity should be possible by NodeMCU. On the off chance that the interior room and external temperature is up to typical or not equivalent NodeMCU naturally produces the flag to run the engine to open the window. At the purpose when the temperature is ordinary naturally it shut. If the humidity level is high naturally fan will turn. Within the meantime relating responsible will get the info through Wi-Fi. A flip switch is used to activate or deactivate the felony security method. An inaudible sensing element is used to

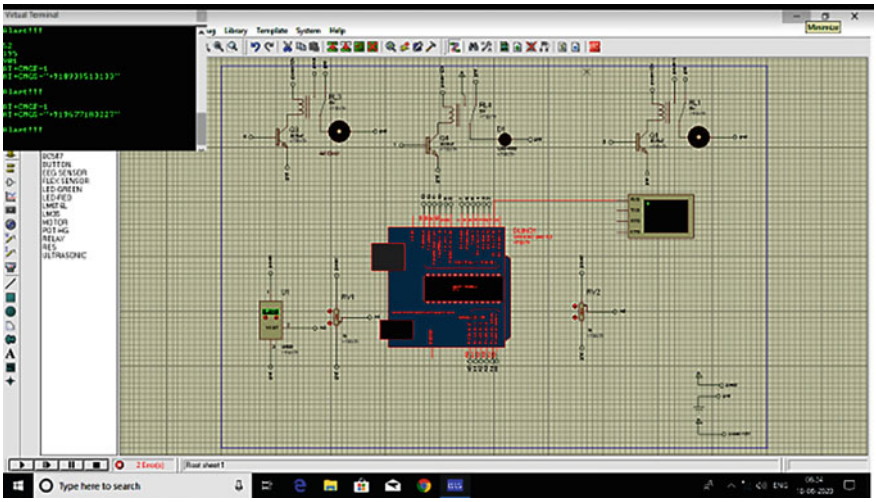


Fig. 5 Simulation result in ideal working condition

create a reverberate movement to avoid species generation (like a feline, mouse, then on) and vertebrate (like Central bearded Drago, Ground agamid lizard, then forth).

Moreover, it's utilized to differentiate the individual passage within the store. Here we've composed one concealed entryway for security reasons. If the individual passage to the shop consequently the info is shipped to the comparison on top of things. The in-charge can initiate the entryway by utilizing the WI-FI. Figure 6 shows the snapshot of the proposed system.

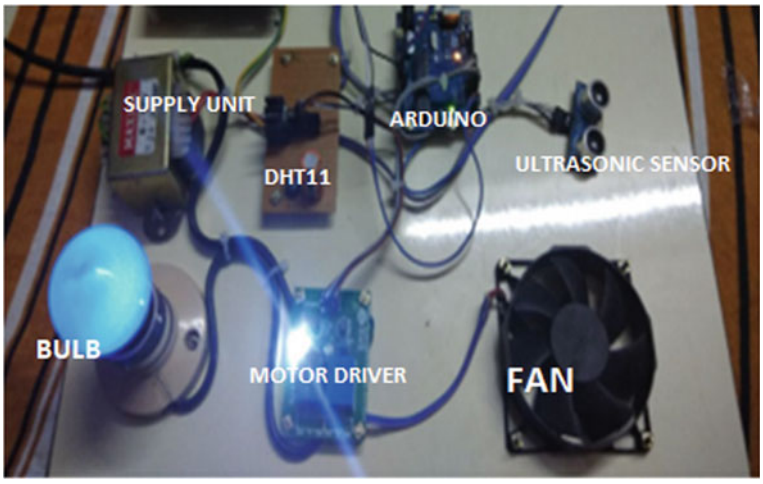


Fig. 6 Snapshot of the proposed system Hardware

5 Conclusion

The need for the hour is to take care of the hygienic practice inside the storage systems. It is shown from the planning carried out in the current work that the data transmission over long networks with lower power consumption is efficient, quantifiable, reliable and safe. The storage of grains is monitored manually in the existing system. There's no proper inspection of the storage grains at periodic levels. Because of this, the grains may get damaged. There's an automatic attribute control facility to take care of an optimized temperature. Hence, there's a requirement to design an automatic system that monitors and maintains the storage grain's quality. Here, chemical tablets are used for grain damage and it affects human health. In future the condition to be uploaded into the cloud in order that the corresponding admin can view the status at any location at an equivalent time to regulate the animals we used the ultrasonic sensor with AM modulation.

References

1. V.S. Suryawanshi, M.S. Kumbhar, Real time monitoring & controlling system for food grain storage. IJIREST **3**(3), 734–738 (2014)
2. L. Xinchun, W. Na, The design of granary environmental monitoring system based on ARM9 and ZigBee, in *International Conference on Computational Problem-Solving* (2010), pp. 395–398
3. X. Zhang, X. Li, J. Zhang, Design and implementation of embedded monitoring system for grain storage, in *2nd IEEE International Conference on Information Management and Engineering* (2010), pp. 197–200. https://doi.org/10.1109/ICIME.2010.5477469
4. S. Kumar, R. Ramaswami, K. Tomar, Localization in wireless sensor networks using directionally information, in *3rd IEEE International Advance Computing Conference (IACC)* (2013), pp. 577–582. https://doi.org/10.1109/IAdCC.2013.6514291
5. M. Jun, C. Zhi-Yan, Design on intelligent node of industrial Ethernet based on ARM, in *Second International Conference on Intelligent Computation Technology and Automation*, (2009), pp. 123–125. https://doi.org/10.1109/ICICTA.2009.497
6. D. Peng, H. Zhang, K. Zhang, H. Li, F. Xia, Research of the embedded dynamic web monitoring system based on EPA protocol and ARM Linux, in *2nd IEEE International Conference on Computer Science and Information Technology* (2009), pp. 640–644. https://doi.org/10.1109/ICCSIT.2009.5234869
7. Y. Lu-gang, L. Yong, A design of communication equipment simulator based on ARM and finite-state machine, in *Third International Symposium on Intelligent Information Technology Application* (2009), pp. 312–315. https://doi.org/10.1109/IITA.2009.24
8. D.P. Bovet, M. Cassetti, A. Oram, *Understanding the Linux*, Kernel,"O' Reilly& Associates, Inc., Sebastopol, CA (2000)
9. L. Zhong, N.K. Jha, Graphical user interface energy characterization for handheld computers. 232–242 (2003)
10. H. Jie, Z. Gen-bao, High-performance embedded processor technology, in *2010 International Conference On Computer Design and Applications* (2010), pp. V1-87–V1-89. https://doi.org/10.1109/ICCDA.2010.5541076
11. N.N. Srinidhi, S.M. Dilip Kumar, K.R. Venugopal, Network optimizations in the Internet of Things: a review. Eng. Sci. Technol. Int. J. **22**(1), 1–21(2019). https://doi.org/10.1016/j.jestch.2018.09.003

12. P. Armstrong, R. Hotel, Wireless data transmission of networked sensors in grain storage. Soc. Eng. Agric. Food Biol. Syst. (2003)
13. L.P. Sivakumar, S.B. Ron Carter, K. Premkumar, T. Thamizhselvan, M. Vishnu Priya, Iot sourced real time PV, wind and fuel cell models for micro and Nano grids, International J. Sci. Technol. Res. **8**(12), 988–993 (2019)
14. P. Sarafoji, V. Mariappan, R. Anish, K. Karthikeyan, J. Reddy, Performance study of solar photovoltaic cold storage system using phase change materials. Mater. Today Proc. **46**(Part 19), 9623–9629 (2021). https://doi.org/10.1016/j.matpr.2020.07.116
15. H. Ikram, A. Javed, M. Mehmood, M. Shah, M. Ali, A. Waqas, Techno-economic evaluation of a solar PV integrated refrigeration system for a cold storage facility. Sustain. Energy Technol. Assessments **44** (2021), ISSN 2213–1388. https://doi.org/10.1016/j.seta.2021.101063
16. S. Adhya, D. Saha, A. Das, J. Jana, H. Saha, An IoT based smart solar photovoltaic remote monitoring and control unit, in *2016 2nd International Conference on Control, Instrumentation, Energy & Communication (CIEC)* (2016), pp. 432–436. https://doi.org/10.1109/CIEC.2016.7513793

Comparative Study of Neural Networks (G/C/RNN) and Traditional Machine Learning Models on EEG Datasets

Gautam Kumar Baboo, Shraddha Dubey, and Veeky Baths

Abstract Background: EEG provides researchers with an opportunity to study neural correlates in terms of temporal connectivity. This connectivity can shed light on the possible network topology between a healthy person versus a patient or help differentiate between two different groups (experts and non-experts). Purpose: With the help of machine learning models, the difference in network topology can be used to understand the neural correlations between healthy control and a patient with ease compared to traditional EEG analysis. Further, a comparative analysis between the different spectral connectivity measures provides the best suitable measure for the study. Methods: EEG data from a meditation study ($n = 31$) and Parkinson's study ($n = 24$) containing the resting-state EEG recordings are utilized here. The EEG data is converted to spectral connectivity: coherence, which becomes the input for the machine learning models, support vector machine, k-means clustering, deep convolution neural networks, recurrent neural networks, and graph neural networks. Results: Classification accuracies of SVM and RNN are 56.585 and 56%, whereas D-CNN provides an accuracy of 59.5%. Both (~ 7%) k-means and GNN failed in the off-the-shelf approach. Conclusion: The comparative study shows the application capabilities of neural networks machine learning with commonly used machine learning models and the impact the various connectivity measures have on model accuracy.

Keywords EEG · Meditation · Parkinson's disease · Classification · Machine learning · Support vector machine · k-means clustering · Graph neural networks

G. K. Baboo · V. Baths (✉)
C-115, Cognitive Neuroscience Lab, Department of Biological Sciences, BITS, Pilani -K.K. Birla Goa Campus, NH-17B, Goa 403726, India
e-mail: veeky@goa.bits-pilani.ac.in

G. K. Baboo
e-mail: p20130404@goa.bits-pilani.ac.in

S. Dubey
Hebbal, Bengaluru 560024, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_17

1 Introduction

The oscillatory nature of electroencephalography (EEG) provides dynamic network coordination between the various regions of the human brain. Concerning the noninvasive EEG recordings, the electrodes and electrode positions are the nodes of a network. The synchronization between the electrodes provides a network, which is noticeable as action potentials or bursts of network activity. Traditional analysis of EEG consists of several steps such as preprocessing, epoching, and event-related potentials analysis [1, 2].

Resting-state EEG recordings can illustrate the baseline activity that occurs within a group of participants or participants. The resting-state can provide the intrinsic neural activity due to the lack of a stimulus or task-state. Given the complexity of the human brain, both in terms of structure and function, inherent with complex interrelations and flow of information. Providing a resting-state EEG tends to contain less noise and complexity when compared to task-state EEG data. Using such data as input to a machine learning model can provide better results when compared to traditional statistical methods, which would use spectral analysis, event-related potentials, or EEG microstate analysis data [3].

Resting-state studies using functional magnetic resonance imaging (fMRI) have been fruitful in providing a diagnosis/difference among groups in a study. Increased activity in visual cortex and low default mode network (DMN) activity of long-term meditators are compared to control under visual memory task conditions. The difference among the groups could be observed in the behavioral performance, wherein the long-term meditators were significantly different faster [4]. The deterioration in DMN in P.D. using fMRI under working memory task provided valuable insights into the cortical deactivations in P.D. patients [5]. The superior spatial resolution of fMRI has shed light on the various aspects of network properties of the patients and long-term mediation practitioners. It does not provide information on the temporal activity that takes place for such participants or patients.

The commonly used feature-based machine learning entails spectral power analysis, functional connectivity, and complexity measures. The spectral analysis includes studying the spectral profile by examining the various rhythm bands, signal power, and spectral edge frequency [6]. Functional connectivity study consists of probing the synchronization or desynchronization of the EEG bands. Two approaches to this include the directed (e.g., Granger causality, mutual information, etc.) or undirected (Pearson's correlation) [1]. Complexity measures include the use of entropy-related techniques such as approximate entropy, permutation entropy, or Kolmogorov–Chaitin complexity [7].

Meditation studies have provided results to indicate a link to mental and physical health benefits; one of the ways this has been achieved is with EEG or fMRI. Various meditative practices and their impact on mental health have been investigated with the help of EEG, which can provide neural correlations in terms of temporal resolution. The methods used on the EEG collected in meditation studies range from spectral analysis to multimodal studies; based on method(s) employed, differences among

the control groups and meditation groups have been reported. The output from the various methods is provided as input to machine learning models, which have been fruitful in differentiating between meditators and novice/non-meditators [8]. Changes in terms of strengthening the attention and executive networks have been observed in meditators versus control groups, moreover, reduced activity in the DMN of long-term meditators versus control groups [9, 10].

Three properties of EEG, viz. permutation entropy order index and complexity in EEG bands, were studied to understand the temporal brain activity differences between P.D. patients and healthy controls [11]. This study comparing the differences between Parkinson's patients and healthy controls revealed that the gamma, beta, and alpha bands had a significantly reduced complexity compared to the theta and delta bands. Further, the study suggested that this complexity may be early cortical or subcortical disease progression markers in P.D. patients. This study focuses on EEG analysis and has scope for network study, thereby providing a more comprehensive picture of the P.D. and warrants the application of machine learning models especially graph neural networks, on such EEG data.

1.1 Literature Survey

Machine learning study on EEG data to classify the pathological state using RNN, SVM, and random forest provided accuracy in the range of 81–86%. The study further provides a review on how to use the EEG data and the tools that would help improve the accuracy when used in a clinical setting [12]. Previous work using correlation matrices from EEG data as inputs to four machine learning models, linear regression, SVM, random forest, and recurrent neural networks showed the utility of using connectivity matrices to classify the datasets [13]. In the study, correlation matrices from five different EEG datasets, three OpenNeuro datasets and two from in-house recordings, were used. Previous work on the efficiency of SVM on EEG data using frequency-domain features on emotion recognition [14] has demonstrated the ability of the learning model to perform well even from a single-channel EEG system. SVM is also suitable for a sparse dataset [10], like the two datasets used here. It has a lower complexity computationally. The essential quality of SVM is that it is not a black box method.

The application of graph neural networks (GNN) is steadily gaining steam, fields such as classifying disease percolation, predicting protein interface, or learning molecular fingerprints. GNNs are neural models that study the data in terms of nodes and edges. Variations of GNNs such as graph convolutional network (GCN), graph attention network (GAT), and graph recurrent network (GRN) have been deployed in diverse deep learning tasks [15].

Here, an off-the-shelf approach using the EEG data and machine learning deployment is made using SVM, k-means clustering, RNN, and GNN. The EEG data from the repository is downloaded as such, and the various epochs are extracted. The meditation study EEG data contains task-related information between meditative

states, whereas the other dataset contains resting-state EEG data in the Parkinson's disease study. The various epochs are then used to calculate the connectivity matrices, which become the input for the classification study with the machine learning aspect. The spectral connectivity function is explored here to understand the changes that occur between the frequencies rather than over a period of time. By studying the changes in the frequency within the various networks, an attempt to understand the brain's functioning is accomplished. The datasets have two different groups within the participants and thus a binary classification problem for the machine learning models.

2 Materials and Methods

2.1 EEG Datasets

Two EEG datasets from the OpenNeuro repository are accessed for the study.

- EEG Meditation Study—OpenNeuro Accession Number: ds001787
- U.C. San Diego Resting-State EEG Data from Patients with Parkinson's Disease—OpenNeuro Accession Number: ds002778.

Written consent from participants was obtained from the respective study groups and has been mentioned in the original articles published by the respective research group(s)/lab(s). Please read the original articles published for details regarding the experimental paradigm and implementation of the study protocol.

EEG Meditation Study

The experiment contains 24(F-12) participants. Participants were meditating and were interrupted about every 2 min to indicate their concentration and mind wandering level. Two EEG recording sessions were available for each participant; EEG acquisition was carried out using a 64-channel Biosemi system with a 10–20 head cap montage at a 2048 Hz sampling rate. The participants are divided into two categories: experts and novice [16]. The participants are interrupted every 30–90 s to answer three questions on a scale of 0–3:

- Please rate the depth of your meditation?
- Please rate the depth of your mind wandering?
- Please rate how tired you are?

The meditators were all practitioners of the Himalayan Yoga tradition. Participants who have had an average of 2 hours of daily practice for a year or longer were grouped as expert meditators. Participants provided written consent to participate in the study and provided details regarding their meditative practices (Table 1).

Table 1 Distribution of participants from the meditation EEG study

Participant ID	Gender	Age	Group
Sub-001	M	32	Expert
Sub-002	M	35	Expert
Sub-003	F	41	Expert
Sub-004	M	29	Expert
Sub-005	F	34	Expert
Sub-006	M	32	Expert
Sub-007	M	32	Expert
Sub-008	M	32	Expert
Sub-009	M	43	Expert
Sub-010	M	33	Expert
Sub-011	M	62	Expert
Sub-012	M	65	Expert
Sub-013	F	47	Novice
Sub-014	F	52	Novice
Sub-015	F	78	Novice
Sub-016	M	77	Novice
Sub-017	F	32	Novice
Sub-018	F	N.A	Novice
Sub-019	F	42	Novice
Sub-020	F	41	Novice
Sub-021	F	41	Novice
Sub-022	F	31	Novice
Sub-023	M	50	Novice
Sub-024	F	38	Novice

U.C. San Diego Resting-State EEG Data from Patients with Parkinson's Disease

The experiment contains 31(F-17) participants, of which 15 are patients with Parkinson's disease [17]. The participants were recruited from the Scripps Clinic in La Jolla, California, and controls were recruited from the local community or the patients' spouses. Of the 15 patients, two EEG recordings were carried out, in which in one session the patient was either on or off a prescribed medication [18]. The recording was carried out for three minutes. The participants were instructed to focus on a fixation point on the center of the screen. The EEG acquisition was carried out using a 32-channel BioSemi ActiveTwo system, sampled at 512 Hz. Eight EOG channels were also used for recording eye blinks and movements (Table 2).

EEG processing for connectivity matrices

The following ways of connectivity matrices are commonly used to study the networks based on EEG data.

Table 2 Distribution of participants (patients) from the U.C. San Diego EEG Parkinson's disease study

Participant Id	Age	Gender	Handedness	MMSE[a]	NAART[b]	Disease Duration (Y)	Group
Sub-HC1	54	F	R	30	48	N.A	Healthy control
Sub-HC2	50	F	R	30	55	N.A	Healthy control
Sub-PD3	52	F	R	29	44	9	Patient
Sub-HC4	50	F	R	30	52	N.A	Healthy control
Sub-PD5	67	F	R	29	48	2	Patient
Sub-PD6	62	F	R	30	42	8	Patient
Sub-HC7	54	F	R	26	48	N.A	Healthy control
Sub-HC8	71	F	R	29	41	N.A	Healthy control
Sub-PD9	55	F	R	28	50	12	Patient
Sub-HC10	59	F	R	29	51	N.A	Healthy control
Sub-PD12	74	M	R	26	53	1	Patient
Sub-PD11	71	F	R	30	52	1	Patient
Sub-PD13	62	M	R	29	47	2	Patient
Sub-PD14	63	M	R	28	35	2	Patient
Sub-PD16 lePara>	74	M	R	29	56	2	Patient
Sub-PD17	55	M	R	30	40	2	Patient
Sub-HC18	57	M	R	30	57	N.A	Healthy control
Sub-PD19	69	F	R	29	43	6	Patient
Sub-HC20	68	M	R	30	58	N.A	Healthy control
Sub-HC21	72	M	R	28	33	N.A	Healthy control
Sub-PD22	47	M	R	30	51	6	Patient
Sub-PD23	66	F	R	29	34	3	Patient
Sub-HC24	61	F	R	30	57	N.A	Healthy control
Sub-HC25	74	F	R	30	49	N.A	Healthy control
Sub-PD26	71	F	R	29	51	3	Patient
Sub-PD28	61	M	R	29	46	9	Patient

(continued)

Table 2 (continued)

Participant Id	Age	Gender	Handedness	MMSE[a]	NAART[b]	Disease Duration (Y)	Group
Sub-HC29	60	F	R	29	45	N.A	Healthy control
Sub-HC30	69	M	R	29	52	N.A	Healthy control
Sub-HC31	82	M	R	28	50	N.A	Healthy control
Sub-HC32	60	M	R	29	37	N.A	Healthy control
Sub-HC33	75	M	R	30	53	N.A	Healthy control

[a] *MMSE* Mini-mental state examination
[b] *NAART* North American Adult Reading Test score

- Coherence
- Phase lag index [19]
- Phase locking value
- Directed phase transfer entropy
- Lagged linear connectivity
- Global field synchronization
- Weighted network
- Minimum spanning tree
- Corrected imaginary PLV [20]
- Weighted phase lag index [21].

These matrices are used as the input for various machine learning models. In this study, the following methods have been implemented. The low pass is set to 0.01 Hz and the high pass to 45 Hz. Apart from this preprocessing set, the EEG data is used as such from the repository.

- Coherence [22]

This is achieved by using the functions made available by the MNE-Python package [22]. The events from each EEG session are extracted based on the events; epoch extraction is carried out [23]. The epoched data is then used for calculating the four spectral connectivity matrices. Following these steps, the data is then converted into a data frame for the machine learning deployment.

2.2 Machine Learning Models

The performance evaluation of the different classifiers is examined using a confusion matrix, whose components are T.P., TN, F.P., and F.N. Further, the accuracies are calculated using these measures, using the formula:

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}) * 100 \quad (1)$$

T.P. True Positives
T.N. True Negatives
F.P. False Positives
F.N. False Negatives.

Overfitting/underfitting: In this study, the problem of overfitting did not pose an obstacle in these datasets. Hence, the results were not unusually accurate. Regularization factors are applied to reduce overfitting. Finally, for underfitting, varying the hyperparameters over a large range is carried out, and the best fitting set of values is appropriated.

Principal Component Analysis (PCA)

To reduce the dimensionality, PCA identifies the hyperplane that lies closest to the data and then projects the data onto it. To reduce to the right dimensions (in order to preserve max variance), the following code is implemented:

```
    pca = PCA()
pca.fit(con_res_hc['coh'])
cumsum = np.cumsum(pca.explained_variance_ration_)
d = np.argmax(cumsum >= 0.90) + 1
```

{d = 12 dimensions}
This reduces the training set and uses less computation space for processing.

Support Vector Machine

The data frame is split into three classes for Parkinson's EEG data:

1. Healthy controls (HC)
2. Parkinson's patients with medication on (PD_On)
3. Parkinson's patients with medication off (PD_Off).

and, in the case of meditation study data:

1. Expert meditators
2. Novice(non-expert) meditators.

By implementing SVM, the aim is to use less computational power and at the same time, produce a significant classification accuracy. The input data is mapped to

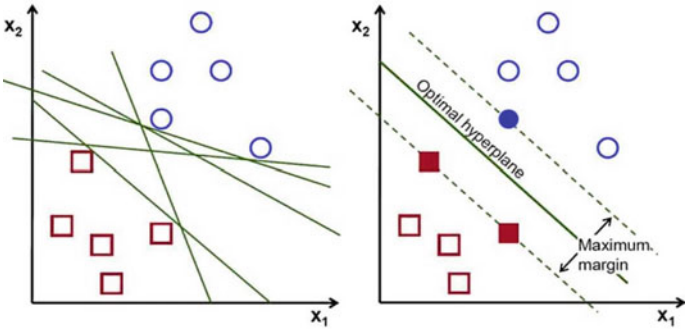


Fig. 1 Support vector machine algorithm with the construction of different hyperplanes that separates the different classes. The most optimal hyperplane is the one that maximizes this separation

a higher dimensional vector space using a linear kernel function to find a hyperplane for classification.

$$w * z - b = 0 \tag{2}$$

w: normal vector

b: bias of separation of the hyperplane (Fig. 1).

k-Means clustering

k-means clustering is an unsupervised machine learning model; one of the recent applications of this has been on artifact removal, precisely the eye-blink artifact in EEG data [24]. The data frame is split into three clusters.

(Centroids: araay ([[− 9.16842105, − 7.07663091], [11.51282051, − 0.27005616], [− 8.45674044, 8.5610777]])).

Deep Convolutional Neural Networks (D-CNN)

Convolutional neural networks (CNNs) performance on EEG data has provided good accuracy as it contains the ability to learn features from the receptive fields. With other machine learning models, preprocessing of the EEG data tends to provide improved accuracy, whereas CCNs are designed to use minimal preprocessing. And with this in mind, we provide the EEG data with only the connectivity matrices that are computed [25]. The input to the D-CNN is a 3D connectivity matrix (coherence matrix). The D-CNN is made up of two different layers, a convolutional layer and a pooling layer, and are alternated. The pooling layer takes care of subsampling, and the last stage is a fully connected layer. The first convolutional stage is with the rectified linear unit (ReLU) activations followed by max pooling. This network learns 20 convolutional filters, each with a size of 5 × 5.

The max pooling operation implements a sliding window that slides over the layer and takes the maximum of each region with a step of 2 units both vertically

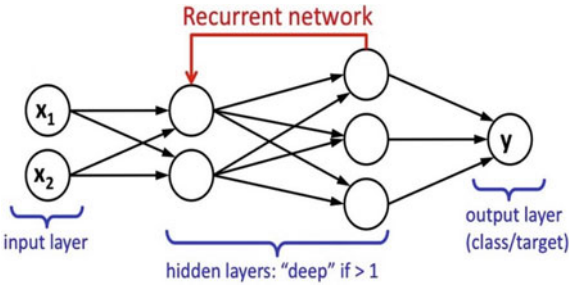


Fig. 2 Recurrent neural network (RNN) represents the skeleton on which every RNN is built. The output of each layer acts as the input to the next and modifies the hyperparameters of the layer in each epoch, thus implementing the learning part of the algorithm

and horizontally. The second stage consists of ReLU activations, followed again by a max pooling layer. Here we increase the number of convolutional filters 50 from the previous 20.

Recurrent Neural Networks

Previous work on the implementation of neural networks on EEG signals has been fruitful, which provided accuracy in the range from 81 to 94% [26]. RNN was a good model for studying both working memories [27, 28] and emotional state [29] EEG data when compared to other models such as SVM or deep belief networks [30]; on that note, the following RNN model is implemented. The RNN is implemented through a long short-term memory (LSTM) model [31, 32], producing exemplary results on sequential data, such as EEG data. A sequential model is used to build the LSTM, which is a linear stack of layers. The first layer is an LSTM layer with 256 memory units, and it defines the input shape. This is done to ensure that the next LSTM layer receives sequences and not just randomly scattered data. The next layer is a dense layer with a "sigmoid" activation function. A dropout layer is applied after each LSTM layer to avoid overfitting the model. The model is then trained and monitored for validation accuracy using loss as "binary cross-entropy", optimizer as "adam", and metrics as "accuracy" (Fig. 2).

$$H(q) = -1/N \sum y_i * \log(p(y_i)) + (1 - y_i) * \log(1 - p(y_i)) \tag{3}$$

$H(q)$ binary cross-entropy

$p(y_i)$ the probability of belonging to class y_i.

3 Result

Classification Comparison on Individual Groups

An increase in the number of electrodes provides better classification accuracy, observed from the Tables 3, 4 and 5.

Table 3 Accuracy of classifying patient category from Parkinson's EEG dataset

Group	SVM	D-CNN	RNN	k-means	GNN
Healthy control	52.17%	60%	56%	~ 10% (failed)	Could not implement (failed)
Parkinson's patients on medication					
Parkinson's patient off medication					

Table 4 Accuracy of classifying expert and non-expert meditators from meditation study EEG dataset

Group	SVM	D-CNN	RNN	k-means	GNN
Expert meditators	61%	59%	56%	~ 6% (Failed)	(Failed)
Non-expert meditators					

Table 5 Overall classification accuracy of the machine learning models on the datasets

Machine learning model	Accuracy (%)
SVM	56.585%
D-CNN	59.5%
RNN	56%
k-means	Failed
GNN	Failed

Classification Comparison of the Machine Learning Models

The performance of the learning models, when compared between the two distinct EEG datasets, provides accuracy within the same range except for the RNN model.

As shown in Fig. 3, the k-means machine learning model fails to provide any useful results with the off-the-shelf approach. The red dots indicate the centroids respective to the healthy controls and patients with medications off and medication on.

4 Discussion

The main objectives of this study are the comparison of the five machine learning models on the coherence connectivity matrix and the impact on the accuracy provided between the machine learning models implemented.

Some studies have highlighted the importance of the depth and size (number of participants, experimental paradigm, etc.) of the EEG data used. Sleep and epileptic study data is numerous and from different sources since hospitals that record them have or are willing to provide access to them. And machine learning models that

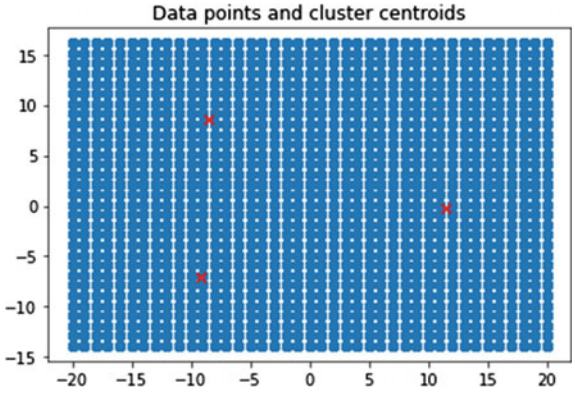


Fig. 3 k-means clustering, red dots indicate the centroids

are implemented to study these conditions perform well thanks to the dimensions of the available data compared to other clinical or psychological/cognitive conditions, which are usually collected in academic settings and tend to lack a certain amount of diversity. Thus with the currently available data, a comparative study makes sense [33].

Comparison of the Learning Models

The input for the machine learning models is the coherence connectivity matrices, a spectral connectivity matrix. This provides a sense of uniformity to the machine learning models in terms of the input data. The performance of the models based on classification accuracy is observed in Table 5. From this comparison, it is clear that SVM seems to perform better.

Phase coherence, phase-locking value, or pairwise phase consistency methods of EEG analysis, which transform the EEG data to a matrix form for network construction, require adding steps to the analysis. Functional connectivity using EEG data is accomplished by using the signal's frequency, time domain, or phase properties, which can be categorized as static or dynamic. The various methods under each of these categories have their advantages and disadvantages, followed by tools/methods that can strengthen or weaken the said functional connectivity analysis methods.

Limitations

With an increase in the number of electrodes, the accuracy increases but so does the complexity in the learning models and the time required for the models to provide suitable results. The implementation of k-means and graph neural networks failed for distinct reasons. In the case of the k-means clustering approach, the model could not identify the clusters as the connectivity matrices are in 0 and 1 s and do not provide significant and useful results, whereas in the case of the hardware requirement which is comparatively larger, there are instances of memory overflow.

5 Conclusion

Using a spectral connectivity measure such as coherence on EEG data and widely used machine learning models provides a good classification of the participants/patients in the respective studies. The off-the-shelf approach to this study provides preliminary evidence on the application of the respective machine learning models for EEG data with minimal preprocessing both on the EEG data and the input data for the various machine learning models.

Acknowledgements We would like to thank OpenNeuro for providing access to the EEG datasets and citations for the datasets have been included within the main text.

Declaration The authors declare no conflict of interest.

References

1. A.M. Bastos, J.M. Schoffelen, A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Front. Syst. Neurosci. **9**, 1–23 (2016). https://doi.org/10.3389/fnsys.2015.00175
2. M. Soufineyestani, D. Dowling, A. Khan, Electroencephalography (EEG) technology applications and available devices. Appl. Sci. **10**, 1–23 (2020). https://doi.org/10.3390/app10217453
3. Z. Li, L. Zhang, F. Zhang, et al., Demystifying signal processing techniques to extract resting-state EEG features for psychologists. Brain. Sci. Adv. **6**:189–209 (2020). https://doi.org/10.26599/bsa.2020.9050019
4. A. Berkovich-Ohana, M. Harel, A. Hahamy et al., Alterations in task-induced activity and resting-state fluctuations in visual and DMN areas revealed in long-term meditators. Neuroimage **135**, 125–134 (2016). https://doi.org/10.1016/j.neuroimage.2016.04.024
5. T. van Eimeren, O. Monchi, B. Ballanger, A.P. Strafella, Dysfunction of the default mode network in Parkinson disease. Arch. Neurol. **66**, 877–883 (2009). https://doi.org/10.1001/archneurol.2009.97
6. S. Corchs, G. Chioma, R. Dondi et al., Computational methods for resting-state EEG of patients with disorders of consciousness. Front. Neurosci. **13**, 1–7 (2019). https://doi.org/10.3389/fnins.2019.00807
7. Y. Bai, X. Xia, X. Li, A review of resting-state electroencephalography analysis in disorders of consciousness. Front. Neurol. **8**, (2017). https://doi.org/10.3389/fneur.2017.00471
8. C.S. Deolindo, M.W. Ribeiro, M.A. Aratanha et al., A critical analysis on characterizing the meditation experience through the electroencephalogram. Front. Syst. Neurosci. **14**, 1–29 (2020). https://doi.org/10.3389/fnsys.2020.00053
9. A. Barrós-Loscertales, S.E. Hernández, Y. Xiao, et al., Resting state functional connectivity associated with Sahaja Yoga meditation. Front. Hum. Neurosci. **15**, 1–11 (2021). https://doi.org/10.3389/fnhum.2021.614882
10. Z. Guo, Y. Pan, G. Zhao et al., Detection of driver vigilance level using EEG signals and driving contexts. IEEE Trans. Reliab. **67**, 370–380 (2018). https://doi.org/10.1109/TR.2017.2778754
11. G.S. Yi, J. Wang, B. Deng, W.X. Le, Complexity of resting-state EEG activity in the patients with early-stage Parkinson's disease. Cogn. Neurodyn. **11**, 147–160 (2017). https://doi.org/10.1007/s11571-016-9415-z

12. L.A.W. Gemein, R.T. Schirrmeister, P. Chrabąszcz, et al. Machine-learning-based diagnostics of EEG pathology. Neuroimage **220**, 117021 (2020). https://doi.org/10.1016/j.neuroimage.2020.117021
13. B. Prakash, G.K. Baboo, V. Baths, A novel approach to learning models on eeg data using graph theory features—a comparative study. Big Data Cogn. Comput. **5**, (2021). https://doi.org/10.3390/bdcc5030039
14. Z. Wei, C. Wu, X. Wang, et al., Using support vector machine on EEG for advertisement impact assessment. Front. Neurosci. **12** (2018) . https://doi.org/10.3389/fnins.2018.00076
15. J. Zhou, G. Cui, S. Hu et al., Graph neural networks: a review of methods and applications. AI Open **1**, 57–81 (2020). https://doi.org/10.1016/j.aiopen.2021.01.001
16. T. Brandmeyer, A. Delorme, Reduced mind wandering in experienced meditators and associated EEG correlates. Exp. Brain. Res. **236**, 2519–2528 (2018). https://doi.org/10.1007/s00221-016-4811-5
17. N. Jackson, S.R. Cole, B. Voytek, N.C. Swann, Characteristics of waveform shape in Parkinson's disease detected with scalp electroencephalography. eNeuro **6**,1–11 (2019). https://doi.org/10.1523/ENEURO.0151-19.2019
18. J.S. George, J. Strunk, R. Mak-Mccully, et al., Dopaminergic therapy in Parkinson's disease decreases cortical beta band coherence in the resting state and increases cortical beta band power during executive control. NeuroImage Clin. **3**, 261–270 (2013). https://doi.org/10.1016/j.nicl.2013.07.013
19. C.J. Stam, G. Nolte, A. Daffertshofer, Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. Hum. Brain. Mapp. **28**, 1178–1193 (2007). https://doi.org/10.1002/hbm.20346
20. R. Bruña, F. Maestú, E. Pereda, Phase locking value revisited: Teaching new tricks to an old dog. J. Neural. Eng. **15**, (2018). https://doi.org/10.1088/1741-2552/aacfe4
21. M. Vinck, R. Oostenveld, M. van Wingerden, et al., An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. Neuroimage **55**, 1548–1565 (2011) . https://doi.org/10.1016/j.neuroimage.2011.01.055
22. A. Gramfort, M. Luessi, E. Larson et al., MEG and EEG data analysis with MNE-Python. Front. Neurosci. **7**, 1–13 (2013). https://doi.org/10.3389/fnins.2013.00267
23. S. Appelhoff, M. Sanderson, T. Brooks, et al., MNE-BIDS: organizing electrophysiological data into the BIDS format and facilitating their analysis. J. Open Source Softw. **4**, 1896 (2019). https://doi.org/10.21105/joss.01896
24. A.K. Maddirala, K.C. Veluvolu, Eye-blink artifact removal from single channel EEG with k-means and SSA. Sci. Rep. **11**, 11043 (2021). https://doi.org/10.1038/s41598-021-90437-7
25. X. Lun, Z. Yu, T. Chen et al., A simplified CNN classification Method for MI-EEG via the electrode Pairs signals. Front. Hum. Neurosci. **14**, 338 (2020)
26. R. Dhanapal, D. Bhanu, Electroencephalogram classification using various artificial neural networks. J. Crit. Rev. **7**, 891–894 (2020). https://doi.org/10.31838/jcr.07.04.170
27. Z. Jiao, X. Gao, Y. Wang et al., Deep Convolutional Neural Networks for mental load classification based on EEG data. Pattern Recognit. **76**, 582–595 (2018). https://doi.org/10.1016/j.patcog.2017.12.002
28. S. Kuanar, V. Athitsos, N. Pradhan, et al., Cognitive analysis of working memory load from eeg, by a deep recurrent neural network, in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2018), pp. 2576–2580. https://doi.org/10.1109/ICASSP.2018.8462243
29. M. Bilucaglia, G.M. Duma, G. Mento, et al. Applying machine learning EEG signal classification to emotion-related brain anticipatory activity. F1000 Res. **9**, 173 (2020). https://doi.org/10.12688/f1000research.22202.1
30. A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review. J. Neural Eng. **16** (2019).https://doi.org/10.1088/1741-2552/ab0ab5

31. S. Chang, W. Dong, H. Jun, Use of electroencephalogram and long short-term memory networks to recognize design preferences of users toward architectural design alternatives. J. Comput. Des. Eng. **7**, 551–562 (2020). https://doi.org/10.1093/jcde/qwaa045
32. A.V. Medvedev, G.I. Agoureeva, A.M. Murro, A long short-term memory neural network for the detection of epileptiform spikes and high frequency oscillations. Sci. Rep. **9**, 1–10 (2019). https://doi.org/10.1038/s41598-019-55861-w
33. Y. Roy, H. Banville, I. Albuquerque, et al., Deep learning-based electroencephalography analysis: asystematic review. arXiv (2019)

Artificial Intelligent Models for Automatic Diagnosis of Foetal Cardiac Anomalies: A Meta-Analysis

M. O. Divya and M. S. Vijaya

Abstract The foetal anomaly scanning is one of the most challenging areas where accuracy of diagnosis much fluctuating with respect to the expertise of the radiologist and the mental equilibrium of the radiologist at the time of scanning. Amongst the various anomalies, foetal heart anomaly diagnosis expects precise and sensitive intellectual presence since perilous congenital heart diseases are one of the common causes resulting in the major population of infant mortality or into permanent natal faults. The accuracy of manual diagnosis of foetal cardiac abnormalities from the ultrasound scan images vary based on the human expertise and the presence of mind. Therefore, the scope of computer-assisted judgement can produce accurate diagnosis irrespective of the operator's profile. Numerous researches are going on to explore the scope of computer-assisted judgement of abnormalities using ultrasound imaging technique (USIT), specifically using machine learning and deep learning models. This work exploits the opportunities of computer-assisted diagnosis in foetal cardiac anomaly diagnosis as this is one of the most sensitive areas where appropriate diagnosis can save a life and a wrong diagnosis may lose a life unnecessarily.

Keywords Congenital heart disease · Early diagnosis · Ultrasound images · Image processing · Machine learning · Deep learning · Foetal cardiac anomaly

1 Introduction

Ultrasound examination is one of the fundamental medical imaging techniques used in clinics on a regular basis. This is a non-invasive and non-radioactive technique to understand the internal body structure; thus becomes a universally accepted medical

M. O. Divya (✉)
Research Scholar, Department of Computer Science, PSGR Krishnammal College for Women, Coimbatore, India
e-mail: divyammo@gmail.com

M. S. Vijaya
Associate Professor, Department of Computer Science, PSGR Krishnammal College for Women, Coimbatore, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_18

imaging mechanism for preliminary examination and diagnosis of various biological disorders. The challenge with the ultrasound imaging technique (USIT) is that it purely depends on the operator's perception and expertise.

There has been a recent explosion in the use of artificial intelligence (artificial intelligence), which is now part of our everyday lives. Uptake in medicine has been more limited, although in several fields there have been encouraging results showing excellent performance when artificial intelligence is used to assist in a well-defined medical task. Depending on the problem scenario, data availability and other constraints the researcher can decide which technique or combination of techniques are appropriate. The subsets of AI are mentioned in Fig. 1. In medical field, most of the research has been performed using retrospective data and there have been few clinical trials published using prospective data [1]. This review focuses on identifying the potential uses of machine learning and deep learning models in the field of foetal cardiology.

Congenital Heart Defects (CHD) are one of the most common forms of birth malformations in newborns and infants. They occur at a prevalence of 6–8 per 1000 live births. The prevalence of these defects may be higher in prenatal scans. About 1

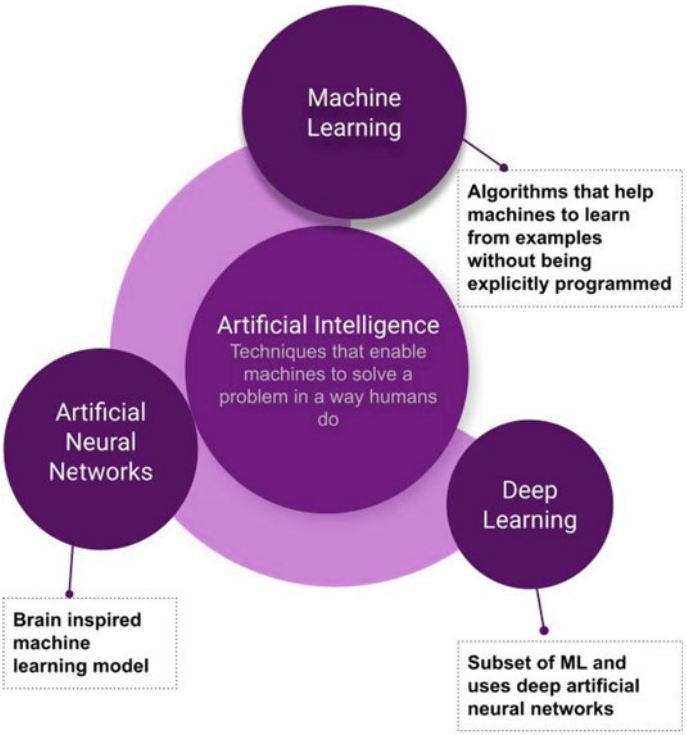


Fig. 1 Relation between AI, ML, DL

in 4 babies with a CHD have a critical CHD. CHDs are a leading cause of birth defect-associated infant illness and death. CHD forms the 5th most common cause of infant deaths on a global level infants with critical CHDs generally need surgery or other procedures while in the uterus or during their first year of life. In India, it is estimated that around 200–250,000 infants are born with CHD every year of which 100,000 are deemed critical requiring early surgical intervention. These surgical procedures are very critical and need thorough setting of environment before the baby is born. Hence identification of these anomalies during the second trimester or third trimester has a value in life.

In the study conducted by Parikh et al. [2] in UK shows that 68.2% of the CHD cases were been diagnosed postnatally and 8.4% of the cases reported with neonatal death. This shows the sensitivity of diagnosis of CHD prenatally and the relevance of USIT. Ultrasound of the foetal heart is highly specific and sensitive in experienced hands, but despite this, there is significant room for improvement in the rates of prenatal diagnosis of congenital heart disease in most countries. USIT can help to improve the present statistics of prenatal diagnosis rate of foetal cardiac anomalies. This research paper projects the possible breakthroughs in automated diagnosing of foetal cardiac anomalies with the help of machine learning and deep learning models.

2 Background Study

Prenatal diagnosis of major forms of CHD is feasible during early and mid-trimester ultrasound scans of pregnancy. The optimal timing of this evaluation is 18–22 weeks of gestation. Several international associations have laid down clear guidelines for the conduct of the foetal heart evaluation during mid-trimester anomaly scans. In the simplest form, this screening involves two views—the four-chamber view and the outflow tracts using the three-vessel view. In the event of a failed screen, the patient needs a more thorough evaluation by an expert for a possible heart defect. Prenatal diagnosis of major CHDs offers more options for management to the expectant families including termination of pregnancy (if diagnosed early < 20 weeks) for complex defects and planned peri-partum care for critical correctable CHDs. Studies from India have shown that prenatal diagnosis of critical CHDs improves the preoperative clinical status and in-hospital outcomes and results in lower costs of in-hospital care due to shorter duration of ICU stay.

However, despite the availability of ultrasound equipment and clear protocols for foetal heart screening, the overall pick-up rates of major CHDs is very low in prenatal scans, especially in the low-middle-income countries (LMICs). Lack of awareness and familiarity with imaging protocols and difficulty in interpretation of images are stated as personal reasons for this. Since the imaging views for the basic screening (two views) are very standard and the image patterns are stereotyped, it is feasible to develop an artificial intelligence (artificial intelligence) based algorithm for a basic level evaluation.

Figure 2, graph explains how minimal is the diagnosis rate of various structural anomalies and the major part of it is unspecified. It shows the importance of intervention of technology so that the device auto diagnoses from the scan image which can very well improve the diagnosis rate. Figure 3 shows the mortality rate due to CHD over the specified period. This is not because of the medical incapability, but it is due to the inaccurate prenatal diagnosis of CHD.

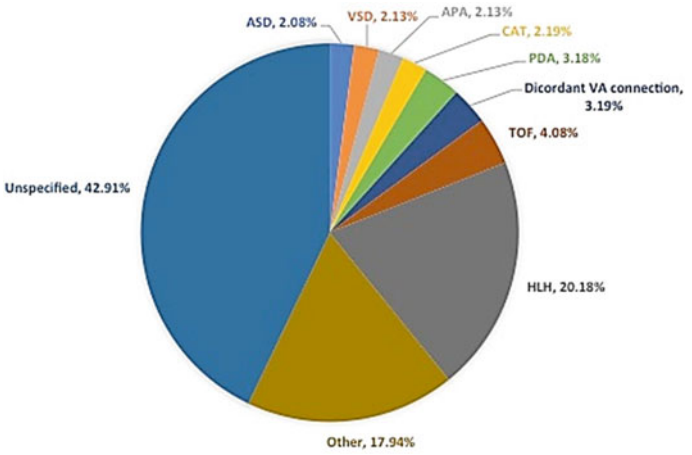


Fig. 2 CHD mortality trends by diagnosis

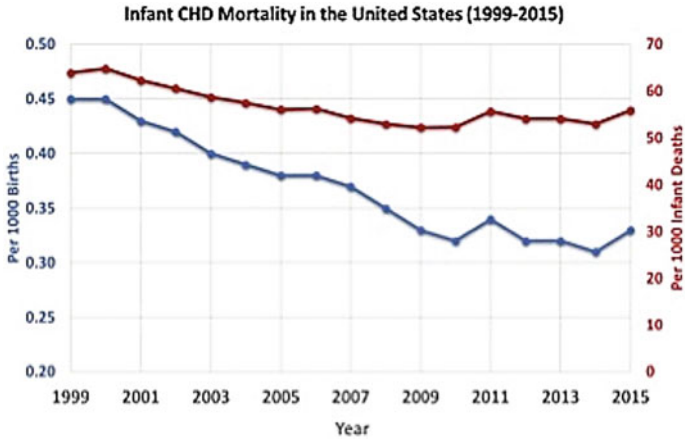


Fig. 3 CHD mortality trends in the US

3 Literature Review

There are numerous researches happening in medical image processing and diagnosing diseases from USIT, with minimum manual interventions. Majority of the work revolves around the opportunities in breast cancer diagnosis and thyroid abnormalities. Those existing research works throw light on the automated diagnosis of diseases affecting other organs as well. For this meta-analysis research, two categories of literatures were taken into consideration. The first set of literatures belongs to the research publication from medical practitioners who have suggested the technological interventions required in foetal cardiac anomaly diagnosis. The second category of literatures constitute the published research work that addresses the diagnosis of various diseases from USIT using machine learning and deep learning [3].

3.1 Medical Literatures Suggesting Technological Interventions

1. Parikh et al. [4] in their research identified how the early diagnosis of CHDs can result in the best-attempted route for the delivery. The research shows that the mode of delivery for a foetus with cardiac structural anomalies is very critical because if it gets unnoticed can cause neonatal morbidity based on the severity of CHD. In the current system, most of the CHDs are diagnosed postnatally. In neonates with connatural defects, 8.4% of 107 neonatal death happened and in 54.2% of 83 neonates, grave respiratory morbidity happened with left ventricular outflow tract defects.
2. Holland et al. [5] The research was a meta-analysis, where the difference in mortality rate was analyzed when the cardiac anomalies were identified prenatally and postnatally. The study indicates that the importance of prenatal diagnosis of CHDs makes a lot of difference and reduces the postnatal mortality rate.
3. Suard et al. [4] did an evaluation on the accuracy rate of prenatal diagnosis of CHDs in South France. This observational study considered 249,070 deliveries into consideration. The diagnosis rate of Group 1 CHD (Where no possibilities of anatomical repair is possible) was 97.8%, for Group 2 (anatomical repair was possible but requires neonatal cardiologic attention and management) was 6.3% and Group 3 (no emergency anatomic procedure was required) was 65.9%. Which shows the Group 2 and Group 3 diagnosis rate could be improved for easy anatomic corrections prenatally and postnatally.
4. Changlani et al. [6], The study shows, short-term consequences of infants where prenatal diagnosis of CHD delivered in a tertiary cardiac care facility. 552 foetuses were diagnosed to have CHD. The study revealed that the prenatal finding of CHD in infants will help for planned delivery in a cardiac facility showed satisfactory fast outcomes, specifically among those receiving dedicated postnatal cardiac care.

5. Vijayaraghavan et al. [7] in their observational study revealed that the planned peri-partum care is an unexplored concept for care of neonates with critical CHDs in low-middle-income countries. The study was a comparison of outcomes of prenatal and postnatal diagnosis of CHDs.

3.2 Review of Articles Related to Diagnosing Diseases Using Machine Learning/Deep Learning and Ultrasound Scan Images

Ding et al. [8], in this research paper explain how the features can be extracted from ROIs. Then these features were used for the diagnosis of breast tumours. As the breast ultrasound images lack clarity due to speckle noise, Multiple-instance learning (MIL) method is more appropriate to classify breast tumours into benign and malignant, using BUS images, for which a novel MIL method is proposed in this research paper. First, a self-organizing map is used to map the instance space to the concept space. The experimental results show better performance, accuracy is 0.9107 and the area under receiver operator characteristic curve is 0.96 ($p < 0.005$).

Wei et al. [9] in their study, texture and morphological features are used for breast tumour classification. This combination is selected because the texture feature is supposed to be very aggressive due to which the low dimensional features will get unnoticed. The study uses, LBP, HOG, GLCM and morphological (i.e. shape complexities) features of breast ultrasound images. To classify, SVM classifier on texture features and a naive Bayes classifier on morphological features is applied. Then the classification results are fused to get the final classification. This method shows accuracy of 91.11%, sensitivity of 94.34% and specificity of 86.49%.

Song et al. [10], in their study, a hybrid multi-brach CNN with feature cropping technique is used for classification of ultrasound images of thyroid nodules. The convolutional neural network for extracting the shared feature maps with a global branch classifier. A feature cropping branch for reducing the impact of similar local features of benign and malignant images, a feature cropping branch also is there. To train the system, a weighted cross-entropy loss function was used. The method has attained accuracy − 96.13%, precision − 93.24%, recall − 97.18% and F_1-measure 95.17% in public and local dataset.

Liyang et al. [11] intended to propose a statical analysis to identify benign and malignant lesions from the breast ultrasound scan images. This research was conducted on 85 cases where 35 malignant and 50 benign ultrasound images were collected. From the images, a fractal dimensional image was produced from the ROI by using box-counting method. From the FD and ROI images, features were extracted including mean, standard deviation, skewness and kurtosis. Statistical tests were conducted on these features. The statistical analysis revealed that the mean texture of images performed the best in differentiating benign versus malignant tumours. The sensitivity, specificity, accuracy, positive predicted value (PPV), negative predicted

value (NPV) and Kappa for the mean was 0.77, 0.84, 0.81, 0.77, 0.84 and 0.61 respectively.

Chi et al. [12], their study proposed a deep learning model for classifying the thyroid gland nodule as benign and malignant. They have used a trained GoogLeNet model and the experimental results exhibit excellent performance, accuracy – 98.29%, sensitivity – 99.10% and specificity – 93.90% for the images in an open access database, while classification accuracy – 96.34%, sensitivity – 86% and specificity – 99% for the images in their local region health database.

Abdel-Nassar et al. [13], this research focuses on reducing the speckle noise of breast ultrasound images and thereby increase the accuracy of the model. For this, multiple images of the same targets are used as input to the classifier. The system works in four phases which include, super-resolution computation, extraction of ROI, feature extraction and then classification.

Gao et al. [14] This study explores the eight viewpoints of the heart using the echo videos which intern assist for the diagnosis of cardiologic disorders. The proposed architecture gives the best classification results with 92.1% accuracy rate whereas 89.5% is achieved using only single spatial CNN network.

Reviews on research articles cited in authenticated repositories including Scopus, Web of Science, Medline, PubMed and EMBASE to explore the opportunities in applying artificial intelligence techniques to diagnose foetal cardiac anomalies. The review revealed that a single model might not be enough rather a combination of models are appropriate for accurate diagnosis. The results are based on the guidelines provided by PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).

4 Methodology

Congenital cardiac disease is seen in 2–6.5 of 1000 live births and is a major cause of morbidity and mortality, with half of these cases being lethal or requiring surgical correction. Finding of CHDs is very sensitive and can be troublesome. Detection of anomalies alters the obstetric course and outcome, including reassurance, termination, foetal therapy, mode of delivery and postnatal referral to a tertiary care canter with advanced expertise in management of these patients, which is the major difference that could save/change the baby's life after birth. The conventional diagnosis is done manually using the USIT captured by foetal cardiac ultrasound performed between 18 and 22 weeks. This examination is a targeted study where the heart structure and its function. During the first phase (Phase 1), two views of the foetal heart evaluation will be done manually. The four-chamber view and the three-vessel view. In cases where there are critical anomalies found, a detailed echocardiography is performed. Name it as Phase 2 which will include the six views analysis, recommended for an extended basic evaluation of the foetal heart including the situs, four-chamber view, LV outflow tract, RV outflow tract, three vessel and three vessel tracheal view.

The same phases could be implemented with the deep learning/machine learning model as well [3]. There are common classifiers like Support vector machine (SVM), k-nearest neighbours (KNN), etc. in ML and Convolutional neural network (CNN), Recurrent neural networks (RNN), Self-organizing maps (SOM), etc. in deep learning. The most recent research by Liu et al. [15] has proved their research that the deep learning model Stacked Sparce auto encoder (SSAE) based model shows 100% accuracy for classifying medical images.

4.1 Phase 1: Model for Preliminary Classification and Screening

The first phase of proposed system is to develop an algorithm to train the model using labelled images. Here, the model will be trained to evaluate the two views of the foetal heart, the four-chamber view and the three-vessel view. The model will be trained to recognize all possible permutations and combinations of the normal four-chamber and three vessel views (apical, basal and lateral views). This could be done by looking at the patterns (features) extracted from the images and by comparing it to the patterns generated for the normal USIT.

Based on these two views, the model will be trained to report as:

- Normal: If both the views are normal
- Abnormal: If either of the views are not normal

Only the abnormal cases (where CHD is found) need to be taken into consideration for further screening. If the image is diagnosed with anomaly, next-level screening will be done. Here, the model evaluates the six views together with the situs, four-chamber view, LV outflow tract, RV outflow tract, three vessel and three vessel tracheal view. All these will be implemented using the feature vectors extracted from the images. The algorithm will be trained to recognize the patterns [16] in the most common forms of major CHDs found in the UTIs and it will offer a possible diagnosis. Table1 lists the CHDs and gives an overview of the frequency of prevalence of each one.

4.2 Phase 2: Diagnostic Model

The Phase 2, Diagnostic model will be the actual clinical diagnosis system where the live images could be supplied to the trained model and the model will report (Predict) the exact anomaly (ies).

Figure 4 explains the overall workflow of the proposed artificially intelligent system. The major advantage of this automated system will be, once the system is ready after clinical experiments, there might not be changes in the model until unless

Table 1 List of Heart anomalies and its incidence

S. No.	Defect	Incidence	Occurance
1	Ventricular septal defect	Most Common	1.5–3.5 per 1000 live births
2	Atrial septal defect	Fifth most common	1 of 1500 live births
3	AV septal defect	2–7% of All CHD cases	0.19–0.56 per 1000 live births
4	Tetralogy of Fallot	5–10% of All CHD cases	0.24–0.56 per 1000 live births
5	Truncus Arteriosus	1–2% of All CHD cases	NA
6	Transposition of great arteries	1% of All CHD cases	NA
7	Single ventricle	2% of All CHD cases	NA
8	Double outlet RV	less than 1% of All CHD cases	0.08–0.16 per 1000 live births
9	Hypoplastic right heart syndrome	Rare	1.1% of stillbirths
10	Hypoplastic left heart syndrome	2–4% of All CHD cases	0.16–0.25 per 1000
11	Aortic coarctation or hypoplasia	6–7% of All CHD cases	
12	Aortic Atresia or stenosis	Rare	5.2% in newborns
13	Pulmonary stenosis	Rare	7.4% of newborns
14	Ebstein anomaly	less than 1% of All CHD cases	1 per 20,000 live births
15	Ectopia cordis and pentalogy of Cantrell	Rare	14–18 days gestational age
16	Cardiomyopathies	8–11% of All CHD cases	8–11%
17	Arrhythmia	2% of All CHD cases	NA

a new CHD is discovered in the medical field. The logic could be attached with any imaging unit with minor customizations.

5 Results and Discussions

There are many researches happening with regard to medical images majorly for diagnosing, for guiding interventions, for planning a surgery, etc. Table2, lists the latest researches conducted for diagnosing of diseases or structural abnormalities from ultrasound scan images using various ML/DL models. For all the listed researches, the image dataset are ultrasound images.

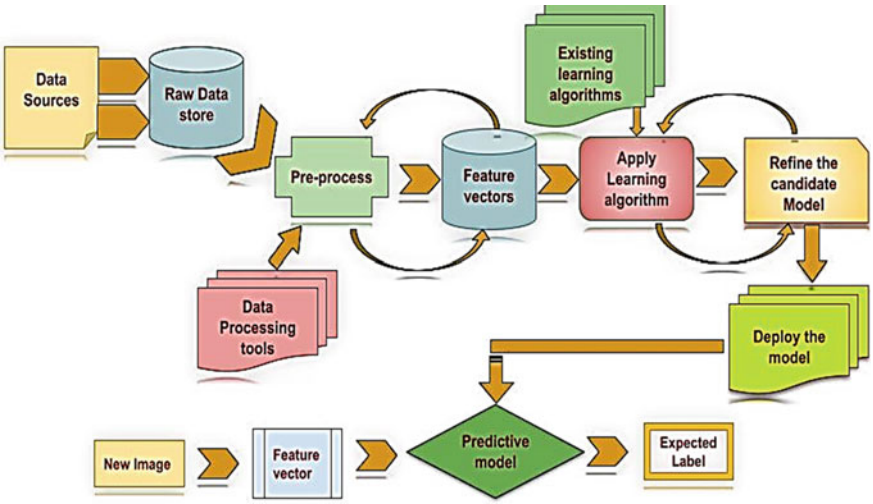


Fig. 4 Proposed CHD diagnosis model building and its implementation

Wei et al. in their research [9], a combination of SVM classifier, naïve bayes classifier and KNN are used which resulted in 91.1% accuracy. Liang et al. [11] using GLCM feature, statistical tastings were done to diagnose breast cancer. The accuracy for the method was 81%. Chollet et al. [17] in their research to detect thyroid nodule, CNN and RCNN were used which has an accuracy of 96.16%. Chi et al. [12] using random forest classifier for thyroid nodules classification got an accuracy of 98.29%. Abdul et al. for classifying breast tumours used co-occurrence matrix which showed an accuracy of 99%. Abdel et al. [13] used GLCM and co-occurrence matrix for classifying breast tumours which showed 99% accuracy. Komatsu et al. [18] in their research used CNN for identifying the structural anomalies with accuracy 70%. Song et al. in his research used CNN for identifying and recognizing thyroid nodules which was showing 98% accuracy. Ding et al. in their research used self-organizing maps along with CNN and bag of words concepts together for breast tumour classification. This showed an accuracy of 91.07%. Virmanj et al. [19] in their research SVM classifier was used with texture feature classifying liver ultrasound images. This showed an accuracy of 88.8%.

The literatures considered for this review includes two categories, the medical articles which prove the requirement of technological interventions in foetal cardiac anomaly diagnosis and the second set where different models were suggested for similar diagnosis using ultrasound scan images. While scanning through published articles that suggests the different models for different diagnosis, most of them fall under the following categories.

- Breast tumour classification
- Thyroid nodule classification

Table 2 The overview of performance of various ML/DL models used in different research used for various diagnosis

S. No.	Date of publishing	Reference index	Title	organ or body location	Application	Technique used	Accuracy
1	Jun-21	[8]	Breast ultrasound image classification based on multiple instance learning	Breast	CNN SOM	Shape	0.9107
2	Jan-21	[18]	Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning	Heart	CNN	Supervised object detection with Normal Data only	0.7
3	Oct-20	[9]	A benign and malignant breast tumour classification method via efficiently combining texture and morphological features on ultrasound images	Breast tumor diagnosis	Classification LBP, HOG, GLCM, morphological feature	SVM classifier, naïve bayes classifier, KNN, DT, LDA	91.11
4	Apr-20	[11]	Classification of Breast ultrasound tomography by using textural analysis	Breast tumor diagnosis	GLCM, Statical methods, t-test, mean, average etc.	Laplacian of Gaussian (LoG) method, texture images were calculated using box-counting	0.81
5	Mar-20	[17]	Thyroid nodule ultrasound image classification through hybrid feature cropping network	Thyroid	Hybrid feature cropping network, CNN	CNN, RCNN, Boundary feature cropping (RFC),Random feature cropping (RFC)	96.13

(continued)

Table 2 (continued)

S. No.	Date of publishing	Reference index	Title	organ or body location	Application	Technique used	Accuracy
6	Jul-17	[12]	Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network	Thyroid	GoogleNet model	Random forest classifier	98.29
7	Aug-16	[13]	Breast tumour classification in ultrasound images using texture analysis and super-resolution methods	Breast	GLCM, LBP, phase congruency based LBP, HOG, Pattern lacunarity system	Co-occurrence matrix	0.99
8	Aug-15	[1]	Multi task cascade convolution neural networks for automatic thyroid nodule detection and recognition	Thyroid	CNN	Texture	0.98
9	Oct-12	[19]	SVM based characterization of liver ultrasound images using wavelet packet texture descriptor	Liver	SVM	Texture	0.888

- Brain tumour classifications
- Liver-related diagnosis.

Though the medical literatures clearly prove the requirement of technological inventions in foetal cardiac anomaly detection, the interesting part noticed here was after long search for literatures suggesting models/Algorithm which propose any intervention in foetal cardiac anomaly detection were very few. There can be another fact that keep researchers away from this because there is no standard medical database available for this research. Getting the real patient images are also a challenge as there will be long procedure to undergo. The ethical issues related to the confidentiality of patient data can be a hurdle to get data from clinics. Even then it can be collected with not stop efforts and if the researcher could prove his/her genuine interest to solve a social issue. There are concepts like crowdsourcing which could also help developing the dataset. The advancements in IoT can supplement the implementation of the model in smart gadgets for live monitoring of patients. This projects the strong scope of this area.

6 Conclusion

This review paper projects the importance of accuracy in the cardiac anomaly diagnosis and the positive revolutions that artificial intelligence can bring in this field. Any compromise on the prenatal diagnosis may result with the death of the baby or with a permanent damage that affects the whole family throughout their lives. The various research quoted above has modelled several mechanisms for diagnosing various abnormalities in different organs, majorly breast and thyroid. The outcome of these research shows the possibilities of similar models for diagnosing CHDs as well. The outcome is going to be revolutionary because the postnatal treatments can be well planned in advance and will result in promising healthy life for humankind.

References

1. W. Song, S. Li, J. Liu, H. Qin, B. Zhang, S. Zhang, A. Hao, Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition. IEEE J. Biomed. Health Inform. **23**(3), 1215–1224 (2018)
2. L.I. Parikh, K.L. Grantz, S.N. Iqbal, C.C. Huang, H.J. Landy, M.H. Fries, U.M. Reddy, Neonatal outcomes in fetuses with cardiac anomalies and the impact of delivery route. Am. J. Obstet. Gynecol. **217**(4), 469-e1 (2017)
3. G. Carneiro, J.C. Nascimento, A. Freitas, The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. IEEE Trans. Image Process. **21**(3), 968–982 (2011)
4. Suard, C., Flori, A., Paoli, F., Loundou, A., Fouilloux, V., Sigaudy, S., F. Michel, J. Antomarchi, P. Moceri, V. Paquis-Flucklinger, C. D'ercole, F. Bretelle, Accuracy of prenatal screening for

congenital heart disease in population: a retrospective study in Southern France. PloS one, **15**(10), e0239476 (2020)
5. B.J. Holland, J.A. Myers, C.R. Woods Jr., Prenatal diagnosis of critical congenital heart disease reduces risk of death from cardiovascular compromise prior to planned neonatal cardiac surgery: a meta-analysis. Ultrasound Obstet. Gynecol. **45**(6), 631–638 (2015)
6. T.D. Changlani, A. Jose, A. Sudhakar, R. Rojal, R. Kunjikutty, B. Vaidyanathan, Outcomes of infants with prenatally diagnosed congenital heart disease delivered in a tertiary-care pediatric cardiac facility. Indian Pediatr. **52**(10), 852–856 (2015)
7. A. Vijayaraghavan, A. Sudhakar, K.R. Sundaram, R.K. Kumar, B. Vaidyanathan, Prenatal diagnosis and planned peri-partum care as a strategy to improve pre-operative status in neonates with critical CHDs in low-resource settings: a prospective study. Cardiol. Young **29**(12), 1481–1488 (2019)
8. J. Ding, H.D. Cheng, J. Huang, J. Liu, Y. Zhang, Breast ultrasound image classification based on multiple-instance learning. J. Digit. Imaging **25**(5), 620–627 (2012)
9. M. Wei, Y. Du, X. Wu, Q. Su, J. Zhu, L. Zheng, G. Lv, Zhuang, J., A benign and malignant breast tumour classification method via efficiently combining texture and morphological features on ultrasound images. Comput. Math. Methods Med. (2020)
10. B. Liu, H.D. Cheng, J. Huang, J. Tian, X. Tang, J. Liu, Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. Pattern Recogn. **43**(1), 280–298 (2010)
11. C.Y. Liang, T.B. Chen, N.H. Lu, Y.C. Shen, K.Y. Liu, S.Y. Hsu, C.J. Tsai, Y.M. Wang, C.I. Chen, W.C. Du, Y.H. Huang, Classification of breast ultrasound tomography by using textural analysis. Iran. J. Radiol. **17**(2), (2020)
12. J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, M. Eramian, Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. J. Digit. Imaging **30**(4), 477–486 (2017)
13. M. Abdel-Nasser, J. Melendez, A. Moreno, O.A. Omer, D. Puig, Breast tumor classification in ultrasound images using texture analysis and super-resolution methods. Eng. Appl. Artif. Intell. **59**, 84–92 (2017)
14. X. Gao, W. Li, M. Loomes, L. Wang, A fused deep learning architecture for viewpoint classification of echocardiography. Inf. Fusion **36**, 103–113 (2017)
15. Liu, J. E., & An, F. P. (2020). Image classification algorithm based on deep learning-kernel function. *Scientific programming*, *2020*.
16. F. Chollet, Xception: deep learning with depth wise separable convolutions, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 1251–1258
17. R. Song, L. Zhang, C. Zhu, J. Liu, J. Yang, T. Zhang, Thyroid nodule ultrasound image classification through hybrid feature cropping network. IEEE Access **8**, 64064–64074 (2020)
18. M. Komatsu, A. Sakai, R. Komatsu, R. Matsuoka, S. Yasutomi, K. Shozu, A. Dozen, H. Machino, H. Hidaka, T. Arakaki, K. Asada, R. Hamamoto, Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning. Appl. Sci. **11**(1), 371 (2021)
19. J. Virmani, V. Kumar, N. Kalra, N. Khandelwal, SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors. J. Digit. Imaging **26**(3), 530–543 (2013)

DLoader: Migration of Data from SQL to NoSQL Databases

Kanchana Rajaram, Pankaj Sharma, and S. Selvakumar

Abstract Data is increasing exponentially in the modern world which requires more proficiency from the available technologies of data storage and data processing. This continuous growth in the amount of structured, semi-structured, and unstructured data is called as big data. The storage and processing of big data through traditional relational database systems are not possible due to increased complexity and volume. Due to improved expertise of big data solutions in handling data, such as NoSQL caused the developers in the previous decade to start preferring big data databases, such as Apache Cassandra, MongoDB, and NoSQL. NoSQL is a modern database technology designed for fast read and write operations and provides horizontal scalability to store large amount of voluminous data. Large organizations face various challenges to shift their relational database framework to NoSQL database framework. In this paper, we proposed an approach to migrate the data from a relational database to the NoSQL database. We have specifically done transformation for Cassandra and MongoDB from MySQL database. The experiments show that the proposed approach successfully transforms the relational database to a big data database, and the performance analysis of such transformed databases shows that Cassandra database requires less storage space and offers a better performance.

Keywords Big data · Horizontal scalability · Relational database · Transformation · Cassandra · MongoDB

K. Rajaram (✉)
Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India
e-mail: rkanch@ssn.edu.in

P. Sharma · S. Selvakumar
IIIT Una, Una, Himachal Pradesh, India
e-mail: pankajk27897@gmail.com

S. Selvakumar
e-mail: director@iiitu.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_19

1 Introduction

Relational databases have been the top choice of organizations in the last decade for storing, processing, and analyzing the data generated in the organizations. Relational databases store structured data and support structured query language (SQL) to access the database [1]. When the data is in structured format and of low volume, relational databases can be comfortably used. Even the unstructured or semi-structured data can also be stored in relational databases after using ETL tools to convert into structured format. However, recently, the data is increasing at an exponential rate having huge volume, high velocity of data generation, varied variety of the data as big data [2]. With this, it has become almost impossible for relational databases to store and process this huge amount of big data generated by large organizations and social media. Moreover, social networking and cloud computing paradigms have become popular which require data stores to manage massive amount of data generated per second. In particular, the cost of storing and querying big data from relational databases is very high, and they cannot serve the requirement of millions of users at the same time.

The big data storage and processing are not much suitable with relational databases as it has structured, semi-structured, and unstructured data, whereas relational databases can store only structured data and provide very little support to unstructured and semi-structured data [3]. Considering the challenges of big data for relational databases, the modern larger organizations which have large data storage requirements are rapidly switching from existing relational databases to NoSQL databases. NoSQL organizes data in different formats such as key-value, columnar, documents, and graphs. Cassandra stores data in columnar format, which offers considerably fast write operations [4], whereas MongoDB [5] stores data in document format. Every data type of NoSQL database uses different data structures as per the requirements.

The successful handling of big data complexity is a great challenge for the conversion of relational databases to NoSQL databases. NoSQL databases are designed according to the application specific access patterns and queries without using a normalization process, and they do not support any join operation and foreign or primary key relations. It is challenging for the organizations to migrate data from the relation databases to NoSQL databases because they completely work on different technologies. To overcome this problem, we have proposed a framework, namely DLoader, to migrate data from relational database such as MySQL database to NoSQL databases such as Cassandra and MongoDB. The proposed approach DLoader involves extraction of data from the MySQL database, preprocessing the data by applying transformations and standardizations on the data, and finally mapping the columns in the MySQL tables into the fields of NoSQL databases.

The rest of the paper is structured as follows: Review of the related work is presented in Sects. 2 and 3 describes the proposed work; Sect. 4 describes about the experimentation with Cassandra and MongoDB. Lastly, Sect. 5 concludes the work and suggests a future work.

2 Literature Review

NoSQL is a modern database technology to store and process big data. Most of the utilities developed were developed for converting one form of SQL database to another SQL database, and no specific utility is available for NoSQL to NoSQL conversion. A data adapter system was proposed to promote hybrid database architecture including both SQL and NoSQL [6]. Many frameworks have been developed to solve the data migration problem with a different solution, characteristics, and properties which migrates data from one NoSQL database to another but not between SQL to NoSQL [7]. It provided algorithm migration and migration schemes to migrate data between NoSQL databases in actual operating environment. It is a challenging task for business organizations to migrate or transform their data from existing relational database to NoSQL databases due to the complexity in relational data.

An approach consists of two modules: Data transformation and data cleansing modules were proposed which transforms relational database to big data database [8]. To handle the complexity of automatic transformation of existing relational database into a NoSQL database, a bi-fold transformation consisting of schema-to-schema and data-to-data transformation approach was proposed which dealt with heterogeneous and complex data [9]. Heterogeneous data exchange and conversion of data across kinds of database systems are achieved through relationship schema mapping [10]. It specifically focused on exchanging XML heterogeneous, solving field attribute changing during migration problem.

A study revealed that NoSQL is faster than RDBMS in case of big data. A methodology was presented for data migration from MySQL to MongoDB as NoSQL database [11]. A study evaluated the performance of data insertion and retrieval speeds up by making comparison between MongoDB as NoSQL and MySQL as a relational database which showed NoSQL database is faster than the MySQL considering the parameters used in the study for huge amount of data [12]. An approach that used data and query features to migrate data from relational to NoSQL databases preserved the features in the source data in the relational database and queries for accessing the data on the source data as well for arranging the data in the target database [13, 14]. This system works for all NoSQL databases but requires meta data information of the source and target databases.

NoSQL layer can also be used as an interface resides between the source database and application to migrate the data between relational to NoSQL without changing the application code which gives high performance for MongoDB [15]. Other approaches include data adapter approach which integrates relational database and NoSQL databases which support hybrid database architecture. Data may not be consistent in this approach always [16]. Content management system for schema de-normalization works with the Hadoop framework [17]. In a study, document-oriented data schema was proposed covering all data types in databases, further it overcomes the issue of managing the relationships of a complex database. Two stages of the approach include designing the document-oriented data schema and migrating the ER model to the document-oriented data schema [18]. Data migrated from relational to NoSQL

data models or between different NoSQL platforms needed to be validated in order to find the errors during transformation during migration of the data. Data can be validated using denormalized schema structures and bloom filters, and other approaches can also be implemented to validate the data [19].

3 Proposed Work

The concepts of relational databases and NoSQL databases and the differences between these two databases are discussed in the next two sub-sections. Thereafter, the proposed approach for transforming the data from relational databases to NoSQL databases is elaborated.

3.1 Relational Databases

Relational or structural databases are based on relational model. Data stored in these tables is related and can be accessed through the relationships between them. The relationships can be one-to-one, one-to-many, and many-to-many. Data stored in the relational databases is structured only. Relational databases consist of tables, and the tables consist of rows and columns. There can be number of rows in a table, and all the rows are of same type that contains different data. Relational database management system (RDBMS) is used for creation and management of the relational databases. Each row in a table is unique. There are mainly two types of keys in RDBMS: primary key and foreign key. Primary key consists of one or more columns which can uniquely identify all the records in a table, and foreign key consists of one or more columns of a table that refers to the primary key of another table. Relational databases also follow ACID properties (atomicity, consistency, isolation, and durability) for making transactions. For querying the data from the relational models, structured query language (SQL) is used. These databases should have a predefined schema which is not easy to alter if structure of the data coming into the database changes with time. Hence, schema must be strictly designed keeping the future requirements in the mind.

3.2 NoSQL Databases

NoSQL databases have been designed to facilitate big data storage and processing. NoSQL databases store data in various formats such as key-value, column store, documents, and graphs. They have their own query mechanism and do not have any specific query language. Key-value databases consist of items where each item contains keys and values. A value can be simply retrieved by referring its key. Columnar databases

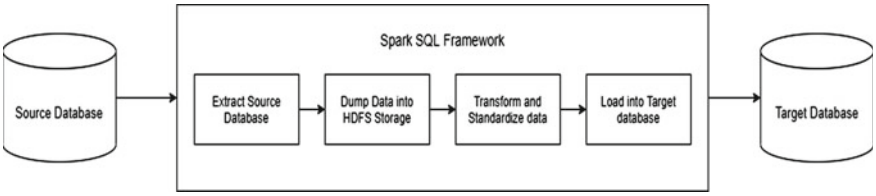


Fig. 1 SQL to NoSQL data migration

consist of tables which consist of rows and dynamic columns. They have flexible schema and number of columns can be increased with time without creating a new schema. The column names need not be predefined in NoSQL databases, and hence, the structure is not fixed. Columns in a row are kept in a sorted order according to their keys which include partition keys and clustering keys. Document type NoSQL databases store data in documents which is same as JSON objects. The collection of documents is called a collection. Each document contains key-value pairs where value can be of different types. Graph databases store data in nodes and edges. Nodes generally hold information about entities, whereas edges store information about the relationship between the nodes. NoSQL databases have denormalized structure as they do not have primary and foreign key concept and hence do not support join operations. They provide features like horizontal scalability, data replication, data availability, and data consistency at low cost. With the increasing amount of data, the NoSQL database can be easily scaled, and more data storage can be accommodated [20].

3.3 SQL to NoSQL Transformation

Data stored in the relational database in the tables must be mapped to NoSQL database tables column to column. We have used Spark framework for migrating data from SQL to NoSQL database [21]. Hadoop cluster is serves as a middle layer between the source and target database. NoSQL database such as Cassandra also supports collection data types such as set, map, and list, and tables are denormalized, so columns in the SQL database can also be mapped to collection data types. The detailed process is briefed in the further sub-sections. The steps in migrating the data from a relational database to a NoSQL database are shown in Fig. 1.

3.3.1 Data Loading

Data in the relational database stored in MySQL holds a specific schema in which various tables are linked through primary and foreign key references. Each column in a table has a defined data type. We have made use of timestamps to get the data

from the MySQL database. Timestamp is used to make sure that only recently added data gets retrieved. JDBC is used to establish connection with the source database. The data from the MySQL database is loaded into the Hadoop distributed file storage (HDFS) [22] using Spark SQL framework. HDFS is a clustered storage consisting of one name node and N data nodes. This Hadoop cluster acts as a middle layer between the source database (MySQL) and the target database (NoSQL). SQL like queries can also be made on the data in the Spark data frames by creating temporary tables by using Spark SQL. Spark SQL also supports join and aggregate functions as supported in the SQL. The incoming data form the source database is dumped in the HDFS keeping the schema, the same as that of source database. Data is temporarily stored in the HDFS as staging tables before applying processing the source data.

3.3.2 Data Transformation and Mapping

HDFS holds large amount of data dumped from the relational databases (MySQL) in delta file format without altering the schema of the source database. This data is loaded in a data frame of the Spark SQL framework, one table at a time. Transformations are applied on the table data contained in the data frame, keeping the target database schema in mind. Since the target NoSQL database does not support joins, the MySQL source tables are denormalized to enable migration of data from the relational database. Every column of a table in the source database is transformed according to the column schema of the target database. Different concepts in relational databases are mapped to NoSQL databases as shown in Fig. 2.

Transformation includes changing columns names, changing data type of columns according to target database, altering the tables by adding, or removing the columns according to the target database schema. Standardization includes ensuring consistent date formats, uniform format for name values, uniform abbreviations, for example, for gender column value, consistent format for mobile number values, etc. Finally, the transformed data frame is loaded into a particular table of target NoSQL database in append mode using Spark SQL. The algorithm for preprocessing and mapping the source data to target is given below:

Algorithm 1 Data preprocessing and mapping

```
Algorithm Data preprocessing mapping
Input:        Relational database
Output:       NoSQL database
1. Establish a connection with source database.
2. Create table object for source.
3. Dump the table objects in the HDFS storage.
4. Load the data from all the tables into HDFS storage
5. For each column of each of the tables, if transformation rule
is applicable: Transform the columns according to the target table
columns.
6. Standardize the content of the columns.
7. Map the HDFS table objects into target table objects.
8. Map the HDFS column objects into target column objects.
```

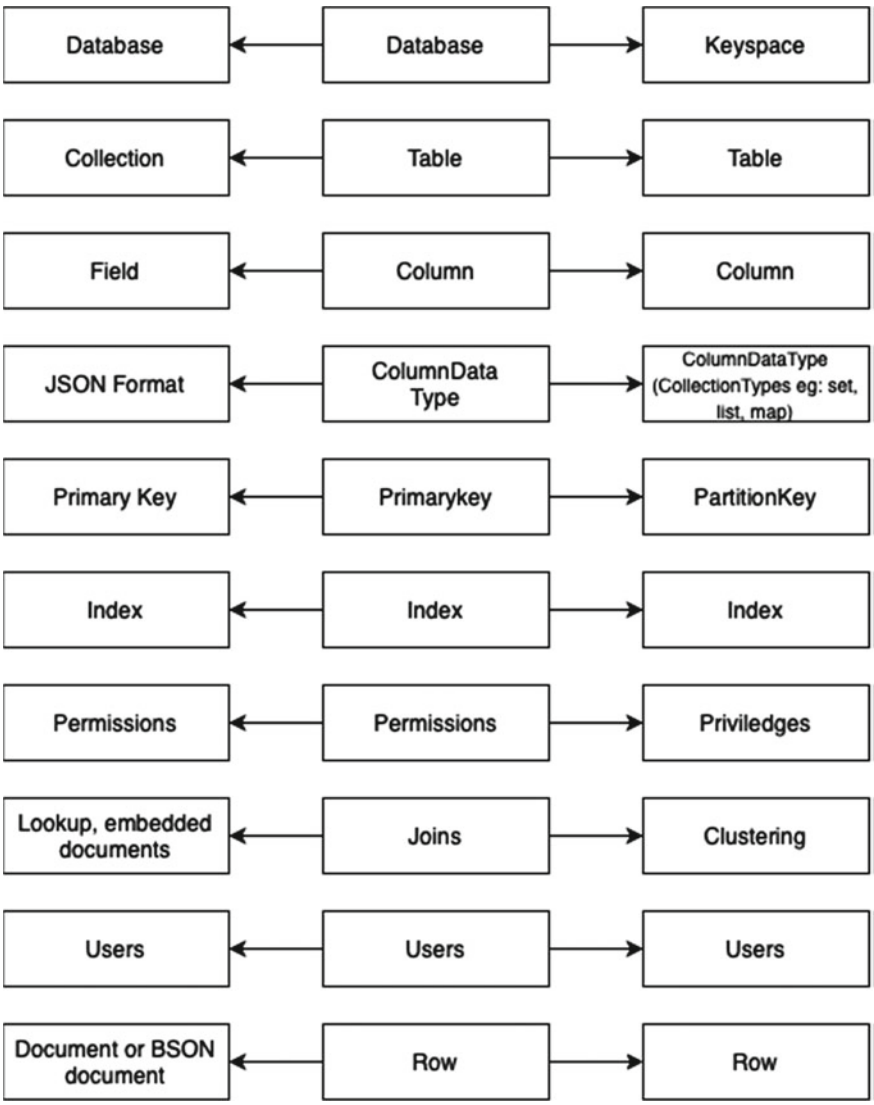


Fig. 2 Concept mapping between relational database and NoSQL database

9. Construct each target table.
10. Construct target object of each column.
11. Select partition key and clustering key for each target table..
12. Compare partition key of the target tables to keep track of the duplicate records.

4 Experimentation

For experimentation purpose, a test bed comprising of a HDFS-based cluster with a name node and three data nodes has been set up. The name node is a Intel Xeon server 3.3 GHz, 32 GB RAM, 4 Cores with 2 TB HDD, and the data nodes are Intel i-4 core workstations with 16 GB RAM and 1 TB HDD. Our own dataset has been generated consisting of immunization data related to children of Tamil Nadu state. The average population of Tamil Nadu state is 7.7 crore in 2020 [23], and around 30% of them are children. The number of children approximately in a state could be 2.3 crores. The immunization database consists of tables for storing details of children, their mothers, and immunization details. The data is generated using a tool called DbGen [24] and using MS Excel. DbGen is a Windows-based tool that can be configured based on the schema to generate data.

The immunization data in the MySQL database is migrated using our proposed approach of DLoader into two NoSQL databases such as Cassandra and MongoDB. For a given number of children records, the associated mother records and immunization records are also considered to be migrated from the source database. When loading the same number of records, the database sizes of Cassandra and MongoDB are different. The number of records used in the dataset along with the varied size of Cassandra and MongoDB databases are shown in Table 1. The number of children records has been varied from 10 lakh to 2.5 crores. It is observed that Cassandra offers good compression of data and stores the same number of records in the source database in storage space of 41% lesser when compared to MongoDB.

As a second experiment, the performance of both NoSQL databases is analyzed in terms of response time and throughput. The load testing has been done using Jmeter tool [25]. The following two queries have been considered.

Q1. Retrieve the details of a particular child whose *child_id* is given.

Table 1 Varied sizes of NoSQL databases for storing given relational data

No. of children	Cassandra DB				MongoDB			
	Child master DB Size (GB)	Mother master DB size (GB)	Imm. DB size (GB)	Total DB size (GB)	Child master DB Size (GB)	Mother master DB size (GB)	Imm. DB size (GB)	Total DB size (GB)
1,000,000	0.07	0.1	0.1	0.3	0.07	0.2	0.2	0.5
2,000,000	0.1	0.3	0.1	0.7	0.1	0.4	0.3	0.9
3,000,000	0.2	0.5	0.2	1.0	0.2	0.7	0.7	1.3
5,000,000	0.4	0.9	0.4	1.5	0.3	1.0	1.0	2.1
10,000,000	0.9	1.8	0.8	3.5	0.7	2.0	2.0	4.7
15,000,000	1.2	2.7	1.2	5.0	1.1	3.0	3.0	7.1
20,000,000	1.6	3.6	1.8	6.8	1.3	4.0	4.0	9.3
25,000,000	2.1	4.5	2.0	8.0	1.8	5.0	5.0	11.8

Table 2 Performance analysis of Cassandra and MongoDB

S. No.	Cassandra DB					MongoDB				
	DB size (GB)	Throughput		Response time (s)		DB size (GB)	Throughput		Response time (s)	
		Q1	Q2	Q1	Q2		Q1	Q2	Q1	Q2
1	0.3	0.44	0.34	2.29	2.95	0.5	0.39	0.17	2.54	5.98
2	0.7	0.44	0.33	3.06	3.01	0.9	0.33	0.15	3.09	6.60
3	1.0	0.43	0.33	3.40	4.42	1.3	0.31	0.14	4.01	9.57
4	1.5	0.33	0.32	4.08	5.01	2.1	0.29	0.14	4.80	11.02
5	3.5	0.29	0.26	8.07	10.05	4.7	0.22	0.10	9.32	20.04
6	5.0	0.21	0.19	10.80	18.02	7.1	0.19	0.06	13.50	32.00
7	6.8	0.15	0.12	15.00	22.05	9.3	0.12	0.03	20.04	40.30
8	8.0	0.10	0.08	20.97	26.45	11.8	0.09	0.02	30.50	50.02

Q2. Retrieve immunization details of a children in a particular routine immunization session.

The data load is varied from 0.3 to 8 GB of source MySQL database and migrated to Cassandra as well as MongoDB. For each of the queries Q1 and Q2, the throughput and response time are computed and shown in Table 2. The queries were executed three times under different network loads. The average throughput and maximum response time in seconds for Cassandra and MongoDB databases among three runs are shown in the table.

The performance of Cassandra and MongoDB by varying the data load for two different queries is shown in Fig. 3. The following observations are made from this experiment.

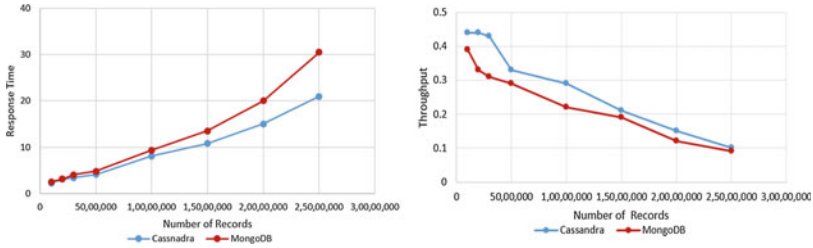
For query Q1

- For 26-fold increase in Cassandra database size, there is only ninefold increase in Response time and throughput decreases by 77%.
- For 26-fold increase in MongoDB size, there is only 12-fold increase in response time and throughput decreases by 77%.

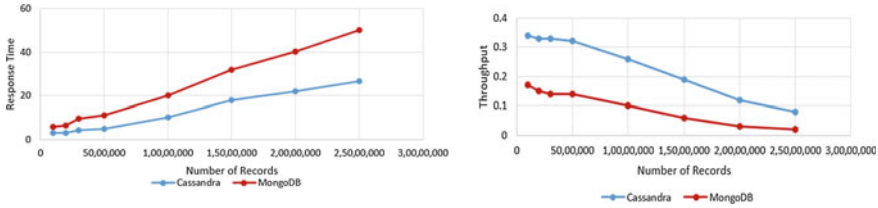
For query Q2

- For 26-fold increase in Cassandra database size, there is 12-fold increase in response time and throughput decreases by 76%.
- For 26-fold increase in MongoDB size, there is eightfold increase in response time and throughput decreases by 88%.

The rate of increase in response time is more in case of Cassandra for both the queries when compared to MongoDB. However, in Cassandra, the response time for queries Q1 and Q2 for any database size is lesser by 16% and 50%, respectively, as compared to MongoDB.



(a) For Query 1



(b) For Query 2

Fig. 3 Performance analysis of Cassandra and MongoDB

5 Conclusion

The proposed system DLoader for migration of data from SQL to NoSQL database is the generalized system which is capable to migrating any schema in SQL to NoSQL database. Spark framework which provides fast performance through its inbuilt parallel processing is used for data extraction, transformation, standardization, and data loading. Spark cluster-based HDFS storage is used in the middle layer to handle large amount of data at a time and makes the data readily available during the transformation and mapping process. The proposed approach has achieved column to column mapping from source to target database. The proposed approach is tested for migrating relational data to NoSQL databases such as Cassandra and MongoDB. Cassandra consumes less storage space and offers better performance as compared to MongoDB.

Acknowledgements This work was supported by Grand Challenges India (GCI) for Immunization Data: Innovating for Action (IDIA) funded by BIRAC and jointly funded by Department of Biotechnology and Bill & Melinda Gates foundation.

References

1. S. Ramzan, I.S. Bajwa, B. Ramzan, W. Anwar, Intelligent data engineering for migration to NoSQL based secure environments. IEEE Access **7**, 69042–69057 (2019)
2. A. Katal, M. Wazid, R.H. Goudar, Big data: issues, challenges, tools and good practices, in *2013 Sixth international conference on contemporary computing (IC3)* (IEEE, 2013), pp. 404–409
3. M. Potey, M. Digrase, G. Deshmukh, M. Nerkar, Database migration from structured database to non-structured database. Int. J. Comput. Appl. **975**, 8887 (2015)
4. V.D. Jogi, A. Sinha, Performance evaluation of MySQL, Cassandra and HBase for heavy write operation, in *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)* (IEEE, 2016), pp. 586–590
5. MongoDB (2021) https://www.mongodb.com/
6. S. Ghule, R. Vadali, Transformation of SQL system to NoSQL system and performing data analytics using SVM, in *2017 International Conference on Trends in Electronics and Informatics (ICEI)* (IEEE, 2017), pp. 883–887
7. Y.S. Wijaya, A.A. Arman, A framework for data migration between different datastore of NoSQL database, in *2018 International Conference on ICT for Smart Society (ICISS)* (IEEE, 2018), pp. 1–6
8. A.E. Lotfy, A.I. Saleh, H.A. El-Ghareeb, H.A. Ali, A middle layer solution to support ACID properties for NoSQL databases. J. King Saud Univ.-Comput. Inf. Sci. **28**(1), 133–145 (2016)
9. S. Ramzan, I.S. Bajwa, R. Kazmi, An intelligent approach for handling complexity by migrating from conventional databases to big data. Symmetry **10**(12), 698 (2018)
10. N. Li, B. Xu, X. Zhao, Z. Deng, Database conversion based on relationship schema mapping, in *2011 International Conference on Internet Technology and Applications* (IEEE, 2011), pp. 1–5
11. M. Hanine, A. Bendarag, O. Boutkhoum, Data migration methodology from relational to NoSQL databases. World Acad. Sci. Eng. Technol. Int. J. Comput. Electr. Autom. Control Inf. Eng. **9**(12), 2369–2373 (2016)
12. A. Abdullah, Q. Zhuge, From relational databases to NoSQL databases: performance evaluation. Res. J. Appl. Sci. Eng. Technol. **11**(4), 434–439 (2015)
13. S. Ghotiya, J. Mandal, S. Kandasamy, Migration from relational to NoSQL database, in *IOP Conference Series: Materials Science and Engineering*, vol. 263, no. 4 (IOP Publishing, 2017), p. 042055
14. D. Liang, Y. Lin, G. Ding, Mid-model design used in model transition and data migration between relational databases and nosql databases, in *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, (IEEE, 2015), pp. 866–869
15. L. Rocha, F. Vale, E. Cirilo, D. Barbosa, F. Mourão, A framework for migrating relational datasets to NoSQL. Procedia Comput. Sci. **51**, 2593–2602 (2015)
16. Y.T. Liao, J. Zhou, C.H. Lu, S.C. Chen, C.H. Hsu, W. Chen, M.F. Jiang, Y.C. Chung, Data adapter for querying and transformation between SQL and NoSQL database. Future Gener. Comput. Syst. **65**, 111–121 (2016)
17. C.H. Lee, Y.L. Zheng, SQL-to-NoSQL schema denormalization and migration: a study on content management systems, in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, (IEEE, 2015), pp. 2022–2026
18. S. Hamouda, Z. Zainol, Document-oriented data schema for relational database migration to NoSQL, in *2017 International Conference on Big Data Innovations and Applications (Innovate-data)* (IEEE, 2017), pp. 43–50
19. A. Goyal, A. Swaminathan, R. Pande, V. Attar, Cross platform (RDBMS to NoSQL) database validation tool using bloom filter, in *2016 International Conference on Recent Trends in Information Technology (ICRTIT)* (IEEE, 2016), pp. 1–5
20. Y. Huang, T.J. Luo, Nosql database: a scalable, availability, high performance storage for big data, in *Joint International Conference on Pervasive Computing and the Networked World* (Springer, Cham, 2013), pp. 172–183
21. K. Atkotiya, P. Shukla, Migration from relational database like MySQL to nosql database like Cassandra is necessary and how to migrate it using spark

22. Apache Hadoop (2021) http://hadoop.apache.org/
23. Rural Health Statistics https://hmis.nhp.gov.in/downloadfile?filepath=publications/Rural-Health-statistics/RHS%202019-20.pdf
24. DbGen Database Management Tool (2021) https://www.bcdsoftware.com/iseries400solutions/dbgen/
25. Apache Jmeter (2021) https://jmeter.apache.org

Maximum Boost Control of Quasi-Z-Source Inverter with DSTATCOM for a Wind Energy System

P. V. S. S. A. Parimala and Ribu Mathew

Abstract Maximum boost control method of a single-phase enhanced quasi-Z-source inverter with distribution static synchronous compensator (DSTATCOM) for a 0.5 kW wind energy system is presented in this paper. Generation of electrical energy from wind being renewable source is very effective as far as efficiency, reliability, and cost are considered. A DSTATCOM is used to improve power quality of the system. A 0.5 kW permanent magnet synchronous generator (PMSG) employed wind energy system is used as a source. The enhanced quasi-Z-source converter has common ground with the inverter and source with which it draws continuous current, eliminates starting inrush current problem (which is one of the major concerns), and also reduces the stress of voltage at the capacitors. The voltage of the system has been increased with the help of quasi-Z-source inverter. The implemented method of maximum boost control provides maximum gain in voltage for a given M (modulation index), and it does not produce ripple of lower frequency.

Keywords Wind energy system · Enhanced quasi-Z-source inverter · DSTATCOM · Renewable energy

1 Introduction

Extraction of electrical energy from wind is an economical methodology and an alternative to fossil fuel-based energy. The recent advancements of wind energy technology have proven to be reliable and sustainable source of energy that provides a significant contribution in fulfilling the local electrical energy demand. On the other hand, penetration of renewable energy sources leads to stability, flexibility, and adequacy issues. Stability which belongs to power quality has been addressed using

P. V. S. S. A. Parimala (✉)
School of Electrical and Electronics Engineering, G. Narayanamma Institute of Technology and Science, Hyderabad, India
e-mail: parimala2019@vitbhopal.ac.in

R. Mathew
School of Electrical and Electronics Engineering, VIT Bhopal University, Bhopal, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_20

DSTATCOM. The DSTATCOM has the ability to provide continuous and quick inductive and capacitive mode compensation. The synchronous reference frame theory (SRFT)-based DSTATCOM is used to reduce harmonics which is a major issue of concern in each and every electrical system, reactive power compensation, improve power factor. The management of reactive power leads to enhancement of power factor. To maintain current constant, hysteresis control is employed, and to maintain a constant voltage, the controller such as PI is used.

The energy obtained from the renewable energy sources is usually very less, so generally, a transformer is used to increase the voltage value which may increase the space occupied and also increases the amount incurred in erecting the system. In this paper, an enhanced boost quasi-Z-source inverter with maximum boost control strategy is implemented to eliminate the transformer for increasing the voltage. In the literature, many pulse width modulation techniques have been discussed for traditional voltage source converters, where in the output, voltage gain is dependent on the modulation index. Strength and quality of the signal are generally determined by the modulation index. The stress in voltage at the capacitor is reduced by maximum boost control strategy providing required gain in the voltage.

2 Operation of Maximum Boost Control of Single-Phase Enhanced Quasi-Z-Source

The schematic representation of the system under study is shown in Fig. 1. Wind energy system of 0.5 kW with PMSG is connected to the rectifier which converts the alternating current (AC) output of wind energy system to direct current (DC). This obtained DC from rectifier is fed as input to enhanced boost converter which increases the value of input, by making use of control strategies like simple boost control and maximum boost control (both the strategies have been implemented in this work). This boosted output is fed to the non-linear load. To improve the power factor and compensate reactive power of the system, a synchronous reference frame theory control-based DSTATCOM is used.

2.1 Wind Energy System

The heart of the wind energy system is wind turbine without which the conversion of wind energy is not possible. The rotor in the wind turbine captures kinetic energy from wind as and when it spins and converts it in to rotary motion which drives the generator. The classification of wind turbines is of two types: fixed speed wind turbine and variable speed wind turbine [1, 2]. A variable speed wind turbine is mostly considered rather than a fixed speed wind turbine as it reduces fluctuations in power and improves power quality [3, 4]. Annually, there is 5% more energy capture

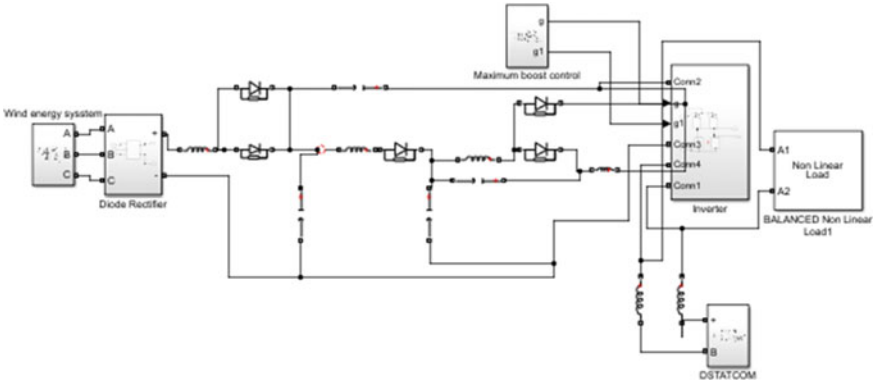


Fig. 1 Schematic of maximum boost control of single-phase enhanced quasi-Z-source inverter with DSTATCOM for a wind energy system

Table 1 Parameters of wind energy system

Parameters	Ratings
Wind speed	12 m/s
λ (constant)	8.1
Radius of machine	1.3 m

from varying speed technology than fixed speed technology [5]. The involvement of power converters in variable speed turbines may increase its cost. In variable speed turbine, either a permanent magnet synchronous generator (PMSG) or doubly fed induction generator (DFIG) is used. A wind energy system with PMSG of 0.5 kW power has been implemented. The obtained electrical energy from wind is rectified by using universal bridge and fed to the enhanced boost converter (Table 1).

2.2 Enhanced Boost Quasi-Z-Source Inverter

Conventionally, there are two converters: voltage-fed and current-fed converter. Figure 2 shows the conventional single-phase voltage-fed converter. A large capacitor supported DC voltage source acts as the main source for the two bridge arms. The output of the voltage-fed converter is less than the DC rail voltage making it a buck converter [6].

Figure 3 shows the conventional current-fed converter. A DC current source feeds the bridge arms. The DC current source can be realized by a DC voltage source to which a large inductor is connected in series. The output of a current-fed converter is usually greater than its DC input making it a boost converter. In this, it is mandatory to maintain one of upper device and one of the device connected at the bottom of

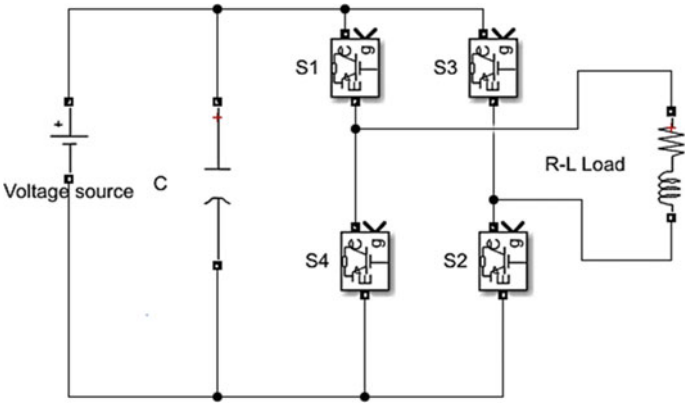


Fig. 2 Voltage-source inverter

the leg in conduction, if not, it may damage the circuit as it leads to open circuit of inductor.

The voltage source and current source converter are prone to electro-magnetic interference, and the output of these converters is limited to more or less than input DC voltage.

A better way to handle the issues of voltage-fed converter and current-fed converter is to implement impedance-source converter. The impedance-source converters with an appropriate control strategy can be used to implement conversions such as AC-to-DC, DC-to-AC, AC-to-AC, and DC-to-DC. These converters have the capability of either increasing or decreasing the voltage when compared to input, by applying suitable control strategy.

The enhanced boost quasi-Z-source inverter (QZSI) consists of four inductors, four capacitors, and five diodes. Figure 4 shows the enhanced boost QZSI. The operation of enhanced boost QZSI is divided in to two modes: shoot through state

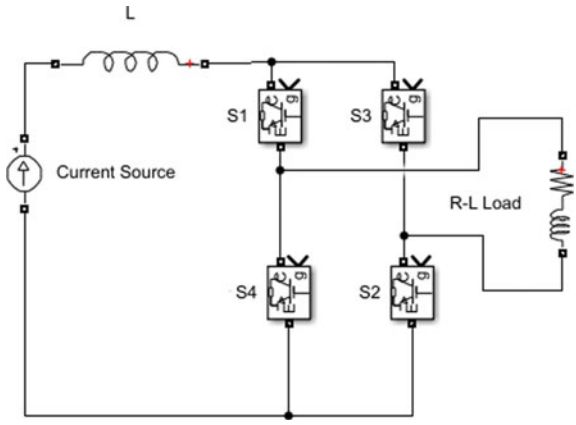


Fig. 3 Current-source inverter

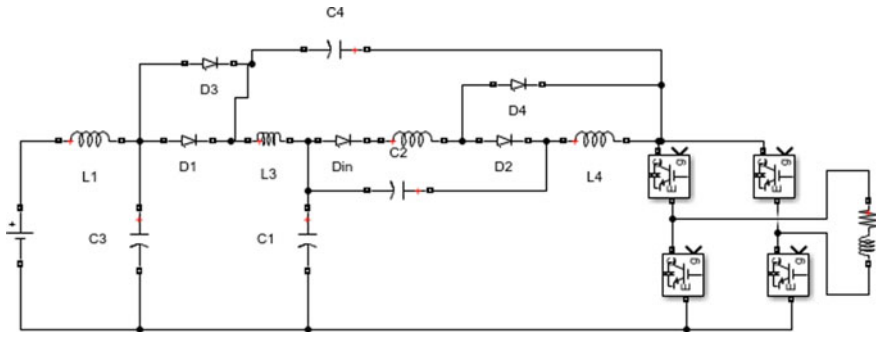


Fig. 4 Enhanced boost QZSI

and non-shoot through state. Shoot through state is a state where in both upper device and lower device of either one leg or both legs of a single-phase inverter are turned ON. During this mode, the inductors store energy from supply and provide it to the load during non-shoot through state [7, 8].

In this paper, a simple boost control strategy and maximum boost control strategy have been implemented using Simulink/ MATLAB environment, and *comparison* is provided about voltage stress across capacitors of enhanced quasi-Z-source inverter.

Simple Boost Control In this, a sinusoidal waveform is compared with triangular waveform to provide normal triggering pulses to non-zero and active states of the single-phase inverter. A comparison is also made between triangular wave and a constant voltage to produce control signals for shoot through states. In this method, the maximum shoot through state is confound to (1-M) (where M is the modulation index) so that the shoot through state is inserted only into conventional zero/states. The modulation index has to be reduced for large shoot through duty cycle, and a low-modulation index results in increase of total harmonic distortion (THD) and reduction in overall efficiency of DC-to-AC inversion [9].

Maximum Boost Control In this method, two control sinusoidal waveforms $+ \mathrm{V_C}$ and $- V_C$ having amplitude of M are compared with a triangular waveform V_t of high frequency to generate triggering pulses to non-shoot through states of the inverter. A sinusoidal waveform having twice the line frequency of voltage magnitude 'd' is compared with another triangular waveform having half the amplitude and twice the frequency as that of V_t to generate control signals for shoot through state [10–14]. This method does not introduce low-frequency ripples with respect to output frequency.

$$d = M - A + A \sin(2\omega t - (\pi/2)) \tag{1}$$

where 'A' is half the peak-to-peak value of the 'd'.

2.3 DSTATCOM

DSTATCOM is a shunt compensating device used at distribution level for the compensation of real and reactive power. It is commonly used for compensating reactive power, mitigating harmonics, and balancing the load. SRF theory-based control is simple in implementation with reduced computations. The SRFT method for a single-phase system is implemented, in which the control system obtains reference current waveform which will be injected to achieve the desired task. A hysteresis current controller is a closed loop control used for generation of pulses for voltage source converter of DSTATCOM. This hysteresis current control ensures that the current at the output terminals of inverter 'I' follows the intended current 'I_1' by maintaining the error always in the specified hysteresis band [15].

3 Simulation Results

Wind energy system as a source whose output is rectified by using universal bride with diode as switch and fed as an input to enhanced boost quasi-Z-source inverter. The triggering pulses of the inverter are controlled by maximum boost control method, and a DSTATCOM to improve power quality has been implemented in Simulink/MATLAB environment for the following specifications (Figs. 5, 6, 7, 8, 9 and 10; Tables 2, 3 and 4).

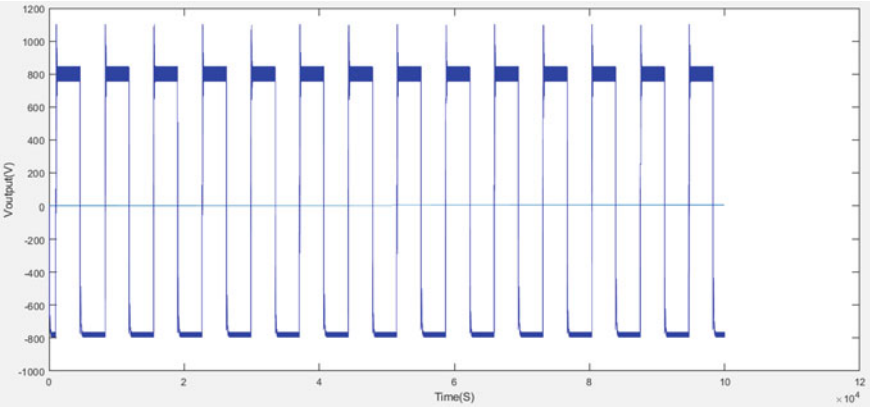


Fig. 5 Output voltage ($V_O = 800$ V) across the inverter of the enhanced boost QZSI with maximum boost control

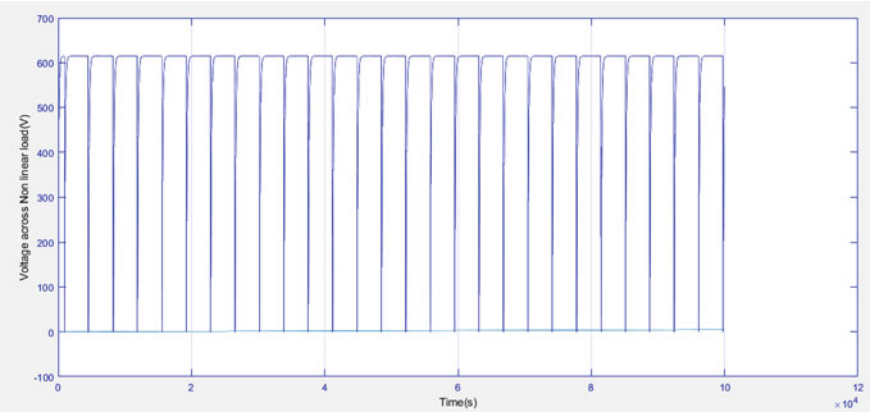


Fig. 6 Output voltage ($V_{NL} = 750$ V) across the non-linear load with maximum boost control

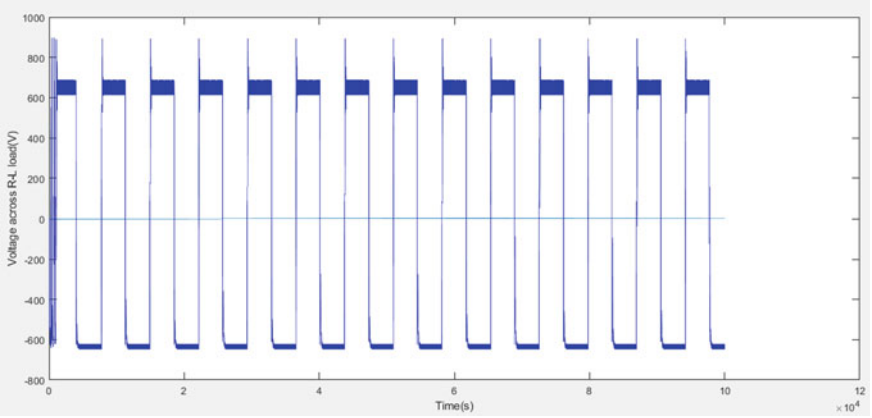


Fig.7 Output voltage ($V_{RL} = 800$ V) across R-L load with maximum boost control

4 Conclusion

In this paper, maximum boost control method for a single-phase enhanced boost QZSI with DSTATCOM for wind energy system is implemented using MATLAB. A comparison is made between maximum boost control strategy and simple boost control. It is observed that the maximum boost control strategy has a higher boosting factor for the same modulation index and lesser starting inrush current when compared to simple boost converter.

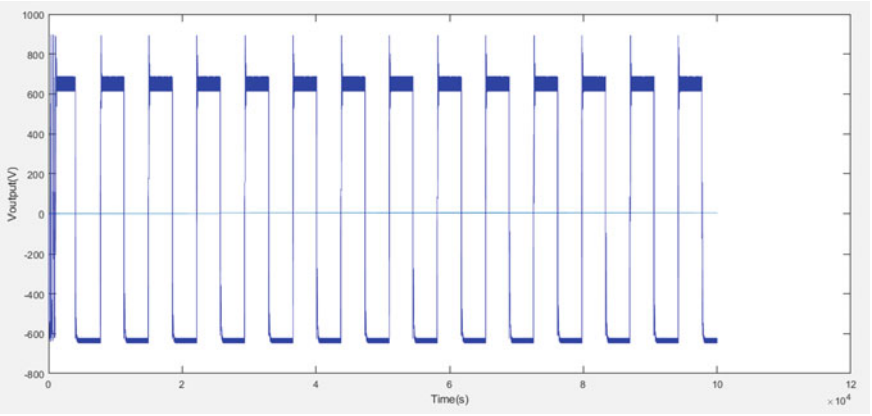


Fig. 8 Output voltage ($Vo = 708$ V) across inverter with simple boost control

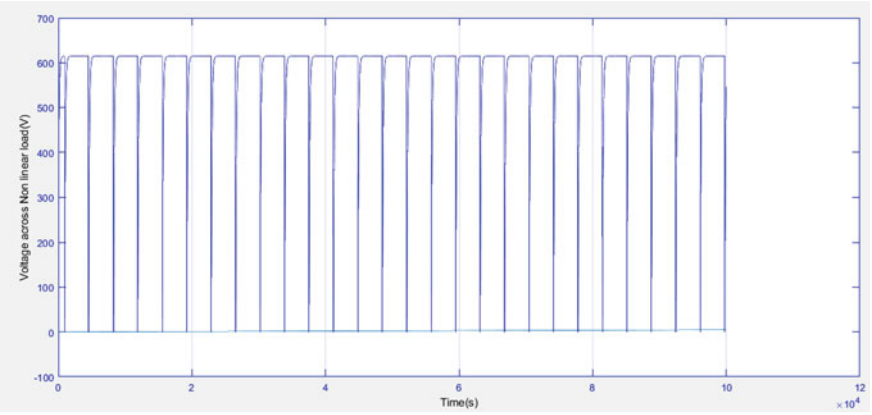


Fig. 9 Output voltage ($V_{\mathrm{NL}} = 610$ V) across non-linear load with simple boost control

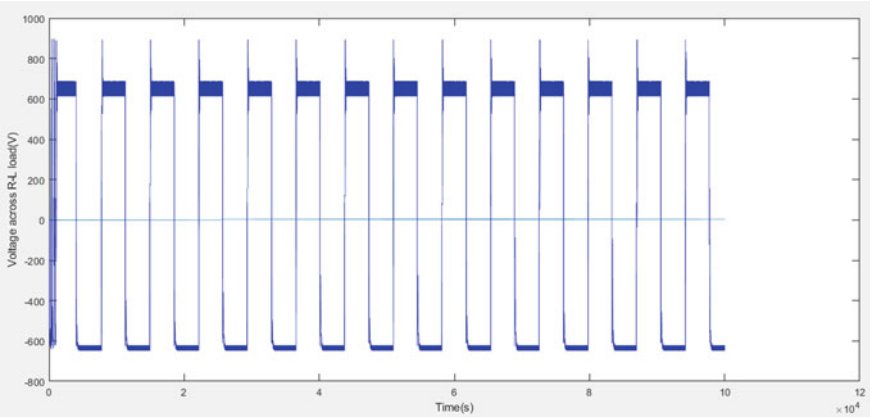


Fig.10 Output voltage ($V_{RL} = 698$ V) across *R-L* load with simple boost control

Table 2 Operating parameters of implemented wind energy system for enhanced quasi-Z-source inverter system with DSTATCOM

Parameters	Values
Power rating of wind energy system	0.5 kW
Output DC voltage of rectifier = Input of enhanced quasi-Z-source inverter	100 V
$L_1 = L_2 = L_3 = L_4$ of enhanced boost converter	1.2 mH
$C_1 = C_2 = C_3 = C_4$ of enhanced boost converter	250 μF

Table 3 Voltage across the capacitors of enhanced boost converter

Parameter	Maximum boost control	Simple boost control
Voltage across C_1	248.4 V	278 V
Voltage across C_2	84.95 V	96 V
Voltage across C_3	115.6 V	127 V
Voltage across C_4	161.7 V	169 V

Table 4 Comparison of power factor with and without DSTATCOM for simple boost control and maximum boost control

Power factor	Without DSTATCOM	With DSTATCOM
cosø for maximum boost control	0.854	0.99
cosø for simple boost control	0.81	0.96

References

1. E. de Vries, Wind turbine drive system: a commercial overview, in *Electrical Drives for Direct Drive Renewable Energy systems* (2013)
2. A. Rashad, S. Kameel, F. Jurado, The basic principles of wind farms. Distrib. Gener. Syst. (2017)
3. W. Cao, Y. Xie, Z. Tan, Wind turbine generator technologies. Adv. Wind Power **1**(1), 177–204 (2012)
4. M.M. Hossain, M.H. Ali, Future research directions for the wind turbine generator system. Renew. Sustain. Energy Rev. **49**, 481–489 (2015)
5. P.W. Carlin, A.S. Laxson, E.B. Muljadi, The history and state of the art of variable-speed wind turbine technology
6. T.B. Lazzarin, G.A.T. Bauer, I. Barbi, A control strategy for parallel operation of single-phase voltage source inverters: analysis, design and experimental results. IEEE Trans. Ind. Electron. **60**(6), 2194–2204 (2013)
7. V. Jagan, J. Kotturu, S. Das, Enhanced-Boost quasi Z-source inverters with two switched impedance networks. IEEE Trans. Ind. Electron. **64**(9), (2017)
8. S. Mishra, B. Dwivedi, A. Tripathi, Performance analysis Of PV fed enhanced boost quasi Z source inverter using two switched impedance and maximum boost control. Int. J. Adv. Sci. Technol. (2021)
9. F.Z. Peng, Z-source inverter. IEEE Trans. Ind. Appl. **39**(2), (2003)
10. M.-K. Nguyen, Y.-O. Choi, Maximum boost control method for single-phase quasi-switched-boost and quasi-Z-source inverters. *MDPI Energies*
11. M. Shen, J. Wang, A. Joseph, F.Z. Peng, L.M. Tolbert, D.J. Adams, Maximum constant boost control of the Z-source inverter, in *Conference Record of the 2004 IEEE Industry Applications Conference*, *2004. 39th IAS Annual Meeting*
12. K. Rajesh, Design and analysis for various controlling methods of a Z-Source inverter. Int. J. Electr. Eng.
13. M.M.K. Nguyen, Y.O. Choi, Maximum boost control method for single-phase quasi-switched-boost and Quasi-Z-Source inverters. Energies
14. R. Palanisamy, K. Vijayakumar, Maximum boost control for 7-level Z-source cascaded H-Bridge inverter. Int. J. Power Electronics Drive Syst.
15. N. Sampath, P.V.S.S.A. Parimala, D-STATCOM control using SRFT Method for PQ improvement in a PV system. IJITEE (2021)

Design and Development of Multi-copter Drone Incorporating with Multispectral Sensor for Agricultural Application

M. Kalaiselvan and K. Senthil Kumar

Abstract In modern years, the remote sensing by using unmanned aerial vehicles (UAV) is deployed into the agricultural sector, for improving productivity and has been vastly developed in precision agriculture for various applications. Nowaday's vegetation and crop field have been successfully monitored and done through the UAV systems. Meantime the UAV- based agriculture remote sensing offers us a high spatial image and is more flexible compared to the satellite-based remote sensing data. In this work, the main criteria to fly a UAV with autonomous flight control that can carry a payload (Multispectral Sensor) and easy way to acquire aerial images. Drone with multispectral sensor captures the unhealthy plants and monitor the health status of the crop field, and also used to identify the diseases of crops in order to reduce the human work. The radiometric calibration by using the calibration reflectance panel (CRP) relates the raw pixel value from the captured image to absolute reflectance. The drone captured aerial images and CRP techniques were used in the agricultural framework to collect multispectral images and to assess various vegetation indices (VI) such as chlorophyll vegetation index (CVI), normalized different vegetation indices (NDVI), enhanced vegetation index (EVI), soil adjusted vegetation index (SAVI), etc. Moreover, the model builder has been created for calculating the different vegetation indices in the single toolbox for analysis of the crop health status and the vegetation greenness. The major advantage of this methodology is that mapping or finding out resources in remote areas can be made easy and at a low cost.

Keywords Agriculture UAV · Drone · Multispectral sensor · CRP · Vegetation indices · NDVI

M. Kalaiselvan (✉) · K. Senthil Kumar
Department of Aerospace Engineering, Madras Institute of Technology, Anna University, Chennai, Tamil Nadu, India
e-mail: kalai.madesh@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_21

1 Introduction

In this rapid moving technology, agriculture development plays a crucial role in the UAV platform. The UAV technologies are stepped in the agriculture field to improve the development and help the farmers to enhance more data and information about their crop field without the field inspection. Nowaday's Framers prefer UAV-based remote sensing data over agriculture gives them more accurate data and saves time to overcome the field visits. The UAV is typically defined as a self-powered aerial vehicle that carries no humans, uses aerodynamic forces for vehicle lift, fly autonomously and carry a payload. UAVs can be deployed quickly, operated in remote areas and have relatively low operating cost [1]. The developed micro UAV with autonomous flight control and longer endurance was taken into consideration UAV platforms coupled with imaging, ranging, and positioning sensors, which can collect multispectral images at centimeter level resolution and provides such a great possibility in precision agriculture. The UAV survey enables us to deal with drone aerial images with very small pixel sizes ground sample distance (GSD), sometimes in cm level, a value that significantly improves the normal resolution of an aerial platform [2]. The multi-spectral images can be captured using a UAV-Quad copter at the Nadir position (90°) for the high-resolution images to get better quality images at low-resolution levels with different band ranges helps us to find the crop stress/plant health analysis for monitoring and calculation of vegetation indices. In general, multispectral sensor can collect the reflectance, remittance or backscattered energy from an object or area in multiple bands of spectrum that is useful for discriminating among land-cover classes or substances like, e.g., water, crop, forestry, landslide, urban area, etc. [3].

Remote sensing through the satellite data can also be done for the calculation of plant health analysis and monitoring but the accuracy of the data will not be up to date as compared to the UAV data. Moreover, satellite imagery data was too expensive with a big temporary resolution suffering by a high percentage of cloudiness and the pure conventional aerial mission was also expensive. At present, the study provides the overview of the UAV techniques to derive the vegetation indices from the multispectral image for precision agriculture application. Different VI was applied in the model builder, which is an ArcGIS application toolbox that can create, edit, and manage the different VI formulas. This work presents a platform and gives us a solution for calculating the different vegetation indices (Table 1).

Table 1 Data comparison of different platform

Data	Resolution	Field of view (FOV)
UAV	0.5–10 cm	50–500 m
Air-Borne	0.1–2 m	0.5–5 km
Satellite	1–30 m	10–50 km

2 AIMS and Objectives

This paper exhibit some critical consideration for agriculture mapping using unmanned aerial vehicle platform, multispectral mapping image, and different photogrammetric methods are performed in this study. The multispectral images over an agriculture field evaluate the crop health status by calculating the vegetation indices. In particularly work focused on the development of the UAV system and by creating a toolbox-model builder for calculating the different vegetation indices.

3 UAV System and Payload Description

3.1 UAV Platform

In this work, UAV is developed based on the configuration at the Centre for Aerospace Research lab and the system X8 quad model-multi-rotor UAV (see Fig. 1) which is used in the agriculture platform for capturing the field images. The system developed with the adoption of semi-autonomous control and can be operated from the pilot view and designed for longer endurance to acquire data. Similarly, the UAV is incorporated with the multispectral camera (payload) for data acquisition. The global positioning system (GPS) helps us in identifying and mapping the vast expense of the field's. The main advantage of this UAV is low cost with flexible and easy design, etc., since the UAV Quadcopter has Autonomous flight control (i.e.) it can be operated via ground control station (GCS) with a pilot view. The specification of X8 Quadcopter model is listed in the following Table 2.

(a) Isometric view (b) Bottom view

Fig. 1 X8 quad model with payload

Table 2 Specification of X8 Quadcopter

Specification—X8 rotor model	
Model name	X8 rotor
Weight without payload	4 kg
Payload weight	150 g
Maximum flying height	300–400 m
Normal flight time	35 min
Data radio range	2.5 km
Frequency	2.4 GHz
Speed	30–60 km/h
Structure	Carbon fiber and composite material
Autopilot	Open-Source

3.2 Sensor Platform

The Mica Sense-Red edge camera is the multispectral sensor used in this research. The sensor consists of five bands (blue, green, red, red edge, and near-infrared) Multispectral camera designed especially for agriculture and it is lightweight (i.e., 150–180 g), optimized for use in our UAV system. The main feature of this camera, simultaneously capture of five discrete spectral bands optimized for crop health data gathering and all bands captured once per second enables faster flight speeds. It helps farmers to manage crops, yields, fertilizing, and irrigation more effectively. The multispectral camera has more benefits in the agriculture sector to reduce and minimize the use of fertilizers, pesticides and saves water for irrigation and at the same time, it increases the yield of the harvesting field. The visible and infrared bands captured by the multispectral camera were used to calculate vegetation indices (Fig. 2 and Table 3).

4 Proposed Methodology

4.1 Study Area

In this study, the field test site of the crop field consists of cultivation land were located in the Coimbatore, Tamil Nadu, India (Lat: 11° 0′ 13.35″ N/long: 76° 55′ 38.69″ E). The multispectral image was acquired with a UAV system on 5th February 2018. The entire surface area is covered by the paddy and scrub field of area 6 hectares. In order to monitor the crop field, the study focuses on the static analysis of the crop health status by using multispectral imaging (Fig. 3).

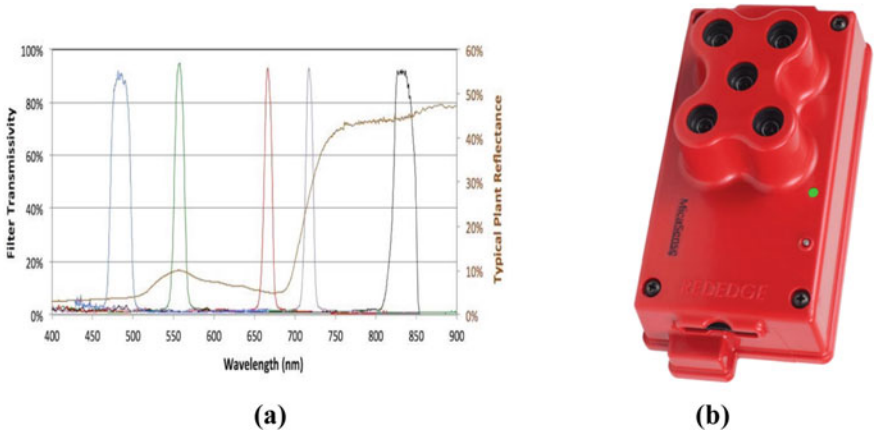


Fig. 2 **a** Spectral response curve for mica sense sensor, **b** multispectral camera

Table 3 Specification of mica sense multispectral sensor

Payload specification	
Weight	150 g
Dimensions	12.1 cm × 6.6 cm × 4.6 cm (4.8″ × 2.6″ × 1.8″)
Power	5.0 V DC, 4 W nominal
Spectral bands narrowband:	Blue, Green, Red, Red Edge, Near-IR
GSD	8.2 cm/pixel (per band) at 120 m (400 ft.) AGL
Capture speed	1 capture/second (all bands), 12-bit RAW

5 Process of UAV Data Collection

The UAV data collection contains the mission planning and captures the multispectral image (agriculture field). UAV system follows the flight pattern that can be the initial setup in the mission planner software. Mission planner is the open source software in which the pilot or the framer can control and monitor the operating maneuvers of the Quad copter from GCS. Through the mission planning software UAV moves over the waypoints of the test field. Either the UAV is to fly in (north-south) or (east-west) directions to get the perfect overlapping images and impact resolutions. The image acquired (flying phase) by the UAV Quad copter must recommend 60% of side overlap and 70% of forward overlap for better image alignment. For the accuracy of the field location, we collect some of the ground control points (GCP) for geo-referencing, scaling, and analysis. The GCPs were measured at the survey of the study field. If the survey field is very large in terms of square kilometres UAV should

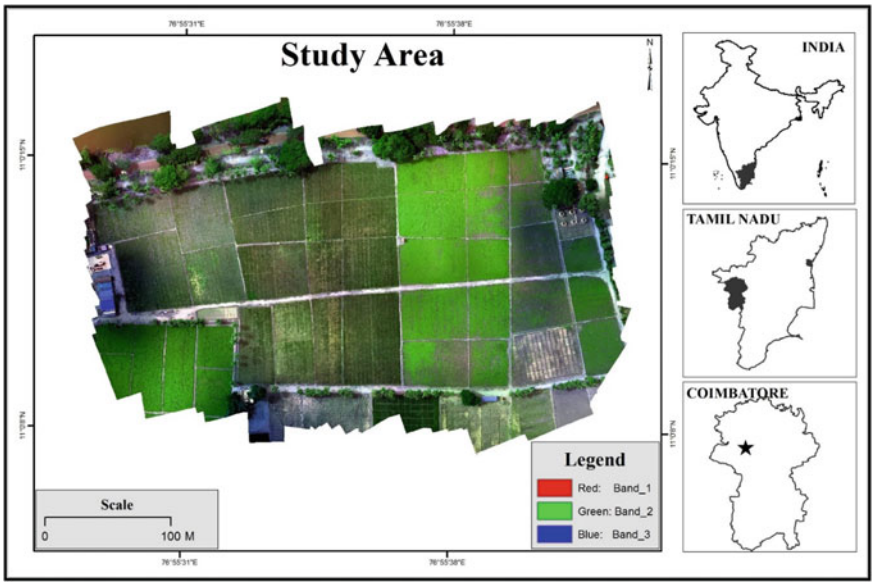


Fig. 3 Study area

have very high endurance and speed, so that image acquisition will be quick. For small area field UAVs with low endurance will suffice. Next, the main stage of the UAV data collection is to capture the reflectance panel, for accurate data to merge with the reflectance value. Subsequently, the image of the CRP is to be captured immediately before and after the time of each flight. The image capturing should be straight up to the reflectance panel and there should not be shadow fallen on the reflectance panel. These captured images are referred to with the panel images for similar light conditions at the time of flight. The figure represents the image capturing of the reflectance panel without a shadow falling on the panel (Fig. 4).

6 UAV Data Processing

6.1 Multispectral Calibration Reflectance Panel for Radiometric Calibration

The radiometric calibration aims to convert at-sensor radiance into absolute surface reflectance values, accounting for variation caused by differences in ambient light due to weather and sun, between sensors types and units [4]. The calibration reflectance panel is to calibrate the reflectance value of the atmosphere and the main purpose of radiometric calibration is to rectify the captured image data for various radiometric

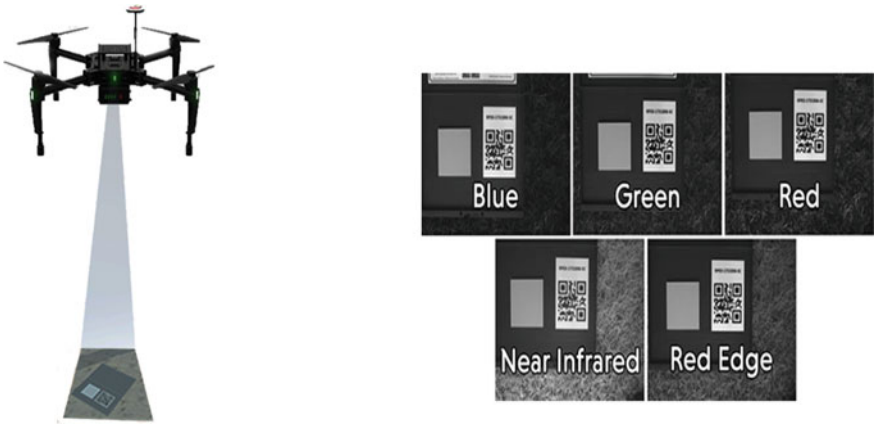


Fig. 4 Procedure of calibrating

distortions. While capturing the reflectance panel before and after flight avoid the casting of shadow over the target may reduce the contamination of light scattered from the object. Therefore, the calibration is done by manually selecting the area of the reflectance target on the calibration image and assigning the known reflectance value of the target. Once the calibration is done, the panel database is required for the system process for generating the othromosaic image (Fig. 5).

Fig. 5 Multispectral calibrating reflectance panel

6.2 Creating Toolbox for Calculating Vegetation Indices

The creating of the toolbox is done in ArcGIS 10.4. Initially, the toolboxes were using the model builder according to the different vegetation indices based on its formulae. Different spectral bands are used for the preparation of tools. The different indices calculated as they follow (CVI, NDVI, ENDVI, DVI, EVI, SAVI, RVI, GDVI, GNDVI, NDREI, NGRDI, and VI Green).

- CVI—Chlorophyll Vegetation Index: This vegetation index can be used to determine and estimate the chlorophyll concentration in an agricultural field.

$$\mathrm{CVI} = (\mathrm{NIR} * \mathrm{Red})/(\mathrm{Green}^2) \tag{1}$$

- DVI—Difference Vegetation Index: It explains the difference between soil and plants. The difference between reflectance and radiance caused by the atmosphere or shadows is not taken into account.

$$\mathrm{DVI} = \mathrm{NIR} - \mathrm{Red} \tag{2}$$

- EVI—Enhanced Vegetation Index: To enhance the vegetation signal with improved sensitivity in high biomass regions and also for better vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmosphere influences.

$$\mathrm{EVI} = 2.5 * [(\mathrm{NIR} - \mathrm{Red})/(\mathrm{NIR} + 6 * \mathrm{Red} - 7.5 * \mathrm{Blue} + 1)] \tag{3}$$

- NDVI—Normalized Difference Vegetation index: This vegetation value of -1 to $+1$ indicate more "greenness" vegetation and photosynthetic activity.

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{Red})/(\mathrm{NIR} + \mathrm{Red}) \tag{4}$$

- SAVI—Soil Adjusted Vegetation Index: The spectral indices might be calibrated in such a way that the variations of soils are normalized and do not influence measurements of the vegetation canopy.

$$\mathrm{SAVI} = [(\mathrm{NIR} - \mathrm{Red})/(\mathrm{NIR} + \mathrm{Red} + 0.5)] * (1 + 0.5) \tag{5}$$

- RVI—Ratio Vegetation Index: It denotes the amount of vegetation and show high for vegetation & Low for soil, water, etc. Reduces the effects of atmosphere and topography.

$$\mathrm{RVI} = \mathrm{NIR}/\mathrm{Red} \tag{6}$$

6.3 Data Processing

The RAW images from the multispectral camera of each flight were pre-processed and exported in a single tri-band TIFF image. The quality of the photo information was assessed in this way whether it had an impact on the photogrammetric processing and vegetation indices [1]. The geo-tagged images were uploaded to Agisoft Photoscan (photogrammetric) software for generating tie points, dense point cloud, and orthophoto. The UAV aerial photos are combined in the initial process for a better photo alignment. GPS/IMU details are used and compared to generate the point for combing the raw images. For better alignment, the ground control points are matched during the process. The detailed processing work flow is shown in below Figure 6.

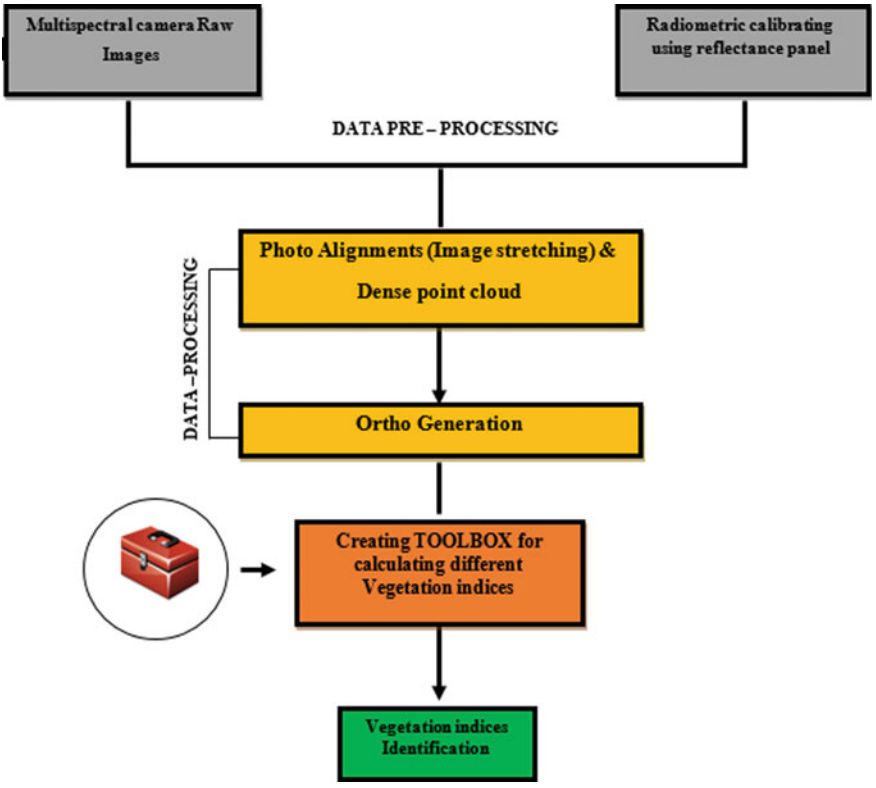


Fig. 6 Flowchart of data processing

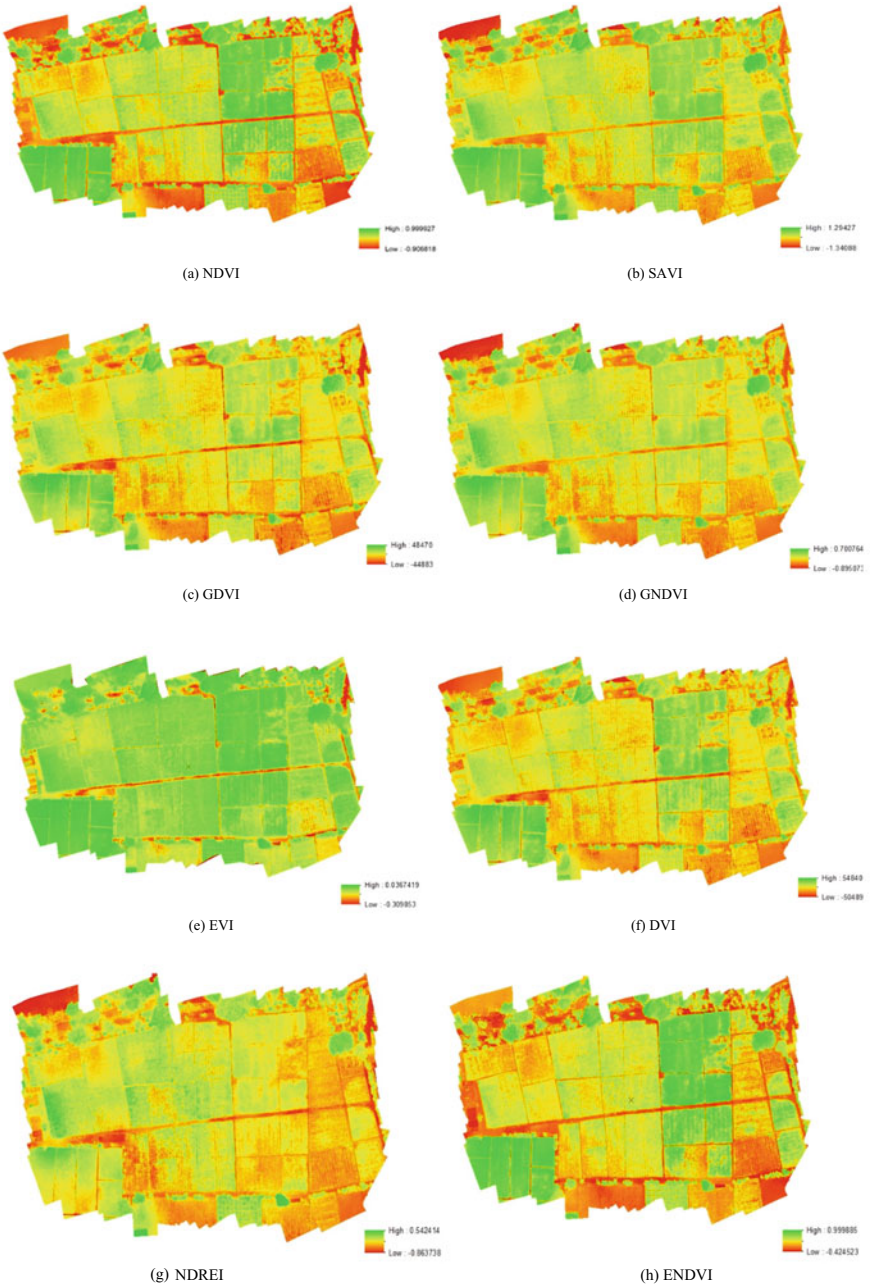
7 Results and Discussions

7.1 Vegetation Analysis for Different Vegetation Indices

The collected multispectral images were processed, and hence the orthomosaic image was chosen to see the health status of the crop field, projected changes in the vegetation. This can be achieved through a specific vegetation indices toolbox created using a model builder in ArcGIS 10.4. This tool separates the orthomosaic layers into blue, green, red, NIR, and Rededge to create vegetation indices based on the parameters in the selected indices formula as explained before para 6.2. The generated model builder toolbox is a collection of familiar vegetation indices module to particularly monitor crop health precisely using aerial remote sensing method to reduce manual work and increase the agriculture yield by taking timely decisions. The accuracy of the generated indices is cross verified with field samples in different radiometric conditions (Fig. 7).

8 Conclusion

In this paper, the designed UAV and the multispectral sensor were performed for collecting agricultural field data. UAV remote sensing data gives better high spatial images at 10 cm resolution compared to the satellite Land Sat images. As per the results, the UAV has been examined over a field and the integration of a multispectral camera helps us to monitor the crop. NDVI calculation was able to detect the health and unhealthy statuses of the agriculture field and interpret the specific area and indicate the crop health condition to the farmer. The analyzed results and the data information represent an important source for framers to make an appropriate decision to improve the quality and maintain the management strategies of their crop field on time. These results show that the methodology is accurate and cost-effective as the affected areas are identified and located precisely. It has also proven the concept of reduced time consumption and safe workers.



(a) NDVI (b) SAVI

(c) GDVI (d) GNDVI

(e) EVI (f) DVI

(g) NDREI (h) ENDVI

Fig. 7 Results of different vegetation indices

References

1. S. Candiago, F. Remondino, Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. Remote Sens. **7**, 4026–4047 (2015). https://doi.org/10.3390/rs70404026
2. J. Wijitdechakul, Y. Kiyoki, C. Koopipat, UAV-based multispectral image analysis system with semantic computing for agricultural health conditions monitoring and real-time management. Int. Electron. Symp. (IES) (2016). 978-1-5090-1640-2/16
3. Mica Sense red edge Manual. https://support.micasense.com/hc/enus/articles/215261448-RedEdge-User-Manual-PDF-Download
4. J. Kelcey, A. Lucieer, Sensor correction of a 6-band multispectral imaging sensor for UAV remote sensing. Remote Sens. **4**, 1462–1493 (2012). https://doi.org/10.3390/rs4051462

Mitigation of RSSI Variations Using Frequency Analysis and Kalman Filtering

Arunkant A. Jose, P. H. Rishikesh, and Shilpa Shaju

Abstract Several indoor locating systems rely on the groundbreaking use of the received signal strength indication (RSSI). RSSI is a cost-effective and straightforward solution that does not necessitate the purchase of any additional hardware. Variations in RSSI of Bluetooth low energy (BLE) devices in indoor positioning degrade the precision of indoor positioning systems. This paper provides a filtering strategy for processing the RSSI signal to solve the aforementioned difficulty. The RSSI of the BLE node is captured. The raw RSSI signals are then filtered using various filtering methods and their combinations. This work employs a variety of filters including frequency analysis, Kalman filtering, frequency Kalman filtering and Kalman frequency filtering. The standard deviation value after applying various filters was compared under various test cases. The results suggest that applying frequency filters followed by Kalman filtering can reduce RSSI variability in BLE devices, resulting in more consistent and reliable RSSI measurements.

Keywords BLE · Frequency analysis · Indoor positioning · RSSI · Kalman filter

1 Introduction

Smartphones make it easier for people to complete their daily tasks. As mobile devices and technologies have improved, navigation apps have become more popular. In an unknown environment, the navigation system assists the user in navigating. Outdoor navigation is now possible because of the availability of global positioning system (GPS) navigation on cellphones. When the users enter the building, however,

A. A. Jose (✉) · P. H. Rishikesh · S. Shaju
Muthoot Institute of Technology and Science, Varikoli, Kerala, India
e-mail: arunkant@mgits.ac.in

P. H. Rishikesh
e-mail: rishikeshph@mgits.ac.in

S. Shaju
e-mail: shilpashaju@mgits.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_22

the circumstance changes, while indoors, GPS satellite signals deteriorate, causing GPS navigation to fail. Interior navigation can enable dependable access to a variety of services within indoor settings where GPS signals are not available. Indoor navigation has a wide range of applications in a variety of domains, including hospitals, shopping malls, colleges and multi-floor buildings. The proliferation of smartphones necessitates smartphone-specific indoor navigation solutions.

GPS gives individuals a new perspective on navigation. This has caused users to use the GPS in indoor environments, which has failed. As a result, the need for an indoor position services system arose.

Navigation, location-based applications and asset tracking all require precise location estimation of the indoor. Bluetooth, WIFI, acoustics signals, ultra-sound, radio frequency identification (RFID) and ultra-wide band (UWB) are the technologies used in indoor services. The methods like RSSI, time difference of arrival (TDoA), time of flight (ToF) and angle of arrival (AoA) are used in positioning technologies [1]. Many indoor-based services use a range of approaches and technologies to increase precision, dependability and minimize the overall system's processing complexity and energy usage. Apart from RSSI, the installation of the others necessitates the use of dedicated hardware. Because of its simplicity, RSSI is a widely used indoor navigation. However, it lacks precision, according to [2]. Even though they are in the same place, RSSI values shift due to barriers and noise.

In this study, a new filtering method based on a combination of frequency analysis and Kalman filtering is employed to reduce the fluctuations of RSSI of BLE devices. The remaining sections of the paper were structured as: Sect. 2 work provides an overview of related works, and Sect. 3 describes the work's preliminary study. The system design is described in Sect. 4. Section 5 describes the suggested method's experimentation and analysis. Finally, Sect. 6 wraps up the paper and discusses its future prospects.

2 Related Work

To determine the specific indoor location, different methods were used. The RSSI method [3] stands out among these methods. Most wireless devices have a built-in RSSI circuit [4–6] in them, and thus, there is no need of additional hardware, thereby reducing the hardware cost as well as the power consumption. The distance between a transmitter and receiver can be calculated using RSSI method. However, the major shortfall of the RSSI system is that the RSSI signals readings are unreliable. The RSSI signals are affected by different environmental factors and fluctuate due to the multi-path problem [7].

Based on the previous works, various researchers tried a variety of approaches to reduce RSSI fluctuation and thus produce a better indoor localization system. Various filtering methods are used to mitigate the fluctuations. Here is a rundown of some of the works. A. Booranawong et al. implemented a span thresholding filter in [8] to lessen fluctuations, claiming the peaking errors. Since the span thresholding filter is

using RSSI from the preceding sample for position estimation, it is highly impaired that a human is blocking its radio signal route. It is also impaired if the room is filled with people for a prolonged duration.

By combining the proposed hard fluctuations removal filter with the Kalman filter, Mouhammad et al. [9] demonstrated that the variance in RSSI readings caused by interference from other devices is eliminated.

The experiment in [10] looked at a comparison of RSSI filtering algorithms for the exterior environment. The raw data, the methods like weighted average, exponential smoothing, wavelet filtering and the least mean square filtering were selected and tested in to reduce the fluctuation in the received signals. The results were then compared using the root mean square error (RMSE) values. Since the methods listed above were not effective in reducing the fluctuation, those methods were merged to build an ideal filtering method, which has reduced RMSE when compared to raw RSSI data. However, the study is limited to the exterior environment.

In the work [11], the pre-processing step is done to reduce the fluctuations by using Kalman filtering. After the filtering process, a database is created for the offline phase. The filtering process helps in reducing the disturbance of RSSI.

Andrew Mackey et al. [12] explored the indoor navigation system using Kalman filtering and beacons. The experimental result shows that the position accuracy can be improved up to 78.9% using a static Kalman filter. But the proposed system is suitable for large indoor areas. In a similar work, [13], the RSSI signals are smoothed using Kalman filtering before applying the filtered RSSI and beacon weight algorithm. The results show that the accuracy in position estimation has improved to a few centimeters.

FenceBot [14] is an elderly tracking app using the properties of RSSI from BLE nodes. The hindrance and noise affect the RSSI causing inaccurate distance measurements. But FenceBot needs only approximate distance measurement. Hence, the application mapped the RSSI values to the proximity levels. However, the RSSI variations affect the proximity detection and proper estimation of user location.

In the work [15], the enhanced indoor localization method using RSSI was highlighted. The smoothing of the RSSI signals is done by using the Kalman filter. The proposed system shows a 50% enhancement in the distance estimation accuracy. Several indoor experiments are done inside the laboratory environment but not in real-time scenarios. If the error of the latest measured RSSI data is fairly substantial, the Kalman filter has a major impact on the location result. To eliminate RSSI fluctuations, Huang et al. [16] presented a constant speed Kalman filtering (CVKF) fusion approach by experiments. The experimental result shows that the CVKF filter's standard deviation is much lower than that of a single Kalman filter, for the fluctuation reduction.

The location estimation using the fingerprinting method has two-phase; one is online phase, and another is offline. In the later phase, the reference positions are matched with RSSI acquired from the nearby access points using finger printing. After creating a database for offline phase, which is a critical task, the match position is determined by comparing online RSSI collection measures to fingerprints in database. Fingerprinting approach is part of the classification paradigm [17], and

different authors have employed a variety of algorithms. For the construction of an efficient RSSI fingerprint database, it is better to remove the fluctuations in RSSI.

3 Preliminary Study

3.1 RSSI

The RSSI detected by the android phone fluctuates, owing to a variety of disturbances, which has a significant impact on location accuracy. It is impossible for a BLE device to transmit signals at a constant signal strength because of the time varying properties of RSSI. Furthermore, the indoor environment is complex, with fading and other noise to contend with.

3.2 RSSI Filtering

Mean Filtering: The result is determined as the average of collected RSSI measurements. When the RSSI number fluctuates a lot, the results' accuracy suffers.

Median Filtering: It arranges the different RSSI based on their values. The RSSI value in the middle position is the output value. It can efficiently overcome the signal interference caused by accidental factors; however, the filtering is not perfect in the case of strong pulse interference and small samples.

Maximum Filtering: The maximum filter uses the RSSI value that corresponds to the received signal's line of sight (LOS) component [9]. However, if the greatest value is the fluctuation value, it is not accurate.

Kalman Filtering: Kalman filter is a linear optimal state estimation method which uses Eqs. 1 and 2 [18].

$$x_k = Ax_{k-1} + Bu_{k-1} + w_k \tag{1}$$

$$z_k = Hx_k + v_k \tag{2}$$

In Eq. 1, x_k is the RSSI value to be estimated at time k based on the previous RSSI value x_{k-1}, the control signal u_{k-1} and the process noise w_k. In Eq. 2, the value z_k is the observed RSSI value at time k from a BLE device and v_k is the measurement noise. A and B are the state transition matrix. H is the observation matrix. In our work, we assume w_k and w_k as zero.

The two steps in the Kalman filtering are the predication step and measurement update as shown in Eqs. 3 and 4.

For the predication step:

$$\begin{aligned} \hat{x}_k^- &= A\hat{x}_{k-1} + Bu_k \\ P_k^- &= AP_{k-1}A^T + Q \end{aligned} \tag{3}$$

For the measurement update:

$$\begin{aligned} K_k &= P_k - H^T\left(HP_k - H^T + R\right)^{-1} \\ \hat{x}_k &= \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \\ P_k &= (1 - K_kH)P_k^- \end{aligned} \tag{4}$$

where $\hat{x}_k^-$ indicates the prior estimate at time k, P_k^- is the prior error covariance, and K_k is the Kalman gain, the smoothed RSSI value indicated as $\hat{x}_k$ is the posterior estimate at time k, P_k is the posterior error covariance. Section 4 explains the parameter selection criteria for the work.

Frequency Analysis Filtering: To reduce RSSI fluctuations, this method can be used. It is to count the number of times the RSSI value appear and select the one that is equal or greater than a predefined threshold value [19].

4 System Design

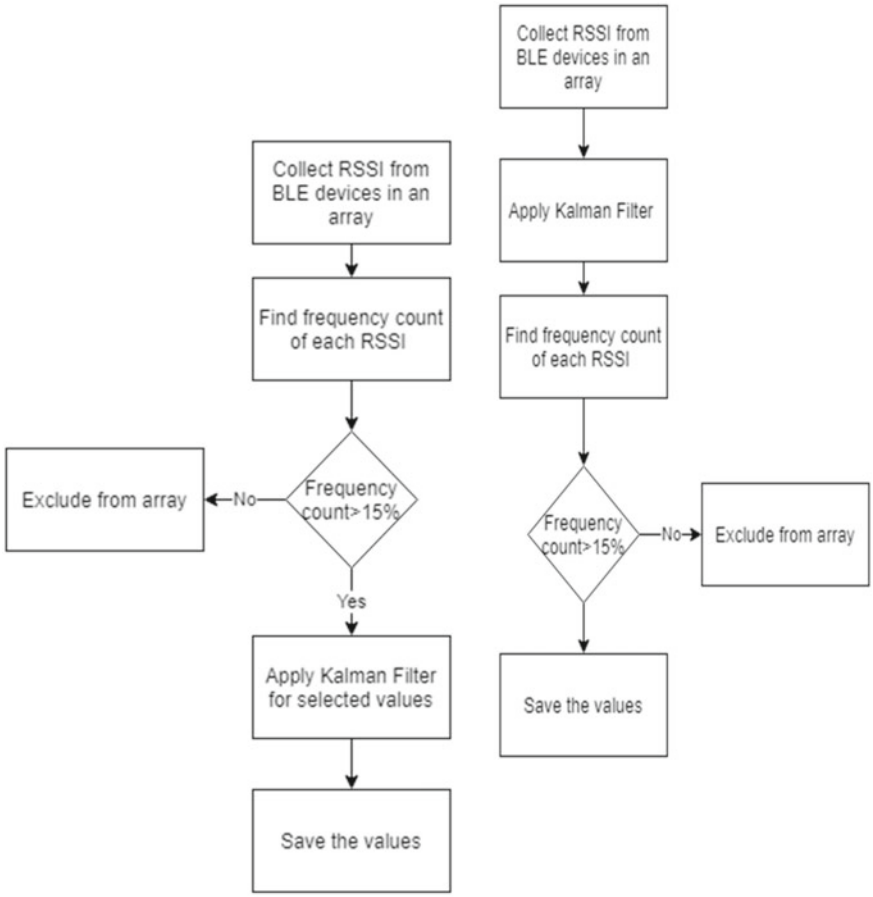
The fluctuations in raw RSSI values are undesirable. Mitigating the RSSI variations is a primary step for every RSSI-based approach. In this work, we use the combination of frequency analysis and Kalman filtering as shown in Fig. 1.

The data was gathered with the help of an android app designed to create fingerprints. The application employs a grid-based user interface. In the application, the user can choose the appropriate grid location. The user can start BLE scanning by clicking on the grid.

During scanning, the RSSI value from the BLE device is obtained. The RSSI values received are saved in an array. The filtering approach is used to the gathered RSSI values due to the variations in the RSSI values. In this work, the combination of frequency analysis filtering and Kalman filtering is used. In this work, frequency–Kalman filtering uses frequency analysis before Kalman filtering as shown in Fig. 1a.

This work also uses Kalman–frequency filtering, which uses Kalman filtering before frequency analysis as shown in Fig. 1b.

In this work, for Kalman filtering, we have set values as 1 for the transition matrices A and B in Eq. 3 and 4. We choose the observation matrix H as 1. The standard deviation of a sample of RSSI values at each distance as in Tables 1 and 2 determines the value of R, and Q is set as half of R for each distance. The initial value for P is set as 0. Since there is no control input, we take its value as zero. For frequency–Kalman filtering, the RSSI, the frequency count of each value in the array



(a) Frequency-Kalman Filter (b) Kalman-Frequency Filter

Fig. 1 Flowchart for new filters

is calculated. As an initial step, select that RSSI whose frequency count is equal to or greater than 15% of the RSSI considered in the array. The rejected values are excluded from the array. The accepted values are given to the Kalman filter. When the filtering process of RSSI values finishes, the average value of the output array is considered.

Kalman filtering is used to the raw RSSI signals as the first stage in Kalman–Frequency filtering. The results are then subjected to frequency analysis. When the RSSI values have been filtered, the average value of the resultant array is taken into account.

Table 1 Statistics of data samples at various distance in test 1

Distance (meter)	Filters	Mean	Variance	Std. deviation
1	No filters	− 61.9830	5.64378	2.37566
	Frequency analysis filter	− 60.8437	1.81933	1.348827
	Kalman filter	− 61.6779	4.455	2.1108
	Frequency–Kalman filter	− 60.5625	0.808	0.89921
	Kalman–frequency filter	− 61	2	1.4142
2	No filters	− 58.7323	34.5621	5.8789
	Frequency analysis filter	− 55.8387	0.973985	0.986907
	Kalman Filter	− 58.239	22.1257	4.7038
	frequency–Kalman filter	− 55.419	0.2434	0.49345
	Kalman–frequency filter	− 55.3714	2.233	1.4944
3	No filters	− 62.0833	38.28472	6.1874
	Frequency analysis filter	− 59.7575	19.6382	4.4315008
	Kalman filter	− 61.416	29.1597	5.39997
	Frequency–Kalman filter	− 58.96	17.2415	4.15228
	Kalman–frequency filter	− 59.656	19.538	4.42
4	No filters	− 68.423	7.43	2.7259
	Frequency analysis filter	− 68.2857	6.20408	2.4907
	Kalman filter	− 67.7966	6.67	2.5827
	Frequency–Kalman filter	− 67.8	3.874	1.9683
	Kalman–frequency filter	− 67.847	7.5637	2.7502
5	No filters	− 65.2666	4.195555	2.0483
	Frequency analysis filter	− 64.9	4.2899	2.071
	Kalman filter	− 64.8666	3.3155	1.82
	Frequency–Kalman filter	− 64.3	2.41	1.5524
	Kalman–frequency filter	− 64.454	2.4297	1.55876
6	No filters	− 68	29.4117647	5.4232
	frequency analysis filter	− 66.153	0.745562	0.863459
	Kalman filter	− 67.4705	− 16.955	4.117647
	Frequency–Kalman filter	− 65.692	0.213	0.46153
	Kalman–frequency filter	− 65.923	0.5325	0.729756

5 Experiment and Analysis

This section will describe the experimental setup and its details with results.

Table 2 Statistics of data samples at various distance in test 2

Distance (m)	Filters	Mean	Variance	Std. deviation
1	No filters	− 52.089	2.009	1.4177
	Frequency analysis filter	− 51.8695	1.63516	1.27873
	Kalman filter	− 51.3928	1.8099	1.3453
	Frequency–Kalman filter	− 51.1956	0.9399	0.96952
	Kalman–frequency filter	− 51.8372	1.20605	1.0982
2	No filters	− 58.6274	3.21414	1.7928
	Frequency analysis filter	− 58.2592	0.9327846	0.9658
	Kalman filter	− 58.1764	2.8119	1.6769
	Frequency–Kalman filter	− 57.6296	0.2331	0.4829
	Kalman–frequency filter	− 58.4615	0.2485	0.498
3	No filters	− 60	1.619	1.27241
	Frequency analysis filter	− 59.7096	0.91571	0.95692
	Kalman filter	− 59.5	1.0119	1.005934
	Frequency–Kalman filter	− 59.3548	0.228928	0.478464
	Kalman–frequency filter	− 59.6857	0.61551	0.78454
4	No filters	− 71.111	8.8606	2.9766
	Frequency analysis filter	− 70.6071	0.2385	0.48838
	Kalman filter	70.666	6.7301	2.5942
	Frequency–Kalman filter	70.1071	0.0956	0.30929
	Kalman–frequency filter	70.4814	0.24965	0.4996
5	No filters	− 69.32	6.2975	2.5095
	Frequency analysis filter	− 67.7575	0.6685	0.81762
	Kalman filter	− 68.88	5.34559	2.31205
	Frequency–Kalman filter	− 67.3939	0.4205	0.6485
	Kalman–frequency filter	− 67.5161	0.7658	0.8751
6	No filters	− 67.0625	55.9752	7.4816
	Frequency analysis filter	− 63.7647	3.9446	1.98611
	Kalman filter	− 66.6666	58.305	7.6358
	Frequency–Kalman filter	− 63.323	2.218	1.489
	Kalman–frequency filter	− 63.068	6.0642	2.4625

5.1 Setup for Experiments

The experimental venue was a laboratory room of the Department of Electronics and Communication Engineering at Muthoot Institute of Science and Technology located in Kerala, India. This test field is 15 tiles by 19 tiles in size where each tile is of dimension 2 * 2 ft^2. The reference node is placed at a fixed location. The RSSI receiver is placed in LOS with the BLE device. RSSI readings from BLE devices are

received every two seconds to save battery life and avoid introducing a significant delay in the system.

For the experimental study, we study the RSSI of the BLE device kept at a static location. The RSSI values are collected using an android phone at a fixed distance of 1 m. The BLE device and the android phone are in LOS and at the same level. This is repeated for distances of 2, 3, 4, 5 and 6 m. All the collected data are saved to a common database. For each scanning, we collected the time stamp, unique ID, RSSI values, position and android phone model. The time interval for each scan is set as two seconds. From each positions, we did scanning for multiple times.

We consider two test cases. In the first test case, the RSSI values are collected in the morning from 10.30 AM to 11.30 AM for each distances as in Table 1. In the second test case, the RSSI values are collected in the afternoon from 3.30 PM to 4.30 PM for each distances as in Table 2.

5.2 *RSSI Fluctuation*

To investigate RSSI fluctuations, we measured RSSI values from a BLE device placed at a fixed distance to an android phone with each scanning interval being two seconds. The RSSI measurements are shown in Fig. 2. The readings vary between − 72 dBm and − 62 dBm. It is also observed that the number of RSSI values received in each scan is different and shows fluctuations. This experiment demonstrates that raw RSSI measurements are insufficient for accurate distance estimation. As a result, filters are necessary to improve RSSI measurement stability.

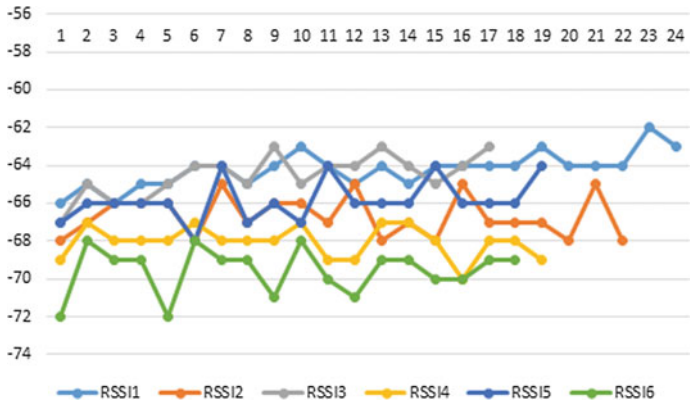


Fig. 2 RSSI measurements from same BLE device at different locations

5.3 *Analysis of Various Filters*

The data collected during the experiment is analyzed. In the analysis, we compare different filtering methods for test case 1 and test case 2. In both the test cases, it is observed that filtering applied to the RSSI readings reduces the fluctuations. Tables 1 and 2 show the mean, variance and standard deviations for different filtering methods in different test cases. The frequency–Kalman filter's accuracy is determined by the values of R and Q. By manually changing the measurement noise R and process noise Q, two test scenarios were explored. We observed that variation of R and Q has a considerable impact on Kalman filtering algorithms.

In test case 1, the standard deviation obtained at distance 1 m for the frequency–Kalman filter is 0.89921 where the raw signal shows a standard deviation of 2.37566. The standard deviation after applying the frequency–Kalman filter for 2 m, 3 m, 4 m, 5 m, 6 m is 0.49345, 4.15228, 1.9683, 1.5524, 0.46153, respectively. But the standard deviation of the raw signals with*o*ut applying the filters at 2 m, 3 m, 4 m, 5 m, 6 m is 5.8789, 6.1874, 2.7259, 2.0483, 5.4232, respectively. The standard deviation after applying the filters such as frequency analysis filter, Kalman filter and Kalman–frequency filter also decreases the standard deviation versus the standard deviation of raw signals. Standard deviation obtained for several filtering algorithms at various distances is depicted in Fig. 3 for test case 1.

Figure 3 shows that for each distance 1, 2, 3, 4 m, 5 and 6 m the longest vertical bar is for the raw signals where no filtering is applied. The maximum standard deviation is obtained at 3 m with a value of 6.1874, and the minimum standard deviation value is obtained at 5 m with a value of 2.0483 for the raw signal. The shortest vertical bar is for the frequency–Kalman filter. The maximum standard deviation is obtained at 3 m with a value of 4.15228, and the minimum standard deviation value is obtained at 6 m with a value of 0.46153 for the raw signal.

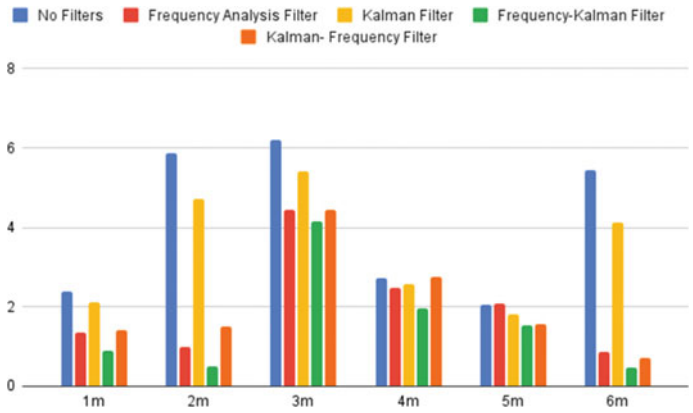


Fig. 3 Standard deviation of different filters in test case 1

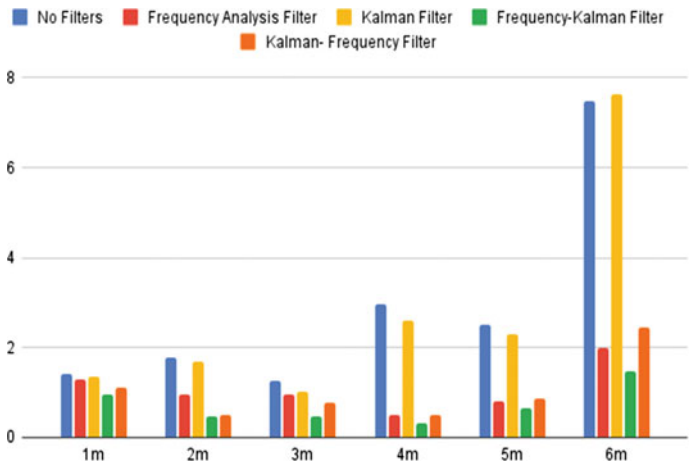


Fig. 4 Standard deviation of different filters in test case 2

In test case 2, the frequency–Kalman filter has a standard deviation of 0.96952 at a distance of 1 m, while the raw signal has a standard deviation of 1.4177. After applying the frequency–Kalman filter, the standard deviation is 0.4829 at 2 m, 0.478464 at 3 m, 0.30929 at 4 m, 0.6485 at 5 m and 1.489 at 6 m. However, the standard deviation of raw signals without applying filters is 1.7928 at 2 m, 1.27241 at 3 m, 2.9766 at 4 m, 2.5095 at 5 m and 7.4816 at 6 m. The standard deviation after filtering with frequency analysis filter, Kalman filter and Kalman–frequency filter decreases when compared with the raw signals.

For test case 2, the standard deviation among several filtering techniques at varying distances is shown in Fig. 4. The longest vertical bar in Fig. 4 is for the Kalman filter for each distance of 1, 2, 3, 4, 5 and 6 m. For the raw signal, the greatest standard deviation is 7.4816 at 6 m and the minimum standard deviation is 1.005934 at 5 m. The frequency–Kalman filter has the shortest vertical bar.

For the raw signal, the maximum standard deviation is obtained at 3 m with a value of 4.15228 and the minimum standard deviation is obtained at 6 m with a value of 0.46153. The frequency–Kalman filter has the shortest vertical bar in both test situations.

The values away from the mean are considered as the disturbance in the RSSI. In this work, the standard deviation value can show how close the RSSI are to the mean. As a result, the standard deviation value in this experiment reveals the disturbance in the RSSI data. It is logical to assume that if the sample's standard deviation is lower the disturbance will be lower as well. In both test cases, it is noticed that frequency–Kalman filtering reduces noise better than other filtering methods.

6 Conclusion

In this paper, the raw RSSI signals from BLE devices are filtered using different filters for the mitigation of RSSI variations in BLE devices. The frequency filter, Kalman filter and its combination filter were investigated. Our experimental results suggest that using frequency in conjunction with a Kalman filter produces the best outcomes. When using a frequency filter before a Kalman filter, the standard deviation is lower than when using a Kalman filter before a frequency filter. So, we concluded from the analysis that the performance of frequency–Kalman filtering was found to be best among the other filtering methods used in this experiment.

In this work, one challenge we faced was that during scanning for a fixed time interval to record the RSSI values, the number of samples received was not consistent.

In future work, we will apply the filtering process discussed above in the test cases with real-time environmental factors such as human movement considered. The RSSI can be affected by the presence of more number of mobile users and devices in the vicinity of the BLE devices.

Acknowledgements The authors would like to appreciate the funding agency, DST-NGP division (formerly NRDMS), Department of Science and Technology, Government of India, with File No: NRDMS/UG/NetworkProject/e-13/2019(G) P-2 for granting the fund for the proposal. The authors would like to extend the gratitude for the efforts of our institute management and the officials.

References

1. R. Yamasaki, A. Ogino, T. Tamaki, T. Uta, N. Matsuzawa, T. Kato, TDOA location system for IEEE 802.11b WLAN, in *IEEE Wireless Communications and Networking Conference*, vol. 4 (2005), pp. 2338–2343. https://doi.org/10.1109/WCNC.2005.1424880
2. A.K.M.M. Hossain, W.-S. Soh, A comprehensive study of bluetooth signal parameters for localization, in *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications* (IEEE, 2007)
3. Q. Dong, W. Dargie, Evaluation of the reliability of RSSI for indoor localization, in *2012 International Conference on Wireless Communications in Under- ground and Confined Areas*, (2012), pp. 1–6
4. F. Soldovieri, G. Gennarelli, Exploitation of ubiquitous Wi-Fi devices as building blocks for improvised motion detection systems. Sensors **16**(3), 1–13 (2016)
5. J. Wilson, N. Patwari., See through walls: motion tracking using variance-based radio tomography network. IEEE Trans. Mob Comput. **10**(5), 612–621 (2011)
6. Q. Lei, H. Zhang, H. Sun, L. Tang., A new elliptical model for device-free localization. Sensors **16**(4), 1–12 (2016)
7. A. Booranawong, N. Jindapetch, H. Saito, Adaptive filtering methods for RSSI signals in a device-free human detection and tracking system. IEEE Syst. J. **13**(3), 2998–3009 (2019). https://doi.org/10.1109/JSYST.2019.2919642
8. A. Booranawong, K. Sengchuai, N. Jindapetch, Implementation and test of an RSSI-based indoor target localization system: human movement effects on the accuracy. Measurement (2018). https://doi.org/10.1016/j.measurement.2018.10.031

9. C.S. Mouhammad, A. Allam, M. Abdel-Raouf, E. Shenouda, M. Elsabrouty, BLE indoor localization based on improved RSSI and trilateration, in *2019 7th In- ternational Japan-Africa Conference on Electronics, Communications, and Computations, (JAC-ECC)* (2019), pp. 17–21. https://doi.org/10.1109/JAC-ECC48896.2019.9051304
10. Q. Dong, F. Zhu, Y. Cai, L. Fang, M. Lu., Analysis of RSSI feasibility for sensor positioning in exterior environment, in *2021 Wireless Telecommunications Symposium (WTS)* (2021), pp. 1–7. https://doi.org/10.1109/WTS51064.2021.9433708
11. S. Huang, K. Zhao, Z. Zheng, W. Ji, T. Li, X. Liao, An optimized fingerprinting-based indoor positioning with kalman filter and Universal Kriging for 5G Internet of Things. Wirel. Commun. Mob. Comput. **2021**, 9936706 (2021)
12. A. Mackey, P. Spachos, K.N. Plataniotis, Enhanced indoor navigation system with beacons and kalman filters, in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (2018), pp. 947–950. https://doi.org/10.1109/Global-SIP.2018.8646581
13. L. Alsmadi, X. Kong, K. Sandrasegaran, G. Fang, An improved indoor positioning accuracy using filtered RSSI and beacon weight. IEEE Sens. J. **21**(16), 18205–18213 (2021). https://doi.org/10.1109/JSEN.2021.3085323
14. J. Massollar, A.C.B. Garcia, FenceBot: an elderly tracking app for mitigating health risk contacts, in *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (2021), pp. 1009–1014. https://doi.org/10.1109/CSCWD49262.2021.9437612
15. M. Ibrahim, O. Moselhi, Enhanced localization for indoor construction, in *Proceedings of the Creative Construction Conference 2015, Krakow, Poland, 21–24 June* (2015), pp. 241–249
16. Z. Huang, X. Zhu, Y. Lin, L. Xu, Y. Mao, A novel WIFI-Oriented RSSI signal processing method for tracking low-speed pedestrians, in *2019 5th International Conference on Transportation Information and Safety (ICTIS)* (2019), pp. 1018–1023. https://doi.org/10.1109/ICTIS.2019.8883759
17. A. Chatzimichail, A. Tsanousa, G. Meditskos, S. Vrochidis, I. Kompatsiaris, RSSI fingerprinting techniques for indoor localization datasets, in *Internet of Things, Infrastructures and Mobile Applications. IMCL 2019. Advances in Intelligent Systems and Computing*, vol. 1192, ed. by Auer M.E., Tsiatsos T. (eds) (Springer, Cham). https://doi.org/10.1007/978-3-030-49932-7_45
18. N.S. Duong, V.T.A. Trinh, T.T.M. Dinh, Bluetooth low energy based indoor positioning on iOS platform, in *2018 IEEE 12th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)* (2018), pp. 57–63. https://doi.org/10.1109/MCSoC2018.2018.00021
19. P. Suwannawach, S. Chivapreecha, Reduce RSSI variance for indoor localization system using frequency analysis. Int. J. Future Comput. Commun. **8**(2), 34–38 (2019)

A Novel Approach for High Authentication in Digital Watermarking Technique

Mulatu Gebeyaw Astarkie, Swapna Gangone, Bhoomeshwar Bala, G. J. Bharat Kumar, and Yagnam Nagesh

Abstract Digital image watermarking is a fascinating area of research as they protect multimedia information from unauthorized access. It has become very important in our Internet community. Digital watermarking techniques are used to protect electronic media against illegal flow in the form of bitmaps, sound and movies or any other network, watermark into the server information with copyright protection, access management, broadcast tracking, owner identification, etc. A watermark can be almost any form, label, tag or electronic signal. A host can be multimedia elements such as pictures, video or audio. Since images are the building blocks of multimedia, developing a productive watermark will be critical method for graphics. The most obvious way to embed data in multimedia information is through electronic watermarking. The digital watermark is a powerful alternative and plays an important role in copyright security.

Keywords Digital watermarking · Copyright · Authorized access · Data hiding

1 Introduction

Today, many of these images on the web are not watermarked; it becomes quite easy for anyone to download and modify these images, which could result in the founder losing the rights to his image. The profitability of gambling applications, high-quality digital artwork and video streaming around the world has improved significantly with advances in technology. Digital media offers you few mixed advantages over analog—the caliber of electronic sound, graphics and video signals is

M. G. Astarkie · B. Bala (✉) · Y. Nagesh
Department of Information Technology, Debre Tabor University, Debre Tabor, Ethiopia
e-mail: drbhoom08@dtu.edu.et

Y. Nagesh
e-mail: nageshyagnam1@dtu.edu.et

S. Gangone · G. J. Bharat Kumar
Department of Computer Science, Mettu University, Mettu, Ethiopia

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_23

greater than that of analog counterparts. Editing becomes easy as you can make changes in specific places that need to be changed. The data is straightforward with no lack of quality. One approach to protecting multimedia information from illegal recording and retransmission is to embed a character known as a digital copyright or signature tag or watermark that fully characterizes who is applying it.

The term "electronic watermark" was coined in 1993 when Tirkel introduced two watermarking methods to embed the watermark in images. Media (e.g., text, image, sound and video) that is watermarked can later be detected or extracted to recognize the item. An electronic watermark system integrates data directly into a file or over a network. The digital watermark alone cannot stop files from being copied, modified and even redistributed. However, if security and copy protection are neglected, the watermark allows conclusions to be drawn about the rightful owner and the purpose of unauthorized use. Digital watermarking requires components from various fields such as telecommunications and signal processing, cryptography, psychophysics and regulation.

2 Related Survey

The digital watermark can be distinguished as the activity of hiding a message, articles, mark or logo in an image, audio recording, video or other type of networking. The field of electronic watermarks is comparatively young and has only gained popularity as a research topic since the late 1990s. Watermarks can be seen everywhere, just like images on banknotes, and the watermark is hidden in the press. The watermark can be useful for physical things; illustrations include materials, garment labels and merchandise packaging that can be watermarked with extraordinary and imperceptible colors and inks or even digital signs.

Cryptography is the search for ways to send messages in other ways so that the target audience of this medium can get rid of the disguise and carry and capture the message. The plan of altering over plain text to cipher text is popularly called enciphering or encryption, and also the converse procedure is called decryption. Encryption protects information during transmission within the transmission medium. But after the material is received and received, it is no longer protected. The newspaper's watermark was first displayed in Italy in 1282. It all started with the inclusion of a small wire in the type of paper that had a simple print inside the paper (which made the newspaper identifiable or used as a signature) as an understanding of the forms in which the newspaper sheets were made, or simply as registered trademarks to differentiate the paper maker. On the other hand, they could have been spoken in cryptic signals or introduced as an advance. It started functional in Europe and America as well in twentieth century. They were used as trademarks to document the date the paper was made and to prove the invention [1]. Steganography, or cipher writing, could be seen in ancient Indian literature, such as in Kautilya's Arthsastra and Vatsayana's Kamasutra. They are made of thin sheets of silk and paper. From

the twentieth century onwards, Emil Hembrooke of this Muzac Corporation discovered the electronic watermark of 1954. Over time, watermarking techniques gained critical focus and became an emerging part of this study.

In comparison with cryptography, steganography works differently in that it hides the presence of the message. In the general regulations, the covert communications of spies and the messages exchanged between drug dealers by e-mail in an incomprehensible manner are all types of steganography. The steganography and watermarking techniques use covert communication techniques, but the steganography techniques are by no means robust. The message has been hacked or tampered with, or changed between the wires, and then, the message may not be seen as intended on the receiving end. We evaluate steganography using the watermark method; the watermark process is a subset of steganography. In steganography, the hidden message and the displayed message have no connection at all. The hidden information or the watermark is the owner's authentication key for his cover image or server in the watermark process. The watermark strategy helps a person keep rightful possession of their media. In order to evaluate cryptography with the cryptography watermarking method, the information is converted into an individual, non-identifiable form so that only the uninstalled program can change it. But after the message has been encrypted, the message is locked for all types of attacks, as it can currently be fully read by anyone. It makes cryptography more reliable and less vulnerable. By using the watermarking method, all of these problems are overcome and more robustness is gained.

3 Proposed System

Hybrid domain watermark algorithms and current trends in watermarking hybrid domain watermark algorithms are often viewed as a combination of plasma and transform domain computations. These algorithms guarantee both robustness and improved data compilation properties. Several studies on hybrid methods have been completed. These studies show the present watermark trends. Reference authors [2] have merged the location and frequency domain names, such as the image watermark, as an example. For example, more watermark information can be embedded in the server image. This technique increases the capacity of the host image and splits the watermark into two components, which improves security. The spatial domain replaces the LSB bits of the server image with the parts of the server image watermark image. On the other hand, the frequency range name inserts data into the low-frequency elements of the server image.

In addition, a random permutation of the watermark can be used to increase robustness against various signal processing attacks, including image archiving. Another strategy [3] was launched to secure digital rights to a hybrid domain name. The scheme protects electronic content by transmitting it to an unsecured station using the least significant parts as well as a wavelet transform (DWT and SVD) in which the server image is divided into sub-bands (LL, HL, HH and LH) with the frequency range, name conversion process. To extract the watermark, the decoding algorithms

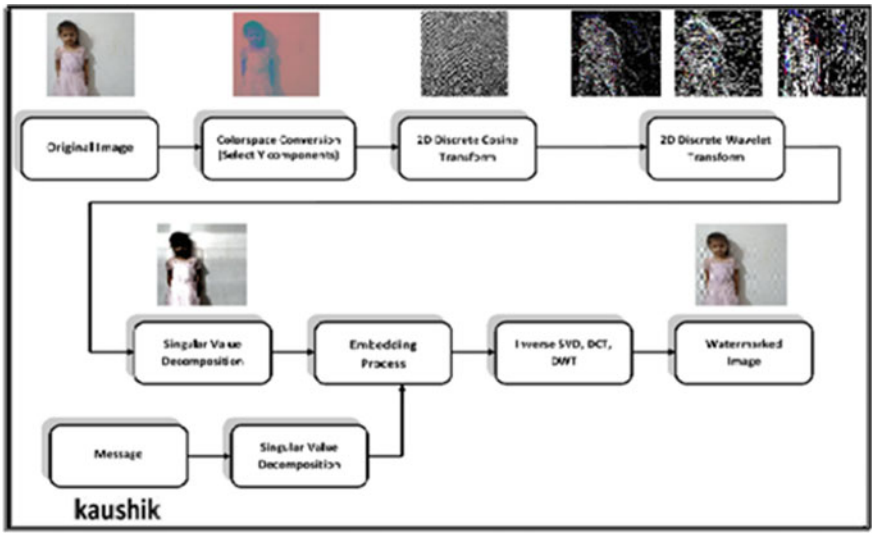


Fig. 1 Hybrid watermarking approach

defined above are used; this hybrid scheme offers higher quality and robustness against various attacks, for example Gaussian noise and JPEG compression; a hybrid of DWT and SVD has been proposed, which enhances the robustness and imperceptibility. Current watermarking algorithms are slightly less powerful than geometric attacks. In order to withstand geometric attacks, a multi-watermarking algorithm for medical imaging was therefore proposed, which depends on the dual tree complex wave transformation (DTCWT), DCT together with a Henon map [4–6]. This algorithm can be used for medical protection, protection authentication and cloud storage along with cloud transmission (Fig. 1).

Then the *Y* element is selected for the watermark. The explanation for choosing the *Y* element for the watermark is that it is extremely safe and the human eye is more sensitive to the adjustments introduced in your watermark. Then the discrete cosine transformation is implemented in *Y* part of the picture. Then, three different degrees of wavelet transforms are implemented. In the latter case, the image is decomposed using the singular value decomposition system. The watermark that will be added to the image is selected. The singular value decomposition is placed over the watermark. The value of the changed image is changed according to the unique values of this current SVD watermark image. Inverse SVD is inserted into the resulting image. Then, a DWT inverted by three degrees is used. Finally, DCT is implemented. The color space conversion of the watermarked image from YCbCr to RGB then takes place. In this way, the embedding of a watermark in an image using the proposed watermarking technique is completed (Fig. 2).

Initially, a 2D watermarked image is selected, then the color space conversion from RGB to YCbCr is overlaid on the selected image, and then the *Y* element is selected to extract the watermark. Then, the discrete cosine transformation is implemented

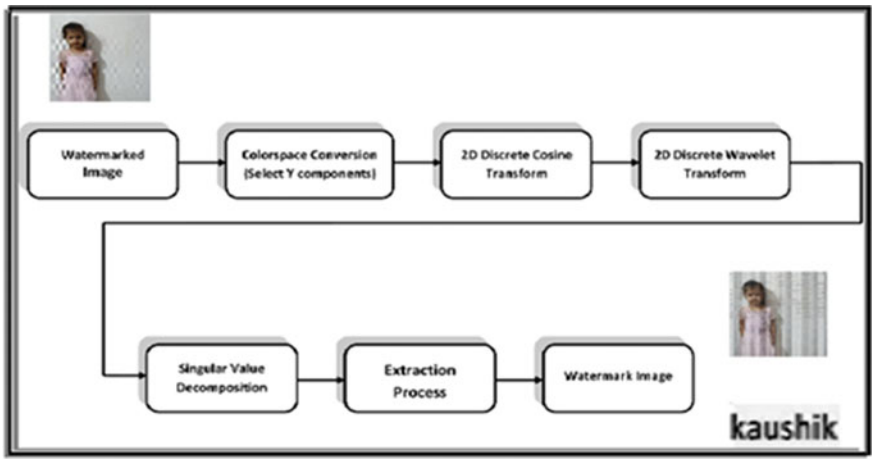


Fig. 2 Hybrid approach extraction process

in this AND part, after which three different degrees of wavelet transformations are implemented. At this point, the singular value decomposition is placed on the image. The value of the changed image will be changed to keep the restored content. Then, the color space conversion is completed on the retrieved watermark image from YCbCr to RGB. The watermark of an image is completed along with the extraction process for the projected watermark.

Some calculations degrade the image quality, and others, distort it severely. Some are difficult to reach, and some are exceptionally robust to standard image processing operations, but not resistant to the elements. On the other hand, others are exceptionally robust to geometric impacts, but quite sensitive to various types of noise. Some of them guarantee much better PSNR and NCC levels, but others result in a reduced PSNR depending on the programs you want, some are strong.

4 Conclusion and Future Work

At present, the information could be easily reproduced, thanks to the interactive and electronic communication of multimedia information. The topic of electronic image watermarks creates a considerable area of study. In this review, we look at the most current and dominant watermarking methods; it is argued that DWT is a powerful and high-quality method for image watermarking because of its multi-resolution attributes. Robustness, imperceptibility and capacity are the key requirements for designing an efficient watermarking system. However, it is almost impossible to meet all of these requirements at the same time. Hence, fantastic compensation needs to be maintained that encompasses all three of these requirements. But security remains a major hurdle in electronic image watermark technology,

and hosting IoT and blockchain-based authentication approaches poses a hurdle for researchers. Hence, future work will be expanded to include the three technique names by combining different techniques in different areas Main requirements to be met. In order to promote robustness in addition to safety, researchers must also concentrate on developing innovative and innovative processes.

2D discrete cosine transform
2D discrete wavelet transform
Singular value decompision

References

1. H. Tao, L. Chongmin, J.M. Zain, A.N. Abdalla, Robust image watermarking theories and techniques: a review. J. Appl. Res. Technol. **12**, 122–138 (2014)
2. F.Y. Shih, S.Y. Wu, The spatial and frequency domains. J. Pattern Recognit. Soc. **36**, 969–975 (2003). Combinational image watermarking
3. Sridhar, Hybrid frequency domain. Int. J. Eng. Technol. **7**, 243–248 (2018). Robust digital image watermarking
4. N.A.K. Abdulrahman, S. Ozturk, Novel hybrid DCT, and DWT based robust watermarking algorithm for color images. Multimed. Tools Appl. **78**, 17027–17049 (2019)
5. Y. Zhang, Digital watermarking technology: a review, in *Proceedings of the ETP International Conference on Future Computer and Communication, Wuhan, China, 6–7 June* 2009, pp. 250–252
6. H.M. Yang, Y. Liang, X.D. Wang, S.J. Ji, A DWT-Based evaluation method of imperceptibility of watermark in watermarked color image, in *Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2–4 November 2007*, pp. 198–203

Exploring the Impact of Indian Revenues During COVID-19 Using Social Network Analysis

Suriyakrishnan Sathish, Aravinda Boovaraghavan, S. Hashwanth, and L. Jani Anbarasi

Abstract The second wave of the COVID-19 pandemic affected economy resulting in job loss for many throughout the world. The main objective is to analyze the Indian revenues before COVID-19, i.e., before December 2019 and the ongoing COVID-19 pandemic. A comparison is done between the revenues that are collected from various departments sectors for the years 2019–2020 and 2020–2021 to find out how the COVID-19 has affected the Indian economy as well as jobs for millions of people. Studying the network through social network analysis helps us to understand clearly how the Indian economy has changed during the COVID-19 period. The network is studied using an average path along with a weighted path to ascertain the small word characteristics. Centrality measures helps us to understand the characteristics of the dataset. The tools used to analyze them are designed using Gephi.

Keywords COVID-19 · Economy · Revenues · Indian economy · Social network analysis

1 Introduction

In recent times, we have heard a lot of increase in COVID-19 cases in India. This deadly disease not only affects human beings but also affects the whole economic system in India. Because of this pandemic, in the second wave of COVID-19, more than 15 million Indians have lost their employment. This has a significant impact on both the Indian economy and the lives of Indian citizens. The COVID-19 pandemic has had a significant economic impact in India where the growth declined to 3.1% in the fourth quarter of financial year 2020, according to the Ministry of Statistics. This decrease is primarily due to the impact of the Corona virus pandemic on the Indian economy. All the major sectors are affected due to the lockdowns as the factories, manufacturing companies, transportation, small-scale industry, construction, food

S. Sathish (✉) · A. Boovaraghavan · S. Hashwanth · L. J. Anbarasi
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India
e-mail: suriyakrishnan.s2019@vitstudent.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_24

wastage increased due to affected supply chains, affecting small farmers and other departments. As a result, the daily wage workers, migrant workers, and household workers are severely hit in this pandemic.

Consumer demand, which was already slowing prior to the pandemic, has now collapsed, accounting for nearly 55% of the economy as household incomes and jobs have declined. As a result, it is important to understand the various reasons the economy has been affected. Similarly, it is seen that there is a major effect of COVID-19 on Indian revenue. There is no existing research for the Indian revenue data before and after COVID-19. So, an analysis was performed in this area and presented an outline between these two periods. Janssens et al. [7] investigated economic effects on low-income households in rural Kenya. An analysis using weekly financial household data was performed. World Development, 138, 105,280. According to the World Bank, the COVID-19 problem has the potential to force between 40 and 60 million people into extreme poverty, the majority of whom are in Sub-Saharan Africa [12]. According to another research, if income and consumption decrease by 20%, between 420 and 580 million people will be forced into poverty, reversing decades of decreasing poverty trends [11]. However, for most low-and middle-income nations, the immediate implications of the COVID-19 crisis at the household level are missing. The majority of the evidence comes from wealthy countries. In the United States, the lockdown policy reduced time spent outside the home [5], contributed to a significant drop in employment [9], and resulted in a significant drop in job vacancy postings [3].

Social network analysis has always been a prominent field of research. With the increased popularity of online socializing, a great amount of data and information are gathered. The amount of information available is enormous, and Big Data [2, 6] has resulted in numerous issues. Furthermore, such information provides opportunities for researchers in a variety of sectors. Social network analysis is one such field. With social network analysis, you may learn a lot about how people interact and socialize with one another. Such social data can be used to track the user behavior of an individual as well as a community. Based on the metrics, an analysis is performed with the Indian data to study the effects of the economy before and after COVID-19.

The paper is structured as follows: Sect. 2 elaborates the data and the method used for analysis, Sect. 3 details the related work along with all the metrics used for evaluation. Section 4 details the proposed work followed by observations and discussions. Section 6 concludes the work with its future scope.

2 Data and Methods

In our research, two communities are formed out of which one community refers to the 3 quarters of 2019 (i.e.) Pre-COVID period and the other community refers to the 3 quarters of 2020. Data from 8 major departments is collected in order to form a network, and the same are segregated as per their year. This resulted in values for

Table 1 Data collected from various departments

List of departments analyzed	Generated dataset in network form
Agriculture and allied activities	
Finance, insurance, real estate and business services	
Trade, hotels, transport and communication	
Community, social and personal services	
Transport and communication [6, 11]	
Mining and quarrying	
Electricity, gas and water supply	
Mining and quarrying	

the eight departments in Pre-COVID-10 data as well as during COVID-19 data. The eight departments that are considered for this research are given in Table 1a.

From the main data, only the revenue data for the years 2019 and 2020 is collected. Then the data is segregated into separate csv files, forming the nodes, resulting in 2 datasets, 2019 and 2020. With the help of this data in the proposed work edges are manually formed, by connecting the 3 quarters of a year to each of the eight departments. The 3 quarters represent a specific month range in a year. Here, Quarter 1 denotes April to June, Quarter 2 denotes July to September, Quarter 3 denotes October to December, and finally, Quarter 4 denotes January to March. For the research process 2019–20 Quarter 1, 2, 3 is taken for pre-COVID-19 data and 2019–2020 Quarter 4 along with 2020–2021 Quarter 1 and 2 for During COVID-19 data. The data has been segregated as per the mentioned quarters and then it is linked from those quarters to the departments mentioned. The nodes representing the quarters are considered as the source, and then the target nodes are nodes that represent each department [1, 4, 8]. The graph type that is declared is an undirected graph and importing the datasets (both node and edge datasets) in such a form as shown in Table 1b.

3 Related Work

This section details the measure of modularity, Girvan-Newman Community Detection, average degree and weighted degree, which are used to measure the proposed

work. To understand the economy in different quarters, degree, betweenness, and eigenvector centrality were performed. To comprehend the network's complexity and the adhesiveness of the nodes inside it, the clustering coefficient and average shortest path [10] were determined.

Modularity, which measures the structure of graphs, is used to assess the strength of a network's partition into modules. High modularity refers to dense connections between nodes within modules, and low modularity refers to weak connectivity among nodes and is commonly utilized in optimization techniques for identifying community. The modularity of the network model is computed as given in Eq. (1).

$$\text{Mod} = \frac{1}{M} \sum_{i,j} \left[M_{ij} - \frac{M_i M_j}{M} \right] \delta_{c_i c_j} \tag{1}$$

M_i and M_j are the input and output of the nodes 'i' and 'j'. δ is the Kronecker's delta where $\delta_{c_i c_j} = 1$ refers that nodes 'i' and 'j' from same community or 0 otherwise.

By progressively eliminating edges from the original network, the Girvan-Newman algorithm identifies the communities. This identifies the edges that are most likely "between" communities, rather than measuring the vital edges. By deleting these edges, the groups are separated from one another, showing the network's underlying community structure.

Average degree refers to the average number of links present for each node. It refers to the number of links compared to the number of nodes in an undirected graph. The straightforward computation of the average degree is computed as follows (Eq. 2).

$$\text{Average Degree} = \frac{\text{Total Edge}}{\text{Total Nodes}} \tag{2}$$

The term "degree centrality" refers to how many alters an ego is linked with. The centrality of a node is determined by the number of neighbor nodes that are directly linked to it. It relates to the degree to which departments are involved in this study Eq. (3).

$$C_D(i) = \sum_{j=1}^{n} d_{ij} \tag{3}$$

Betweenness centrality assesses a node's ability to act as an intermediary in a network. The centrality of betweenness is expressed in Eq. (4) where g_{jk} is the number of the shortest distance links between j and k and $g_{jk}(i)$ is the number of times the path crosses these two nodes.

$$C_B(i) = \sum_{j<k}^{n} \frac{g_{jk}(i)}{g_{jk}} \tag{4}$$

The nodes which influences more in a network is identified using eigenvector or eigen centrality. It shows which node has had the most impact on the other nodes.

The eigenvector centrality (also known as eigen centrality) of a node in a network is a measure of its influence.

Closeness centrality is used to determine a node's reach in a network. The minimal hops required to move from one node to the other is the connectedness of a pair of nodes in undirected and unweighted networks. The closeness centrality of a node u is defined as follows in Eq. (5)

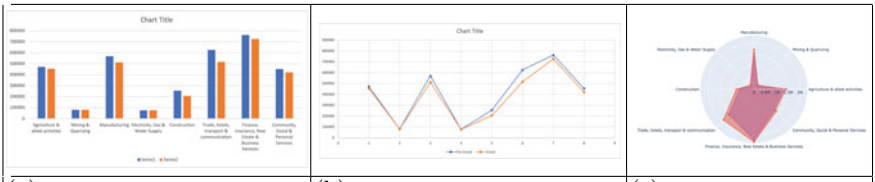
$$C_u(u) = \frac{n-1}{\sum_{\forall v} d(u,v)} \tag{5}$$

4 Proposed Work

In this research, two datasets before and after COVID-19 are collected and compared across both networks to determine the impact of COVID-19 on our Indian revenue. From the results achieved, it can be seen that there is a strong effect of COVID-19 on many industries and departments. The nodes are the three quarters and the 8 departments. In both the segregated data, the departments are the same except for the quarters. The analysis and graph generation were performed using Gephi. The clustered column graph in Table 2a shows department vs revenues before the pandemic and during the pandemic. In Table 2b, it shows the revenue which was collected before the pandemic is slightly higher than in the COVID-19 period.

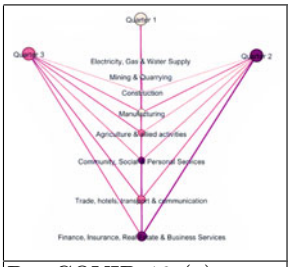
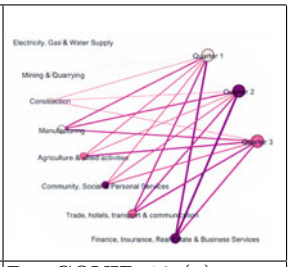
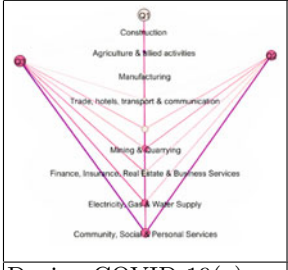
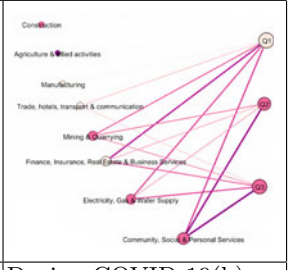
The social network model is generated for both the datasets to gain knowledge about the revenue status during the pre-COVID-19 period. The edge ranking, node color, and node size based on calculated values are defined. Edge ranking is based on weight , so the more the weight, the darker the edge is visible. Also, for node size, it is defined on the basis of weight. The minimum size is given as 1, and the maximum size of the node is given as 4. The node color is generated with respect to modularity. Once the average modularity is computed, the node color is displayed based on the calculated value. The higher the value of modularity, the darker the node color. With the help of these, the nodes and edges are defined so that a clear view of the network model is described. During the pre-COVID-19 period, Finance, Insurance, Business services hold the highest revenue values, followed by others. Table 3a, b shows the network model generated for pre-COVID-19 data.

Table 2 Revenue generated before and after COVID-19



(a) (b) (c)

Table 3 Social network model for pre-COVID-19 and during COVID-19 data

	
Pre COVID-19 (a)	Pre COVID-19 (a)
	
During COVID-19(a)	During COVID-19(b)

The graph is generated from the dataset using Gephi, to attain knowledge about the revenue status during the COVID-19 period. The same procedure is followed for edge ranking, node color and node size which is based on calculated values. Edge ranking is based on weight, so the more the weight, the darker the edge is visible. Based on average modularity the node color is displayed based on the calculated value. The higher the value of modularity, the darker the node color. With the help of these, the nodes and edges can be defined so that it gives us a clear view of the network. It is seen that during the COVID period, Community, Social, and Personal services hold the highest revenue value, followed by Electricity, Gas and Water Supply. So if listed down in ascending order, like the higher revenue value node holds the first position followed by the next value, results in: Community, Social and Personal Services followed by Electricity, gas and Water Supply followed by Finance, Insurance, Real Estate and Business services followed by Mining and Quarrying followed by Trade, hotels, transport and communication, followed by Manufacturing, followed by Agriculture and Allied activities, and construction. Table 3a, b shows the network model generated during COVID-19 period. So, from the two formed graphs of both the pre-COVID-19 and during COVID-19 period. It can be easily seen that there is a tremendous effect on the revenue table. Top departments have been shattered during this COVID period. A comparison of the revenue data between pre-COVID-19 and during COVID-19 data is shown in Table 4.

Table 4 Metrics used to evaluate Pre COVID-19 and During COVID-19 data

Measures	Pre Covid-19	During Covid-19
Average degree	4.364	4.364
Average weighted degree	1,800,740.364	1,636,174.182
Eigen vector centrality	0.0038	0.0038
Graph density	0.436	0.436
Modularity	0.031	0.032
Betweenness centrality	9.333333	9.333333
Closeness centrality	0.833333	0.833333

From the table, the effect of COVID on the departments can be visualized where Finance sectors had a greater impact. Pre-covid, it was the hottest department with respect to revenue, but this covid had a greater effect on that department. It has been pushed down from first place to third place. At the same time, it can be seen that Community, Social, and Personal Services have been raised during this COVID period. It simply shows the fatal truth of how people were affected by the COVID wave, both financially and personally. If personal expenses increase, then financial problems arise. After this, it can be seen that Electricity, Gas and Water supply have the greatest increase. Due to quarantine, it can definitely said that more electricity will be needed and more water supply will be needed. So there has been massive growth in the revenue of these departments, which made them get top profit during this COVID period. Also, the trade, hotels, transportation, and communication sectors has experienced a significant drop in the lane. Mining and Quarrying may seem to have increased, but by checking with values, it is seen that there has been no change in their revenue status.

5 Observation and Discussion

Basic parameters like average degree and average weighted degree can be used to analyze the network's topology. The average degree of the network trained for the Pre-COVID-19 is 4.364, and the average weighted degree is 1,800,740.364. The average degree of the network constructed during COVID-19 is measured to be 4 and the average weighted degree of the same is measured as 1,406,244.727. The values clearly shows the downfall of the economy between these periods. The diameter can show us the largest shortest path between the pair of the nodes in the graph. The diameter values obtained for Pre-COVID-19 and during COVID-19 shows a good difference. Modularity refers to the network's partition. The modularity is obtained by randomizing the values with resolution as 1 and including the weights of the edge. The modularity obtained is 0.29 and 3 communities are formed for Pre-COVID-19 data. The modularity obtained for during COVID data is 0.037 with the same number

of communities. The values clearly depicts that the economy fall during the period of COVID-19. Similarly, the communities were analyzed using the Girvan-Newman scheme, and only one community was formulated for both the datasets. The centrality and the density measures doesn't show much difference in the values obtained for the network model.

6 Conclusion

The COVID-19 pandemic has changed all our lifestyles. It is necessary to adapt ourselves to the "new normal". This research will help us in understanding the problems of a new recession and economic crisis in a variety of ways. A comprehensive social-economic development strategy that includes sector-by-sector plans to help businesses succeed with stable and long-term business models during the pandemic. The social network model was created for different sectors. These sectors are analyzed for both pre-COVID-19 and during COVID-19 data. This social network data is analyzed using various metrics in the proposed work. The modularity, average path, centrality measures obtained for both the duration shows a clear impact for both pre-COVID-19 and during COVID-19 data.

References

1. A. Agrawal, Sustainability of airlines in India with Covid-19: challenges ahead and possible way-outs. J. Rev. Pricing Manag. **20**(4), 457–472 (2021)
2. D. Dhanya, R. Vidhya, N. Modigari, A. Davamani, L. Anbarasi, in *Consistency and Reliability of Services in Cloud Environment*, pp. 747–751 (2017). https://doi.org/10.1109/CTCEEC.2017.8455064
3. E. Forsythe, L.B. Kahn, F. Lange, D. Wiczer, Labor demand in the time of Covid-19: evidence from vacancy postings and UI claims. J. Publ. Econ. **189**, 104238 (2020)
4. B. Goswami, R. Mandal, H.K. Nath, Covid-19 pandemic and economic performances of the states in India. Econ. Anal. Policy **69**, 461–479 (2021)
5. S. Gupta, T.D. Nguyen, F.L. Rojas, S. Raman, B. Lee, A. Bento, K.I. Simon, C. Wing, *Tracking public and private responses to the Covid-19 epidemic: evidence from state and local government actions*. Technical Report (National Bureau of Economic Research, 2020)
6. S.L. Kumar et al., Predictive analytics of Covid-19 pandemic: statistical modelling perspective. Walailak J. Sci. Technol. (WJST) **18**(16), 15583–14 (2021)
7. S. Kumar, V. Maheshwari, J. Prabhu, M. Prasanna, P. Jayalakshmi, P. Suganya, B.A. Malar, R. Jothikumar, Social economic impact of Covid-19 outbreak in India. Int. J. Pervasive Comput. Commun. (2020)
8. V. Kumar, Indian tourism industry and Covid-19: present scenario. J. Tour. Hosp. Educ. **10**, 179–185 (2020)
9. L. Montenovo, X. Jiang, F.L. Rojas, I.M. Schmutte, K.I. Simon, B.A. Weinberg, C. Wing, in *Determinants of Disparities in Covid-19 Job Losses*. Technical Report (National Bureau of Economic Research, 2020)

10. N. Prabhakar, L.J. Anbarasi, Exploration of the global air transport network using social network analysis. Soc. Netw. Anal. Min. **11**(1), 1–12 (2021)
11. A. Sumner, C. Hoy, E. Ortiz-Juarez, in *Estimates of the Impact of COVID-19 on Global Poverty*. No. 2020/43, WIDER Working Paper (2020)
12. World Bank, in *Monitoring Global Poverty* (2020)

Fingerprint Image Enhancement for Crime Detection Using Deep Learning

Konakanchi Anusha and P. V. Siva Kumar

Abstract Fingerprint has been most popularly used in many commercial applications for person identification. Latent fingerprints are produced largely via the finger sweat or oil left outs by the suspects accidentally. These impressions are generally blurred plus not observed by naked eye. These fingerprint images in crime prospects are significant evidence to resolve sequential cases. The latent prints are low quality, corrupted by noise and exhibit minor details. Image enhancement is compulsory in latent prints to transform the latent image into superior quality image. To rectify these issues, an automated latent fingerprint identification system is presented here with the aid of convolution neural network (CNN) of deep machine learning algorithm. The images are generally imperfect and complicated to categorize. Therefore, appropriate enhancement processes are made for pre-processing the fingerprint images; i.e., the minutiae features are extracted from the fingerprint images. These features are given to the CNN network as input for training as well as testing. The performance evaluation is done by calculating precision, recall, f1-score and accuracy. The experimental results are made by implementing in python where the proposed achieves a high accuracy rate of 99% recognition rate.

Keywords Convolutional neural network · Fingerprint · Minutiae · Criminal identification system · Image pre-processing

K. Anusha (✉) · P. V. Siva Kumar
VNRVJIET, Hyderabad, India
e-mail: konakanchi3anusha@gmail.com

P. V. Siva Kumar
e-mail: sivakumar_pv@vnrvjiet.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_25

1 Introduction

The fingerprint in the crime prospect acts as a significant function to recognize the criminal concerned in the crime. Crime scene images (CSI) are images captured from the crime area [1]. Fingerprint is a necessary forensic method in the examination of the criminal proof since of its uniqueness as well as persistence [2]. The fingerprint recognition issue is assembled into three sub-domains such as fingerprint employment, authentication as well as fingerprint recognition. Additionally, other than manual work for fingerprint detection by specialists, the fingerprint detection is termed as automatic fingerprint recognition system (AFRS), that is program grounded [3].

Even though AFRS attains talented accuracy on plain as well as rolled fingerprints, the experimental results are moreover not fulfilled for recent fingerprint images [4]. In recent fingerprint recognition, the person(s) or owner(s) of fingerprints identified at the crime prospect is recognized from unclear images containing single or multiple fingerprints which are not copyright to the naked eye, and therefore, an extraordinary procedure is insisted [5]. To effectively classify the fingerprint images, several techniques comprising machine learning as well as deep learning algorithms are employed. But those techniques had their unique limitations, and hence, this work is commenced to rectify those limitations.

The rest of the paper organization is Sect. 2 reviews some existing researches related to criminal identification. The CNN architecture is detailed in Sect. 3. The CNN results with discussions are made in Sect. 4. Lastly, Sect. 5 concludes the work followed by the references.

2 Literature Survey

The existing algorithms for criminal identification system worked with image preprocessing with classical Otsu method (for image de-noising or filtering) and some minutiae marking algorithms. Further, they also utilized several deep learning and machine learning algorithms in succession to work with criminal identification. Hence, in this work other than the classical filtering algorithm, a recent filtering algorithm is used since image enhancement is the primary and important step in criminal identification system. After this image enhancement, a suitable efficient deep learning algorithm is used to employ classification with better recognition rate. The comparative results of the existing criminal identification systems are stated in Table. 1.

As shown in Table 1, several existing algorithms comprising two-class classification [6], reconstruction algorithm [7], fingerprint recognition [8], CNN with LSTM algorithm [9], fast Fourier transform variation [10], minutiae-based approach [11], CNN deep learning algorithm [12], CNN grounded patch learning method [13], CNN method [14], automated DCNN and FFT [15] and fingerprint fusion algorithm [16]

Table 1 Comparison results of existing criminal identification system methodologies

References	Methodology	Advantages	Dis-advantages	Database	Parameters
[6]	Two-class classification problem	Achieve elastic distortion	Do not work with camera/projector lens distortion correction	FVC2004 DB1, Tsinghua distorted fingerprint database, and NIST SD27 database	Matching score
[7]	Reconstruction algorithm	Achieve partial fingerprint reconstruction	–	FVC_2002 and NIST SD4	Matching score
[8]	Fingerprint recognition	Fingerprint security	Less accuracy	Nigeria country database	Matching score
[9]	CNN and long short-term memory (LSTM)	Solve performance issues	Less accuracy	FVC2002 dataset	Accuracy, error rate detection, peak signal to noise ratio
[10]	Variation of fast Fourier transform	Reduce manual efforts	–	MySql database	Matching score
[11]	Minutiae-based approach	Enhances the image	–	–	Matching score
[12]	CNN deep learning algorithm	Complete fingerprint identification analysis	Worked only with less datasets	Manual dataset	Accuracy
[13]	A deep CNN grounded patch learning approach	Identifies the fingerprint pattern	–	Manual dataset	Equal error rate
[14]	CNN approach	Complete fingerprint identification analysis of crime scenes	Precise matching is not achieved	Manual dataset	Accuracy
[15]	"Automated deep CNN (DCNN)" and 'fast Fourier transform (FFT)' filters	Achieves robustness in fingerprint matching	High computational cost	FVC 2002/2004, NIST SD27,	Accuracy, precision

(continued)

Table 1 (continued)

References	Methodology	Advantages	Dis-advantages	Database	Parameters
[16]	Fusion of fingerprint matching algorithm	High identification rate	–	NIST SD27	Identification rate

were proposed. The advantages of the proposed algorithms are achieving elastic distortion, achieving partial fingerprint reconstruction, attaining fingerprint security, solving performance issues, and reducing manual efforts, enhancing the image, complete fingerprint identification analysis, recognizing the fingerprint pattern, identification of fingerprint in crime scenes, achieving robustness in matching the fingerprint images and attaining better identification rate. Further, the datasets utilized by these algorithms comprise the FVC2004 DB1, Tsinghua distorted fingerprint database, NIST, SD27 recent fingerprint database, FVC_2002, NIST SD4 Nigeria country database, FVC2002 dataset, MySQL database and some manually created datasets. By exploiting these datasets the performance measures such as Matching Score, Accuracy, Error Rate Detection, Peak Signal to Noise Ratio, Precision and Identification Rate are obtained.

Though several approaches are proposed with different performance metrics, nevertheless it has some issues like it cannot work with camera/projector lens distortion correction, has less accurate classification results and works only with less datasets. Hence in this work a method is proposed and following contributions are made:

- Along with a deep learning classification algorithm, a filtering algorithm and an edge detection algorithm are employed that initially enhances the image and also clearly detects all the edges of the fingerprint.
- Several performance metrics parameters such as the precision, recall, F1-score as well as accuracy are measured for the proposed criminal identification system and are also compared with existing algorithm to validate the proposed method effectiveness.
- Further, the proposed fingerprint criminal identification system is validated with three open source datasets such as the FVC2002, FVC2004 and NIST SD27.

3 Proposed Methodology

The fingerprints are the abrasion impression done via the human fingertip. In this research, the fingerprint identification is done to recognize the criminals. To increase the issues identified in the existing algorithm, with three different datasets like the FVC2002, FVC2004 and NIST SD27 which are employed for both training and testing. The proposed system generally centers on enhancing the accuracy of the criminal matched fingerprint. Initially, the fingerprint is collected from the crime

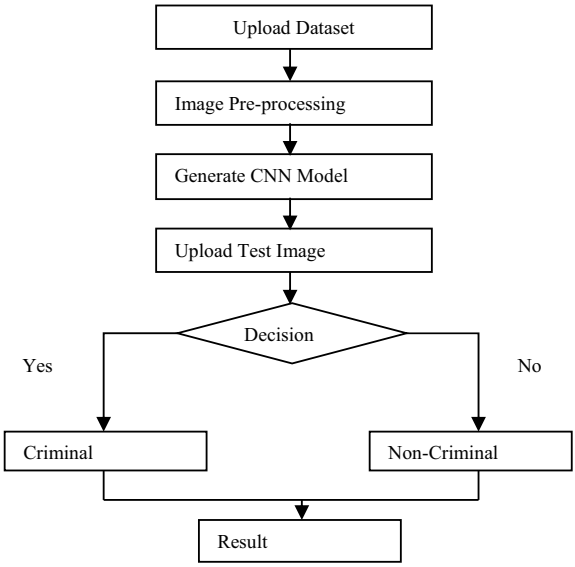


Fig. 1 Flowchart for proposed algorithm

place, and this image is converted to grayscale image and further filtering is done with aid of the Gaussian filter. After filtering the edges of the image, the minutiae of fingerprint image is attained. By these, the features of the fingerprint images are extracted. These features are utilized by the proposed CNN to categorize the criminal fingerprint image and non-criminal fingerprint image. These processing is pictorially illustrated in Fig. 1.

3.1 Image Pre-processing

The input fingerprint images attained from the crime prospect are considered as a major proof in criminal cases. But the fingerprint image data directly gathered from the crime prospect are difficult to categorize. This is due to the low quality of the images since it may be affected by some noise. To work with this issue, image pre-processing is done; that is, image filtering is implemented with the input fingerprint image by utilization of a Gaussian filter [17]. The input fingerprint image is formerly converted to a grayscale image, and then, filtering is done. The gray scale conversion [18] is approved out with the subsequent Eq. (1).

$$0.2989 * R + 0.5870 * G + 0.1140 * B \tag{1}$$

Subsequently, Gaussian filter employed here mostly will eliminate the blurriness of the gray image. After de-noising, the images edges are obtained via employing canny edge detection algorithm, and thus, the fingerprint image features are extracted.

The fingerprint features are the macro-information of the fingerprint ridge stream shape, minutiae features which aid in recognition and finally the fingerprint pore features. Here, minutiae feature is considered as an important feature since it establishes fingerprint individuality. An edge is termed as the boundary among two regions with comparatively different gray-level possessions.

3.2 Image Classification

The output features from feature extraction is sent as input to the proposed CNN model. The leftover features are also extracted by the proposed CNN to discover the equalization function among input fingerprint image pixels. The proposed CNN has two convolution layers, two maxpooling layers, single flatten layer and two dense layers. The CNN exploits consecutive convolutional layers which gather the image data; further, it improves the image pixel properties like pixels values sharpening. The two convolutional layer employed comprises of 32 kernels, 3 padding and 3 strides. This convolution layer moreover comprises of a nonlinear ReLU function meant to accumulate the image features with definite dimension.

Maxpooling layers are employed for down sampling. The two maxpooling layers incorporated here comprise a pool size of (2, 2). The fully connected layer is meant to concatenate the input by a matrix by ReLU activation function (first dense layer) as well as softmax activation function (second dense layer) and appends to a bias vector comprising the feature map. The 'Adam' optimizer is also utilized which offers the filter weight update at the declared learning rate. The deep network explains an extremely nonlinear function parameterized via a huge variables count. At every stride, the training algorithm attempts to bump these parameters; hence, the nonlinear function contains the preferred features from the training set images. The output from the final classification layer or the dense layer is the classified output. The proposed architecture is illustrated in Fig. 2, and the input features obtained are utilized to classify criminal and non-criminal images [19].

The labeled data is separated into two sets, such as the training set and test set. The CNN is trained in advance, with the labeled data; as well as consequently, the trained CNN can recognize the different classes. At the training stage, the training set is employed for CNN training. At the testing stage, the test set is exploited to compute the classification performance of the proposed system. Lastly, the identification accuracy is computed along with parameters like precision, recall along with f1-score. Training set as well as test set are composed of 70, 30% of total samples individually [20]. Finally, the crime detection results are attained.

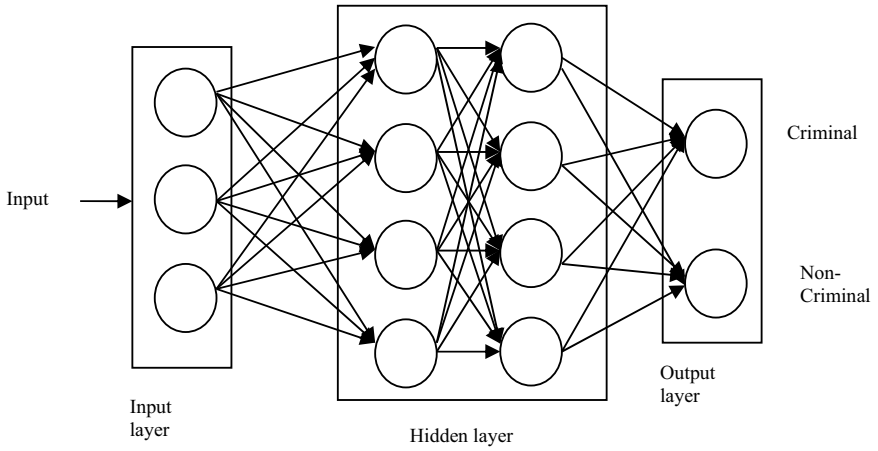


Fig. 2 Proposed architecture

4 Results and Discussion

The experimental results and the performance evaluation of proposed fingerprint identification algorithm by CNN are described in this section. The configuration made to estimate the proposed algorithm performance is—Intel(R) Celeron(R) CPU N3060 @ 1.60 GHz processor, 4 GB RAM, 64-bit windows operating system and Python environment with keras framework. Figure 3 describes the input fingerprint gray image. Figure 4 illustrates the filtered image obtained after performing filter operation with Gaussian filter. Then, the output image obtained from canny edge detection algorithm is illustrated in Fig. 5. And finally, the final classified output is illustrated in Fig. 6.

The performance of the proposed criminal identification system is determined quantitatively with aid of the parameter like the accuracy, precision, recall as well as f1-score and are equated in (2), (3), (4) and (5), correspondingly.

Fig. 3 Input fingerprint image

Fig. 4 Filtered image

Fig. 5 Edge detection image

Fig. 6 Final classified image

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{2}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$F1 - \text{Score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

where TP, FP, TN and FN are the count of true positives, false positives, true negatives and false negatives, correspondingly. Accuracy determines the global performance of the criminal classifier. Recall determines the reliability of the classifier at providing positive recognitions as well as the precision determines the system reliability at providing negative recognitions. F1-score is the harmonic mean of precision along with recall and offers a highly appropriate computation of analytical performance than the plain percentage of exact predictions in this criminal identification system.

Table 2 details the precision, recall, f1-score and accuracy values of the proposed criminal identification system for the two different classes, i.e., criminal and non-criminal. When considering with the criminal class, the precision value is identified as 100, recall as 96.42, f1-score as 98.67 and the accuracy is 99.33. Consequently, when considering with the non-criminal class, the precision value is identified as 97.78, recall as 100, f1-score as 100 and the accuracy is 99.33. Thus, the higher values for all performance measures confirm the better performance of the proposed system. Also, the performance measures such as the accuracy and loss of the proposed algorithm are taken and is illustrated pictorially in Fig. 7. Thus, from Fig. 7, it is obvious that as the accuracy increases the loss decreases, thus enhancing the performance of the proposed criminal identification system.

The performance of the proposed algorithm is compared with vgg16 algorithm and is pictorially depicted in Fig. 8. This states that the proposed criminal identification system performance improves drastically when compared to the existing vgg16 algorithm. All the proposed and the existing results are tested and implemented in Python keras environment.

As mentioned that the proposed algorithm is executed with three different open source datasets, such as the FVC2002, FVC2004 and NIST SD27, their results are shown in Table 3. From Table 3, the precision value for the proposed criminal identification system is 98.89, recall is 98.21, f1-score is 99.33, and accuracy is 99.33

Table 2 Proposed criminal identification system parameters

	Precision	Recall	F1-score	Accuracy
Criminal	100	96.42	98.67	99.33
Non-criminal	97.78	100	100	99.33

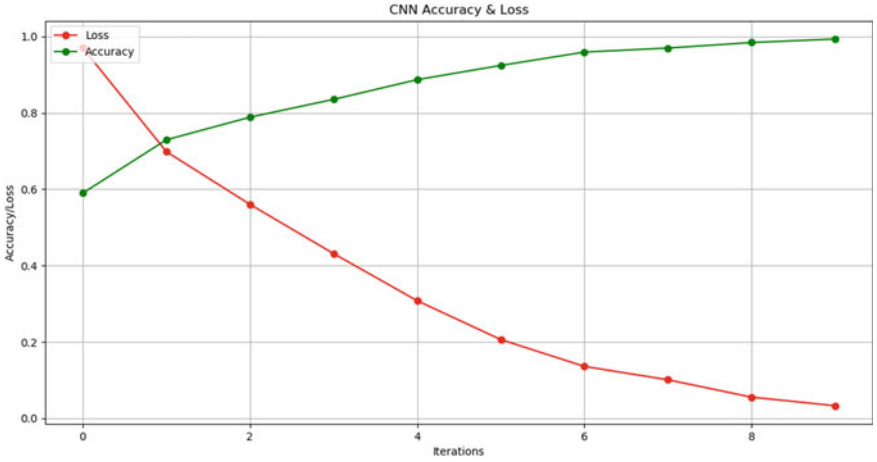


Fig. 7 Accuracy and loss of proposed method

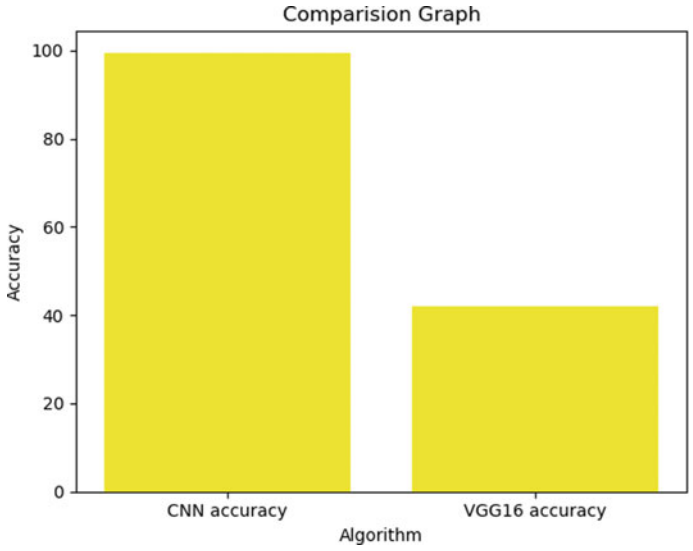


Fig. 8 Comparison graph

for FVC2002 dataset. Also, the precision value for the existing VGG16 algorithm is 40.87, recall is 40.25, f1-score is 41.86, and accuracy is 41.89 for FVC2002 dataset. Then, for FVC2004 dataset the precision value is 98.23, recall is 98.23, f1-score is 99.35, and accuracy is 99.34 for proposed criminal identification system, and alternatively the precision value is 40.89, recall is 40.29, f1-score is 41.87, and accuracy is 41.92 for the existing VGG16 algorithm. Lastly, for the NIST SD27 dataset, the precision value is 98.85, recall is 98.18, f1-score is 99.32, and accuracy is 99.31

Table 3 Comparison with existing algorithm

Parameters	Proposed system	Existing VGG16
FVC2002 dataset		
Precision	98.89	40.87
Recall	98.21	40.25
F1-score	99.33	41.86
Accuracy	99.33	41.89
FVC2004 dataset		
Precision	98.93	40.89
Recall	98.23	40.29
F1-score	99.35	41.87
Accuracy	99.34	41.92
NIST SD27 dataset		
Precision	98.85	40.85
Recall	98.18	40.21
F1-score	99.32	41.86
Accuracy	99.31	41.85

for proposed criminal identification system, and alternatively, the precision value is 40.85, recall is 40.21, f1-score is 41.86, and accuracy is 41.85 for the existing VGG16 algorithm. From the comparison, it is obvious that the proposed algorithm shows an improvement of more than 43% than the existing vgg16 algorithm on working with three different datasets. This proves that the proposed criminal identification system provides better results when compared with existing VGG16 algorithm for all the three datasets. Thus, the proposed automated criminal identification system can be utilized to identify the criminals accurately.

5 Conclusion

Here, the main work is the development of criminal identification system which is executed in two steps fundamentally as image processing and recognition. In this work, three datasets such as the FVC2002, FVC2004 and NIST SD27 are employed and the results are evaluated. Using the proposed classifier, an enhanced fingerprint recognition accuracy of 99% is attained. This is because initially the input criminal image is enhanced by using a Gaussian filtering algorithm. By this, the additional noise in the image is removed and thus the exact features are extracted by employing canny edge detection algorithm.

References

1. R. Pavithra, K.V. Suresh, Fingerprint image identification for crime detection, in *2019 International Conference on Communication and Signal Processing (ICCSP)*, (IEEE, 2019), pp. 0797–0800
2. K.N. Win, K. Li, J. Chen, P.F. Viger, K. Li, Fingerprint classification and identification algorithms for criminal investigation: a survey. Future Gener. Comput. Syst. **110**, 758–771 (2020)
3. B. Roli, S. Priti, B. Punam, Minutiae extraction from fingerprint images—a review. Int. J. Comput. Sci. **8**(5), 1694–0814 (2011)
4. X. Huang, P. Qian, M. Liu, Latent fingerprint image enhancement based on progressive generative adversarial network, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, (2020), pp. 800–801
5. C. Megha, M.K. Shukla, K.K. Ravulakollu, Bagging-and boosting-based latent fingerprint image classification and segmentation, in *International Conference on Innovative Computing and Communications*, (Springer, Singapore, 2021), pp. 189–201
6. S. Paul, V. Dalvi, A. Chavare, A. Chitari, Detecting and improving distorted fingerprints using rectification techniques. Int. J. Innovative Res. Sci. Eng. Technol. **6**(3), 4019–4035 (2017)
7. P. Chhajer, O. Kadam, D. Agarwal, N. Pandey, S. Das, Criminal identification using partial fingerprint reconstruction method. Int. J. Innovative Res. Sci. Eng. Technol. **6**(5), 8891–8899 (2017)
8. A. Tukur, Fingerprint recognition and matching using Matlab. Int. J. Eng. Sci. (IJES) **4**, 01–06 (2015)
9. A. Tamrakar, N. Gupta, Low resolution fingerprint image verification using CNN filter and LSTM classifier. Int. J. Recent Technol. Eng. (IJRTE) **8**(5), 3546–3549 (2020)
10. H. Aghalte, M. Shinde, N. Hadap, Crime investigation system using biometrics. Int. Res. J. Eng. Technol. (IRJET) **03**(04), 642–645 (2016)
11. A.B. Kanbar, Fingerprint identification for forensic crime scene investigation. Int. J. Comput. Sci. Mob. Comput. **5**(8), 60–65 (2016)
12. T.S. Deepthi Murthy, Fingerprint image recognition for crime detection. Turk. J. Comput. Math. Educ. (TURCOMAT) **12**(12), 2230–2237 (2021)
13. I. Goel, N.B. Puhan, B. Mandal, Deep convolutional neural network for double-identity fingerprint detection. IEEE Sens. Lett. **4**(5), 1–4 (2020)
14. T.S. Deepthi Murthy, Fingerprint Image recognition for crime detection. Turk. J. Comput. Math. Educ. (TURCOMAT) **2**(12), 2230–2237 (2021)
15. U.U. Deshpande, V.S. Malemath, S.M. Patil, S.V. Chaugule, End-to-end automated latent fingerprint identification with improved DCNN-FFT enhancement. Front. Robot. AI **7**, 166 (2020)
16. D. Valdes-Ramirez, M.A. Medina-Pérez, R. Monroy, An ensemble of fingerprint matching algorithms based on cylinder codes and mtriplets for latent fingerprint identification. Pattern Anal. Appl. **24**(2), 433–444 (2021)
17. K. Gong, J. Guan, C.-C. Liu, J. Qi, PET image denoising using a deep neural network through fine tuning. IEEE Trans. Radiat. Plasma Med. Sci. **3**(2), 153–161 (2018)
18. R. Anusha, K.R. Deepak Raj, N.V. Deepak, E. Lakshmi, S. Manoj, Forgery detection by biometric images using SVM classifier. Int. J. Eng. Res. Technol. (IJERT) **6**(13), 1–5 (2018)
19. D. El Hamdi, I. Elouedi, A. Fathallah, M.K. Nguyuen, A. Hamouda, Combining fingerprints and their radon transform as input to deep learning for a fingerprint classification task, in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, (IEEE, 2018), pp. 1448–1453
20. C. Fan, H. Gong, M. Cheng, B. Ye, L. Deng, Q. Yang, D. Liu, Identify the device fingerprint of OFDM-PONs with a noise-model-assisted CNN for enhancing security. IEEE Photonics J. **13**(4), 1–4 (2021)

Semantic Segmentation of Retinal Vasculature Using Light Patch-Based Dilated CNN

Nisha R. Wankhade, K. K. Bhoyar, and Ashutosh Bagde

Abstract Automatic vessel segmentation is in corking demand as it requires expertise and experience of ophthalmologist. Presently CNN is more popular in the field of computer vision, due to its higher accuracy, but it requires large data. On the other side medical imaging has limited high resolution samples in the dataset. The proposed model utilizes patch-based input, to avoid the issue of down sampling and overfitting. As retinal vasculature varies in shape and width throughout its length, the proposed model adopted the different size kernel for dilated convolution to extract different size features without using pooling layer. This results into deduction in computational cost, nearly five times than the conventional neural network model as well it projects a notable improvement in the segmentation process.

Keywords Vessel segmentation · CNN · Dilated convolution · Medical imaging · Retinopathy · Semantic segmentation

1 Introduction

Fundus image prominently show the structure of receptive field of eye, which consists optic disk, macula, fovea, and vasculature. Vasculature is a tree like structure of artery, vein, and its branches [1]. This vessel structure of retinal image contains ample number of geometric features like vessel width, length, curvature, angle, etc. Ophthalmologist observes retinal structure carefully and critically to analyses these features, which reflects the sign of retinopathy [2, 3]. Vessel features also used to identify vessel related diseases such as diabetic and diabetic Maculopathy (MD) [1].

N. R. Wankhade (✉) · K. K. Bhoyar
Department of Information Technology, Yeshwantrao Chavan College of Engineering, Nagpur, India
e-mail: nisha.wankhade@gmail.com

A. Bagde
Jawaharlal Nehru Medical College, Belagavi, India

Datta Meghe Institute of Medical Sciences, Wardha, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_26

Accurate identification and timely treatment of retinal diseases required to avoid blindness [4]. Lower contrast, uneven illumination, varying size and shape of vessels, and presence of different diseases makes the fundus image complex [5], due to these complexities different experts generate different segmentation output [6]. All of these realizes the need of automated system for vessel segmentation.

Deep learning, especially convolutional neural networks (CNNs), has gained much attention for image analysis [7, 8]. Deep learning methods, automatically learn features by using massive data with less human inference. They have better generalization ability and recognition capability; hence they can learn different level patterns automatically and can be used by wide applications.

In this paper, we have devised a Light patch-based dilated convolution model for vasculature segmentation. The model accepts the patches as an input of high resolutions images. It extracted the features using dilated convolution layer instead of traditional CNN. We evaluated the proposed model on the standard DRIVE [9] retinal image datasets. The proposed automatic retinal vascular segmentation framework has achieved good results, and dramatically reduces the computational cost of model.

2 Dataset

The Diabetic Retinopathy screening program was organized in Netherland. This program launches the DRIVE dataset. 400 subjects between the age of 25–90 were evaluated in this program out of which 40 were selected. From these 40 subjects 7 shows the diabetic retinopathy sign and 32 were normal subjects. (http://www.isi.uu.nl/Research/Databases/DRIVE/).

3 Methodology

3.1 Patched Input

Popularity of the Deep Convolutional Neural network is very high in the field of imaging [10]. Deep network is a layered architecture that is being learned through multiple layers of abstraction. The important point is, the general-purpose learning algorithm is used in layer by layer feature learning [9]. The complex functions can be learned by composing modules of previous to next layers. These modules represent the input data. An image consists of array of pixels and processing this array of pixels in the form of features requires, a lot of computations. Applying high resolution images to the CNN results comprises high computational cost. Another limitation of applying high resolution images to the deep networks is, to require down sampling of images by large factor, which eliminates some of the credential features such as

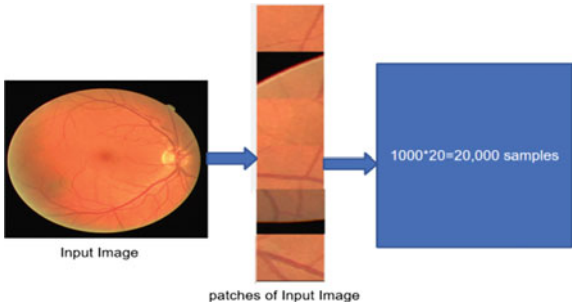


Fig. 1 Process of patches creation

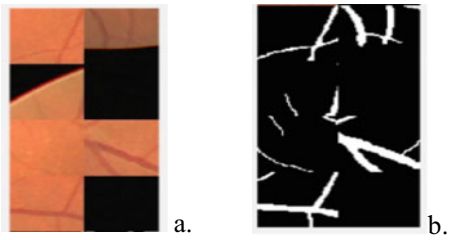


Fig. 2 **a** Sample patches of input image, and **b** respective ground truths

fine vessels [11]. On the other hand, deep networks may learn from, one of these credential features out of the multiple features resulting in data inefficiency [12].

The proposed method handles both issues by training the deep network using the image patches. It also provides the solution for small dataset of retinal images as a single image, which is divided into number of small patches and increases the training dataset. The procedure is shown in Fig. 1. And Fig. 2 predicted the patches and its respective ground truths.

3.2 Dilated Convolution

The aim is to categorize each pixel into one of the given sets. Most of the researchers have used Convolutional neural networks for semantic segmentation. The Convolutional network has a series of convolution and pooling layers. These convolutional and pooling layers together, extracts most of the primitive features such as local texture and lines of an image. In pooling layer, local features are collected and transformed into scale invariant form. The pooling layer reduces the size and therefore the cost. Such CNN is highly dependent on the spatial pooling, but unbounded pooling results into the information loss [13] as there is no specific control on pooling in convolutional networks. Hence the proposed work has adopted the dilated convolutional layers for dense (pixel wise) prediction, addressing two important issues. First, it achieved local, pixel level accuracy such as very fine vessels of retinal vasculature. Secondly, it incorporates the perception of global context, as vasculature is a single

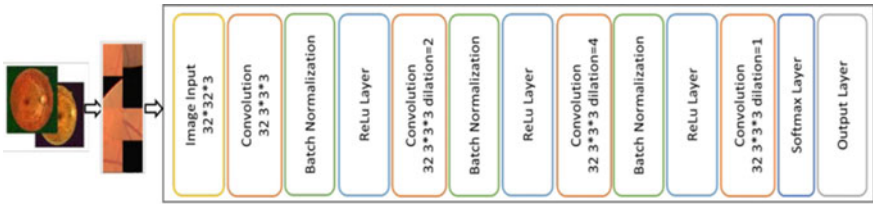


Fig. 3 Structure of semantic segmentation with dilated convolutional layers

entity spread throughout the image. The Dilated network also handles the multi-scale feature problem efficiently by utilizing depth wise dilated convolutional layers without increasing the number of parameters. The Dilated convolution collects the features, by enlarging the receptive field without changing resolution and maintaining the same kernel parameters [13] as shown in the Eq. (1).

$$D_{\mathrm{N}} = \sum_{n=1}^{N} X[i + d * n] * P_n \tag{1}$$

where D_{N} and $X[i]$ is output and input signal, d is dilation rate, P_n is nth parameter of filter, N is filter size. When value of d set is 1, it becomes standard convolution kernel. By increasing the value of d, the receptive field of convolution kernel is increased. Larger receptive field can be achieved by assembling multiple layers [13]. The Structure of proposed network with the dilated convolution is shown in Fig. 3.

3.3 Light Patch-Based Network

The structure of semantic segmentation consists of a patch input layer, convolution layer, batch normalization layer, activation layer, softmax layer and output layer. Most of the structure is covered by the alternative layers of convolution, and activation layers.

This light patch-based dilated CNN model were implemented from scratch. The input to the structure is 32 * 32 * 3 pixel size patch of input image. 1000 such patches are extracted from single input image. The DRIVE dataset has 20 such images, so the total number of patches is 1000 * 20 = 20,000. All 1000 patches are extract randomly. The striking thing of random patch extraction is, the overlapping patches extracted, and more information were taken, that leads in improvement for the accurate dense prediction. Whereas single patch may contain less information results in poor prediction. Next, we use zero center normalization process to ensure all the patches from input images that has same luminance and contrast.

Each convolution layer is connected with receptive field of previous layer without pooling layer. The dilated convolution expands the receptive field and learns the

essential features without losing local information with a smaller number of computations. In retinal vasculature structure the vessels are bright, thick at root level and gradually get fainted and thin at rare side hence become difficult to visualize. So, we use small kernel to extract the feature of local, fainted, and fine vessels and larger kernel learns the more global information to extract the complete connected vasculature. That's why our model consists of four convolution layers with 1, 2, and 4 dilation rates. We have used depth wise (layered) convolution with increasing dilation factor, so that the receptive field will grow linearly, which allows us to extract different sized features with same 3 * 3 kernel size. Last convolution layer with the dilation rate 1 holds the feature of current scale.

Resulting features of convolution layer were normalized with batch normalization layer. This layer prevents network from overfitting problem, also helps to increase the leaning rate. ReLu [14] activation function is used for nonlinear transform. Probability distribution of each pixel was found out at softmax layer. Finally, classification layer was added to classify each pixel with cross entropy loss function.

The Light Patch-based Semantic Segmentation model were trained on Intel i7 processor with 24 GB RAM with GeForce GTX GPU card. Training is done with SGDM optimizer with initial learning rate of 0.001 for 50 epochs.

4 Result and Discussion

The proposed light patch-based model was evaluated on the following measures.

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

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

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

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

We achieved Accuracy value 0.92, Precision value 0.90, recall value 0.92, and F1_Score = 0.90. From Table 1 it is proved that by adding dilation factor in convolution layer the accuracy of segmentation is improved nearly 5% whereas Table 2 shows the parameters required by the network with and without dilation convolution. Total learnable parameter of dilated convolutional network is reduced from 137,170 to 22,432 that means only 16.35% learnable parameters need to be trained as compared to conventional convolution neural network, this reduces the computational cost of network more than 83% for retinal vasculature segmentation. Table 3. shows the

Table 1 Confusion matrix with and without dilation convolution

	With dilation		Without dilation	
	Vessels	Background	Vessels	Background
Vessels	0.90	0.097	0.85	0.14
Background	0.075	0.92	0.077	0.92

comparison of proposed dilated model with some of the state of the art models on the basis of f1_score and Accuracy.

Figure 4 shows the sample segmented output images of with and without dilation convolution network. From Fig. 4 it is observed that there is no need of post processing for removal of small segmentation error, e.g., boundary artifact, small dot like structure. The dilated convolutional network itself improve its segmentation.

Table 2 Learnable parameters of network

Name	Parameters of without dilated CNN	Total	Parameters of dilated CNN	Total
Input	32 * 32 * 3	3072	32 * 32 * 3	3072
Conv layer 1	3 * 3 * 3 * 32	864	3 * 3 * 3 * 32	864
Conv layer 2 ($d = 2$)	7 * 7 * 32 * 32	50,176	3 * 3 * 32 * 32	9216
Conv layer 3 ($d = 4$)	9 * 9 * 32 * 32	82,994	3 * 3 * 32 * 32	9216
Output layer	1 * 1 * 32 * 2	64	1 * 1 * 32 * 2	64
Total		137,170		22,432

Table 3 Comparison with state of the art algorithm

Sr. No.	Model	F1 score	Accuracy	Paper	Year
1	RV-GAN	0.8690	0.9790	RV-GAN: segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial net [15]	2021
2	Study group learning	0.8316		Study group learning: improving retinal vessel segmentation trained with noisy labels [16]	2021
3	SA-UNet	0.8263	0.9698	SA-UNet: spatial attention UNet for retinal vessel segmentation [17]	2020
4	Dilated CNN	0.90	0.92	Our proposed model	–

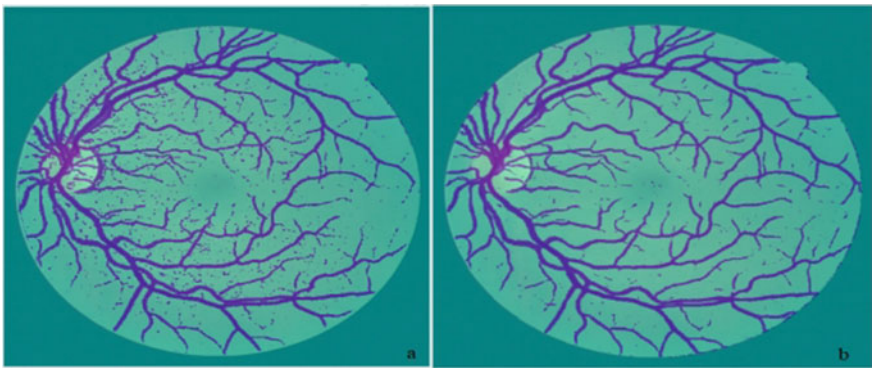


Fig. 4 Output image of **a** without dilation network, **b** with dilation network

5 Conclusion

The present work proposes the light patch-based dilated CNN for semantic segmentation that handles the issue of down sampling of high resolution images and overfitting of network due to small dataset, using patch-based input. Through the experimentation results it is proved that the model reduces the computational cost nearly 83% and handles the issue of high computational cost of the deep networks by adopting the dilation convolution. Final output shows that dilation convolutional network itself improves the quality of segmentation. In future work we will try to improve the performance of network and make it generalized.

References

1. S.W. Franklin, S.E. Rajan, Computerized screening of diabetic retinopathy employing blood vessel segmentation in retinal images. Biocybern. Biomed. Eng. **34**(2), 117–124 (2014)
2. Early treatment diabetic retinopathy study research group, "Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12." Ophthalmology **98**, 823–833 (1991)
3. J.W. Yau, S.L. Rogers, R. Kawasaki, E.L. Lamoureux, J.W. Kowalski, T. Bek, S.J. Chen, J.M. Dekker, A. Fletcher, J. Grauslund, S. Haffner, Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care **35**(3), 556–564 (2012)
4. M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, M.D. Abramoff, Comparative study of retinal vessel segmentation methods on a new publicly available database. Proc. SPIE **5370**, 648–656 (2004)
5. M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, S.A. Barman, Blood vessel segmentation methodologies in retinal images—a survey. Comput. Methods Programs Biomed. (2012). https://doi.org/10.1016/j.cmpb.2012.03.009
6. P. Dai, H. Luo, H. Sheng, Y. Zhao, L. Li, J. Wu, Y. Zhao, K. Suzuki, A new approach to segment both main and peripheral retinal vessels based on gray-voting and Gaussian mixture model. PLoS ONE **10**(6), e0127748 (2015)

7. G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
8. S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M.P. Reyes, M.L. Shyu, S.C. Chen, S.S. Iyengar, A survey on deep learning: algorithms, techniques, and applications. ACM Comput. Surv. **51**(5), 1–36 (2018)
9. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature **521**, 436 (2015)
10. K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, in *ICCV* (2015)
11. Y. Jiang, N. Tan, Retinal vessels segmentation based on dilated multi-scale convolutional neural network. IEEE Access. https://doi.org/10.1109/ACCESS.2019
12. L. Hou, D. Samaras, Patch-based convolutional neural network for whole slide tissue image classification, in *IEEE Conference on Computer Vision and Pattern Recognition* (2016). https://doi.org/10.1109/CVPR.2016.266
13. F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions. Published as a conference paper at ICLR 2016. arXiv:1511.07122v3 [cs.CV]. 30 Apr 2016
14. G. Hinton, V. Nair, Rectified linear units improve restricted Boltzmann machines, in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (Haifa, Israel, 2010)
15. S.A. Kamran, K.F. Hossain, A. Tavakkoli, S.L. Zuckerbrod, K.M. Sanders, S.A. Baker, RV-GAN: segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network. arXiv:2101.00535v2 [eess.IV]. 14 May 2021
16. Y. Zhou, H. Yu, H. Shi, Study group learning: improving retinal vessel segmentation trained with noisy labels, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. MICCAI 2021. Lecture Notes in Computer Science, vol. 12901. Springer, Cham. (2021). https://doi.org/10.1007/978-3-030-87193-2_6
17. C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, C. Fan, SA-UNet: Spatial Attention U-Net for Retinal Vessel Segmentation, Image and Video Processing (eess.IV); Computer Vision and Pattern Recognition (cs.CV), ICPR (2020). https://doi.org/10.48550/arXiv.2004.03696

A Novel Study on Tools and Frameworks for Mitigating Bias in Multimodal Datasets

Venkata Naresh Mandhala, Debnath Bhattacharyya, and Divya Midhunchakkaravarthy

Abstract Bias in the data always affects the performance of the system. Our need of finding the accurate results in a system using any Machine Learning algorithms will be degraded when bias is present in the data that we choose to analyze. Mitigating bias in Artificial Intelligence, mainly deals with the dataset development from an existing datasets which reduces the bias by comparing and validating different samples of training data and algorithm development with an established policies and practices to enhance the performance of the algorithms at later stages of the development of the framework. To detect the bias from the given datasets, and mitigate the bias, various tools and technologies are proposed and in consequences of this work shows that not only bias, but also helps to decrease without forfeiting model performance rate and improves the performance of the model. This paper focuses on exhibiting various tools and frameworks associated to mitigate the bias present in various databases choose to different applications intended to verify and validate.

Keywords Artificial intelligence · Bias · Frameworks · Machine learning · Tools

1 Introduction

The common bias problems that occur in most of the AI and Machine Learning models are gender-bias, racial-bias and few other are addressed by various tool and frameworks. To mitigate the bias that are occurred in the datasets are been trying to

V. N. Mandhala (✉) · D. Midhunchakkaravarthy
Department of Computer Science and Multimedia, Lincoln University College, Kuala Lumpur, Malaysia
e-mail: vnmandhala@lincoln.edu.my

D. Midhunchakkaravarthy
e-mail: divya@lincoln.edu.my

D. Bhattacharyya
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, A.P, India
e-mail: debnathb@kluniversity.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_27

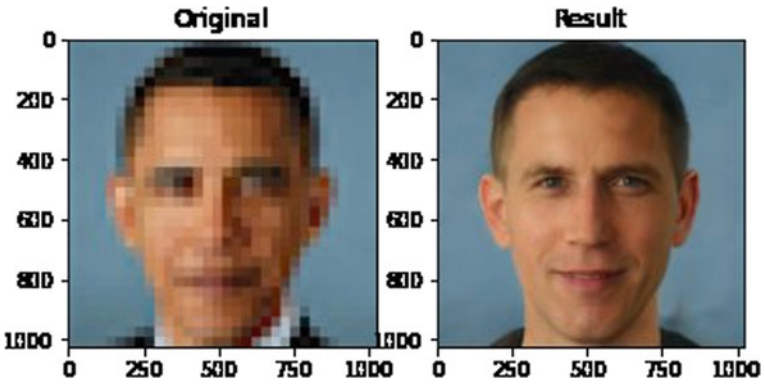


Fig. 1 Blurred image as pixilated input and unpredictable output

fix by creating the tools. In Artificial Intelligence Bias is the potential field for the researchers. The bias present in various datasets occur due to the human data when given as an input to the machine learning algorithms where the concept of implicit bias comes into the picture. The main outcome of this study mainly focuses on how to eliminate the human bias and to address this challenging task for improving the performance of the ensemble learning methods.

Most of the time, when the human generated data is given to the machine learning models it generates the high quality realistic output when a pixelated input was provided [1]. However when the input as a blurred image was given to the model, it produced an errorless output which is not an expected to be for the input provided as shown in Fig. 1. And so far no proper solution has been identified to overcome this problem. This had not provided the basis for the improvement of the study though the above said statement was made to bring the attention to the researchers as the issue by bringing the age-old excuse for the dataset chosen to claim it.

Though, at later point of time, Luca Massaron a data scientist also said that, Yann LeCun made a correct point, and concluding that the data plays an important role to be a sensible [2].

If the Algorithms that were proposed were biased it would automatically leads to a problem and also a challenging task for the researchers to mitigate them. If we neglect and made the algorithm hidden which could lead to more adverse situation [3].

The data collection for the dataset generated by humans plays a crucial role as there is no one-stop solution to mitigate the bias in the domain expertise [4].

2 Need of Identifying Bias in Datasets

Collection of data from various sources is an important aspect with respect to the dataset creation or preparation. When large volumes of data are collected for analysis from different sources there is a possibility of bias presence in the data. Morstatter

[5] proposed that bias can be formed in the social media themselves and mainly age-bias is one of the major aspect with the data. API bias is the bias which is used by the researchers to collect the data from API. Data is collected from the similar topic from Twitter streaming APIs. Second is Population bias, which deals with the different categories of the human based in the usage of the social media like young users and elder age group user. This leads to identifying the new types of biases on the data when chosen based on the problem statement.

FaiRecSys is an algorithm which can be used to mitigate the bias in the algorithms by performing the post-processing. As we are using a large volumes of data for recommendations there is a chance of bias existing in the recommendation systems and which will enable the users to trigger in return which is based on the algorithms that are used for recommendations. Edizel et al. [6] mainly focused on finding the algorithmic bias in the recommender systems on features that are sensible and biased results.

Bias sometime becomes goon while taking decisions. Bias is a now a days' a common word that we hear while talking about the data collection and while performing some analysis on large amounts of data. Bountouridis et al. [7] surveyed on 12 datasets and established a framework which helps to identify the bias that is present in the datasets. During the compilation of the datasets there may be a possibility of data bias which is a major concern while performing the analysis.

Natural Language Processing (NLP) Technology having potential adverse impact fails to work on specific datasets due to the bias present in it. Bender and Friedman [8] proposes the short term and long term helps to develop the dataset representations. The short term concentrate on the how the data will work whereas the long term basically addresses the issues of the bias in the datasets. Datasets like Hate Speech Twitter Annotations, Twitter Hate Speech Short Form are considered for training and testing purpose to mitigate the bias in those datasets. This also helps to identify the vulnerable to emergent bias and define the mechanisms to train and test the data for better representation.

3 Tools and Frameworks for Mitigating Bias in Datasets

A wide variety of tools and frameworks are developed to mitigate the bias from multimodal datasets. They are: REVISE, FairML, IBM AI Fairness 360, Accenture's "Teach and Test" Methodology, Google's What-If Tool and Microsoft's Fairlearn.

The brief descriptions of all the tools are given below.

3.1 REVISE (REvealing VIsual BiaSEs)

This is an open-source tool which mitigates the bias had been developed by Princeton University automatically uncover the potential bias in image datasets. To identify the

potential bias the REVISE uses the statistical methods across the gender, object, and geography-based dimensions. This works by filtering the images and also balancing the images. The existing image annotations are used and object counts, images and the co-occurrence of people and object are measured to study the bias [9]. An Image dataset by name ImageNet has been used by Princeton and Stanford University to address the bias in Artificial Intelligence which consists of 14 million images generally used for Computer Vision Model development which identifies the non-visual and offensive categories among the database. It revolutionized the area of computer vision and which lead to the development of more such type of toll to mitigate the bias in Artificial Intelligence and Machine Learning Models [10].

3.2 FairML

To identify the bias in the Machine Learning Models the FairML is mostly used developed by Julius Adebayo. The working mechanism of this framework is features that are important and having relative significance that are used in the machine learning models mainly to detect the bias in both the linear and non-linear models. As shown in Fig. 2, this framework mainly focuses on such an attributes like race, religion, gender, and others to focus out the type of the data where it is biased. The relative significance of the model's input is quantified by auditing the models which are predictive and it helps to improve the fairness of the model mainly for assessing them.

The unintentional discrimination is one of the potential which was not addressed completely as it is a potential systematic flaw despite of its regular productivity and efficiency gains. The bias on the sexual orientation, gender, race, religion or may be other characteristics which hold the discrimination. It determines the input which has relative significance projected to the black-box predictive model to gain the model's fairness [11].

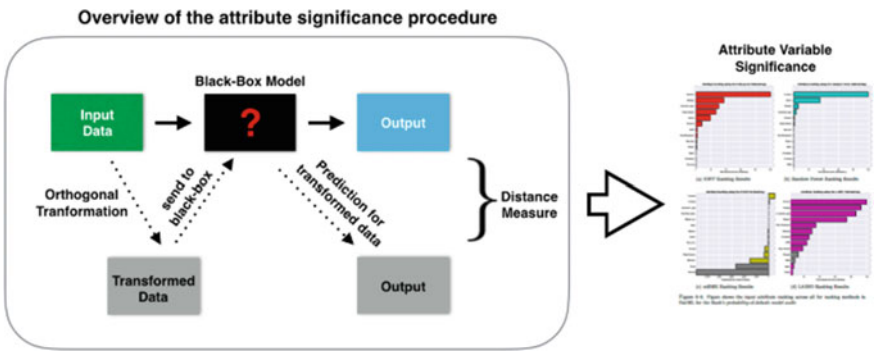


Fig. 2 Architecture and workflow of FairML

3.3 IBM AI Fairness 360

This toolkit is an open-source, developed by IBM Research and later given to the Linux Foundation AI and Data. This toolkit is used to mitigate bias by considering the large volumes of datasets taking 10 bias algorithms and 70 fairness metrics [12]. It works on optimized preprocessing and re-weighting. To identify the fairness and for comparison study with existing models a researcher can use this toolkit. It also examines and produces reports on mitigation discrimination in the wide variety of Machine Learning models. To address the bias in Artificial Intelligence systems there are numerous mitigating algorithms are available to use.

3.4 Accenture's "Teach and Test" Methodology

This Framework is used to overcome the bias or risks of any form in the datasets. This was developed by Accenture to endure the Artificial Intelligence for producing the right decision. It uses the "teach and test" methodology in two phases. The primary focus is to make choice of the algorithms, models that are used to train the ensemble learning for the data that we pick. Then it applies the evaluation and scoring on the specified models of AI. The performance of this methodology on customer recommendations is 85% accurate on the financial services.

As the teach and test are the primary method used in this framework, the "Teach" mainly focuses on the choice of algorithms, models and data used to train the ensemble learning's.

Whereas the "Test" phase, compares the output of the Artificial Intelligence to determine the outcome with the key performance indicators and also assess for the decision on the selected systems are determined properly or not. This phase uses the cloud enabled tools and also novel techniques to achieve best performance by monitoring the system in a continuous basis [13, 14]. This methodology is used on sentimental analysis for improving the performance with a right training data and with a proper solution for faster analysis and better results.

3.5 Google's What-If Tool

This tool is used to analyze the datasets to mitigate the bias, it was developed as an interactive open-source tool by Google. One of the important features that this tolls has besides lot of open-source tools like TensorBoard is it allows the user to visually investigate the ML models, and additionally it also analyses the data that are trained on TensorFlow models. It builds a rich visualization while modeling the different scenarios. The bias was detected manually by editing the samples from the dataset. It detects the features which effects the changes that occur while using the associate

models. This algorithm improves the fairness by analyzing the data which are not identified by the previous methods and helps us to discover the patterns that mitigate the bias in the datasets.

3.6 Microsoft's Fairlearn

The Microsoft had developed an open-source toolkit by name Microsoft's Fairlearn which is used to mitigate the bias in the datasets. To detect the fairness and also to correct the fairness of the Artificial Intelligence systems to contains two components. One an interactive visualization dashboard and the other is a bias mitigation algorithm. It helps to improve the fairness and also improves the performance of the model drastically. The fairness prioritizing is a challenging task this toll can be used to mitigate the biases that are present in the datasets and removes the fairness-related issues at its best [15].

4 Conclusion

This paper provides various tools and frameworks associated to mitigate the bias present in different datasets which affects the performance of the system. Mitigating bias in Artificial Intelligence, deals with the dataset development from an existing datasets which reduces the bias by comparing and validating different samples of training data and algorithm development with an established policies and practices to enhance the performance of the algorithms at later stages of the development of the framework. Bias detection from the given datasets, and to mitigate the bias can be done using various tools and technologies that are proposed and in consequences of this work shows that not only bias, but also helps to decrease without forfeiting model performance rate and improves the performance of the model. This paper focused on exhibiting various tools and frameworks associated to mitigate the bias present in various databases choose to different applications intended to verify and validate.

References

1. Analytics India Magazine. https://analyticsindiamag.com/yann-lecun-machine-learning-bias-debate. Last Accessed 15 Oct 2021
2. Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, F. Huang, A tutorial on energy-based learning, in *Predicting Structured Data 1.0* (2006)
3. Analytics India Magazine. https://analyticsindiamag.com/5-tools-frameworks-that-can-clear-bias-from-various-datasets. Last Accessed 11 Oct 2021

4. Towards Data Science. https://towardsdatascience.com/5-types-of-bias-how-to-eliminate-them-in-your-machine-learning-project-75959af9d3a0. Last Accessed 05 Oct 2021
5. F. Morstatter, Detecting and mitigating bias in social media, in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2016 Aug 18* (IEEE, San Francisco, CA, USA, 2016), pp. 1347–1348
6. B. Edizel, F. Bonchi, S. Hajian, A. Panisson, T. Tassa, FaiRecSys: mitigating algorithmic bias in recommender systems. Int. J. Data Sci. Anal. **9**(2), 197–213 (2020)
7. D. Bountouridis, M. Makhortykh, E. Sullivan, J. Harambam, N. Tintarev, C. Hauff, Annotating credibility: identifying and mitigating bias in credibility datasets, in *ROME 2019-Workshop on Reducing Online Misinformation Exposure* (2019)
8. E.M. Bender, B. Friedman, Data statements for natural language processing: toward mitigating system bias and enabling better science. Trans. Assoc. Comput. Linguist. **6**, 587–604 (2018)
9. Brookings. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms. Last Accessed 06 Sep 2021
10. A. Wang, A. Narayanan, O. Russakovsky, REVISE: a tool for measuring and mitigating bias in visual datasets, in *European Conference on Computer Vision* (Springer, Cham, 2020), pp. 733–751
11. TechTarget SearchEnterpriseAI. https://searchenterpriseai.techtarget.com/news/252493354/Amazon-SageMaker-Clarify-aims-to-mitigate-bias-in-machine-learning. Last Accessed 30 Sep 2021
12. P. Choudhury, E. Starr, R. Agarwal, Machine learning and human capital complementarities: experimental evidence on bias mitigation. Strateg. Manag. J. **41**(8), 1381–1411 (2020)
13. A. Moro, E. Boelman, G. Joanny, J.L. Garcia, A bibliometric-based technique to identify emerging photovoltaic technologies in a comparative assessment with expert review. Renewable Energy **123**, 407–416 (2018)
14. K.T. Hufthammer, T.H. Aasheim, S. Ånneland, H. Brynjulfsen, M. Slavkovik, Bias mitigation with AIF360: a comparative study, in *Norsk IKT-konferanse for forskning og utdanning* (No. 1), (2020)
15. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics **32**(19), 3047–3048 (2016)

Machine Learning Framework for Stagewise Classification of Alzheimer's Disease

N. Sai Srithaja, N. Sandhya, and A. Brahmananda Reddy

Abstract Alois Alzheimer, a German physician, diagnosed the first case of Alzheimer's disease (AD) in the early twentieth century, so it is named after him. The patient experienced loss of memory, depression, and psychological changes. In the autopsy, Alzheimer noticed that the nerve cells in and around her brain were weakening. It is caused medically by abnormal protein build up in and around the brain cells. Amyloid and tau are two proteins that are involved. Few causes of this disease are: age, family history, head injury, etc. AD is one of the different varieties of dementia. It usually affects persons over the age of 60 but now it affects adults in their middle years as well. As a result, experts are focusing on this disease and using various study strategies to try and control it. Initially, they focused on discovering a treatment for this illness, but as time went on, their attention shifted to disease analysis and prediction. For efficient treatment and recovery of AD, precise early-stage identification is important. As a result, accurate AD diagnosis is a significant research challenge. The dataset used in this study consists of 416 medical records. To provide accurate results, a machine learning (ML) algorithm is used to construct the model. In this study, we provide users a stage by stage prediction of Alzheimer's disease. The stages included are: non-demented, mild demented, moderate demented, and demented. It is a challenging disease for which there is no cure; instead, we can only delay the progression of the disease.

Keywords Alzheimer's disease (AD) · Feature engineering · Decision tree (DT) · Pruning · Tuning

N. S. Srithaja (✉) · N. Sandhya · A. B. Reddy
VNR VJIET, Department of Computer Science and Engineering, Hyderabad, India
e-mail: srithaja.nibhanupudi@gmail.com

N. Sandhya
e-mail: sandhya_n@vnrvjiet.in

A. B. Reddy
e-mail: brahmanandareddy_a@vnrvjiet.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_28

1 Introduction

Alzheimer's disease is a degenerative brain disease that impairs memory, thinking skills, and the ability to do even the basic tasks. The majority of patients with the condition develop symptoms in their mid-60 s (those with the late-onset variety). Alzheimer's disease that develops between the ages of 30 and 60 is highly unusual. Around the age of 65, one out of every eight persons will have this condition [4]. The common cause of dementia in elder people is AD. According to scientists, AD is caused by a mix of genes, behavioral, and environmental variables that affect the brain over time in most people. AD is caused by a series of genetic flaws that nearly guarantee that an individual develops the disease in less than 1% of cases. As a result of these infrequent incidences, the disease often manifests in middle life. Damage to the brain's memory-controlling region usually begins years before the first symptoms appear, and the death of neurons spreads to other sections of the brain in a predictable pattern. By the end of the disease, the brain will have shrunk significantly.

Amyloid and tau are the two primary proteins involved in the development of AD. The brain contains a component of a larger protein called beta-amyloid. When these bits cluster together, they appear to have a negative effect on neurons and impede cell-to-cell communication. These clusters come together to form plaques, which contain more cellular debris. Within a neuron's internal support and transport system, tau proteins are involved in the transportation of nutrients and other essential substances. In Alzheimer's disease, tau proteins change shape and clump together to create neurofibrillary tangles. Cells are harmed by the tangles, which disrupt the transport system. Memory loss is the first and foremost of Alzheimer's disease (AD). One of the earliest signs is the capacity to recall recent events or discussions. As the disease progresses, memory issues worsen and new symptoms appear. Apart from memory loss, there are additional signs and symptoms. Concentration and reasoning are hampered by AD, especially when dealing with abstract concepts such as numbers. Multitasking is extremely challenging, as is managing finances, making timely payments, and balancing cheque books. A person suffering from Alzheimer's disease may gradually lose their capacity to deal with numbers.

Alzheimer's disease affects one's ability to make judgments and decisions in everyday situations. Regular activities involving sequential actions, such as preparing and cooking a meal or playing a favorite game, become a problem as the condition progresses. AD has no cure. However, several lifestyle risk factors for Alzheimer's disease can be modified. Modifications in diet, activity, and lifestyle, which have been found to lower the risk of cardiovascular disease, may also decrease the risk of having AD and other dementia-causing illnesses. There is no treatment for AD and the only thing that can be done is to slow down the growth of the condition. Researchers employed several machine learning and deep learning techniques to achieve accurate results. The proposed system deals with the medical records of Alzheimer's patients, which comprise psychological exams and clinical data.

2 Literature Survey

There are numerous ways to determine whether or not a person is suffering from dementia. Researchers used several forms of data to make predictions. They are: images, speech, sensor, psychological tests, or focusing on a single attribute. Researchers collected various types of data from various sources for predicting AD. They are in the form of surveys, clinical data, or conducting test and questionnaires. Few concentrated on only single attribute for predicting dementia like brain volume [13–15]. The below table discusses about the detailed analysis of the related literature work carried out in our study (Table 1).

There are few noticeable factors considered from all of the above papers. They are:

(1) Every researcher tends to build model that predict early diagnosis in order to delay the progression of the disease, since there is no cure for Alzheimer's.
(2) The most commonly used algorithms to build the model are: support vector machine (SVM), decision tree, random forest, and logistic regression.
(3) Many deep learning techniques like CNN, RNN, SCNN, etc., are commonly used for image processing.

3 Proposed Work

The proposed model gathers psychological data from subjects in order to predict which stage of dementia they are in. The stages included are: non-demented (encoded as 0), mild demented (1), moderate demented (2), demented (3) (Fig. 1).

Different modules of the proposed model are described briefly below.

3.1 Dataset

The open access series of imaging studies (OASIS) is an initiative aimed at making brain MRI datasets openly available to the scientific community. OASIS includes both cross-sectional and longitudinal study. The dataset used in this paper is of cross-sectional study [11]. In this study, we performed data augmentation to combine three datasets to create a new dataset that predicts dementia phases based on psychological parameters. The basic goal is to take into consideration as many psychological parameters as possible when generating predictions, because different types of tests yield different test results, for example, if a person is tested as having dementia in the MMSE exam and no dementia in the CDT, then it is tough to figure out which stage they are in, therefore, our system used a combination of psychological screening tests to forecast dementia stage by stage. The dataset contains 416 records of patients ranging in age from 18 to 96 years old, as well as 19 attributes that are a combination of clinical data and psychological tests.

Table 1 Detailed analysis of related work

Related works	Type of data	Techniques	Key factors
Harish et al. [1]	Sensor data	Random forest classifier and support vector machine (SVM) classifier	An assessment model is created from sensor data which uses classifiers to forecast the condition of patients. The result of this model is that whether the older adult is at risk or experiencing difficulties, healthy
Sivapriya et al. [2]	OASIS (Cross-sectional)	Comparative study of SVM, least square SVM (LSSVM), partial swarm optimization (PSO) + SVM, LSSVM + PSO	The proposed methodology deals with cross-sectional data. Feature selection is done using random forest elimination (RFE). Selected features are used for building model
Khan et al. [3]	OASIS (Cross-sectional)	Comparative study of 18 ML classifiers with discretization and without discretization	The proposed study is a twofold application. EDA, feature engineering (10 steps), and classifiers are the techniques used. Discretization is performed on data by using 3 encoding techniques. Finally, comparative study of all classifiers is illustrated
Neelaveni et al. [4]	OASIS (Cross-sectional)	SVM and decision tree (DT) classifier	The proposed method deals with few selected features of cross-sectional study using R tools
Bari Antor et al. [5]	OASIS (Longitudinal)	Comparative study of SVM, logistic regression (LR), DT, random forest (RF) with and without fine tuning	The proposed methodology deals with longitudinal data. Comparative study on classifiers with and without hyperparameter tuning is clearly illustrated
Basheer et al. [6]	OASIS (Longitudinal)	Modified Capsule Network Technique (MCapNet)	The proposed work justified that MCapNet is more efficient than convolution neural network (CNN). Hierarchical analysis is conducted on OASIS longitudinal data

(continued)

Table 1 (continued)

Related works	Type of data	Techniques	Key factors
Sivakani et al. [7]	OASIS (Longitudinal)	**Feature Extraction** (Expectation and maximization algorithm), **Feature Selection** (BFS Algorithm), Comparative study of Gaussian process, regressor, and decision stump by using metrics	The proposed workflow is: feature engineering, feature selection, and building model using classifiers. The metrics used in comparative study are mean absolute error (MAE), root relative squared error (RRSE), relative absolute error (RAE), root mean squared error (RMSE), correlation coefficient (CC)
Manzak et al. [8]	ADNI	Deep neural networks by random forest importance	They proposed fast and efficient model where DNN is used as classifier. RFE is used for reducing the complexity of model and also for better performance
Solale et al. [9]	ADNI	Comparative study of SVM, long short-term memory (LSTM), gated recurrent unit (GRU) with and without feature selection	This paper proposes the implementation of recurrent neural networks (RNNs) to (a) predict future mini-mental state examination (MMSE) scores in a longitudinal study and (b) introduce a multi-class multimodal neuroimaging classification process involving various stages
Khonthapagdee et al. [10]	Psychological tests	Clock drawing test (CDT) application is developed based on Montreal cognitive assessment (MoCA)	The purpose of this study is to establish an application for a clock drawing test (CDT) that is frequently used as part of screening tests for Alzheimer's, such as the Montreal cognitive assessment (MoCA) test. The complete score for CDT is three points: first point is the drawing of a rounded outline. Second point is to draw the numbers of the clocks in the correct order. Third point is from correctly drawing the hands of the clock according to the test command

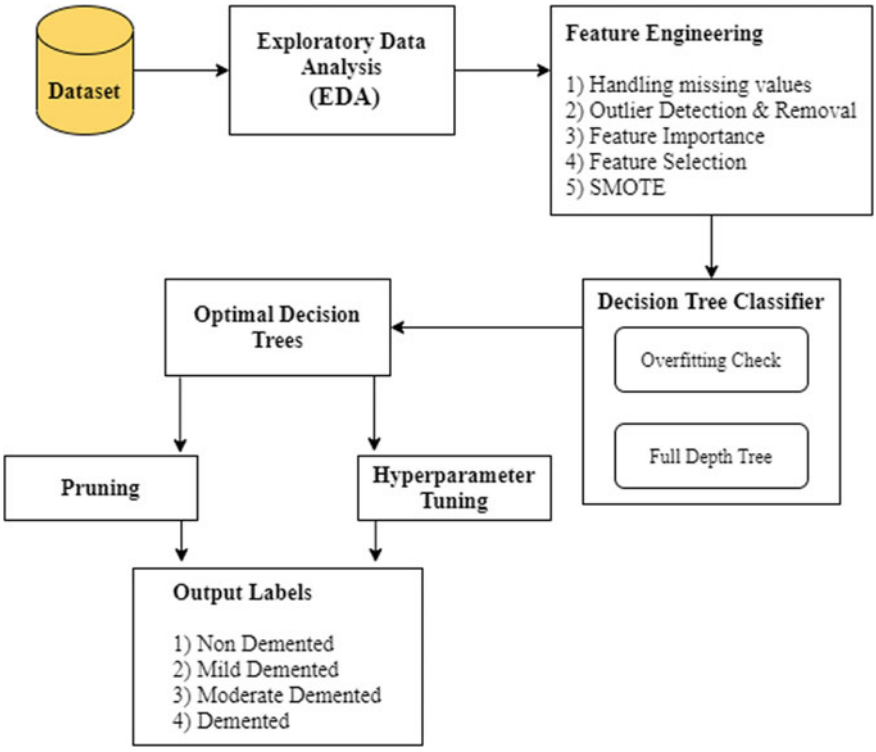


Fig. 1 Detailed view of system architecture

3.2 EDA (Univariate, Bivariate, Multivariate Analysis)

Exploratory data analysis (EDA) is a preliminary data analysis that uses statistics and visualizations to find correlations between data measures and gain insight into trends, relationships, and patterns among different entities in the data collection. EDA is divided into two categories, each of which is either non-graphical or graphical. The methods are then classified as univariate, bivariate, or multivariate. In our paper, EDA is performed on individuals and also combinations of attributes to identify the relation between them (Table 2).

3.3 Feature Engineering

In this module, there are five steps performed on dataset. They are:

(i) Handling missing values—Imputing technique used in this step is KNN.

- Method—KNNImputer

Table 2 Types of EDA

Univariate	Bivariate	Multivariate
Uni—one Variate—variable The goal is to derive data, define and summarize data, and examine the pattern that contains in it	Bi—two Variate—variable The goal is to analyze relationship between two variables	Multi—many Variate—variable The goal is to analyze relation between two or more variables

(ii) Outlier detection and removal—Inter Quartile Range (IQR) is applied in order to remove the outliers which are identified in the box plot when comparing the relation among attributes.
(iii) Feature Importance—This refers to methods that assign score to input features based on their usefulness in predicting output variable.
(iv) Feature Selection—Pearson correlation is used to select relevant features for building model.
(v) Synthetic Minority Oversampling Technique (SMOTE)—It is an oversampling technique used for balancing the class labels, i.e., output/target labels.

3.4 Decision Tree Classifier

DT algorithm is a supervised ML algorithm used for both regression and classification. In our paper, it is used as a classifier. The primary goal of this technique is to develop a model that predicts target label value, i.e., also known as output label, and the DT solves this problem by using the tree representation, where leaf node refers to a class label or output label and features or attributes are represented on internal node. Structure of decision tree is give in below figure (Fig. 2).

ID3, CART, C5.0, C4.5, and CHAID are some of the algorithms used in the construction of decision trees, sometimes known as variants of decision trees [12]. They are all useful in the construction of decision trees and have their own techniques for doing so, such as formulas, procedures, and type of data to be used. The obtained results of full depth tree are given in below table (Table 3).

Full Depth Tree Tree without using any pruning or tuning technique.

The above table displays the obtained results of the full depth tree, i.e., without any pruning or tuning techniques applied.

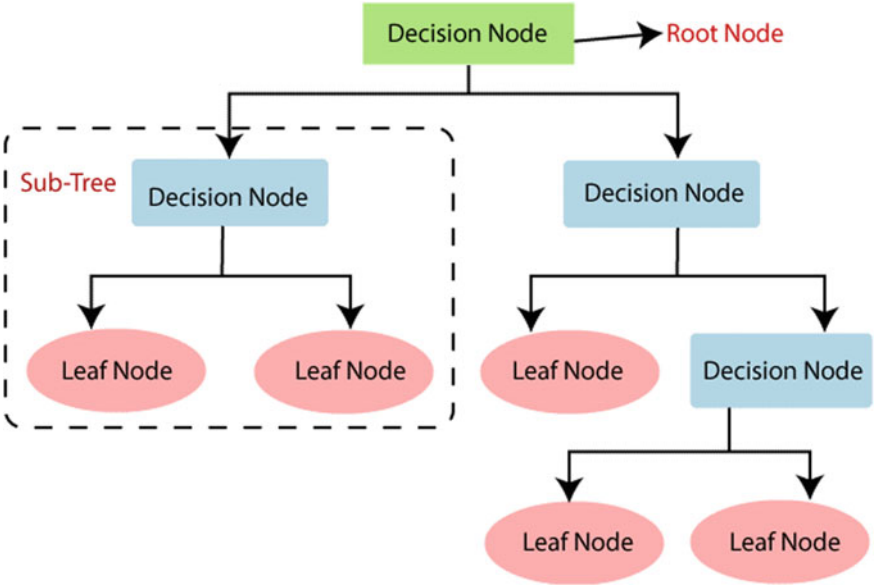


Fig. 2 Structure of decision tree

Table 3 Obtained results of full depth tree

Classifier	Accuracy	Max_depth and leaves
Decision tree classifier	93.2	Depth —8 and Leaves—26

3.5 Optimal Decision Trees

As DT is known to be prone to overfitting, we focused on constructing optimal decision trees through pruning and tuning strategies in our paper. The following are the pruning and tuning techniques used (Table 4):

Table 4 Types of pruning and tuning techniques used in our study

Pruning	Tuning
Cost complexity pruning (**Ccp_alpha** parameter is tuned in this technique)	• Random search • Grid search • Bayesian optimization • Genetic algorithm Hyperparameters used are: (Criterion, max_features, class_weight, min_samples_leaf, max_depth, split, splitter, min_samples)

Table 5 Results of pruned tree using cost complexity pruning technique

Classifier	Accuracy	Max_depth and leaves
Decision tree classifier (ccp_alpha = 0.004379)	94.3	Depth—7 and Leaves—18

Cost Complexity Pruning It is another option for controlling the size of tree for producing good results. This pruning technique uses ccp_alpha as a parameter that needs to be tuned for producing a pruned tree. ccp_alpha is calculated for each node of decision tree, finding the minimal ccp_alpha value is the main goal. Results of Pruned tree using cost complexity pruning technique is given in below table (Table 5).

Procedure The step-by-step procedure of cost complexity pruning is given below.

- **Step-1**: Provide train and test samples as input.
- **Step-2**: Generate ccp_alpha and impurities using "cost_complexity_pruning_path()".
- **Step-3**: Fit generated ccp_alpha's into "DecisionTreeClassifier()".
- **Step-4**: Calculate train and test scores by using accuracy as metric.
- **Step-5**: Select ccp_alpha value of highest test score.
- **Step-6**: Tune that in DecisionTreeClassifier.
- **Step-7**: Obtain accuracy of pruned tree.

It is clear that there is a change in performance of the pruned tree. The no. of leaves and depth of the tree are pruned.

Hyperparameter Tuning It is the process of finding the best combination of hyperparameters that allows the model to perform at its best. Using the appropriate combination of hyperparameters is the only method to obtain the best performance out of models. In our paper, we have considered four optimization techniques to perform hyperparameter tuning (Fig. 3).

Grid Search (GS) It is a method of hyperparameter tuning that involves building and evaluating a model for each set of algorithm parameters provided in a grid. Grid is a Cartesian product of a finite collection of values given by the user (Fig. 4).

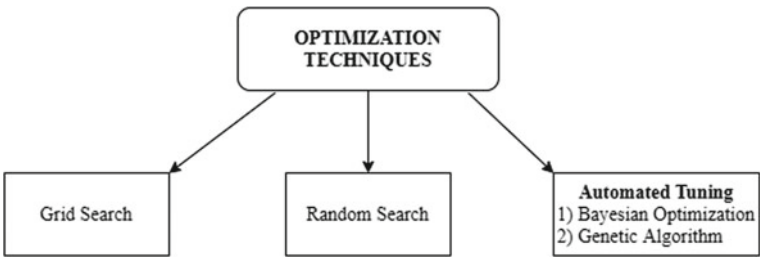


Fig. 3 Types of optimization

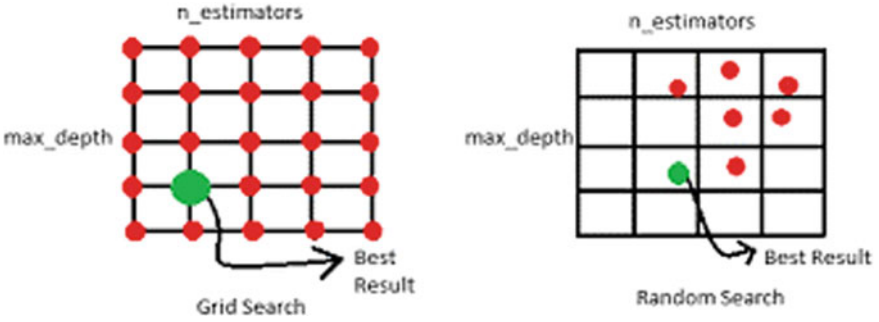


Fig. 4 Visualizing random and grid search techniques

Random Search (RS) It is proposed to overcome some limitations (Curse of dimensionality) of GS. Random combinations of parameters are picked from grid of values. The best solution among them is taken into consideration (Fig. 4).

Bayesian Optimization The iterative Bayesian optimization (BO) approach is often utilized for hyperparameter optimization (HPO) issues. In contrast to RS and GS, BO bases future assessment points on previously attained outcomes. BO employs two important components to select the next hyperparameter configuration: surrogate model and acquisition function. Random forest (RF), Gaussian process (GP), and the Tree Parzen Estimator (TPE) are common surrogate models for BO.

Genetic Algorithms The genetic algorithm (GA) is a popular metaheuristic algorithm based on the evolutionary notion that people with the best survival and environmental adaptability are more likely to survive and pass on their qualities to future generations.

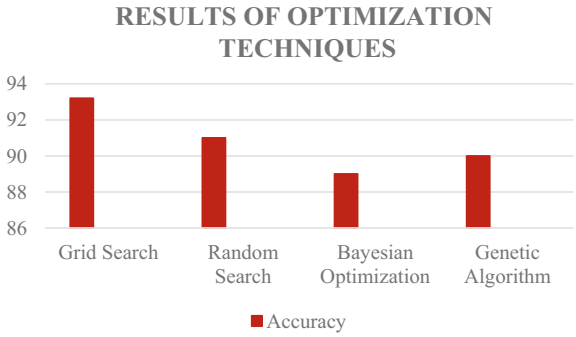
Comparative study of four optimization techniques is given below (Table 6 and Fig. 5).

Grid search shows better performance (i.e., with an accuracy of 93.2%) compared to other techniques.

Table 6 Libraries used for implementation

Type of tuning	Libraries or frameworks
Random search	Sklearn (RandomizedSearchCV)
Grid search	Sklearn (GridSearchCV)
Bayesian optimization	HyperOpt and BayesOpt
Genetic algorithm	TPOT

Fig. 5 Comparative graph



4 Conclusion

AD is the most frequent cause of dementia, and its number is rising globally. The disease's pathology begins years before clinical symptoms appear. Neurological, neuropsychological, and spinal fluid tests can all be used to pinpoint the diagnosis. There is no cure for this disease there can be only a delay in disease progress. This work deals with machine learning algorithms for predicting early detection of AD and delay the progress for an amount of time. In the previous studies when only a few features were considered for prediction and a decision tree classifier was used, the accuracy was around 81%, but when multiple attributes are considered for prediction, the model's performance improved significantly, i.e., 94.3%, especially when optimization techniques are used on decision trees. Pruning produces better results than hyperparameter tuning.

5 Future Work

This study focused on psychological and clinical data, which yielded better outcomes. We intend to use advanced techniques to work on various forms of data in the future, with a particular focus on images, speech, and sensor data. Creating a user interface (UI)-based screening test for all types of tests so that this application may assist users in recognizing their condition and taking appropriate precautions right away.

References

1. S. Harish, K.S. Gayathri, Smart home based prediction of symptoms of Alzheimer's disease using machine learning and contextual approach, in *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)* (IEEE, 2019)
2. T.R. Sivapriya, A.R.N.B. Kamal, V. Thavavel, Automated classification of MRI based on hybrid least square support vector machine and chaotic PSO, in *2012 Third International Conference*

on Computing, Communication and Networking Technologies (ICCCNT'12) (IEEE, 2012)
3. A. Khan, S. Zubair, Expansion of regularized kmeans discretization machine learning approach in prognosis of dementia progression, in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (IEEE, 2020)
4. J. Neelaveni, M.S. Geetha Devasana, Alzheimer disease prediction using machine learning algorithms, in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (IEEE, 2020)
5. M. Bari Antor, et al., A comparative analysis of machine learning algorithms to predict alzheimer's disease. J. Healthc. Eng. (2021)
6. S. Basheer, S. Bhatia, S.B. Sakri, Computational modeling of dementia prediction using deep neural network: analysis on OASIS dataset. IEEE Access **9**, 42449–42462 (2021)
7. R. Sivakani, G.A. Ansari, Machine learning framework for implementing Alzheimer's disease, in *2020 International Conference on Communication and Signal Processing (ICCSP)* (IEEE, 2020)
8. D. Manzak, G. Çetinel, A. Manzak, Automated classification of Alzheimer's disease using deep neural network (DNN) by random forest feature elimination, in *2019 14th International Conference on Computer Science & Education (ICCSE)* (IEEE, 2019)
9. T. Solale et al., Longitudinal prediction modeling of alzheimer disease using recurrent neural networks, in *2019 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)* (IEEE, 2019)
10. S. Khonthapagdee et al., Alzheimer screening using drawing test scores, in *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (IEEE, 2020)
11. D.S. Marcus et al., Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cogn. Neurosci. **19**(9), 1498–1507 (2007)
12. M. Batra, R. Agrawal, Comparative analysis of decision tree algorithms, in *Nature Inspired Computing* (Springer, Singapore, 2018), pp. 31–36
13. A.F. Fotenos et al., Brain volume decline in aging: evidence for a relation between socioeconomic status, preclinical Alzheimer disease, and reserve. Arch. Neurol. **65**(1), 113–120 (2008)
14. A.F. Fotenos et al., Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. Neurology **64**(6), 1032–1039 (2005)
15. B. Magnin et al., Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology **51**(2), 73–83 (2009)

COVID-19 Detection and Remote Tracking System Using IoT-Based Wearable Bracelet

Nasser M. Al-Zidi, Mohammed Tawfik, Talal A. Aldhaheri, Ali Mansour Almadani, Zeyad A. T. Ahmed, and Ali M. Al-Zidi

Abstract A new type of virus was discovered in China in the year 2019, known as COVID-19. One of the main symptoms that are easy to spot is high body temperature. The recent virus outbreak necessitates infrared thermometers used for thermal screening at public places to test the body temperature. However, this protection method still lacks because it requires a significant amount of time to monitor large numbers of people's body temperatures. Moreover, direct contact with people infected with coronavirus may spread it to the person doing the screening. In addition, this method cannot detect the infection early without visiting the infected person to a screening place. This study proposed a new system for automatically detecting the coronavirus in early time through the body temperature with no human interactions using IoT-based wearable bracelets. The body temperature sensor is integrated into the wearable bracelet with IoT technology for monitoring the body temperature and reading the current bodily temperature. The system is additionally equipped with a GPS module. It can capture the location of the person automatically. Suppose the person is suffering from high body temperature. In that case, the system will send it with location through Wi-Fi module or GSM module over the internet to cloud database and notify medical officer at the same moment to do the immediate procedures for that person. Health officers use smartphone applications for monitoring and remote tracking using the application map.

Keywords IoT technology · COVID-19 · Coronavirus · Body temperature · GPS location

N. M. Al-Zidi (✉)
Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad, Hyderabad, India
e-mail: alzidi.nasser@gmail.com

M. Tawfik · T. A. Aldhaheri · A. M. Almadani · Z. A. T. Ahmed
Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India

A. M. Al-Zidi
Faculty of Administrative and Computer Sciences, Albaydha University, Albaydha, Yemen

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_29

1 Introduction

Coronavirus is the newly discovered virus that causes the newly identified disease known as COVID-19. The first wave of the outbreak in China at the end of 2019, coronavirus has now spread to almost all countries. Because of the frequent contact, the coronavirus can easily spread from one person to another [1, 2]. On 30 January 2020, the virus's rate of spread and the entire situation's uncertainties surrounding has led the World Health Organization (WHO) to announce the coronavirus outbreak a 'Global Public Health Emergency'. The coronavirus has spread to almost every region of the world to infect 169,597,415 cases and caused 3,530,582 deaths till 31 May 2021 [3–6].

The main symptoms of COVID-19 are shortness of breath, cough, and fever. The incubation period can vary from person to person after being exposed; the symptoms appear to range between two days to two weeks [7]. One of the main symptoms of coronavirus that can be easily identified is fever and high body temperature. Infrared thermometers have been used since the virus's outbreak for thermal screening at public places to take a temperature tests. Because the thermal screening takes a long time to determine the overall body temperature of many people and the direct contact with the infected people might lead to the spread of the coronavirus to the person who does the screening process, prevention still lacking. In addition, this method cannot detect the infection early without visiting the infected person to a screening place [8]. An alternative modern technology is needed to prevent this flaw. Remote health monitoring of people increases with the popularity of technologies and wearable health-monitoring devices developed to enable remote monitoring [9, 10].

The Internet of Things (IoT) and mobile technologies make it easier to monitor people's health vitals through sharing health information with healthcare officers [11, 12]. The Internet of Things (IoT) is a concept of connecting any physical object capable to transfer the data over the network or internet, such as machines, sensors, and embedded software, to the internet. The associated devices are identified by codes or unique numbers. IoT is now proven and well-established and considered an important technology that combines machine learning, sensory products, instant analytics, and many other tactics to create the future platform. In typical daily functions, the IoT is known and recognized as the utility of the tools or products that meet the various needs of people in the real world, such as health monitoring, smart lighting arrangements, home's security system, and others, many of which are easy to manage on smartphones, smart speakers, etc. Internet of Things (IoT) implementation in some forms of medical treatment assists with improving the infected patient's treatment outcomes. And reducing healthcare costs. Internet of Things (IoT) based healthcare system is a good way to keep track of the symptoms of COVID-19 patients. By employing interconnected sensors and networks to monitor the common symptoms of COVID-19 remotely in real-time [10, 13, 14]. Global Positioning System (GPS) or satellite navigation system was originally developed by the United States Government to handle positioning, timing, and navigation for the military. Still, now anyone who has a GPS receiver can receive the broadcasted signals by the satellites

and determine accurate location and time, anywhere in the world, in any weather, day or night. GPS technology is being used in the mobile phone that helps in tracking the location of people. The most important application for the GPS is satellite navigation in ships, vehicles, and aircraft. One of the most important usages for the GPS is satellite navigation in aircraft, vehicles, and ships [15, 16]. This study aims to design a system for COVID-19 detection and remote tracking using IoT-based wearable bracelet equipped with a microcontroller, temperature sensor, and GPS module. This system transmits the body temperature and GPS location to the cloud database and delivers a real-time notification to healthcare officers' smartphone applications to perform the remote tracking.

This paper is coordinated as follows: Sect. 1 contains the introduction, Sect. 2 transacts with the related works which contain a survey of the literature, the proposed methodology and system components are explained in Sects. 3 and 4 contains the obtained results and the discussion, Sect. 5 provides summarization of the proposed system.

2 Related Works

2.1 IoT with Healthcare

IoT-based healthcare monitoring systems have drawn considerable attention from the research community. As a result, yearly increasing study and development efforts using IoT in health care and remote tracking have been published in the literature. This section includes some of the very recent related works.

Rahman et al. [10] discussed how the Internet of Things (IoT) can help in combat an outbreak of the (COVID-19). The use in combination with a comprehensive digital disease-monitoring system, definitely a requirement to keep this disease under control. Have mentioned some of the smart technologies such as wearable health-monitoring devices, artificial intelligence (AI), web-based tools, mobile computing, mobile sensors, remote health testing on the cloud, and the imperative to apply this technology to predict, detect, or prevent the spread of infectious diseases that are emerging.

Mohammed et al. [17] designed a system that automatically detects coronavirus without human interaction, using smart helmet equipped with a thermal imaging system. The smart helmet is equipped with thermal camera technology and is combined with IoT technology to track the crowd to help authorities monitor those with high body temperature in real-time and send that data to the appropriate parties.

Yang et al. [18] created and implemented an IoT-based healthcare system for ECG monitoring, wearable monitoring node that used for gathering ECG data. The gathered data is transmitted directly to the IoT cloud using Wi-Fi. To provide users with timely and visual ECG data. The HTTP and MQTT protocols are used by the IoT cloud. M. N. Mohammed et al. [19] created a monitoring system that has deployed

smart glasses that collect real-time information on COVID-19 to identify symptoms. This glass used to monitor and notify the responsible authorities of unusually high body temperature in the body of the patient that results shown on mobile app. The system used two cameras for gathering detailed information, called an optical camera and a thermal camera. Body temperature can be detected by using a thermal camera and detected faces can be captured using an optical camera.

Mohammed et al. [8] developed a system for coronavirus detection using a drone or Unmanned Aerial Vehicle (UAV). The drone incorporates a thermal camera to detect the temperature of the ground at various distances from the ground. This drone helps the health officers in the early detection of coronavirus by capturing people's high body temperature through a thermal camera and can notify health officers and send the data to them to be displayed. Fatima et al. [20] implemented an Internet of Things (IoT) approach known as IoTS MCFIS for intelligent and efficient prediction of the coronavirus using Fuzzy Inference System (FIS). The system smartly monitors the humans and predicts if infected by covid-19 or not. The IoTSMCFIS system uses MATLAB for simulation. The simulations indicate promising and attractive results.

2.2 Remote Tracking and Mobile Technology

Remote tracking systems around the world that use mobile technology are used in many fields. These systems have great attention from researchers and developers, and here are some related works.

Pham et al. [21] system designed to use the Global Positioning System to monitor vehicles and used modem of Global System for Mobile Communications (GSM); through the mobile network, it enables users to easily track their vehicles. The system uses GPS to get the vehicle's coordinates and send them to the user's phone via mobile network using a GSM modem. All these items are included in the system: a u-blox NEO-6Q GPS receiver module, GSM module, and Arduino Uno microcontroller. The presented system can be used for personal fleet management, Anti-theft, vehicle security, and applications of various kinds.

Hasan and Sharif Hossen. [22] Developed a smartphone android application-based real-time student bus tracking system. That system used the GPS for positioning and tracking of the buses in the form of longitude and latitude. Students use the map of the android application to locate the exact location of their respective buses so that they arrive at the bus stop in time. They concluded that this application saves student time and helps to arrive at the appropriate time.

Aravind et al. [23] implemented automatic tracking and alert system using a device attached to the animal body. This device is equipped with GSM and GPS technology and would constantly monitor the animal's location concerning the GPS boundaries predefined inside a wildlife sanctuary. When an animal strays out of the predefined GPS boundaries, the system will generate an alarm and uploading the information to a website that can use by officers to monitor the location of the wild animal.

3 Methodology

The proposed system was developed for automatic coronavirus detection early through remote monitoring of a person's body temperature using an IoT-based wearable bracelet. The body temperature sensor is integrated into the wearable bracelet with IoT technology to monitor body temperature to get real-time measures. A GPS module is integrated into the proposed system; it can also capture the person's location automatically. The person wears the bracelet, if there is high a temperature the system can detect it, the GPS module determines the position coordinates of that person, transmits the data wirelessly over the internet to a cloud database, and delivers a notification that includes the temperature and the position coordinates to the health officer. The officer will receive the temperature data for the people and location through a smartphone application to do the immediate procedures and remote tracking on the application map to know what the current location of the person is and what are the places that have been visited by the person to find who was infected with COVID-19 as shown in Fig. 1.

There are three parts to the proposed system. The first component of our proposed system is the input source which consists of the body temperature sensor and GPS module. The development of the microcontroller processor was the second part of system development. In this part, the Arduino IDE software has used for writing the system source code. This software enables the compilation of system source code into the NODEMCU microcontroller processor. Creating a cloud database and developing the smartphone application as an output source using Android Studio IDE was the third part.

3.1 NodeMCU Microcontroller

NodeMCU is a hardware-based on the ESP-12 module and contains one analog I/O pin and nine digital I/O pins with built-in ADC and operates at a 3.3 V power supply. It is equipped with an ESP8266 Wi-Fi module. The features of the device 80 MHz of the system clock, 4 MB of flash memory, an on-chip Wi-Fi Transceiver, and around 50 k of usable RAM [24].

3.2 Temperature Sensor

LM35 temperature sensor is a temperature measuring device that is designed to accurately measure the coldness or hotness of an object and it has a proportional analog output voltage to the temperature (in °C). When the temperature increases, the output voltage also increases. This sensor was formerly used to measure the

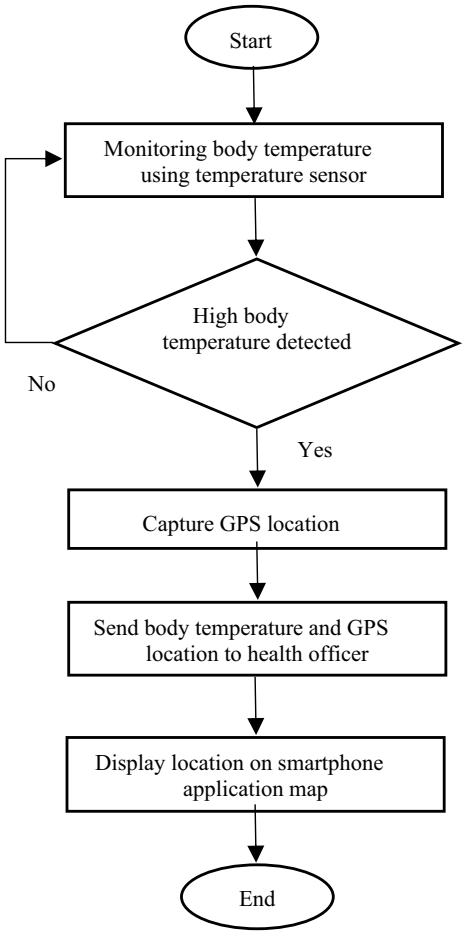


Fig. 1 System flowchart

temperature around it. between −55 and 150 °C and has low self-heating and does not cause temperature rise higher than 0.1 °C in still air [25, 26].

3.3 GPS Module

NEO-6 M is a GPS module that uses serial communication to provide location details and provides a strong capability for satellite search. This module's complete GPS receiver is highly accurate and has indicators for power and signal and four pins for interfacing with the microcontroller. It comes with a built-in ceramic antenna 25 × 25 × 4 mm [24, 27]. Figure 2 shows the system devices.

NodeMCU microcontroller connected to a sensor that detects body temperature and a GPS module that receives GPS location information and transfers it with person

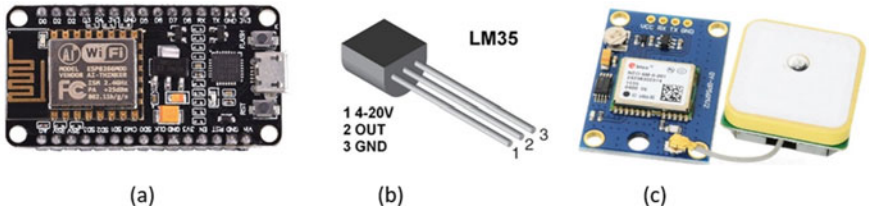


Fig. 2 **a** NodeMCU microcontroller, **b** LM35 temperature sensor, and **c** GPS module

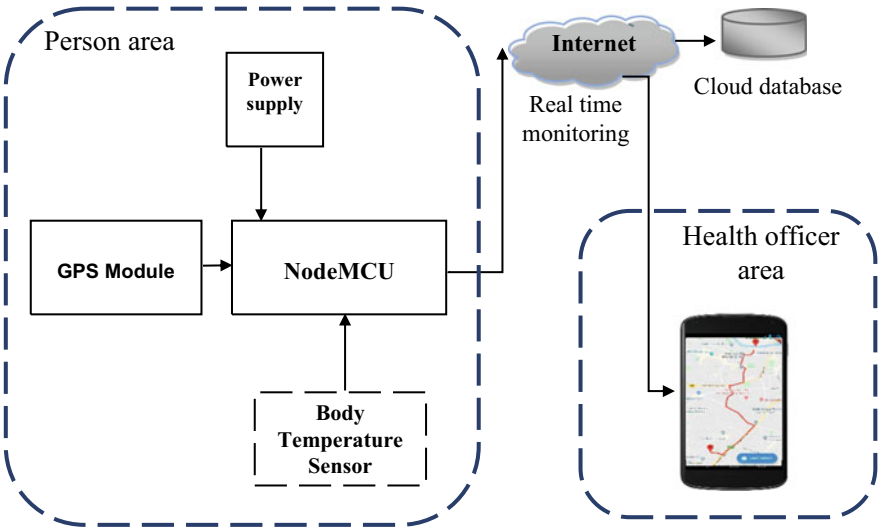


Fig. 3 Block diagram of the proposed system

identifier to Google Firestore database using an embedded Wi-Fi Transceiver. The smartphone application of the health officer is used to retrieve the stored data in the same time for body temperature monitoring and tracking the persons. Figure 3 exposes the system block diagram.

4 Results and Discussion

The proposed system has been tested with a real person suffering from fever after setting the temperature threshold of the system to 37.7 °C. When the person wore the wearable bracelet as shown in Fig. 4, the NodeMCU microcontroller has read the body temperature and GPS location, saves the data in the cloud database, and sent a notification over the web to the smartphone application.

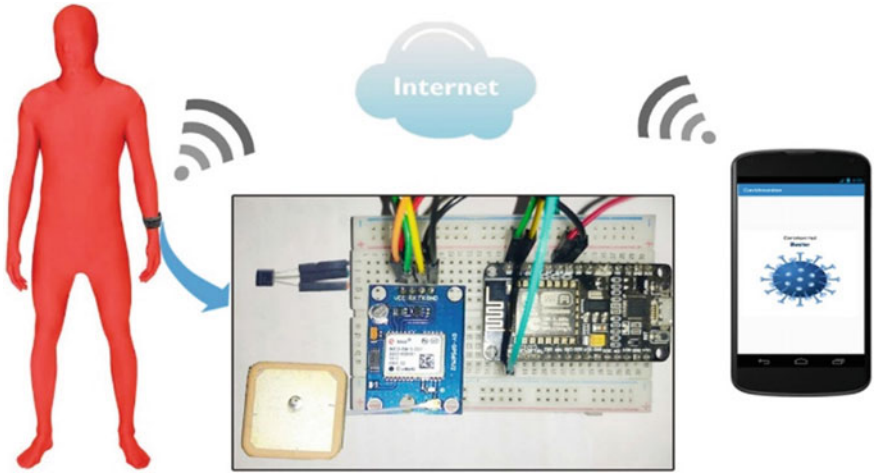


Fig. 4 The wearable bracelet

On the smartphone application side, the notification and data arrived in real-time, and the health officer can track the person through the application map and see what the current location of the infected person is and what has this person visited, as shown in Fig. 5.

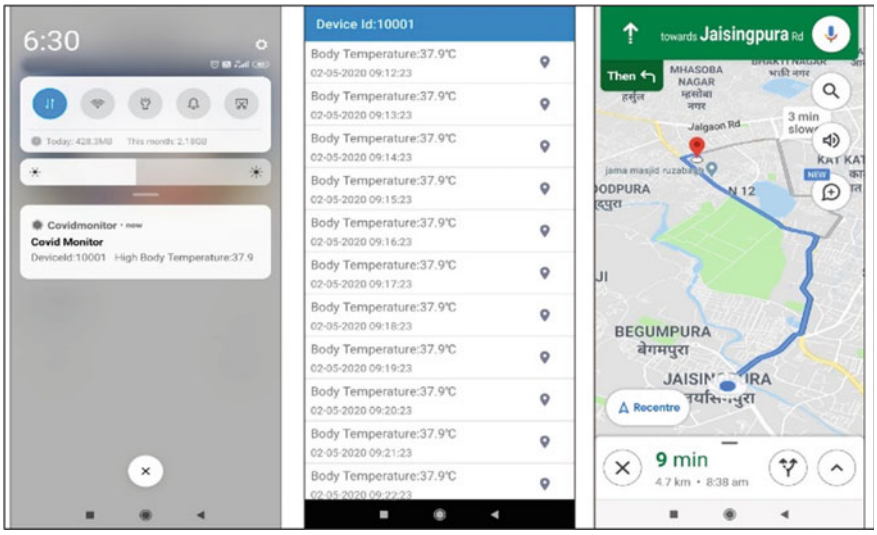


Fig. 5 Smartphone application of health officer

5 Conclusion

Early detection of the coronavirus symptoms is a suitable way to prevent coronavirus from spreading as the high body temperature of a person is one of the very common symptoms. So, the diagnosis of the target people's situation will consume less time and less human interactions that might lead to faster spreading of the coronavirus. It can be concluded that remote tracking technology is a method to prevent coronavirus from spreading. It helps to knowing the current location of the infected person and the places that the infected person has visited.

Detecting the coronavirus early and remote tracking system using a wearable bracelet has been designed. This system is capable of monitoring and identifying body temperature. Then wearable bracelet sends the measured data and GPS location to cloud database to be displayed on a mobile app. When the system detects a temperature higher than normal body temperature will notify a health officer, which would help to control the epidemic.

References

1. C. Wang, P.W. Horby, F.G. Hayden, G.F. Gao, A novel coronavirus outbreak of global health concern. Lancet **395**, 470–473 (2020)
2. L. Bai et al., Chinese experts' consensus on the internet of things-aided diagnosis and treatment of coronavirus disease 2019 (COVID-19). Clin. eHealth **3**, 7–15 (2020)
3. Z. Allam, D.S. Jones, On the coronavirus (Covid-19) outbreak and the smart city network: universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management. Healthc. **8** (2020)
4. WHO: Weekly Operational Update on COVID-19. https://www.who.int/publications/m/item/weekly-update-on-covid-19---16-october-2020. Last Accessed 10 Aug 2021
5. N. Dinesh, *Theatre and War: Notes from the Field*
6. V. Chamola, V. Hassija, V. Gupta, M. Guizani, A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact. IEEE Access **8**, 90225–90265 (2020)
7. W.G. Carlos, C.S. Dela Cruz, B. Cao, S. Pasnick, S. Jamil, Novel Wuhan (2019-nCoV) coronavirus. Am. J. Respir. Crit. Care Med. **201**, P7–P8 (2020)
8. M.N. Mohammed, N.A. Hazairin, S. Al-Zubaidi, S. AK, S. Mustapha, E. Yusuf, Toward a novel design for coronavirus detection and diagnosis system using IoT based drone technology. Int. J. Psychosoc. Rehabil. **24**, 2287–2295 (2020)
9. M. Taştan, IoT based wearable smart health monitoring system. Celal Bayar Üniversitesi Fen Bilim. Derg. **14**, 343–350 (2018)
10. M.S. Rahman, N.C. Peeri, N. Shrestha, R. Zaki, U. Haque, S.H. Ab Hamid, Defending against the novel coronavirus (COVID-19) outbreak: how can the internet of things (IoT) help to save the world? Heal. Policy Technol. **9**, 136–138 (2020)
11. A.M. Ghosh, D. Halder, S.K.A. Hossain, Remote health monitoring system through IoT, in *2016 5th International Conference on Informatics, Electronics and Vision, ICIEV* (2016), pp. 921–926
12. M. Hassanalieragh et al., Health monitoring and management using internet-of-things (IoT) sensing with cloud-based processing: opportunities and challenges, in *Proceedings—2015 IEEE International Conference on Services Computing, SCC 2015* (Institute of Electrical and Electronics Engineers Inc., 2015), pp. 285–292

13. I. Lee, K. Lee, The internet of things (IoT): applications, investments, and challenges for enterprises. Bus. Horiz. **58**, 431–440 (2015)
14. R.P. Singh, M. Javaid, A. Haleem, R. Suman, Internet of things (IoT) applications to fight against COVID-19 pandemic. Diabetes Metab. Syndr. **14**, 521–524 (2020)
15. M.N. Hasan, M. Sharif Hossen, Development of an android based real time bus tracking system, in *1st International Conference on Advances in Science, Engineering and Robotics Technology 2019, ICASERT 2019* (Institute of Electrical and Electronics Engineers Inc., 2019)
16. M. Munir, S. Perälä, K. Mäkelä, Utilization and impacts of GPS tracking in healthcare: a research study for elderly care. Int. J. Comput. Appl. **45**(11), 35–37 (2012)
17. M.N. Mohammed, H. Syamsudin, S. Al-Zubaidi, R.R. AKS, E. Yusuf, Novel COVID-19 detection and diagnosis system using IOT based smart helmet. Int. J. Psychosoc. Rehabil. **24**, 2296–2303 (2020)
18. Z. Yang, Q. Zhou, L. Lei, K. Zheng, W. Xiang, An IoT-cloud based wearable ECG monitoring system for smart healthcare. J. Med. Syst. **40**(12), 1–11 (2016)
19. M.N. Mohammed et al., 2019 Novel coronavirus disease (Covid-19): detection and diagnosis system using IoT based smart glasses. Int. J. Adv. Sci. Technol. **29**, 954–960 (2020)
20. S.A. Fatima, N. Hussain, A. Balouch, I. Rustam, M. Saleem, M. Asif, IoT enabled smart monitoring of coronavirus empowered with fuzzy inference system. Int. J. Adv. Res. Ideas Innov. Technol. **6**, 188–194 (2020)
21. H.D. Pham, M. Drieberg, C.C. Nguyen, Development of vehicle tracking system using GPS and GSM modem, in *2013 IEEE Conference on Open Systems, ICOS 2013* (2013), pp. 89–94
22. S. Balaji, R. Raju, K.S.P. Sandosh, R. Ramachandiran, Smart way tracking to identify individuals location using android system. Int. J. Pure Appl. Math. **119**, 9–15 (2018)
23. D. Aravind, S. Anupriya, A.T. Aarthi, C.S. Kumar, An automatic wildlife tracking system using GPS and wireless sensor networks, in *National Conference on Electronics, Communication and Computing (NCECC-2017)* (2017)
24. B.K. Perumalla, M.S. Babu, An intelligent traffic and vehicle monitoring system using internet of things architecture. Int. J. Sci. Res. (2016)
25. A. Abdullah, A. Ismael, A. Rashid, A. Abou-Elnour, M. Tarique, Real time wireless health monitoring application using mobile devices. Int. J. Comput. Netw. Commun. **7** (2015)
26. P. Gopi Krishna, K. Srinivasa Ravi, S. Adluri, D. Sree Devineni, R. Scholar, Implementation of bi directional blue-fi for multipurpose applications in IoT using MQTT protocol
27. S. Lee, G. Tewolde, J. Kwon, Design and implementation of vehicle tracking system using GPS/GSM/GPRS technology and smartphone application, in *2014 IEEE World Forum Internet Things, WF-IoT* (2014), pp. 353–358

A Novel Approach for Age and Gender Estimation in Surveillance Videos

D. Manju and V. Radha

Abstract Nowadays surveillance cameras are everywhere and it produce a tremendous amount of data every day, making manual detection of abnormal activities a laborious one. So, various automatic abnormal human activities prediction techniques have been proposed. Rain Streaks Removal—Enhanced Object Detection and Tracking—Spatio-temporal Frequent Object Mining (RSR-EODT-STFOM) predicted the abnormal human activity in surveillance videos; STFOM-Face Detection-Pose-Invariant Orthogonal Locality Preserving Projection (STFOM-FD-PIOLPP) recognized face of those persons. In this paper, the STFOM-FD-PIOLPP is extended by detecting the age and gender of the unknown faces. Stacked Trimmed Active Shape Model (STASM) is introduced to locate facial landmarks for efficient face recognition. Speeded Up Robust Features (SURF), OLPP, HOG, Weighted LBP, HFO, and HFD are used to classify the age and gender of the unknown faces using CNN. Estimated age and gender of unknown faces are saved in database for further clarification of human abnormal activity in surveillance videos.

Keywords Human abnormal activity · Age estimation · Gender estimation · Stacked trimmed active shape model · Convolutional neural network · Speeded up robust features

1 Introduction

Nowadays, video surveillance system becomes an intelligent autonomous system and it is capable of capturing all situations. In a situation that may pose potential hazards, the video surveillance system is used to set alarm. Recording videos are extremely difficult to monitor since incidents that need follow-up to have very low probability. The difficulty of ordinary crowding is compounded in crowded scenes. Anomaly

D. Manju (✉) · V. Radha
Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India
e-mail: manju@cit.edu.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_30

detection [1] is the task of finding trends in a given data that are not regular. In general, the local abnormal event and global abnormal event are graded as abnormal.

After the discovery of unusual events in the video surveillance system, it is more required to recognize the human faces to confirm the abnormal event. Face recognition [2] is an automated method to confirm or identify the identity of a living individual on the basis of its physiology. A biometric identification system is typical for the purpose of identifying a person using either physiological features (for example face, fingerprint, or iris pattern) or behavioral patterns (like keystroke, handwriting, or voice patterns). The face recognition system is used to identify a person.

An Orthogonal Locality Preserving Projection (OLPP), Histogram of Gradient (HOG), and *Weighted Local Binary Patterns (WLBP)* were used for [3] face recognition. The face recognition was accomplished by directly matching the corresponding OLPP, HOG and WLBP features of detected face images and faces images in the database through cosine similarity score. The detected faces might have pose variations which may affect the performance of face recognition. So, in order to recognize the faces at different poses, Pose-Invariant OLPP (PIOLPP) was proposed. In PIOLPP, Histogram of Face Orientation (HFO) and Histogram of Face Direction (HFD) features were used along with the OLPP, HOG and WLBP features to improve the face recognition process.

In this article, PIOLPP-based face recognition process is extended by detecting the age and gender of a person who is detected as unknown. A Stacked Trimmed Active Shape Model (STASM) is introduced to improve the face recognition and Speeded Up Robust Features (SURF) are used for age and gender estimation of a person. STASM label the feature points and the SURF is extracted from the image. The extracted SURF, OLPP, HOG, WLBP, HFO, and HFD features are trained and tested by using Convolutional Neural Network (CNN) to estimate the age and gender of the unknown person in the video surveillance system.

2 Literature Survey

For 3-D face recognition with Spherical Harmonic Features (SHF), a competitive model [4] was proposed. At first, 3-D faces are represented as spherical depth maps in a canonical representation. SHF has been calculated from the spherical depth. Then, especially in the presence of facial expression and occlusion, the predictive contribution of the individual SHF characteristic was considered, feature selection methods have been used for improving predictive performance and providing faster and rentable predictions. Nevertheless, when occlusion and expression were considered, the efficiency of a competitive model were decreased slightly.

For video-based face recognition, a Trunk-Branch Ensemble Convolutional Neural Network (TBE-CNN) [5] model was proposed. TBE-CNN was a full template extracted from face images. Through share the convolutional low to medium stage layers between the branch networks and trunk, the TBE-CNN model was efficient in extracting the features. Through adding an enhanced triplet loss function the efficacy

of the TBE-CNN design was further increased. However, CNN is computationally expensive for face recognition.

A Domain-Specific Face Synthesis (DSFS) [6] has been suggested to be recognized for video in one sample per human. Operating Domain (OD) data were used by the DSFS for intra-class representative variation information. In the captured condition space affinity propagation clustering was used to pick a lightweight array of unidentified people's faces that appear on the OD. Synthetic faces were created using the DSFS to create a cross-domain dictionary, based upon a sparsely defined representation. Nevertheless, the computational complexity of this approach increases moderately.

For face identification and recognition of facial expression, a Multi-Task Facial Inference Model (MT-FIM) [7] was proposed. The sparsity variable reflected the data that was shared. The MT-FIM method permitted the rigorous execution of every individual task optimizing the distance between various classes concurrently and increasing the internal dispersion of classes. MT-FIM may misclassify the data often.

A novel approach [8] was proposed using transform and spatio domain for robust expression invariant face recognition. A new approach was designed to fuse the characteristics with the proposed contourlet enhancement algorithm and feature-level fusion process. For feature selection, a discrete domain contourlet transformation was explored that was capable of capturing edge information across multiple lines and multi-resolution. Also, a new coefficient algorithm for contourlet sub-bands has been proposed that enhanced the system's robustness by enhancing skin area. The transform coefficients of the contours vary considerably which affects facial recognition performance.

An adaptive feature descriptor [9] was proposed for face recognition based on Weber's law. A Weber-based auto threshold was built and double channel patterns were analyzed for further data. However, it has low recognition rate.

A new GAN-based method, AgeGAN [10] was proposed that completes facial texture and uses the data to estimate age of the face detected. It uses a Wasserstein-based GAN to restore uv texture presentation. Facial texture map and age characteristics of the face under study are learned simultaneously and age estimation is done. Performance of the method is not considerably better.

A deep learning-based approach for age estimation [11] from a single face image was proposed. This solution avoids using facial landmarks for the estimation and is trained using VGG-16 architecture and ImageNet. However, execution time increases and is computationally expensive.

A novel method is proposed to use factor analysis [12] for dimensionality reduction on the features extracted from facial images. An object is assumed to be made of content factors and style factors. These factors are to be optimally used in feature extraction, and subsequently in reducing the number of features that are necessary for object recognition. This method is computationally intensive and difficult be adapted in real-time scenarios.

An attempt was made to apply divide-and-rule strategy [13] along with CNN to estimate age of face images. The deep learning method employed uses factor analysis model to extract robust features from the input face images. Rank-based age estimation learning methods are studied based on annotated age tags and then a divide-and-rule face age estimator is proposed. Though performance metrics of the age estimates are slightly better than SVM, they still need to be improvised.

3 Proposed Methodology

In this section, the Stacked Trimmed Active Shape Model (STASM) for face recognition and Convolutional Neural Network (CNN) to learn the Speeded Up Robust Features (SURF), OLPP, HOG, WLBP, HFO, and HFD for age and gender estimation of unknown faces are described in detail. Initially, Spatio-Temporal Frequent Object Mining (STFOM) is applied in the video surveillance system which detected the abnormal events. Then, a Viola-Jones edge detector is applied on the abnormal events to detect the faces who are all involved in the abnormal event. A pose-invariant Orthogonal Locality Preserving Projection (OLPP) is applied to recognize the faces. The age and gender of the unknown faces are estimated using SURF, OLPP, HOG, WLBP, HFO, and HFD in CNN.

3.1 Stacked Trimmed Active Shape Model

During the recognition of faces, STASM is applied in the detected faces. STASM produces landmarks and descriptions for human facial features. In most of the pictures, a landmark is a recognizable point. The facial features are situated by establishing landmarks. A variety of landmarks creates a structure representing all the x-points as vectors, followed by all the y-points of the shape points. Line one form (i.e., shape) to another with transforming resemblance which decreases the average Euclidean distance between shape points. The average form is the average of the aligned training forms. The Active Shape Model (ASM) analysis was conducted on the Viola-Jones face detector and found landmarks from the average structure associated with the location and dimensions of the face. After that, the following steps are carried out until convergence

- Adapt the form position by matching a template with an object texture around each point recommend a preliminary form.
- Confirm the preliminary form with the global model.

The matches for each template are inconsistent and the shape model blends the outcomes of the poor pattern matches into a strong classification. The entire search is replicated at every point, from rough to fine resolution, in an image pyramid. The shape model and profile model are the types of sub-models that make up the ASM.

The profile models are utilized by the template matching to identify the relative locations of each landmark. The ASM produces a structured gradient profile with a specified length for a sampling of the object around a line (whisker) orthogonal to the structure boundary at the landmark. Whereas the average profile vector $\overline{g}$ and the covariance matrix S_g are trained on manually labeled faces. The pixel with the profile g has the lowest Mahalanobis range to the median profile $\overline{g}$ is relocated during searches.

$$\text{Mahalanobis Distance} = (g - \overline{g})^{\mathrm{T}} S_g^{-1} (g - \overline{g}) \tag{1}$$

The structure model defines acceptable collection of landmarks. It creates a structure $\hat{x}$ with

$$\hat{x} = \overline{x} + \Phi b \tag{2}$$

In Eq. (2), $\overline{x}$ is the average shape, Φ is a matrix of chosen eigenvectors of the covariance matrix S_s of the points of the aligned training shapes and b is a parameter vector. Then design the structure with the principal components method as much deviation in training. It needs arranging the eigenvalues λ_i of S_s and holding a proper number of the corresponding eigenvectors in Φ.

For the entire ASM, a single shape template is used and it scales for each pyramid level. ASM looks for landmarks in the images according to the profile models and shape models.

STASM is the extension of ASM that enlarges the count of landmarks in the model which also improve the mean fit of this model. In STASM, two-dimensional (2D) profile is used at each landmark by sampling a square area around the landmark. More data about the landmark is found in a 2D profile region and it is widely utilized to more information about the image. The sampling area in directions x and y will be sampled when looking for landmarks, where x and y are orthogonal to the shape at the landmark and tangent to the structure edge correspondingly. It has to depend on an approximately straight face since two-dimensional profiles are centered on the image's edge.

For better face recognition, the STASM search for landmarks in the detected faces. The landmarks are used for alignment and warping. Then, a standard image (video frame) is selected from the database and aligns all landmarks on the image to the landmarks on the standard image. The cosine similarity score of the landmarks, OLPP, HOG, WLBP, HFO, and HFD are calculated for the detected faces and the faces in the database. Based on the similarity score, the unknown faces are recognized.

After this step, a new image is obtained after warping. The SURF is extracted from the new image. SURF is a robust local feature detector that utilizes Hessian matrices determinants to find image's significant points. SURF is the combination of feature extraction and feature description steps. The feature extraction uses Hessian matrices to locate image's significant points. Given a point $x = (x, y)$ in an unknown face image I, the Hessian matrix $\mathcal{H}(x, \sigma)$ in x at scale σ is given as follows:

$$H(f(x, y)) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \tag{3}$$

In Eq. (3), $L_{xx}(x, \sigma)$ is the convolution of the Gaussian second order derivative with the image I in point x, and similarly for $L_{xy}(x, \sigma)$ and $L_{yy}(x, \sigma)$. In order to compute the determinant of Hessian matrix, initially apply convolution with Gaussian kernel, then second-order derivative is applied. Because Gaussian kernel filters are in any situation not optimal, SURF even further improves the approximation with box filters. The determinant of the Hessian is represented as,

$$\det\left(\mathcal{H}_{\text{approx}}\right) = D_{xx} D_{yy} - \left(w D_{xy}\right)^2 \tag{4}$$

In Eq. (4), D_{xx}, $D_{yy,}$ and D_{xy} are the approximations for Gaussian second-order derivatives.

The images are then continuously smoothed with a Gaussian and sub-sampled to reach a better degree of scale space. Since box filters are used, SURF applies filters of whichever size directly to the original image at the same speed. Therefore, upgrading the filter size analyzes the scale space. Non-maximum suppression is applied in the neighborhood to find points of interest in the images. The definition of the SURF features correcting a reproducible orientation according to the circular information around the key point. After that, a square area is created in line with the orientation chosen and deletes the SURF lemma from it.

3.2 Convolutional Neural Network for Age and Gender Estimation

Support Vector Machine [14, 15] is used for age and gender estimation. With only a huge volume of training data, SVM-based gender estimation has high accuracy. Moreover, SVM-based age estimation still needs improvement in terms of accuracy. In order to solve these issues, a CNN is introduced for age and gender estimation. CNN based age estimation is denoted as CNN-AE and the CNN-based gender estimation is denoted as CNN-GE. The SURF, OLPP, HOG, WLBP, HFO and HFD features of unknown faces are given as input to the CNN to estimate the age and gender of the unknown faces. CNN is comprised of convolutional, pooling, and fully-connected layers. The convolutional layer intends to learn SURF, OLPP, HOG, WLBP, HFO and HFD representation. The convolutional layer consists of different convolution kernels used for calculating maps. Each neuron of a feature map is typically linked to a neighborhood neuron region in the previous layer. First, convert the input into learns kernel. After that an elementally nonlinear activation function is used in conjunction with the converted results to obtain the new feature map. The shared by all spatial locations of SURF, OLPP, HOG, WLBP, HFO, and HFD to generate each feature map. By using various kernels, the complete feature maps are

obtained. The value of feature at position (i, j) in the kth feature map of lth layer, $z^l_{i,j,k}$ is calculated as follows:

$$z^l_{i,j,k} = \left(w^l_k\right)^{\mathrm{T}} F^l_{i,j} + b^l_k \tag{5}$$

where, w^l_k is the weight vector of the kth filter of the lth layer, b^l_k is the bias terms of the kth filter of the lth layer, and $F^l_{i,j}$ is the SURF, OLPP, HOG, WLBP, HFO, and HFD features located at position (i, j) of the lth layer. Consider $a(.)$ is the nonlinear activation function that enables to track SURF, OLPP, HOG, WLBP, HFO, and HFD. The activation value $a^l_{i,j,k}$ of convolutional feature $z^l_{i,j,k}$ can be calculated as,

$$a^l_{i,j,k} = a\left(z^l_{i,j,k}\right) \tag{6}$$

The resolution of SURF, OLPP, HOG, WLBP, HFO, and HFD maps is reduced in pooling layer which helps to attain shift-invariance. Generally, pooling layer is located between two convolutional layers. Each SURF map of a pooling layer is linked to its related SURF, OLPP, HOG, WLBP, HFO, and HFD map of the former convolutional layer. For each feature map $a^l_{:,:,k}$ the pooling function pool(.) is given as:

$$y^l_{i,j,k} = \mathrm{pool}\left(a^l_{m,n,k}\right), \quad \forall (m, n) \in \mathcal{R}_{ij} \tag{7}$$

In Eq. (7), $\mathcal{R}_{ij}$ is a local neighborhood around location (i, j). The first convolutional layer processors are designed to separate low-level characteristics, whereas higher-layer kernels are more abstractly encoded SURF, OLPP, HOG, WLBP, HFO, and HFD. With different layers of convolution and pooling, steadily extract representations of higher-level features. After several convolutions and poolings, one or more fully-connected layers may exist. The fully-connected layer performs high-level reasoning. The generation of global semantine data takes all neurons in the previous layer and connected to each neuron in the current layer. The last CNN layer is the layer of output. A softmax operator is used in the output layer where the gender and age classification is carried out. Consider θ denote weight vectors and bias terms of CNN. By reducing suitable loss function, the optimum weight vector and bias terms are obtained for age and gender classification task. CNN has N desired input–output relations $\left\{\left(x^{(n)}, y^{(n)}\right); n \in [1, 2, \ldots, N]\right\}$, where $x^{(n)}$ is the n-th SURF, OLPP, HOG, WLBP, HFO, and HFD, $y^{(n)}$ is its related target label and $o^{(n)}$ age and gender of unknown faces. The loss of CNN is calculated as,

$$\mathcal{L} = \frac{1}{N} \sum_{n=1}^{N} l\left(\theta; y^{(n)}, o^{(n)}\right) \tag{8}$$

By optimizing the above loss function, CNN effectively detects the age and gender of unknown faces.

4 Result and Discussion

The performance of the face recognition methods and age and gender estimation based on SVM and CNN methods are analyzed here.

4.1 Analysis of Face Recognition Methods

The existing STFOM-FD-PIOLPP and proposed STFOM-FD-PIOLPP-STASM face recognition are tested in terms of accuracy, precision, and recall. For the experimental purpose, Multicamera Human Action Video Data (MuHAVi) dataset is used which collects a large body of human action data using 8 cameras [3].

Accuracy

It is measured according to the percentage of the recognized faces per the total number of tested faces of the person.

$$\text{Accuracy} = \frac{\text{True Positive (TP)} + \text{True Negative(TN)}}{\text{TP} + \text{TN} + \text{False Positive(FP)} + \text{False Negative(FN)}}$$

where TP is the number of suspicious faces that are correctly recognized as suspicious faces, TN is the number of non-suspicious faces that are correctly recognized as non-suspicious faces, FP is the number of non-suspicious faces are wrongly recognized as suspicious faces and FN is the number of suspicious faces are wrongly recognized as non-suspicious faces.

Table 1 shows the accuracy of STFOM-FD-PIOLPP and STFOM-FD-PIOLPP-STASM methods.

The accuracy of STFOM-FD-PIOLPP and STFOM-FD-PIOLPP-STASM methods is shown in Fig. 1. When the observation ratio is 60, the accuracy of STFOM-FD-PIOLPP-STASM method is 4.35% greater than STFOM-FD-PIOLPP method. From this analysis, it is proved that the proposed STFOM-FD-PIOLPP-STASM face recognition method has high accuracy than STFOM-FD-PIOLPP method.

Table 1 Accuracy versus observation ratio

Observation ratio	STFOM-FD-PIOLPP	STFOM-FD-PIOLPP-STASM
20	0.74	0.79
40	0.88	0.93
60	0.92	0.96
80	0.87	0.9
100	0.945	0.985

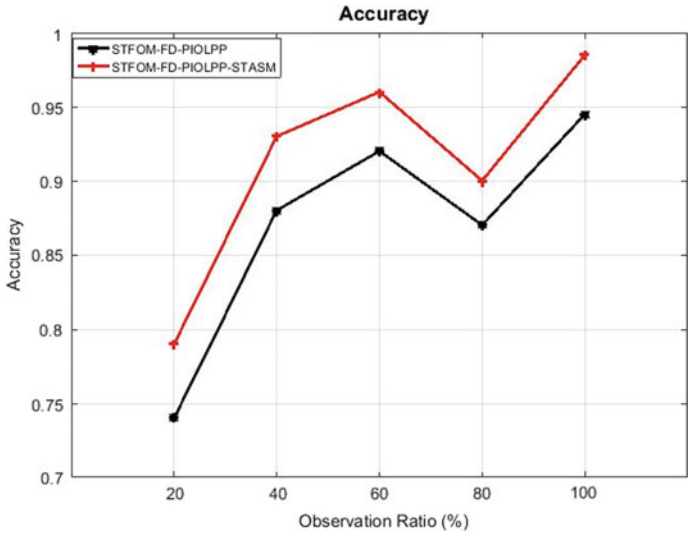


Fig. 1 Accuracy versus observation ratio

Precision

It is the fraction of the number of suspicious faces that are correctly recognized to the sum of the number of correctly recognized suspicious faces and the wrongly recognized suspicious faces.

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

Table 2 shows the precision ofSTFOM-FD-PIOLPP and STFOM-FD-PIOLPP-STASM methods.

The precision of STFOM-FD-PIOLPP and STFOM-FD-PIOLPP-STASM methods is shown in Fig. 2. When the observation ratio is 60, the precision of STFOM-FD-PIOLPP-STASM method is 4.44% greater than STFOM-FD-PIOLPP method. From this analysis, it is proved that the proposed STFOM-FD-PIOLPP-STASM face recognition method has high precision than STFOM-FD-PIOLPP method.

Table 2 Precision versus observation ratio

Observation ratio	STFOM-FD-PIOLPP	STFOM-FD-PIOLPP-STASM
20	0.742	0.793
40	0.882	0.934
60	0.923	0.963
80	0.874	0.903
100	0.95	0.989

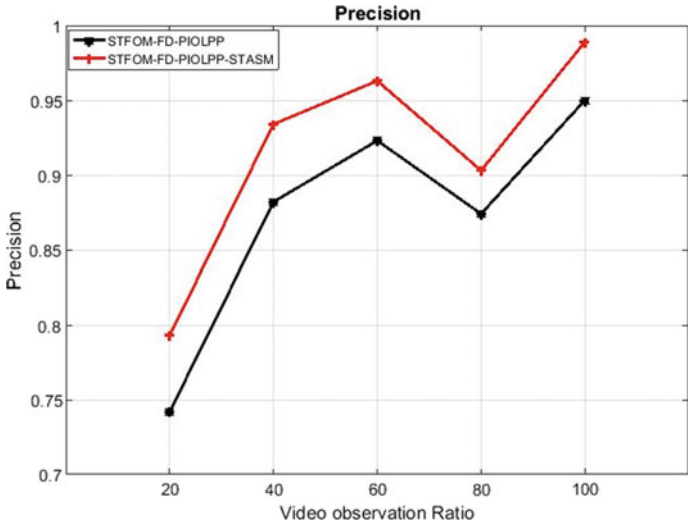


Fig. 2 Precision versus observation ratio

Recall

It is the fraction of the number of suspicious faces that are correctly recognized to the sum of the number of correctly recognized suspicious faces and the wrongly recognized non-suspicious faces.

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

Table 3 shows the recall ofSTFOM-FD-PIOLPP and STFOM-FD-PIOLPP-STASM methods.

Figure 3 shows the recall of STFOM-FD-PIOLPP and STFOM-FD-PIOLPP-STASM methods. The recall of STFOM-FD-PIOLPP-STASM is 2.4% greater than STFOM-FD-PIOLPP method. From this analysis, it is proved that the proposed STFOM-FD-PIOLPP-STASM face recognition method has high recall than STFOM-FD-PIOLPP method.

Table 3 Evaluation of recall

	STFOM-FD-PIOLPP	STFOM-FD-PIOLPP-STASM
Recall	96.2	98.5

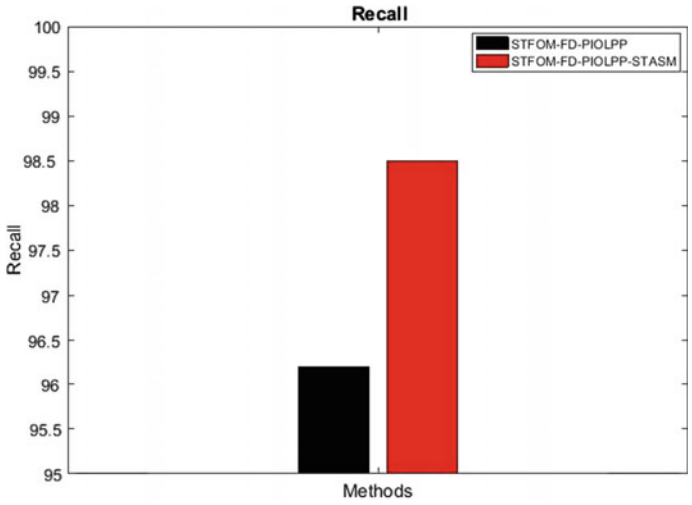


Fig. 3 Evaluation of recall

4.2 Analysis of Age and Gender Estimation Methods

The SVM-AE, SVM-GE, CNN-AE, and CNN-GE are tested in terms of accuracy. For the experimental purpose, a longitudinal image database of normal-adult age progression called MORPH is used. It includes 1742 face images of 515 people. These images depict a variety of people regarding age, sex, and ethnicity. There are 1278 images of decent African-American individuals, 433 of decent Caucasians, and three other images. 294 pictures of women and 1.430 of men are present. 76% have a form of facial hair, generally a mouse, in the male images.

Accuracy

Age and gender estimation accuracy can be measured based on the percentage of the estimated age and gender of a person per the total number of tested faces of the person.

Table 4 shows the accuracy of SVM-based age estimation and CNN-based age estimation methods.

Table 5 shows the accuracy of SVM-based gender estimation and CNN-based gender estimation methods.

Figures 4 and 5 show the comparison of accuracy for age and gender estimation, respectively. The accuracy of CNN-AE is 8.21% greater than SVM-AE and the

Table 4 Comparison of accuracy for age estimation

	SVM-AE	CNN-AE
Accuracy	89.92	97.3

Table 5 Comparison of accuracy for gender estimation

	SVM-GE	CNN-GE
Accuracy	93.43	96.8

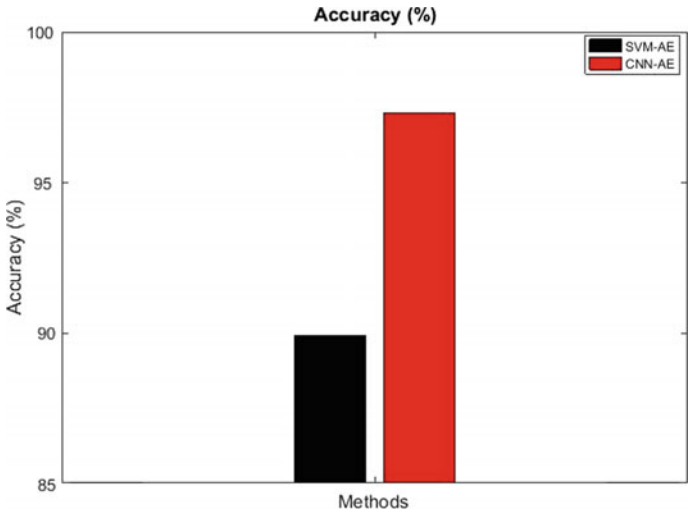


Fig. 4 Comparison of accuracy for age estimation

accuracy of CNN-GE is 3.61% greater than SVM-GE method. From this analysis, it is proved that the proposed CNN-AE and CNN-GE method has better accuracy than other methods.

5 Conclusion

In this paper, the STFOM-FD-PIOLPP-based face recognition is improved by detecting the landmarks in the detected faces. After that, SURF, OLPP, HOG, WLBP, HFO, and HFD features are extracted and these features are trained in CNN to estimate the age and gender of the unknown faces. The CNN used of convolutional, pooling, and fully-connected layers to estimate the age and gender of the unknown faces. It will be helpful for the further clarification of the abnormal activity in the video surveillance system. The experiments are carried out in MuHAVi and MORPH databases. It proves the proposed STFOM-FD-PIOLPP-STASM face recognition method has better accuracy, precision, and recall and the proposed CNN-AE and CNN-GE age and gender estimation method has better accuracy than the other method.

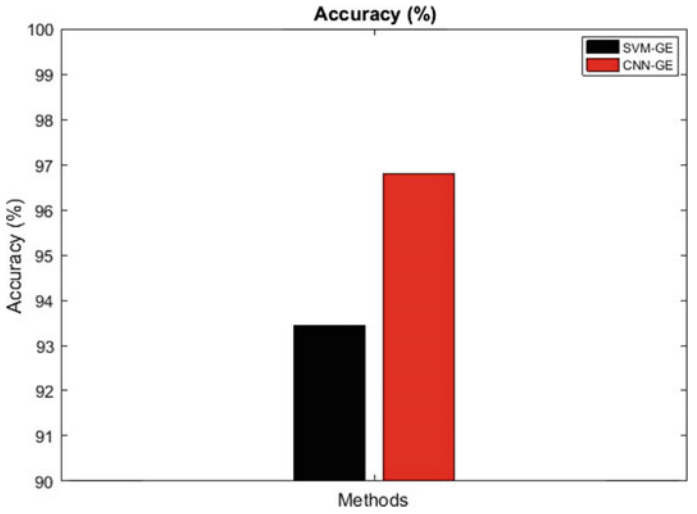


Fig. 5 Comparison of accuracy for gender estimation

References

1. Y. Cong, J. Yuan, J. Liu, Abnormal event detection in crowded scenes using sparse representation. Pattern Recogn. **46**(7), 1851–1864 (2013)
2. C. Liu, The development trend of evaluating face-recognition technology, in *2014 International Conference on Mechatronics and Control (ICMC)* (IEEE, 2014), pp. 1540–1544
3. D. Manju, V. Radha, A novel approach for pose invariant face recognition in surveillance videos. Procedia Comput. Sci. J. **167**, 890–899 (2020)
4. P. Liu, Y. Wang, D. Huang, Z. Zhang, L. Chen, Learning the spherical harmonic features for 3-D face recognition. IEEE Trans. Image Process. **22**(3), 914–925 (2012)
5. C. Ding, D. Tao, Trunk-branch ensemble convolutional neural networks for video-based face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 1002–1014 (2017)
6. F. Mokhayeri, E. Granger, G.A. Bilodeau, Domain-specific face synthesis for video face recognition from a single sample per person. IEEE Trans. Inf. Forensics Secur. **14**(3), 757–772 (2018)
7. H. Zheng, X. Geng, D. Tao, Z. Jin, A multi-task model for simultaneous face identification and facial expression recognition. Neurocomputing **171**, 515–523 (2016)
8. H.Y. Patil, A.G. Kothari, K.M. Bhurchandi, Expression invariant face recognition using local binary patterns and contourlet transform. Optik-Int. J. Light Electron Opt. **127**(5), 2670–2678 (2016)
9. W. Yang, Z. Wang, B. Zhang, Face recognition using adaptive local ternary patterns method. Neurocomputing **213**, 183–190 (2016)
10. J. Han, W. Wang, S. Karaoglu, W. Zeng, T. Gevers, Pose invariant age estimation of face images in the wild. Comput. Vis. Image Underst. **202** (2021)
11. R. Rothe, R. Timofte, L. Van Gool, Deep expectation of real and apparent age from a single image without facial landmarks. Int. J. Comput. Vis. **126**, 144–157 (2018)
12. N. Hao, J. Yang, H. Liao, W. Dai, A unified factors analysis framework for discriminative feature extraction and object recognition. Math. Probl. Eng. 1–12 (2016)
13. H. Liao, Y. Yan, W. Dai, P. Fan, Age estimation of face images based on CNN and divide-and-rule strategy. Math. Probl. Eng. 1–8 (2018)

14. A.L. Ingole, K.J. Karande, Automatic age estimation from face images using facial features, in *IEEE Global Conference on Wireless Computing and Networking (GCWCN)* (2018)
15. C.L. Ku, C.H. Chiou, Z.Y. Gao, Y.J. Tsai, C.S. Fuh, Age and gender estimation using multiple-image features, in *Chinese Conference on Biometric Recognition* (Springer, Cham, 2013), pp. 441–448

Finding Optimal Path for Gas Pipeline Using GIS and RS

Sahil Sawant and Suraj Sawant

Abstract Finding an optimum route for laying natural gas pipelines is a geospatial problem involving decision-making based on the multiple-criterion evaluation technique and computer vision techniques. This study involves finding an optimum route for the gas pipeline from Mangalore to Kochi. Influential features were selected and added into the model of the analytic hierarchy process (AHP). By using random forest (RF) classifier, Sentinel-2 imagery of the study area was classified. The classified image was used in ArcMap 10.7.1 with other geospatial features such as railways, rivers, and roads to produce a route. The results of this study revealed that the land cover and terrain of the region played a significant role in influencing the path of the pipeline as they have a considerable influence on the completion and operating cost of the project. Route 2 is selected to be optimum as it went through minimal crossings of roads and railways with a lesser elevation of the ground.

Keywords Remote sensing · Geographic information system · Random forest · Analytic hierarchy process · Least-cost path analysis · Multi-spectral satellite imaging · Sentinel-2

1 Introduction

Remote sensing (RS) is a field in which we collect data related to an entity involving a process without physically accessing or capturing it with the help of sensors. The data is used to gain valuable information and make new inferences about the object [1]. Satellite imagery is collected using remote sensing technologies. Geographical

S. Sawant (✉) · S. Sawant
Department of Computer Engineering, College of Engineering Pune, Pune, India
e-mail: sawantsj19.comp@coep.ac.in

S. Sawant
e-mail: sts.comp@coep.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_31

information system (GIS) contains geographic information such as roads, rivers, railways, places, and landmarks in the form of a database. For many decades, these satellites observing the earth's surface have been used in various applications. Computer vision has many applications in the geospatial domain [2, 3]. Advancement in thermal multi-spectral systems and imaging spectrometers has made it possible to capture detailed spectroscopic data of physical information of the surface of the earth. Interferometric systems study dynamic processes such as cyclones, thunderstorms, landslides, and changes in sensitive areas such as country borders [4]. We can now create a three-dimensional and very high-resolution model of the earth's surface, which is possible using synthetic aperture radar (SAR) interferometry, laser altimetry, and high-resolution imaging [5]. With the rise of several multi-sensor missions comes an advanced age of image and other data collection, creating a prospect of combining data of high level of heterogeneity from various sensor systems.

The field of RS has seen an exponential increase in the application of artificial intelligence (AI) techniques. At the same time, there is a short supply of literature covering these two broad areas of RS and AI [6]. When the information captured in the data is carefully chosen and effectively utilized in the applications, knowledge is added to the field earth observation (EO) [7]. There is a steady growth in EO products and the improvement of their quality and various other technological parameters [8].

The fusion of GIS with AI has created a new discipline known as geospatial AI (geoAI) [9]. A recent study shows that 80% of data can be georeferenced [10]. Thus, AI is required in GIS to extract meaningful patterns from such a significant share of data.

The process of developing a pipeline is as old as the oil industry. As time passes, the complexity of pipeline projects is increasing due to the continuous addition of factors affecting its implementation. At the heart of pipeline, routing lies the rational reasoning for selecting a single route out of numerous alternatives. The path chosen must be morally and legally defensible to shareholders affected by it during the construction and operation of the project. Land use of the area where the pipeline is constructed plays a substantial part in the project's planning process. State-of-the-art algorithms for classification such as artificial neural networks (ANNs), support vector machines (SVMs), random forest (RF), and classification and regression trees (CART) [11] were previously selected to avoid the problem of expert knowledge avoidance.

The project is divided into three logical parts. The first part deals with processing raw data to make it suitable for further use and classification of different regions in satellite images according to its use. This classification is known as land use/land cover (LULC). The second logical part consists of deriving relative priorities on absolute scales for various factors specific to the project at hand. These factors affect the path carved by the path-finding algorithm. The last part is made of finding the best way to connect two locations. This path is generated using the priorities and weights computed in the second part.

2 Literature Review

GIS is a computer system designed for collecting, saving, transforming, examining, managing, and presenting all kinds of spatial information. Back in 1985, GIS was used to plan a structure, a study by Sperry and Smail [12]. Here, GIS was used for assessing the environmental impact of a high-level repository for managing nuclear waste [13]. GIS applications have grown from inventory and map drawing systems to ones that support methods of analysis [14]. For implementing a pipeline route selection method, given the massive quantity of data required for a project [15], GIS technology works very well. The analytic hierarchy process (AHP) plays a big part in the decision-making aspect of constructing a pipeline route which was introduced into the GIS [16]. AHP is a commonly used optimization technique in GIS [17].

Traditionally, pipeline routes were selected using low-resolution data sets in the early stages of prospective development. There are continuous geographical changes happening which play a predominant part in the selection of pipeline route. These changes cannot be shown in traditional maps. Thus, there is a need for collecting and processing high-quality data for efficient results [18]. The high amount of data processing and efficient interpretation is only possible with high-quality data. The presence of high-quality data promises pipeline route decisions with a better degree of confidence, thus leading to the project's success. Satellite RS images can give new information timely because of qualities such as macro-prospect, direct-perceptibility, and precision. These advantages can assist route selection of pipelines more effectively if they are fully developed. Other than satellite images, we can use aerial images [19]. Aerial images are preliminarily used for mapping, whereas satellite images are used for mapping, environmental monitoring, archaeological surveys, and weather prediction.

Thus, satellite images are found to be suitable for pipeline route selection. Satellite images can be separated into three categories using resolution: low, medium, and high. Medium resolution data is highly used in pipeline route selection, including satellite images like NASA's Landsat 8 and ESA's Sentinel-2 data. Sentinel-2A bands have variable resolutions. The bands B2, B3, B4, and B8 have 10 m of spatial resolution per pixel, while B5, B6, B7, B8A, B11, and B12 have a 20-m spatial resolution per pixels and 60 m of spatial resolution per pixel for other bands. Processing Sentinel data becomes a tedious job. The Landsat Thematic Mapper (TM) images comprise of seven 30 m of spatial resolution spectral bands. These are numbered from 1 to 5 and 7. Spatial resolution for TM is re-sampled to 30-m pixels for band 6 but originally has a spatial resolution of 120 m.

Most of the researches in GIS are done on Landsat 8 [20]. According to Astola [21], Sentinel-2 predicted the forest area better than Landsat 8 and recommended Sentinel-2 multi-spectral instrument (MSI) imagery as a better source for earth observation. Therefore, we have selected Sentinel-2 as our source of data.

Unsupervised methods of classification, such as ISODATA and the K-Means, are unappealing to the RS community. The main reason behind their dislike is that they do not take into account domain/expert knowledge to be inputted into the LULC map generation process. One example of the above is cluster labeling. For classifying multi-spectral and hyper-spectral images, RF and SVM algorithms are widely used in the RS community [22] for classification tasks. These classifiers have better tolerance than traditional parametric approaches when dealing with factors such as distortion and noise and unbalanced data sets of both types discrete/continuous [23]. The accuracy claimed by the models still remains a very disputable issue, and the research in LULC is still going on a large scale [24, 25]. Researchers are continuously trying to improve the predictive accuracy of LULC models with more robust models and techniques.

Different factors such as land cover, railways, rivers, and roads require assigning appropriate weights and priorities specific to a given project. These results influence the outcome of our work in a significant way. This ranking of factors is essential to produce the best possible path for laying gas pipelines [26]. These factors, which are conflicting in nature, are carefully selected by domain experts. This ranking is done after a thorough study and analysis of the region and with help from relevant literature. Concerned people and authorities are consulted to understand their priorities regarding their land usage and their culture. There are various techniques [27–29] for estimating the weights and ranks priorities of these factors when we have such a big project to implement. Few notable techniques are Delphi process [30], scoring [31]; AHP [32–34]; analytic network process (ANP) [35]. There has been much utilization (AHP) to solve many problems related to path-finding and routing [27, 36, 37].

Least-cost path analysis (LCPA) algorithm in GIS is widely used to select suitable patches of land for building linear infrastructures, for finding out the optimal path starting from the source point to sink point over an estimated friction or cost surface (raster maps) [38, 39]. LCPA algorithm is grid-based. It builds on the concept of the shortest path-finding algorithm. LCPA algorithm decides on various criteria which are given related to the pipeline routing. This is done by evaluating the significance (weights) of each factor relative to each other [17]. Afterward, it identifies routes in the middle of two geographical points within the corridors of interest. The most used LCP technique is Dijkstra's algorithm. This algorithm generates the least-cost path between two nodes in a graph structure consisting of weighted links with no negative cycles. These LCP models can be implemented with the help of GIS tools like ArcGIS, ERDAS imagine, etc. LCP has two requirements which include "cost distance surface" and "cost distance backlink surface." The least-cost path will provide an idea of a pipeline route or corridor where potential pipelines could be constructed.

Over the last twenty years, the pipeline community has undertaken considerable efforts to promote the use of GIS for pipeline route selection/alignment. Still, a review of related papers suggested that the technique of automating the route planning process using GIS is still under the exploratory stage [28].

3 Study Area

The study area taken for this research belongs to the state of Kerala in India. A diagram showing its location is present on Fig. 1. It is a region between Kochi and Kozhikode with cloudy weather for most of the year. Kochi, which lies in district Ernakulam of Kerala, is an important port city. Kozhikode, nicknamed as "City of Spices," is a city in Kerala. The study area has a mountainous terrain with dense forests.

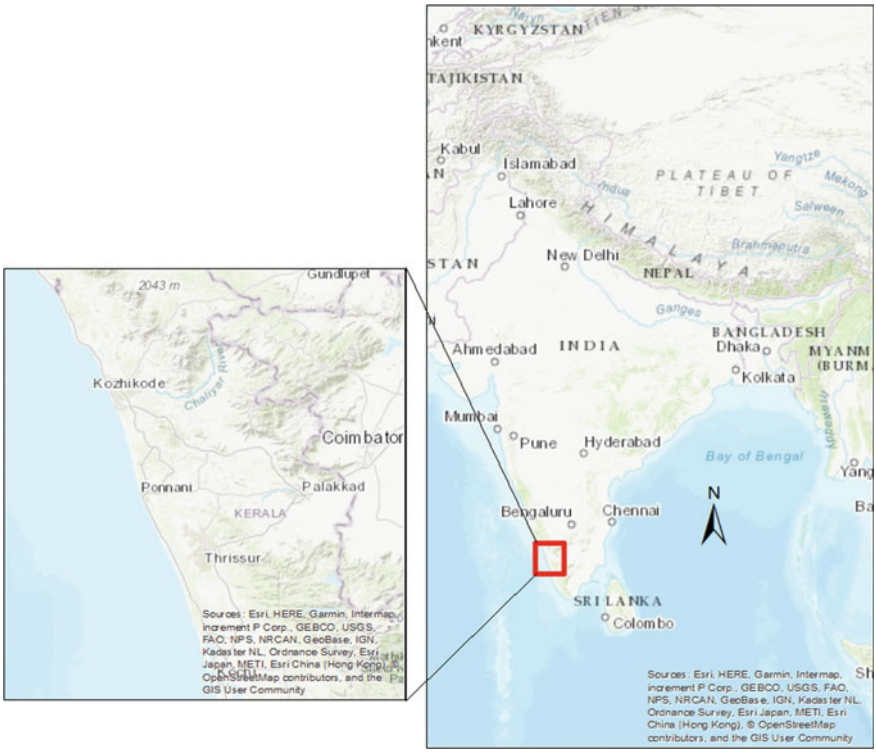


Fig. 1 Location of study area

4 Materials and Methodology

4.1 Data Sets

The project plan contains the consolidation of various data. Data sets used include rivers, railways, roads, digital elevation model (DEM) from Global 30 Arc-Second Elevation (GTOPO30), and multi-spectral images from Sentinel-2 mission.

4.2 Workflow

Following the approach used in published papers [28, 35, 40], criteria and sub-criteria were selected. The AHP model is used by carefully ranking the routing factors for structuring the problem statement into hierarchies. The criteria are grouped into clusters within which sub-criteria are stored.

Geographic-multi-criteria decision-making (G-MCDM) defines geographic factors used to find the optimum route. The factors for obtaining optimal path are land cover, railways, roads, digital elevation model, and rivers. The factors were divided into sub-factors. Land cover was divided into water, agriculture, barren land, dense vegetation, urban area, vegetation. The slope was further divided into ten classes. There is no division of roads, rivers, and railways. Further, it should be taken into consideration that the criteria that could be selected are incomplete. It was noted in the previously published work on this topic by Wang et al. [41]. Their study was about factors that contributed to the corrosion of pipelines. It was based on factors related to geology and soil, geological, and surface hazards.

For the data processing, part Sentinel-2 imagery was upsampled to a common resolution of 10 m per pixel. A Sentinel-2 image consists of 13 bands; using all the bands would not be computationally feasible. Bands 2, 3, 4, 8, 8A, 11, and 12 were chosen for study using optimum index factor (OIF). OIF gives a numerical value depicting variance between all combinations of 3 bands. The combinations with the highest variance were chosen. The bands chosen were stacked to and mosaicked. The mosaicked image was then clipped to produce the study area (Fig. 2).

The study area was then classified using RF classifier. The satellite imagery was classified into five classes water, urban area, dense vegetation, vegetation, agriculture, and barren land. DEM for the study was fetched from the USGS Website, and a slope model was created from it.

The route-finding model is formulated on various optimization methods such as suitability analysis, proximity analysis, distance analysis, and least-cost path analysis.

The DEM image is classified in ArcGIS using the slope function. The slope function produces an image with different pixel values for ranges. The image is then classified reclassify tool. The output from reclassify tool along with reclassified land cover image, railway, rivers, and roads is fed to weighted overlay. The output of

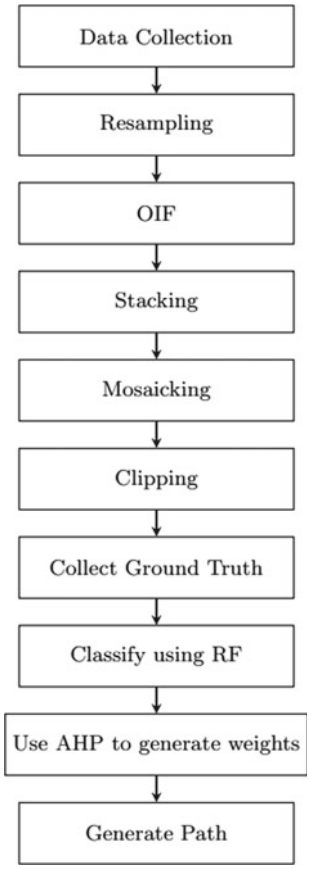


Fig. 2 Framework for the study

weighted overlay and destination is fed to cost distance tool to produce cost surface and backlink. The output of cost distance and origin point is given to produce the final output, i.e., a path which has the least cost according to the given weights (Fig. 3).

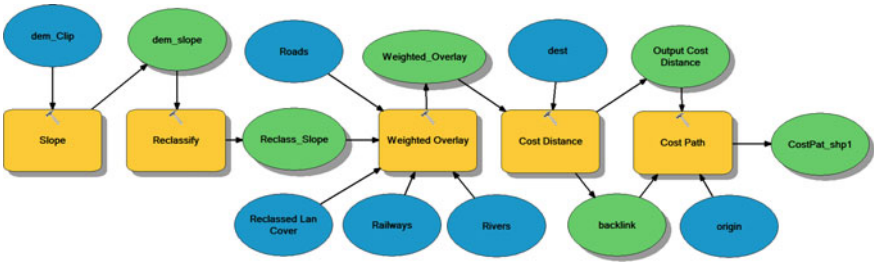


Fig. 3 ArcMap model to obtain path

5 Results

The results are divided into two parts: weights using AHP, GIS model. Table 1 shows the confusion matrix of land cover classification.

1. Weights using AHP

The weights of factors found using AHP are shown in Table 2. The weights of sub-factors of land cover and slope are shown in Tables 3 and 4, respectively. From the results in the table, land cover has the highest priority of over 37.18%. This is in agreement with the recent results [28, 29]. The slope has the second-highest priority of 29.18%. Both land cover and slope have somewhat the same weightage. Under land cover, dense vegetation has the highest weightage, which follows the recent awareness regarding climate change and wildlife protection.

The next important criteria are railways and roads with weightage of 15.17% and 10.43%. The high priority of railways is due to bureaucratic hurdles and engineering costs associated with them. Weightage of roads is for similar reasons. The rivers have the lowest weight of 8.04% since the study area has many rivers and giving it higher importance will produce sub-optimal results.

2. GIS Model

The GIS model includes land cover area. Figure 4a shows land classification obtained from Sentinel-2 imageries.

Table 1 shows the confusion matrix of land cover classification. The overall accuracy is 98.128495%, and Kappa coefficient is 0.94. The maps for railway, rivers, and roads are provided for reference in Fig. 4b (Tables 2 and 3).

6 Conclusion

An observation could be made that the study overlooks the economic aspect of the project. The focus of the study is to produce a path that undertakes environmental, engineering, and land cover factors into consideration. This prioritization ensures that the local shareholders' wishes are respected and reduces conflict. It will eventually lead to the timely completion of the project, thus leading to savings. These results are in acceptance with the previous studies. Disregarding the wishes of local communities and destruction of settlements should be avoided as it might endanger the safety of pipes and lead to cost escalation.

Earlier studies avoided the inclusion of RS images [28], but our results suggest that the classification of Sentinel-2 imageries positively influences the routing process. It corroborates with the use of integrated approaches in finding the optimal path. Such studies have been conducted in Europe, but the current research shows a successful study in India. The findings of this study provide advancement to the existing literature on the logical prioritization of factors and sub-factors. This study will shed

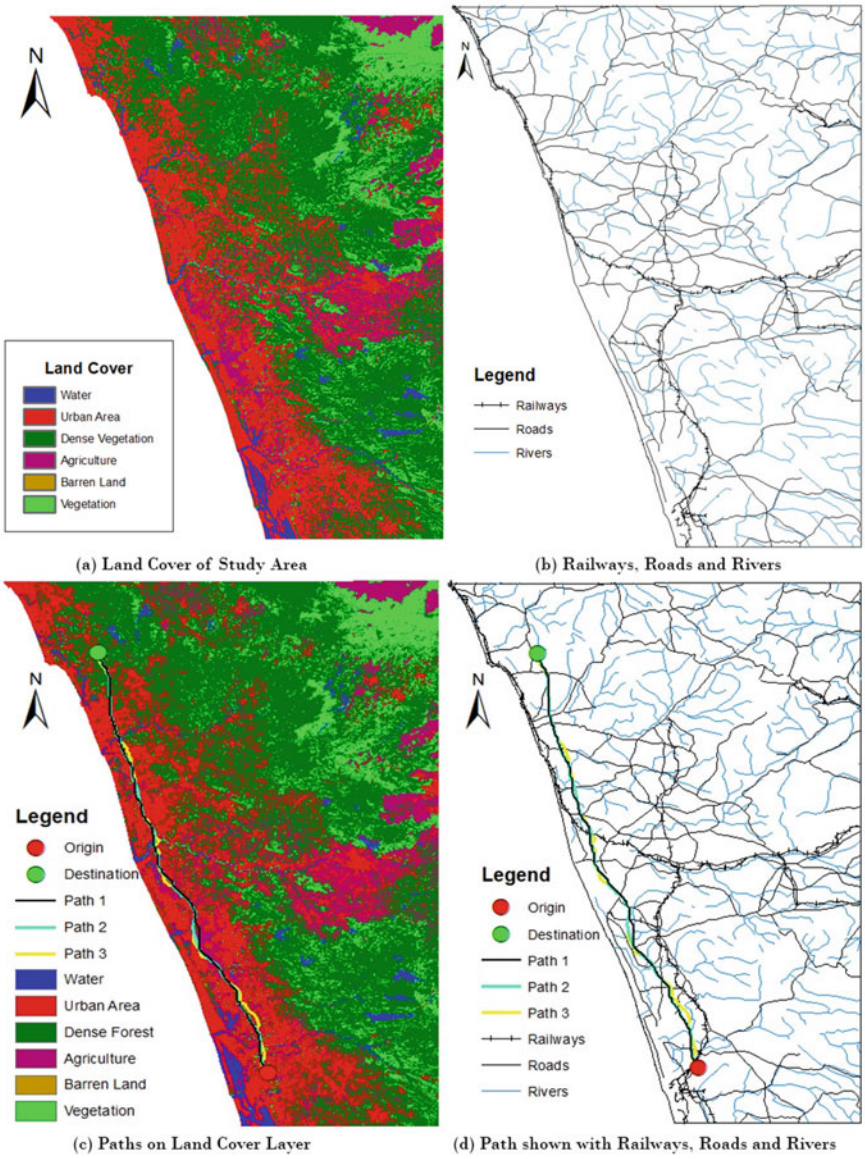


Fig. 4 Resulting maps

light on the potentials of integrated GIS, RS, and ANP systems for the pipeline route selection process (Tables 4 and 5).

Future research on this topic could focus on using a weighting technique that supports cyclic property and uses a classification algorithm that is not biased toward a minority class.

Table 1 Confusion matrix

Classes	Water	Urban area	Dense vegetation	Agriculture	Barren land	Vegetation	All
Water	4278	0	16	0	0	0	4294
Urban area	0	2413	49	0	0	0	2462
Dense vegetation	0	77	3228	0	0	102	3407
Agriculture	0	0	0	2254	53	42	2349
Barren land	0	0	0	134	2957	32	3123
Vegetation	0	206	247	34	187	2749	3423
All	4278	2696	3540	2422	3197	2925	19,058

Table 2 Weights for factors

Factors	Weights
Land cover	37.18
Slope	29.18
Railways	15.17
Roads	10.43
Rivers	8.04

Table 3 Weights for sub-factors of land cover

Factor	Sub-factor	Weights
Land cover	Water	7.24
	Urban area	8.19
	Dense vegetation	10.43
	Agriculture	5.92
	Barren land	1.92
	Vegetation	3.45

Table 4 Weights for sub-factors of slope

	Sub-factors	Weights
Slope		29.18
	0–2.06	7.29
	2.06–5.15	5.94
	5.15–9.07	4.23
	9.07–13.19	3.97
	13.19–17.51	2.91
	17.51–22.05	1.84
	22.05–27.00	1.19
	27.00–33.18	0.91
	33.18–52.56	0.54
	52.56–	0.36

Table 5 Paths generated

Path	Railway	Roads	Rivers	Distance (km)
1	4	17	8	183.12
2	4	16	8	182.10
3	4	14	10	189.66

Acknowledgements We are thankful to the Department of Computer Engineering and IT for providing the high computing GPU server facility procured under TEQIP-III (A World Bank project).

References

1. J.B. Campbell, R.H. Wynne, *Introduction to Remote Sensing* (Guilford Press, New York, 2011)
2. M. Serrão, S. Shahrabadi, M. Moreno, J.T. José, J.I. Rodrigues, J.M.F. Rodrigues, J.M.H. du Buf, Computer vision and GIS for the navigation of blind persons in buildings, in *Universal Access in the Information Society*, vol. 14 (2014), p. 1
3. S. Ardeshir, A.R. Zamir, A. Torroella, M. Shah, GIS-assisted object detection and geospatial localization, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) [Internet] (Springer, Cham) (2014) [cited 2021 July 24], pp. 602–617
4. S.-E. Qian, Hyperspectral satellites, evolution and development history. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. **14** (2021)
5. H. Jung, Z. Lu, A. Shepherd, T. Wright, Simulation of the SuperSAR multi-azimuth synthetic aperture radar imaging system for precise measurement of three-dimensional earth surface displacement. IEEE Trans. Geosci. Remote Sens. **53**(11), 6196–6206 (2015)
6. M.H. Fallahnejad, Delay causes in Iran gas pipeline projects. Int. J. Project Manage. **31**(1), 136–146 (2013)

7. C. Corbane, M. Pesaresi, P. Politis, V. Syrris, A.J. Florczyk, P. Soille, L. Maffenini, A. Burger, V. Vasilev, D. Rodriguez, F. Sabo, L. Dijkstra, T. Kemper, Big earth data analytics on Sentinel-1 and Landsat imagery in support to global human settlements mapping. Big Earth Data **1**(1–2), 118–144 (2017)
8. M. Sudmanns, D. Tiede, S. Lang, H. Bergstedt, G. Trost, H. Augustin, A. Baraldi, T. Blaschke, Big earth data: disruptive changes in earth observation data management and analysis. Int. J. Digital Earth **13**(7), 832–850 (2019)
9. T. VoPham, J.E. Hart, F. Laden, Y.-Y. Chiang, Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. Environ. Health **17**, 1 (2018)
10. S. Li, S. Dragicevic, F.A. Castro, M. Sester, S. Winter, A. Coltekin, C. Pettit, B. Jiang, J. Haworth, A. Stein, Geospatial big data handling theory and methods: a review and research challenges. ISPRS J. Photogramm. Remote. Sens. **115**, 119–133 (2016)
11. M.T.U. Rahman, F. Tabassum, M. Rasheduzzaman, H. Saba, L. Sarkar, J. Ferdous, S.Z. Uddin, A.Z.M. Zahedul Islam, Temporal dynamics of land use/land cover change and its prediction using CA-ANN model for southwestern coastal Bangladesh. Environ. Monit. Assess. **189**(11), 1–18 (2017)
12. S.L. Sperry, H.E. Smail, The geographical information system process and its application to environmental assessment of a high-level nuclear waste repository, in *Proceedings of the Workshop on Geographic Information Systems in Government*, Springfield, VA (1985), pp. 657–676
13. P. Jankowski, L. Richard, Integration of GIS-based suitability analysis and multicriteria evaluation in a spatial decision support system for route selection. Environ. Plann. B Plann. Des. **21**(3), 323–340 (1994)
14. A. Bade, W.A. Mackaness, GIS as a spatial decision support system for offshore pipeline route optimisation. Int. Hydrogr. Rev. **3**(2) (2002). https://journals.lib.unb.ca/index.php/ihr/article/view/20592.
15. J. Luettinger, T. Clark, Geographic information system-based pipeline route selection process. J. Water Resour. Plann. Manag. **131**(3), 193–200 (2005)
16. J. Wan, G. Qi, Z. Zeng, S. Sun, The application of AHP in oil and gas pipeline route selection, in *Proceedings—2011 19th International Conference on Geoinformatics, Geoinformatics* (2011)
17. I.A. Hamid-Mosaku, O.F. Oguntade, V.I. Ifeanyi, A.L. Balogun, O.A. Jimoh, Evolving a comprehensive geomatics multi-criteria evaluation index model for optimal pipeline route selection. Struct. Infrastruct. Eng. **16**(10), 1382–1396 (2020)
18. M. Iqbal, F. Sattar, M. Nawaz, Planning a least cost gas pipeline route a GIS & SDSS integration approach, in *2006 International Conference on Advances in Space Technologies, ICAST* (2006)
19. A.İ. Durmaz, E.Ö. Ünal, C.C. Aydın, Automatic pipeline route design with multi-criteria evaluation based on least-cost path analysis and line-based cartographic simplification: a case study of the MUS project in Turkey. ISPRS Int. J. Geo-Inf. **8**(4), 173 (2019)
20. A. Waldock, D. Corne, Multiple objective optimisation applied to route planning, in *Genetic and Evolutionary Computation Conference, GECCO'11* [Internet] (ACM Press, New York, 2011), pp. 1827–1834
21. H. Astola, T. Häme, L. Sirro, M. Molinier, J. Kilpi, Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. Remote Sens. Environ. **223**, 257–273 (2019)
22. M. Belgiu, L. Drăgu, Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. **114**, 24–31 (2016)
23. L. Breiman, Random forests. Mach. Learn. **45**(1), 5–32 (2001)
24. M.D. Islam, K.S. Islam, M. Mia, An XGBoost Based Approach for Urban Land Use and Land Cover Change Modelling. Authorea. (2021)
25. V.L. Sivakumar, M. Nallanathel, M. Ramalakshmi, V. Golla, Optimal Route Selection for the Transmission of Natural Gas through Pipelines in Tiruchengode Taluk Using GIS – A Preliminary Study. Materials Today: Proceedings, **50**, 576–581 (2022)

26. R.H.G. de Jesus, M.V. Barros, R. Salvador, J.T. de Souza, C.M. Piekarski, A.C. de Francisco, Forming clusters based on strategic partnerships and circular economy for biogas production: a GIS analysis for optimal location. Biomass Bioenerg. **150**, 106097 (2021)
27. S. Bagli, D. Geneletti, F. Orsi, Routeing of power lines through least-cost path analysis and multicriteria evaluation to minimise environmental impacts. Environ. Impact Assess. Rev. **31**(3), 234–239 (2011)
28. A.L. Balogun, A.N. Matori, A.I. Hamid-Mosaku, A fuzzy multi-criteria decision support system for evaluating subsea oil pipeline routing criteria in East Malaysia. Environ. Earth Sci. **74**(6), 4875–4884 (2015)
29. A.L. Balogun, A.N. Matori, A.I. Hamid-Mosaku, D. Umar Lawal, I. Ahmed Chandio, Fuzzy MCDM-based GIS model for subsea oil pipeline route optimization: an integrated approach. Mar. Georesour. Geotech. **35**(7), 961–969 (2017)
30. C.-C. Hsu, B. Sandford, The Delphi technique: making sense of consensus. Pract. Assess. Res. Eval. **12**(1) (2019)
31. G. Mavrommati, C. Richardson, Experts' evaluation of concepts of ecologically sustainable development applied to coastal ecosystems. Ocean Coast. Manag. **69**, 27–34 (2012)
32. X. Delgado-Galván, J. Izquierdo, J. Benítez, R. Pérez-García, Joint stakeholder decision-making on the management of the Silao-Romita aquifer using AHP. Environ. Model. Softw. **51**, 310–322 (2014)
33. T.L. Saaty, M. Sodenkamp, Making decisions in hierarchic and network systems. Int. J. Appl. Decis. Sci. **1**(1), 24–79 (2008)
34. J. Yu, J. Wen, Multi-criteria satisfaction assessment of the spatial distribution of urban emergency shelters based on high-precision population estimation. Int. J. Disaster Risk Sci. **7**(4), 413–429 (2016). Available from: www.springer.com/13753
35. M. Reisi, A. Afzali, L. Aye, Applications of analytical hierarchy process (AHP) and analytical network process (ANP) for industrial site selections in Isfahan, Iran. Environ. Earth Sci. **77**(14), 1–13 (2018)
36. A.S. Aguda, J.O. Uyeh, GIS-based pipeline route mapping for water distribution in Obafemi Awolowo University, Ile-Ife, Nigeria. Ife Res. Publ. Geogr. **11**(1), 83–96 (2016)
37. A. Sadeghi-Niaraki, M. Varshosaz, K. Kim, J.J. Jung, Real world representation of a road network for route planning in GIS. Expert Syst. Appl. **38**(10), 11999–12008 (2011)
38. D.M. Atkinson, P. Deadman, D. Dudycha, S. Traynor, Multi-criteria evaluation and least cost path analysis for an arctic all-weather road. Appl. Geogr. **25**(4), 287–307 (2005)
39. H.A. Effat, O.A. Hassan, Designing and evaluation of three alternatives highway routes using the analytical hierarchy process and the least-cost path analysis, application in Sinai Peninsula, Egypt. Egypt. J. Remote Sens. Space Sci. **16**(2), 141–151 (2013)
40. G. Akman, H. Pışkın, Evaluating green performance of suppliers via analytic network process and TOPSIS. J. Ind. Eng. **2013** (2013)
41. H. Wang, A. Yajima, R.Y. Liang, H. Castaneda, A Bayesian model framework for calibrating ultrasonic in-line inspection data and estimating actual external corrosion depth in buried pipeline utilizing a clustering technique. Struct. Saf. **54**, 19–31 (2015)

Automatic Helmet Detection: A Deep Learning Perspective

Alwin J. Avaran and K. S. Arun

Abstract In India, two-wheelers are one of the most common mode of transportation because it is affordable and economical for regular use. It is highly advisable for bike riders to wear a safety helmet to reduce the risk of head injury in case of road accidents. Therefore, the government has taken steps to enforce helmet for two wheeler riders and is the duty of traffic police to implement this rule. However, this increases the manual labor which is tedious and inadequate. To this end, this paper proposes an automated system which detects bike riders without helmets from surveillance videos. Three prevailing object detection models such as YOLOv4, SSD and Faster R-CNN have been used for the task automatic helmet detection and evaluated the corresponding performance. All these models have been compared and contrasted on the basis of execution time and detection accuracy. From the experimental results, it is evident that the YOLOv4 model exhibits higher detection rate and at the same time achieves consistent mAP scores as opposed to the rest of the models, and thus, it can be applied to real-time helmet detection task.

Keywords Deep learning · Helmet identification · CNN

1 Introduction

The common trend of violating traffic rules and other protective regulations such as wearing a safety helmet often shoots up the high risk and lethal rate in road accidents. According to World Health Organization (WHO), approximately 1.3 million people die each year as a result of road traffic crashes. WHO also reported that correct helmet use can lead to a 42% reduction in the risk of fatal injuries and a 69% reduction in the risk of head injuries. In order to develop a good driving culture and also to value lives, people should abide by these rules. Currently, the traffic police and related

A. J. Avaran (✉) · K. S. Arun
Department of Computer Science and Engineering, Amal Jyothi College of Engineering, Koovappally, India
e-mail: alwinjavaran2020@cs.ajce.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_32

authorities are assigned with the duty to enforce law and catch the violators on road. They monitor this by conducting periodic vehicle checking manually. However, this type of manual inspection is tedious and laborious.

Nowadays, in most of the high traffic areas, packed roads and busy streets are under the coverage of surveillance cameras. On that account, detecting traffic violations by using surveillance footage is viable and cost-effective. Nowadays, features based on deep learning models are found to be competent in certain areas such as image classification and retrieval [2, 3, 8], object detection [11], activity recognition [13], etc. More recently, object recognition models based on Convolutional Neural Network (CNN) is drawing more attention. There are several representative deep learning-based object recognition models available in literature such as Faster R-CNN [15], SSD [12] and YOLO v4 [5].

The accuracy and efficiency of each of these model is the basic criterion for choosing a particular model in assisting the helmet detection task. In this work, we are investigating the feasibility of the above mentioned object recognition model for helmet detection and also evaluated their performance with regard to detection accuracy and detection rate to figure out the best candidate to assist automatic helmet detection.

The rest of the paper is structured as follows: Section 2 presents the related work in the field of automatic helmet recognition. The proposed approach for helmet identification is summarized in Sect. 3. The experimental evaluation of the proposed approach is delineated in Sect. 4. Finally, the paper is concluded in Sect. 5.

2 Related Work

A number of automatic helmet detection models have been proposed in the literature. The rest of this section will briefly summarize the state-of-the-art techniques for automatic helmet detection. An automatic helmet detection system for bike riders in Indian scenario was first investigated by Dahia et al. [6]. In this proposed system, they initially tries to isolate bike riders from real-time videos, and then SVM classifier is utilized to check whether the traveler is having helmet or not. The bike riders are isolated using background subtraction approach, and hand-crafted features are then extracted from this isolated regions for helmet detection. In similar fashion, Wonghabut et al. [18] proposed an automatic helmet recognition system based on HAAR features and SVM classifier. Moreover, number plate recognition of the vehicle is also part of their system if the rider is identified as not wearing the helmet.

More recently, Babu et al. [4] proposed a deep learning-based model to identify the violators and to crop the region of license plate so that helmet violation tickets can be issued by the authorities effectively. The proposed model first identify a moving vehicle and then classify these identified vehicles into a four-wheeler or a two-wheeler. The model used for this classification can also be used to detect whether a rider is wearing a helmet or not. Afzal et al. [1] proposed a deep

learning-based framework for helmet detection. The proposed framework was based on Faster R-CNN and was found to be efficient in situations where images are of low resolution, having blurriness and also for back view images.

Kumar et al. [9] introduced a system for identifying bike riders without helmets. In the initial step, the bike rider needs to be detected from the video frame. The image is then converted to grayscale, and background subtraction is used to separate the objects in motion. After that, feature extraction is performed using HOG, SIFT and LBP. Next, a SVM classifier is trained using these features to recognize bike riders. Finally, the upper 25% part of the person is segmented from it to determine whether the rider is using a helmet or not.

Devadiga et al. [7] developed a model to identify bike riders as well as helmets. In their approach, object recognition is done using the COCO model. If no helmet is identified in the area of interest, the number of the bike is extracted and this helps the authorities to generate a dataset of bike riders without helmets. On the other hand, Prajwal et al. [14] proposed a background subtraction-based helmet detection strategy. They employed kNN as classifier for recognizing both motorbikes as well as helmets. Similarly, Waranusast et al. [17] also proposed background subtraction to identify a motorcycle, and subsequently, an SVM is trained using Circular Hough Transform features extracted from the region of interest.

Later on, Li et al. [10] proposed YOLOv2 to detect helmets on the rider. In this model, feature extraction is done using Histogram of Oriented Gradients (HOG) method. In parallel, the rider count is also taken. They demonstrated that CNN-based models are effective in recognizing helmets from surveillance videos and its localization. In real-life scenario, it is not necessary to focus too much on detection accuracy. However, we need to consider detection rate while performing helmet detection in real time, and there are no relevant studies focusing on real-time helmet identification.

3 Proposed System

The proposed helmet detection system consists of the following three phases:

1. Image Preprocessing
2. Vehicle Classification
3. Helmet Detection.

3.1 Image Preprocessing

Images given as input to the deep learning models are expected to be of good quality for easy assessment. However, it is not always the case that such good quality images can be acquired from our normal CCTV and other surveillance cameras installed on

streets. Even though, there happen to be better quality cameras in certain areas, to adapt to the whole system, the quality of the acquired images should be enhanced. Firstly, the noise is removed from the captured images by means of median blur operation. Then, histogram equalization is performed on those resultant images having extra brightness or dark background.

3.2 Vehicle Classification

Once preprocessing is done, the proposed model initially classify the vehicles in the input image. For this, we have used three different CNN-based deep learning-based object recognition models. Almost the same number of images was gathered for both the classes, two wheelers and four-wheelers. We eliminated the problem of class imbalance by using equal no of cases for both two-wheelers and four-wheelers which in then leads to better performance of the classifier, giving better results. A brief description of the object recognition models employed in the proposed model is provided below:

Faster R-CNN: The Faster R-CNN model unites region proposal for region of interest segmentation and object of interest detection by means of elevated CNNs. The Faster R-CNN model works as follows:

- Initially, numerous candidate region proposals are obtained in the input image by applying the selective search algorithm. After that, a greedy strategy is applied to merge homogeneous regions, and these combined regions form the final region proposals.
- Then, the elevated CNN unit extracts invariant features from the region proposals.
- Finally, the elevated CNN-based features are fed to a SVM classifier for object detection.

Single Shot Detector (SSD): The SSD model employs anchor boxes defined at various aspect ratios and scales for faster computation. The SSD architecture involves two major modules. The first one, the base network, is used for high quality image classification. The second module consists of a set of stacked convolution layers with the size of these layers decline progressively. Each of these convolution layers generate predictions. Since all these layers are at different resolution, they can generate multiple scale predictions.

YOLOv4: The YOLOv4 architecture was conceived on the fact that humans can detect object at single glance. Therefore, object detection can be formulated as a pure regression problem in such way that we can predict the coordinate values of bounding boxes and the associated class labels with the same network.

3.3 Helmet Detection

In this phase, the above mentioned models for vehicle classification are also employed for predicting whether the traveler put on helmet or not. The images used to train these models were acquired through Kaggle and, in the selected dataset, class balance is achieved by maintaining the number of images having riders with helmet equal to the number of images having riders without helmet.

4 Experimental Evaluation

This section depicts the dataset used in our experiments, the performance measures used for evaluation and a detailed discussion of the results obtained with the proposed strategy for the task of helmet detection.

4.1 Dataset Used

To evaluate the state-of-the-art object detection models for the task of helmet detection, we used the SHWD dataset [16]. This dataset contains 9094 helmet wearing instances and 11514 non-helmet wearing instances. The experimental evaluation is performed on a PC workstation equipped with Intel Core i7-6050X CPU, GTX 1080 Ti graphics card and 8 GB RAM.

4.2 Performance Measures

In our experiments, the evaluation results are summarized in terms of precision (pr), recall (re), $F1$ score, mean average precision (mAP) and frames per second (FPS). Here, frames per seconds denote how many frames the proposed model can process in one seconds. The rest of the values can be computed using the following equations:

$$\text{pr} = \frac{m}{m+n} \tag{1}$$

$$\text{re} = \frac{n}{t+n} \tag{2}$$

$$\text{mAP} = \frac{1}{2}\sum_{i=1}^{1} \text{AP}_i \tag{3}$$

where m is the number of true positive, n is the number of false positives and t denotes the number of false negatives. Also, AP_i denotes the average precision of the i-th category and here we have only two classes.

4.3 Results and Discussion

Initially, the performance of vehicle classification is evaluated. Our objective is to figure out the model having highest accuracy and fastest processing speed. The processing speed is quantified in terms of frames per second (FPS). As the number of training instance is not uniform, the F measure is also calculated together to evaluate the performance. For Faster R-CNN architecture, the classification accuracy is found to be comparatively high as the detection is carried out in stages using the CNN architecture. However, the Faster R-CNN model evaluated to have lower processing speed, and hence, it can not be advisable to real-time applications. On the other hand, the SSD model is found to be relatively faster. However, the SSD model has inferior vehicle recognition accuracy as compared to the rest of the models. In the case of YOLOv4 model, it is found to be slower as compared to SSD, but it can correctly recognize all types of vehicles from each video frame. The experimental results are summarized in Tables 1 and 2. By analyzing this performance summary, it can be inferred that YOLOv4 is the ideal model among the rest of the models more suitable for the task of vehicle detection.

Next, the performance of various models for helmet detection is evaluated. The comparative evaluation of all the selected object recognition model is presented in Table 3. By analyzing this results, it can be seen that highest mAP score is reported for Faster R-CNN as compared to the rest of the models. In mAP values respective

Table 1 Comparative evaluation of various models for vehicle detection

	Two-wheeler			Four-wheeler		
	Faster R-CNN	SSD	YOLOv4	Faster R-CNN	SSD	YOLOv4
mAP	92.37	93.72	98.45	85.47	80.28	98.58
TP	254	261	275	151	166	182
FP	35	23	12	14	11	4

Table 2 Comparative evaluation of various models for helmet detection

	Precision	Recall	mAP	$F1$ score
Faster R-CNN	0.94	0.98	0.97	0.95
SSD	0.89	0.89	0.90	0.89
YOLOv4	0.90	0.94	0.93	0.91

Table 3 Comparative evaluation of the detection rates of various models

	Faster R-CNN	SSD	YOLOv4
FPS	35.22	81.19	104.18

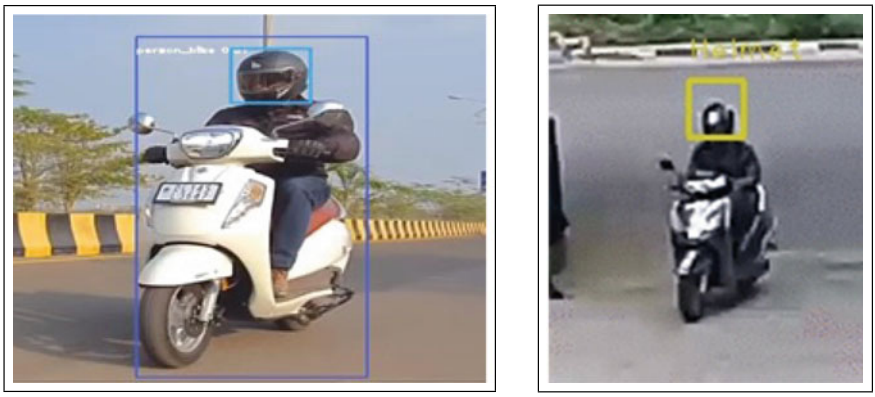


Fig. 1 Sample result obtained with YOLOv4 model for the task of helmet detection

increase of 7% and 4% is reported for Faster R-CNN model as compared with SSD and YOLOv4 model. Similarly, the $F1$ Score of Faster R-CNN model is increased 6% and 4% as compared to those of SSD and YOLOv4 model. From this, it can be concluded that models having two-step processing pipeline has higher detection accuracy in comparison with their counterparts that executes in single stage. On the basis of FPS measure, the YOLOv4, SSD and Faster R-CNN, respectively, detected 104, 81 and 35 frames per second. A sample result set is depicted in Fig. 1.

Hence, if our primary concern is detection efficiency, then YOLOv4 performs best among the three models, and the rest of the models do not fulfill all the demands in a real-time scenario. Therefore, this work draws the following conclusions: when we are only considering detection accuracy, the Faster R-CNN model is more suitable. However, YOLOv4 is more advisable in situations where the prime concern is detection rate and we can tolerate negligible misclassification. Therefore, it can be concluded that YOLOv4 has been an ideal candidate to be applied to assist authorities for the task of automatic helmet detection.

5 Conclusion

This paper investigated the applicability of state-of-the-art CNN-based object detection models for the task of automatic helmet detection. It has been observed that CNN-based deep learning models can derive more discriminative representations for helmet detection, which eliminates false alarms and gives rise to more reliable mod-

els in real-life scenario. It can be inferred that among the most popular CNN-based object detection models, YOLOv4 has the best performance in terms of detection accuracy and speed to be used in real-life applications. However, the detection models used in this work exhibits below par potential if the images obtained from the surveillance cameras contains helmets which are compact, vague and also involves complex background.

References

1. A. Afzal, H.U. Draz, M.Z. Khan, M.U.G. Khan, Automatic helmet violation detection of motorcyclists from surveillance videos using deep learning approaches of computer vision, in *2021 International Conference on Artificial Intelligence (ICAI)* (IEEE, 2021), pp. 252–257
2. K.S. Arun, V.K. Govindan, Optimizing visual dictionaries for effective image retrieval. Int. J. Multimedia Inf. Retrieval **4**(3), 165–185 (2015)
3. K.S. Arun, V.K. Govindan, S.M. Kumar, On integrating re-ranking and rank list fusion techniques for image retrieval. Int. J. Data Sci. Analytics **4**(1), 53–81 (2017)
4. R. Babu, A. Rathee, K. Kalita, M.S. Deo, Helmet detection on two-wheeler riders using machine learning, in *ARSSS International Conference*, vol. 41 (2018), pp. 603–623
5. A. Bochkovskiy, C.Y. Wang, H.Y.M. Liao, Yolov4: optimal speed and accuracy of object detection (2020). arXiv preprint arXiv:2004.10934
6. K. Dahiya, D. Singh, C.K. Mohan, Automatic detection of bike-riders without helmet using surveillance videos in real-time, in *2016 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2016), pp. 3046–3051
7. K. Devadiga, Y. Gujarathi, P. Khanapurkar, S. Joshi, S. Deshpande, K. Devadiga, Y. Gujarathi, P. Khanapurkar, S. Joshi, S. Deshpande, Real time automatic helmet detection of bike riders. Int. J. **4**, 146–148 (2018)
8. J. Joby, R.J. Raju, R. Job, S.M. Thomas, K.S. Arun, Pneumogan: a gan based model for pneumonia detection, in *2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)* (IEEE, 2021), pp. 102–106
9. S. Kumar, N. Neware, A. Jain, D. Swain, P. Singh, Automatic helmet detection in real-time and surveillance video, in *Proceedings of ICM-LIP* (2020), pp. 51–60
10. J. Li, H. Liu, T. Wang, M. Jiang, S. Wang, K. Li, X. Zhao, Safety helmet wearing detection based on image processing and machine learning, in *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)* (IEEE, 2017), pp. 201–205
11. T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 2980–2988
12. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, Ssd: single shot multibox detector, in *European Conference on Computer Vision* (Springer, Berlin, 2016), pp. 21–37
13. L. Pigou, A. Van Den Oord, S. Dieleman, M. Van Herreweghe, J. Dambre, Beyond temporal pooling: recurrence and temporal convolutions for gesture recognition in video. Int. J. Comput. Vis. **126**(2), 430–439 (2018)
14. M. Prajwal, K. Tejas, V. Varshad, M. Mahesh, R. Shashidhar, A review on helmet detection by using image processing and convolutional neural networks. Int. J. Comput. Appl. **182**(50), 52–55 (2019)
15. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. **28**, 91–99 (2015)
16. H. Wang, Z. Hu, Y. Guo, Z. Yang, F. Zhou, P. Xu, A real-time safety helmet wearing detection approach based on csyolov3. Appl. Sci. **10**(19), 6732 (2020)

17. R. Waranusast, N. Bundon, V. Timtong, C. Tangnoi, P. Pattanathaburt, Machine vision techniques for motorcycle safety helmet detection, in *2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013)* (IEEE, 2013), pp. 35–40
18. P. Wonghabut, J. Kumphong, T. Satiennam, R. Ung-Arunyawee, W. Leelapatra, Automatic helmet-wearing detection for law enforcement using cctv cameras, in *IOP Conference Series: Earth and Environmental Science*, vol. 143 (IOP Publishing, 2018), p. 012063

Some Observations on Huffman Trees

K. Viswanathan Iyer, Karthick Seshadri, and Samriddhee Ghosh

Abstract Information theory and coding theory are very important in communication engineering. Data is transferred across the network from source to destination via some communication channel. Data streams can be very lengthy. Moreover, if we consider all characters from 'A' to 'Z' in English, the data stream may or may not contain all possible characters in this character set. Hence, we need not use all the bits used to represent the characters from 'A' to 'Z'. Here, data compression comes into play. Coding theory comes up with various techniques required to create codes for each of the characters in the data stream to represent them with smaller number of bits. Huffman coding is one such important algorithmic technique used to create codes for each of the characters. Apart from compression, they are used in other tasks like cryptography, error detection, and correction. In this paper, our concern is data compression mainly. We recommend to club different techniques like use of special characters, contexts in character stream together with Huffman coding to perform effective compression.

Keywords Huffman tree · Fully skewed Huffman tree · Overlapping Huffman trees · Data compression · Cryptography

1 Introduction

Coding theory deals with the characteristics of codes and their usefulness in various applications. It is used in developing efficient and reliable transmission methods from source to destination. Coding theory is also used for purposes like data compression,

K. Viswanathan Iyer
National Institute of Technology Tiruchirappalli, Tiruchirappalli, India

K. Seshadri (✉) · S. Ghosh
National Institute of Technology Andhra Pradesh, Tadepalligudem, Andhra Pradesh, India
e-mail: karthick.seshadri@nitandhra.ac.in

S. Ghosh
e-mail: mtcs2010@student.nitandhra.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_33

cryptography, error detection, and correction. There are two main types of coding: Data Compression/Source Coding and Error Control/Channel Coding. Characters to be transmitted from source across the channel are represented by bits of 0s and 1s. To make the transmission procedure efficient, we try to reduce the number of bits required to represent each of the characters. For data compression or source coding (source compression), two very important techniques are Huffman coding and Shanon-Fano coding. Huffman coding belongs to the class of lossless compression techniques [1]. The term *lossless* indicates that the original data or character stream can be entirely re-constructed from the encoded data. Huffman coding deals with encoding each of the characters with variable length codes or bit streams. It follows a greedy strategy [2] to generate optimum length codes.

In Huffman coding algorithm, the more frequent character gets the longer length code while the less frequent character gets the shorter length code. For example, if we consider the string "BBBCAABBA", the frequency of the character B is higher than that of A and the character C has the least frequency. So, the length of the code for B is smaller than A, and code for A will be smaller than C. The Huffman codes are represented in the form of a binary Huffman tree. The Huffman tree is transmitted along with the encoded data to the receiver, so that the receiver can reconstruct the original data [3].

In the Huffman tree, the leaf nodes represent the distinct characters and paths from them to the root gives their corresponding codes. We can obtain different Huffman trees for the same character stream (in case there are two intermediate sub-optimal solutions for same character stream). Shanon-Fano coding uses pre-defined word lengths for the codewords of each of the symbols. Unlike Huffman coding, it is not obvious that Shanon Fano will give us codewords of optimal lengths. That is why Huffman coding is always preferred over Shanon Fano coding. A particular scenario called Fano condition can arise in the context of Shanon Fano and Huffman coding. Let's take '*X*' and '*Y*' as two different codewords (represented in binary) used for two different symbols in a character stream. *Fano Condition* denotes the case when another symbol in the character stream has the code '*XY*' [4]; i.e., if '00' is used for '*a*' and '01' is used for '*b*' and '0001' is used for '*c*', then it represents Fano condition. In this paper, we will discuss some significant observations made with respect to Huffman trees.

2 Major Observations

In this section, we describe the major observations along the following headers: (i) fully skewed Huffman trees, (ii) reduced compression techniques using concept of co-occurring characters or contexts, (iii) shortening data stream length with replacement by numeric values, and (iv) overlapping codes with more than one Huffman tree.

2.1 Fully-Skewed Huffman Trees

The Huffman tree creation algorithm is a greedy technique where sub-optimal solutions are clubbed together at the end to get final solution. If at each iteration, the new node obtained at previous iteration is assigned as the left child/right child of the new node created at the current iteration, we get a fully skewed Huffman tree [5]. This procedure is described in Algorithm 1 [6]. Such a structure is considered in Fig. 1.

We will use the set representation as follows. For the Fig. 1, we get

$$Y = (a, X) = (a, X = (b, c)) = (a, (b, c))$$

$H = (l, r)$ denotes the root node of the Huffman tree H, with the left subtree 'l' and the right subtree 'r'.

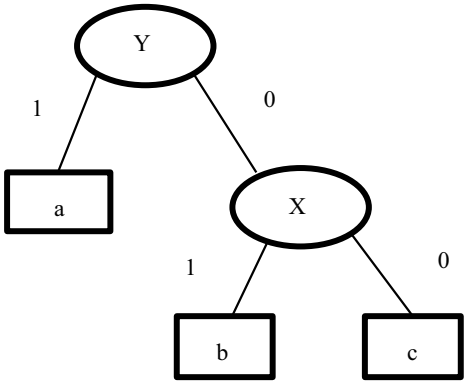


Fig. 1 Fully skewed Huffman tree structure

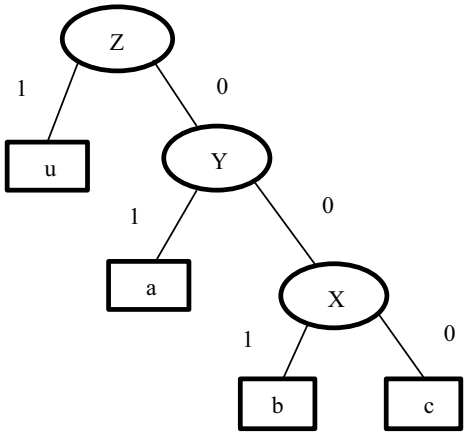


Fig. 2 Fully skewed Huffman tree (by including more characters)

Algorithm 1 Creation of Huffman Tree

1. Creation of a polarity queue for each distinct character
2. Sort all the distinct characters in descending order of frequency
3. For all distinct characters:
   - 3.1 Creation of a new node
   - 3.2 Minimum value is extracted from the queue and assigned to the left child of new node
   - 3.3 Minimum value is extracted from the priority queue and assigned to the right child of the new node
   - 3.4 Sum of these two values is calculated and assigned to the value of the new node
   - 3.5 Insert the new node into the tree
4. return root node

The codes used are as follows $a = 1, b = 10, c = 00$. We get these codes by traversing the path from leaf to the root in Fig. 1. We can interchange b and c (as they have the two least frequencies of occurences, we have to include both of them as the first sub-optimal solution irrespective of which one has greater frequency). We have assumed that

$$f(a) > f(b) > f(c) \tag{1}$$

The condition specified in inequality (1) gives a fully skewed Huffman tree

$$f(b) + f(c)! = f(a) \tag{2}$$

where $f(x)$ = frequency of occurrence of x in the sampled text. If Eq. (2) is not satisfied, then, for more than three characters it will not give us a unique Huffman tree, as it will result in more than one possible solution for a sub-problem in Huffman tree creation procedure [7]. We now assume that we have four alphabets *u*, *a*, *b, and c*, respecting the inequality in Eq. (3). We build the tree as follows

$$f(u) > f(a) > f(b), f(c) \tag{3}$$

Here, 'comma' is used as a and b are the two least occurring characters and we have to combine them together as the first sub-optimal solution irrespective of which one is greater. Now, the tree gets updated as follows by adding new character $Z = (u, Y)$. Z is the new root with the left subtree 'u' and right subtree Y [8]. We need to rename the leaves in Y. The above process can be repeated. It works for cases where the frequencies are all different at all stages from the start. We get a fully skewed tree without running the algorithm. The newly formed Huffman tree is as shown in Fig. 2.

The average path lengths considering the characters only at sub-trees having roots 'X', 'Y', and 'Z' are calculated as follows:

$$\text{AVG}(X) = \frac{\{f(c) + f(b)\}}{L(X)} \tag{4}$$

$$\begin{aligned}
\text{AVG}(Y) &= \frac{\{\{[f(c) + f(b)]*2\} + f(a)\}}{L(Y)} = \frac{\{f(c) + f(b)\} + \{f(c) + f(b) + f(a)\}}{L(Y)} \\
&= \frac{\{f(c) + f(b)\} + L(Y)}{L(Y)} = \frac{f(c) + f(b)}{L(Y)} + 1 \\
&= \frac{L(X)}{L(Y)}\left\{\frac{f(c) + f(b)}{L(X)}\right\} + 1 \\
&= \frac{L(X)}{L(Y)}\text{AVG}(X) + 1
\end{aligned} \tag{5}$$

$$\begin{aligned}
\text{AVG}(Z) &= \frac{\{[f(b) + f(c)] * 3\} + \{f(a) * 2\} + f(u)}{L(Z)} \\
&= \frac{\{\{[f(b) + f(c)] * 2\} + f(a)\}}{L(Z)} + \frac{f(a) + f(b) + f(c) + f(u)}{L(Z)} \\
&= \frac{L(Y)}{L(Z)}\text{AVG}(Y) + 1
\end{aligned} \tag{6}$$

$L(X)$ represents the length of character stream considering only first two characters, $L(Y)$ represents the length for first three characters, $L(Z)$ represents length for first four characters taken in ascending order of occurrence frequencies in the character stream.

From Eqs. (4), (5), and (6), the recurrence relation is formulated as shown in Eq. (7). $L(k)$ represents the length of bit stream considering the first 'k' characters in the non-decreasing order of frequencies.

$$\text{AVG}(k) = \frac{L(k-1)}{L(k)}\,\text{AVG}(k-1) + 1 \tag{7}$$

Solving the above recurrence equation, we get [using Eqs. (1), (2), and (3)]

$$\begin{aligned}
\text{AVG}(k) &= \frac{L(k-2)}{L(k)}\text{AVG}(k-2) + \frac{L(k-1)}{L(k)} + 1 \\
&= \frac{L(2)}{L(k)}\text{AVG}(2) + \frac{L(3)}{L(k)} \\
&\quad + \cdots + \frac{L(k-1)}{L(k)} + 1 = \frac{L(2)}{L(k)} + \frac{L(3)}{L(k)} \\
&\quad + \cdots + \frac{L(k-2)}{L(k)} + \frac{L(k-1)}{L(k)} + 1
\end{aligned}$$

(as AVG(2) = 1 only if two distinct characters are present) (8)

We will have to consider the values of $L(1)$, $L(2)$, …, $L(k-1)$, $L(k)$ which can be obtained from the source character stream. For any arbitrary binary tree with 'n' leaf nodes, we attempt to compute the relation between the average path length and the codes at the leaf nodes [9, 10]. Let l_a be the code length at the a^{th} leaf node.

$$A = \frac{l_1 + l_2 + \cdots + l_n}{n} \tag{9}$$

If c_a is the decimal representation of code at the a^{th} leaf node then, $c_a \leq 2^{l_a} - 1$ (as path length = no. of bits used)

$$\begin{aligned} &\text{i.e.} \quad (c_a + 1) \leq 2^{l_a} \\ &\text{i.e.} \quad \log_2^{c_a+1} \leq l_a \end{aligned} \tag{10}$$

From Eqs. (9) and (10),

$$\begin{aligned} A &\leq \frac{\log_2^{(C_a+1)} + \log_2^{(C_b+1)} + \cdots + \log_2^{(C_n+1)}}{n} \\ A &\leq \frac{\log_2^{\{(C_1+1)(C_2+1)\ldots(C_n+1)\}}}{n} \end{aligned} \tag{11}$$

Thus, from Eq. (11), we can infer that the average path length cannot exceed the average of the logarithms of the code values for all the leaf nodes.

2.2 *Presence of CONTEXTS for Some Characters in the Alphabet Set Based on Semantics*

In this section, we consider the special case when there are two characters p, q with $f(p) = f(q)$ [11]. If we know that p occurs only in the context (a, b) that is "apb" is the only permitted occurrence and q occurs only in the context (a, d) that is "aqd" is the only permitted occurrence, then we can remove say q and retain only p. Here, p and q will be assigned the same code. While decoding if we get a p then we check the context after decoding one more alphabet and change p to q if required as per the context. Thus, the context can be a set of 2-tuples for p and a different set of 2-tuples for q. A table lookup is required. For example, if the contexts for p are "apb", "apc", "apd", "apz", then a 2D array PCON[ ][ ] can be maintained; then we can do a check in PCONN[ ][ ] to find out in which contexts p occurs. Algorithm 2 describes the process of decoding.

Algorithm 2 Decoding with a look-up table

```
if pre-p, post-p in PCON[ ][ ]
    then
        Decode as p
    else if pre-p, post-p in QCON[ ][ ]
    then
        Decode as q
```

In Algorithm 2, "pre-p" and "post-p" refer to the characters occurring before and after p in the specific contexts. We can present the following example as one extreme case where contexts such as the one given or mentioned above will occur naturally. For this, we consider two sets of alphabets SIG1 and SIG2. Here, SIG1 can be lower case English alphabets; SIG2 can be lower case Greek alphabets. In the uncoded (plain) text (source) we specify the following contexts:

Occurrence of English and Greek Alphabets simultaneously: An English alphabet occurs in-between two Greek alphabets and a Greek alphabet occurs in-between two English alphabets. We can consider a situation where character stream consists of alternate sequences of English and Greek alphabets. In other words, we deal with two text files say F1—in English and F2—in Greek. At transmission time, we choose the next (first) character from F1 and F2 alternatively. In such a case, we analyze samples from F1 and F2 separately; we build the Huffman trees T1 and T2 separately for SIG1 and SIG2. During encoding, the table-lookup for SIG1 and SIG2 characters can be from a 2D-array say CODES[ ][ ], where rows represent alphabets and columns 1 and 2 respectively denote the frequencies of characters from SIG1 and SIG2. The trees T1 and T2 can be made to overlap as much as possible [12]. For this, we perform the following—we always take the left subtree with a greater number of nodes at every stage during a merge. Let 'r' be the root of the tree wherein T1 is drawn overlapping with T2. Then at each leaf or may be at an inner node (depending on code lengths for SIG1 and SIG2) we will have some bit fields where one field will indicate if the traversal (based on a global bit SIG1 or SIG2) is with respect to T1 or T2. While decoding we will alternate between the roots of T1 and T2 assuming that we are given the first character from SIG1 or SIG2. Separate Huffman trees are constructed for English and Greek Alphabets. In this scenario, characters are decoded using a look up table at each stage to get the actual sequence of characters. We consider the scenario below:

- a, b, c, d are English alphabets; p, q, r, s are Greek alphabets.
- (a, d) are alphabets appearing in specific contexts of p, and q; $f(a) = f(d)$.
- (p, q) are alphabets appearing in specific contexts of a, b, c, and d; $f(p) = f(q)$.

Table 1 Look-up table

Look-up for Greek (p, q)	Look-up for English (a, b, c, d)
apb	qap
apd	qdq
bqa	
aqc	

Table 2 Successive iterations for decoding

Iteration No.	English	Greek	English	Greek	Right/Wrong	Reason
–	001	001	100	001	–	–
1	b	p			Wrong	'bp' not in Table 1
2	b	q			Right	'bq' in Table 1
3	b	q	d		Wrong	'bqd' not in Table 1
4	b	q	a		Right	'bqa' in Table 1
5	b	q	a	q	Wrong	'qaq' not in Table 1
6	b	q	a	p	Right	'qap' in Table 1

The following look up table (Table 1) is used for two character sets (English and Greek).

The codes for (a, b, c, d) are 100, 001, 000, 100 and code for (p, q) is 001 and it is overlapping with English. We consider some initial part of the data stream. Table 2 shows the iterations for look ups to decode the data stream.

So, "bqap…." is the right sequence. But, this may not be the scenario always. English and Greek may not always occur alternately like this. Moreover, overlapping can be present between the codes of two-character set codes. Thus, it is difficult to distinguish between English and Greek alphabets. A solution to this can be arrived at by adding one more bit to the left of each codeword (0 or 1 will denote whether it is English or Greek).

Use of Contexts in Scientific Articles: Another way to build contexts is by delimiting every equation in a scientific article in English by '$' symbol. The characters in equations can be considered separately to infer relative frequencies for separate Huffman trees [13]. The mathematical equation will contain operators like '+', '−', '*', and '/'. If 'a', 'b', 'c'… lie among these '$' symbols, we can make each of the operators limited to some contexts because an operator is always preceded and succeeded by symbols used as variables. Thus, we can use the same code for two operators having the same frequency and make use of some look up table.

2.3 *Continuous Stream of Single Character Replaced by Numerical Value*

This section deals with replacement of a continuous stream of characters with some numerical value [14]. However, this is applicable in some specific scenario (explained in Ex 1 and Ex 2), where we can get the same character occurring consecutively many times. Let us consider two examples of character streams which will result in fully skewed Huffman trees.

Ex 1: Frequencies of *A*, *B*, *C*, *D*, *E* are 39, 30, 16, 10, 5
Ex 2: Frequencies of *A*, *B*, *C*, *D*, *E* are 85, 8, 4, 2, 1

In these two examples, the frequency of the most frequent character dominates over that of others by a large margin, especially for the second example where the frequency of most frequent character is more than five times the sum of frequencies of other characters. For such a situation, even after encoding the characters by Huffman Coding, large number of bits are required to transfer the entire character stream via a communication channel. To reduce transmission cost, one of the ways can be to replace the continuous stream of the most dominant character by the numerical value of the most dominant frequency followed by the character. For e.g., if we have one character stream of a '*b*', followed by 85 *a*'s and a '*b*', we can write the stream as "b85ab". Earlier, the character stream comprised of only '*a*' and '*b*' and now it comprises of '*a*', '*b*', '8', '5'. All these characters are encoded by Huffman tree. For this stream of characters, we follow the below steps:

- On the text files to be transmitted replace runs of A's with <intx> <A> where intx is an integer. For example, if we see BAAAAAAB we replace it by B6AB.
- Generate the new input text file where SIG′ = union of SIG and {0, 1, …, 9}. Obtain the frequency of occurrences of characters in SIG′.
- Run Huffman coding algorithm with SIG′ and the obtained frequencies.

To reconstruct the original character stream, we perform decoding with respect to SIG′. Whenever a series of numeric characters is encountered, the corresponding decimal values and the next occurring non-numeric character value are calculated to get the original stream. The series of numeric values always appear in between two non-numeric character values. Hence, non-numeric and numeric can be distinguished easily by performing Huffman encoding with the union of the two-character sets. However, this method of achieving a better compression ratio is effective, only if the frequency of occurrence of the most frequent character is several orders of magnitude higher than that of the other characters [15]. If this condition is not satisfied, the data stream will not contain a long, continuous sequence of the same character and this method will not be applicable. For example, instead of 'b85abbb' if we had another stream with same frequencies as 'babababab81a' then, we would still have got a long sequence of continuous character *a*'s since the occurrence of '*a*' is much greater than that of '*b*'. Instead of using one Huffman tree for all the characters (both numeric and non-numeric), we can reduce the bits for code lengths by using two different

Huffman trees. As it can cause overlapping, we can use some special symbol '@' at the beginning and end of numeric character sequences. This character must be assigned some codeword which is not used for any Huffman Tree. In this way, we can make the data more compressed.

3 Conclusion

In this paper, we have attempted to make some observations to comprehend significant properties of Huffman trees using a sound mathematical analysis. Based on the properties, we have tried to utilize Huffman trees in data compression. Different scenarios of lengthy data streams are considered. It is observed that the Huffman trees are used in different ways in different scenarios, namely, Huffman trees in compressing the data by the concept of contexts is much different from that of replacing data stream with numeric values. Our observations are applicable and advantageous in different scenarios. However, we haven't found any generalized use of Huffman tree which can work efficiently in compressing the data stream irrespective of its context. We aim to fulfil this void in our future work.

Acknowledgements The authors would like to thank Minazuddin, Nowseen Pabel and Asiq Rahman of NIT Tiruchirappalli for their discussions and comments and Mrs. K. Sindhu of NIT Andhra Pradesh for her help in formatting the paper.

References

1. M.V. Mahoney, Adaptive weighing of context models for lossless data compression. Technical report CS-2005-16 (Florida Institute of Technology, Melbourne, 2005)
2. A. Shen, *Algorithms and Programming Problems and Solutions*, 2nd edn. (Springer, Berlin, 2010)
3. S.T. Klein, J. Radoszewski, T.C. Serebro, D. Shapira, Optimal skeleton and reduced Huffman trees. Theor. Comput. Sci. **852**, 157–171 (2021)
4. T.J. Richardson, R. Urbanke, *Modern Coding Theory* (Cambridge University Press, Cambridge, 2008)
5. M.A. Weiss, *Data Structures and Algorithm Analysis*, 2nd edn. (Pearson, UK, 2002)
6. J.S. Vitter, Design and analysis of dynamic Huffman codes. J. ACM **34**(4), 825–845 (1987)
7. J.-F. Maurras, T.H. Nguyen, V.H. Nguyen, On the linear description of Huffman trees polytype. Discrete Appl. Math. **164**(1), 225–236 (2014)
8. W. Szpankowski, Average redundancy for known sources: ubiquitous trees in source coding, in *Fifth Colloquium on Mathematics and Computer Science.* Discrete Math. Theor. Comput. Sci. Proc. 19–58 (2008)
9. M. Golin, E. Harb, Speeding up the AIFV-2 dynamic programs by two orders of magnitude using range minimum queries. Theor. Comput. Sci. **865**(14), 99–118 (2021)
10. A. Belal, A. Elmasry, Optimal prefix codes with fewer distinct codeword lengths are faster to construct. Inf. Comput. **268**, 1–14 (2019)

11. D.B. Shmoys, C. Swamy, R. Levi, Facility location with service installation costs, in *15th Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics*, USA (2004), pp. 1088–1097
12. C. Heuberger, D. Krenn, S. Wagner, Analysis of parameters of trees corresponding to Huffman codes and sums of unit fractions, in *Meeting on Analytic Algorithms and Combinatorics* (Society for Industrial and Applied Mathematics, USA, 2013), pp. 33–42
13. O. Manz, Entropy coding, in *Well Packed—Not a Bit Too Much* (essentials) (Springer, Wiesbaden, 2021)
14. C. Elsholtz, C. Heuberger, H. Prodinger, The number of Huffman codes, compact trees, and sums of unit fractions. IEEE Trans. Inf. Theor. **59**(2), 1065–1075 (2013)
15. J.-F. Maurras, T.H. Nguyen, V.H. Nguyen. On the convex hull of Huffman trees. Electron. Notes Discrete Math. **36**, 1009–1016 (2010)

Estimation of Dietary Calories Using Image Processing

Indukuri Harshitha, T. Sunil Kumar, and P. Venkateswara Rao

Abstract According to the WHO, an unhealthy diet is responsible for nearly 20% of all deaths worldwide. Most of the population are living in countries where obesity and overweight cause more fatalities than underweight. The issue here is not a lack of food; rather people are unaware of what is in their diet. Knowing how many calories are in the foods that individuals eat and can assist them maintain their body's health. Knowing how many calories are in the foods that individuals eat, can assist them maintain their body's health by meeting the body's fundamental calorie needs. It would have a variety of beneficial impacts, such as living a healthy lifestyle and providing a suitable amount of energy for regular exercise. Those who do not even worry over their caloric demands, on the other hand, will suffer a variety of health issues, such as obese and increasing ailments such as hypertension and prediabetes. People could easily decide how many calories they want to consume if they could estimate their calorie intake using images of their food. Determining the real quantity of caloric from meal technologically involves the food item's region, size, and weight. Deep learning algorithms can identify the object and calories are estimated based on the object detection method and volume estimation method. If people knew how many calories were in their food, this problem could be mitigated slightly. This is accomplished in three stages: (1) image segmentation to determine each food's contour, (2) image recognition using faster R-CNN, and (3) estimate the food's weight and calories. In this study, the proposed system detects the object of each food's contour using Otsu's method and estimates the calories of each food, with data trained using faster RCNN.

I. Harshitha (✉) · T. Sunil Kumar · P. Venkateswara Rao
Department of Computer Science and Engineering, VNR VJIET, Hyderabad, India
e-mail: harshitha.i281@gmail.com

T. Sunil Kumar
e-mail: sunilkumar_t@vnrvjiet.in

P. Venkateswara Rao
e-mail: venkateswararao_p@vnrvjiet.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_34

Keywords Calorie estimation · Image processing · Deep learning · Faster R-CNN · Otsu method

1 Introduction

Health is one of the most important elements of an individual's life, and in today's society, everyone's life has become exceedingly hectic and stressful. People work very long hours and eat whatever food is available not knowing how many calories they are consuming, resulting in a huge amount of junk in our bodies. This junk has a wide range of negative effects on the body. Rates of obesity are rising as a result, and this is putting people's health at risk. Staying in shape and eating a healthy diet require some effort on the part of the individual. It is now extremely difficult for people to keep track of the calories they consume. Calorie intake is critical to healthy living.

Previously, people would track their calorie consumption using charts or a timetable, or they used to maintain a proper diet in which the meal that had to be consumed, as well as the amount, was fixed, and many alternatives were accessible, including consuming vitamins and a large number of pharmaceutical items. All of these products, however, have a variety of negative side effects on the body, some of which can be dangerous. Workout and exercise are beneficial to some level, but they require a special time commitment. These methods are time-consuming and hard for people to implement correctly. So, rather than opting for such inconvenient methods, one approach to live a healthy life is to take only the amount of nutrition that the body requires, not too much or too little. There is a variety of fitness and diet tracker applications available, but they all need a manual process, which is inconvenient and time-consuming. As a result, most people do not utilize such programmers for long periods. However, even though packaged foods include nutrition and calorie labels, it is still difficult for people to refer to them.

The usage of mobile technologies has opened up new avenues for people to become more conscious of their health and fitness by using diet tracking apps. People may now obtain information anytime they want with the health apps, which can also be used as a tool to help users reach their goals for better health. Furthermore, as people all around the world grow more health concerned, they have begun to adapt to numerous creative techniques to keep fit. One simple method is to assess the nutritional value of the food that people eat.

The following are the project's major contributions:

- Proposing the first food recognition system.
- Advising on a comprehensive and accurate calorie estimation method.

In this research, food recognition is involved in the use of image processing techniques in calorie estimates. The purpose of the research is to figure out how many calories are in a certain amount of food depending on its weight. All of these issues will be addressed by the created system. Based on the image provided by the

user, it will recognize the food product, that the person is eating and displays the calorie amount.

2 Literature Survey

Kasyap et al. [1], this proposed solution uses various algorithms to produce food recognition and detection. These algorithms are CNN, random forest, and SVM, which were used to achieve higher accuracy. The ECUSTFD dataset was used, which is a food image data that also contains information of food quantities and masses. It includes labeled photos as well as true mass and volume for each meal in the image. CNN was utilized for picture recognition. CNN is developed with sequential, convolutional, pooling, and full connections. The CNN model has a 92% accuracy rate. Additionally, the models were trained using data from the dataset. To recognize food items from images, a TensorFlow object was used with the detection API. Naritomi and Yanai [2] presented a "CalorieCaptorGlass," a calorie estimate system based on real size that uses picture recognition and AR/MR glasses technologies based on deep learning using the environment recognition function. In additional to the user experience using AR glasses, an application has been developed that considers the exact size of food that could not be comprehended just through visuals. Mask R-CNN was utilized for meal recognition, with UECFoodPix used for learning, as well as the surroundings recognizing feature of AR glasses was employed for real measurement.

Poply and Angel Arul Jothi [3], by capturing top view photos of multiple-dish food items, this research presents a computer vision and deep learning-based approach for forecasting the calorie contents. The system uses CNNs to carry out a revised image segmentation technique that loosely simulates instance segmentation, in which each pixel is recognized from the instance of objects for each item found in an image, based on recent work in object recognition and semantic segmentation. The pixels that have been detected are in the form of masks/segmentation that is linked to certain food items. Then, using a reference object, these outcome masks are utilized to calculate volume and mass estimates for the identified food items. By assessing the outcomes on previously unknown data, the system established in this study achieves transfer learning. The system has an mean average accuracy rate of 89.30% for object identification and a percentage accuracy of 93.06% for calorie estimation, according to the evaluation.

Parth [4], the purpose of this study is to create a deep learning and computer vision-based method to calculate the caloric count of any meal using its image. Mask R-CNN, a deep learning-based convolutional neural network (CNN), is used to do instance segmentation. The mask R-CNN detects different occurrences of different dietary things and generates a mask for meal items. By using mask, the surface area of the identified meal is calculated. The surface area of the food item, as well as the calorie per square inch number, is utilized to determine the caloric contained in the meal. The created model has an accuracy rate of around 93.7% for food item recognition and an precision of approximately 95.5% for caloric prediction.

Muralidhar et al. [5] present an effective approach for classifying food objects and calculating calorie consumption by utilizing object identification and deep learning neural networks to detect fruit and vegetable. TensorFlow is used to predict the number of calories in a CNN-based food categorization and food detection model, and faster R-CNN is used to recognize nutritious food. Balbin et al. [6] propose a method that provides a fixed-placed camera in a hardware system to calculate the weight and calorie count of a product, especially meat, based on a single picture. The integrated method for recognizing food on a picture was the convolutional neural network (CNN). To identify the regions of the food in the image, graph cut image segmentation was utilized. The system was put to the test for ten trials on each component of the chicken and fish, as well as whether it was cooked or grilled, yielding a food identification accuracy of 91.82% and a calorie estimation accuracy of 88.18%.

Zilic et al. [7] present a mobile application that recognizes food items from a multi-object meal in real-time from a single picture and returns nutrition data with elements and approximated amounts. To accomplish simultaneous multi-object identification and localization with almost 80% average percentage accuracy, they combine a deep convolutional neural network with YOLO, a state-of-the-art detecting approach. Second, they turned the models into a mobile application with a nutritional analysis feature. Yogaswara et al. [8], the deep learning mask region-based convolutional neural network (R-CNN) technique is used to construct a system that uses a computer vision approach to autonomously calculate the amount of dietary protein depends on the amount of the food. The segmented technique employs an instance-aware semantic segmentation method, uniquely identifies every pixel in a food image using instances of items. The precision in calories was calculated to be 97.48%.

Rewane and Chouragade [9] present a system and a mobile app for detecting food items, displaying their calorie counts, and proposing healthy meal selections to the use multiple procedures are utilized in the detection of food items, including extracting food elements from the photo, extracting its characteristics, and classifying it. Lu et al. [10] proposed using a single RGB meal image as input to multi-task learning-based convolutional neural network (CNN) for food analysis, which includes food segmentation, detection, and volume estimate all at the same time.

3 Proposed Methodology

This paper proposes a model for estimating food weight from a single image. This is accomplished in three stages: (1) image segmentation to determine each food's contour, (2) image recognition using faster R-CNN, and (3) estimate the food's weight and calories. The calories are then calculated by comparing this to food data. Furthermore, the concept is to assess the caloric value of the food item being detected. Two major factors influence the precision of the estimation result: the object recognition algorithm and the volume estimation method. In the context of object detection, a deep learning method known as faster R-CNN will recognize particular food in general conditions. In addition, Otsu's technique, which is an

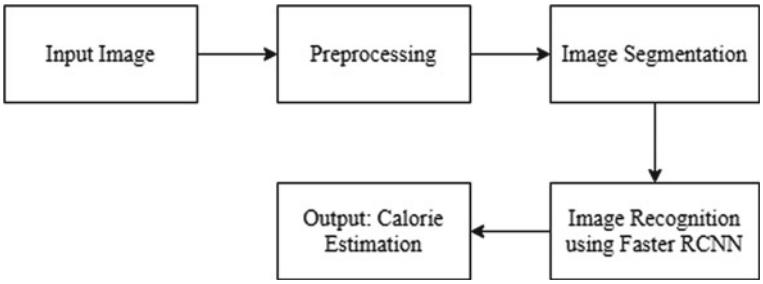


Fig. 1 System architecture

image segmentation method, is used. The proposed approach aids in autonomously recognize food components from an image of a plate and then assess key food characteristics like calories and weight (Fig. 1).

3.1 Faster R-CNN

Faster R-CNN is a deep convolutional network that seems to the user just like singular object identification network. The network can determine the positions of various items correctly and fast. Here, the faster R-CNN method was used to detect images. We feed the image into the convolution network as input and get a convolution feature map as output in this approach. To detect region proposals, the system uses a different network. The ROI pooling layer is utilized to reshape the estimated region suggestions. With their assistance, the bounding box offset values are anticipated, and the photos are also categorized within the specified region (Fig. 2).

RPN, as a region proposal method, and the fast region-based convolutional neural network as a detecting network make up the faster R-CNN architecture. As illustrated in Figure, the fast R-CNN detection also includes a CNN framework, an ROI max pooling, and feed-forward neural network, two consecutive-related branches for classifications and regression problems.

- To obtain the extracted features, the input picture is first processed by the backbone CNN.
- Region proposal network categorization and detection branches are not the same as these. Each of the detection task's classes contains C units in the classification layer. To acquire the classification scores—likelihood of a suggestion class labels—the characteristics are processed through a softmax layer. To enhance the projected bounding boxes, the regression layer values are utilized. Regressor in this case is size agnostic, but it is unique to every class. That is, in the regression layer, each class has its regressor with four parameters that correspond to C * 4 output units.

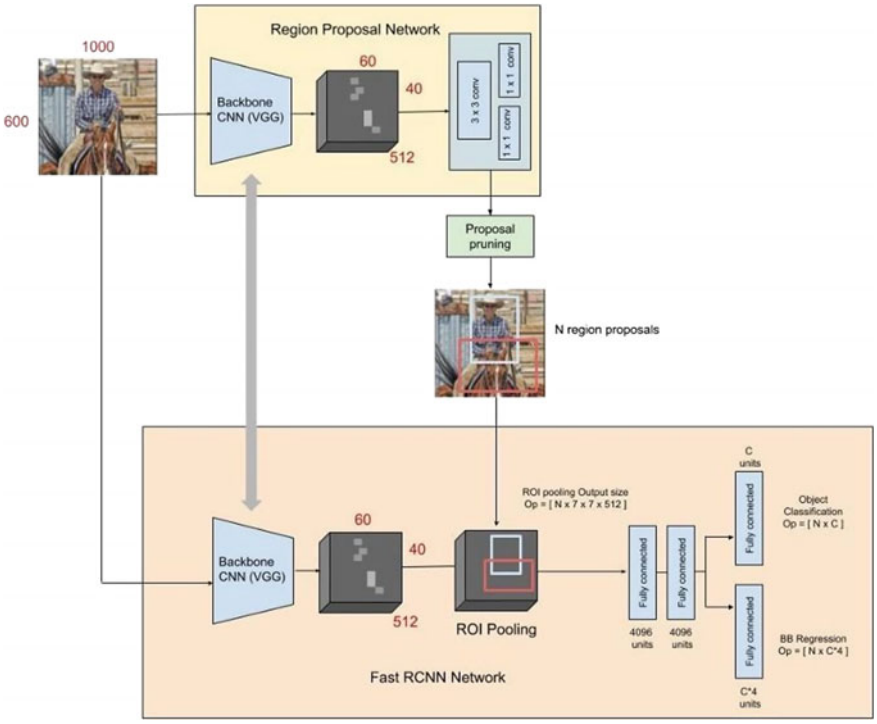


Fig. 2 Architecture of object recognition using RCNN

- A four-step training technique was employed to compel the system to split the parameters of the CNN backbone between both the region proposal network and the detectors.

(a) As previously stated, RPN is trained separately. The backbone CNN for this job is quite well for the object proposals task using parameters from a system trained for an ImageNet classification challenge.
(b) The fast region-based convolutional neural network (RCNN) detector network is trained separately as well. The backbone convolutional neural network for this job is trained with parameters from an ImageNet classification system before being quite well for the item identification challenge. The parameters of the RPN are set, and the RPN's suggestions are utilized to train the faster R-CNN.
(c) RPN is quite well for the object proposals job, using parameters from the faster R-CNN. Weight values in the similar layers between the region proposal network and the detector are fixed this time, and only the region proposal network-specific layers are fine-tuned. The region proposal network is now complete.
(d) Fast R-CNN detector is quite well once again by utilizing updated RPN. Only the layers specific to the detection network are quite well, while the weights of the common layers are constant.

This results in an R-CNN detection system with shared convolutional layers that is faster.

3.2 Otsu

Here for the image segment, Otsu's approach is employed. It automatically implements image thresholding. The technique, in its most basic form, reverts a single intensity threshold that divides images into two classes: foreground and background. Otsu's thresholding method entails replicating throughout all feasible threshold values and computing a scale of scatter for the pixel levels on every corner of the threshold. The goal is to find the threshold value at which the sum of foreground and background spreads is at its smallest binarization level. Otsu's thresholding technique is being used in image processing to make instinctive decisions determined by the shape of the histogram. The method implies the image is divided into two basic classes: foreground and background. The optimum solution threshold value is then calculated by minimizing the set of weights within-class variation of these 2 categories. It is statistically proven that minimizing within-class variation is the same as maximizing between-class variation.

Step 1: Create a histogram for a two-dimensional image.
Step 2: Evaluate the variances for the foreground and background for a single threshold.

(i) Determine the mass of background and foreground pixel value.
(ii) Determine the average of the background and foreground pixel value.
(iii) Determine the variation of background and foreground pixel value.

Step 3: Determine within-class variance

1. Compute mass for background and foreground pixels
   - $W = (w1 + w2 + w3)$/(Total number of pixel)
2. Compute mean
   - Histogram value $= h$
   - $\mu = ((h1 * w1) + (h2 * w2) + \cdots (hn * wn) + (hn * wn))$/(Sum of weights($w$))
3. Compute variance
   - $V = (((h1 - \mu)^{(2)} * w1) + ((h2 - \mu)^2 * w2)\ldots.((hn - \mu)^2 * wn) + ((hn - \mu)^2 * wn))$/(Sum of weights ($W$))
4. Compute within-class variance

   the sum of 2 variables is calculated by multiplying by their random weights variation within classes = Wb * Vb + Wf * Vf (Fig. 3)

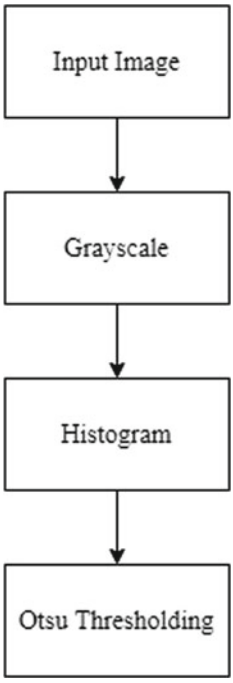


Fig. 3 Flow of Otsu's thresholding in image processing

3.3 Volume and Calorie Estimation

Area of food image:
We have three elements from image segmentation:

- pixel area of foods
- Pixel area of the skin
- The surface area of the skin (skin multiplier).

The estimated area of food based on these criteria is as follows:

- Estimated Food Area = (Foods Pixel Area * Actual Skin Area)/(skin pixel area)

We have three types of food shapes:

Spheres—such as an apple, an orange, a tomato, or an onion, kiwi
Cylinder—such as banana
Irregular shape (other)—such as bread, samosa, and noodles.

Volume Estimation for Sphere

- Estimated Radius = $\sqrt{}$(Estimated Food Area)/π
- Estimated Volume = 4/3 * π * Estimated Radius^3

Volume Estimation for Cylinder

- Estimated Height = Pixel Height ∗ Pix_To_Cm_Multiplier
- Estimated Radius = (Estimated Food Area)/(2 * Estimated Height)
- Estimated Volume = π * Estimated Height * Estimated Radius^2

Volume Estimation for Irregular Shape

- $v = Q \times s \times a^3 \times H$

Weight and Calories Estimation of Food

- Estimated Weight = Actual Density * Estimated Volume
- Estimated Calories = (Estimated Weight * Calories per 100 g)/100.

4 Experimental Evaluation

4.1 Dataset

A food dataset was used in this study, and it is extracted from Kaggle. Images of 14 distinct food products are included in the dataset. The images are trained with the dataset. There are 7110 photos of foods in the dataset. The dataset includes photos taken from various perspectives. Having a diverse range of foods and foods provides a more accurate and trustworthy dataset for calorie food measuring systems. Images of single food items are present in this dataset.

4.2 Preprocessing

Basic preprocessing used here is to clean the dataset, removing any unnecessary or noisy photos which are a random variation of brightness or color information in images. The features have been preprocessed, which contains edge smoothing and noise reduction. Color and adaptive threshold are the preprocessing techniques used in this image, and the features have been preprocessed, which contains edge smoothing and noise reduction. Color and adaptive threshold are the preprocessing techniques used in this image. In its most basic implementation, adaptive thresholding takes a monochrome or color picture as input and outputs a binary image that reflects the segmentation. A threshold must be determined for each pixel in the image. The threshold is determined by interpolating the sub-image results. The results of preprocessing are depicted below (Figs. 4, 5 and 6).

Fig. 4 Original image

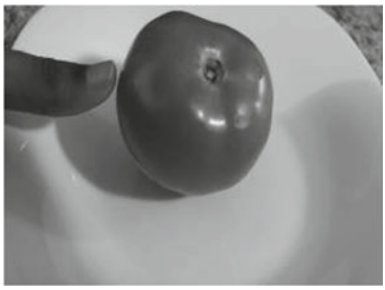


Fig. 5 Grayscale image

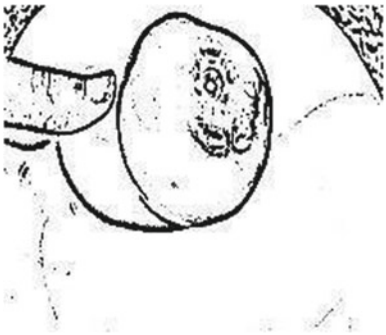


Fig. 6 Adaptive thresholding

4.3 Image Segmentation

Image segmentation is the process of turning an image into a group of pixel areas specified by masks or a labeled image. It divides an image into multiple segments with similar features in image segmentation. The purpose of segmentation is to isolate the relevant features from an image. Image segmentation aids in feature identification and increases classification rates. Unnecessary elements in the picture, such as a plate, dishes, fork, and so on, as well as their color, are omitted, and only that specific food item is evaluated. To determine the contours of the fruit, Otsu's approach was employed to segment the food item. It is used to conduct image thresholding automatically. The threshold used by Otsu's thresholding is chosen to minimize the

Fig. 7 Original image

infraclass variation of the threshold black and white pixels. This method is used to segment an image, after which color, shape, and text features are extracted. The results of image segmentation are depicted below (Figs. 7, 8, 9, 10, 11, 12, 13, 14 and 15).

Fig. 8 Mask image

Fig. 9 Big contours of plate

Fig. 10 HSV image

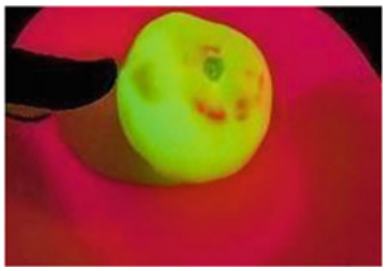


Fig. 11 Mask plate

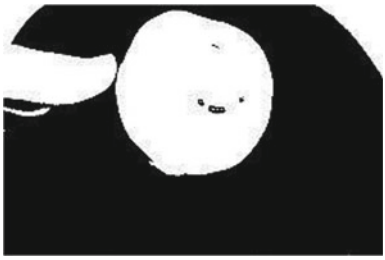


Fig. 12 Mask not plate

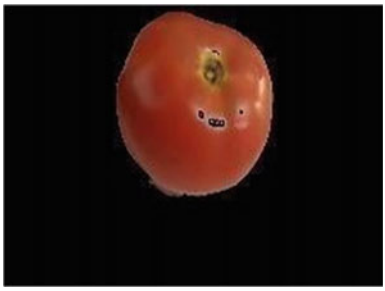


Fig. 13 Food skin

Fig. 14 Erode fruit

Fig. 15 Final image

4.4 Image Recognition

Here, faster R-CNN method is used for image detection. As an input, the image is given to the convolution network, and as an outcome, it gets a convolution feature map. To estimate region proposals, the system uses a different network. The ROI max pooling is used to alter the expected region suggestions. With their assistance, the bounding box offset values are anticipated, and the photos are also categorized within the specified region. Because it is particularly successful for object detection, objects of varied scales, and overlapping objects in an image.

4.5 Calorie Estimation

Total calories are calculated using the expected volume. When the volume is multiplied by the density of the food, an estimate of its mass is obtained. Finally, mass is multiplied by Kcal to calculate total calories.

5 Results

The training and validation images are run for 10 epochs, where after no further progress in precision is noticed, as shown in Figs. 16 and 17.

The percentage of caloric precision is shown in the below table. The following equation is used to calculate the precision of the caloric computation in the table (Fig. 18).

Food item	Actual calories	Obtained calories	Error %	Accuracy %
Apple	45	45.5	1.1	98.9
Banana	61	61.9	1.4	98.6
Bread	77	79.4	3.1	96.9
Chocolate cake	241	256	6.2	93.8

(continued)

(continued)

Food item	Actual calories	Obtained calories	Error %	Accuracy %
French fries	914	915	0.1	99.9
Hot and sour	277	276.5	0.1	99.9
Kiwi	147	147.5	0.3	99.7
Noodles	35	35	0	100
Omelet	203	202.5	0.2	99.8
Onion	244	244.2	0.2	100
Orange	62	62	0	100
Samosa	35	36	1	97.3
Spring rolls	190	187.7	2.3	98.8
Tomato	29	29.2	0.2	99.4

The precision of calorie estimation in the apple item was 98.9%, banana 98.6%, bread 96.9%, chocolate cake 93.8%, french fries 99.9%, hot and sour 99.9%, kiwi 99.7%, noodles 100%, omelet 99.8%, onion 100%, orange 100%, samosa 97.3%, spring rolls 98.8%, and tomato 99.4%. The mean accuracy of calories estimation determined from the result section above is 98.7%.

The findings of calorie estimation are shown here. The calculation is performed for analysis of results by equating it to the precise calorie of the food product available on the Web. If the estimation error is less than $\pm 10\%$, the project is considered a success. The outcome is shown in the above table. This project is in good condition because the calculation error is less than 2%.

6 Conclusion and Future Enhancement

This system allows people to keep a record of their daily food intake. This approach will assist you in keeping your body fit and healthy by consuming food according to the suggestions and recommendations are given. The most efficient methodologies and algorithms are recognizing food items. Deep learning, image processing techniques, volume, and weight metrics are used in this work to determine the calories of various foods. At the end of the process, this work evaluates the calorie content in foods. According to the findings of this project, multiple items or meals may be added to the model in future to allow the classifier to predict a wider range of food items, and it can also detect several count food items in the same meal to assess the calories of the meal more efficiently. Other information, such as BMI, can also be added so that consumers can receive appropriate food and nutrition recommendations.

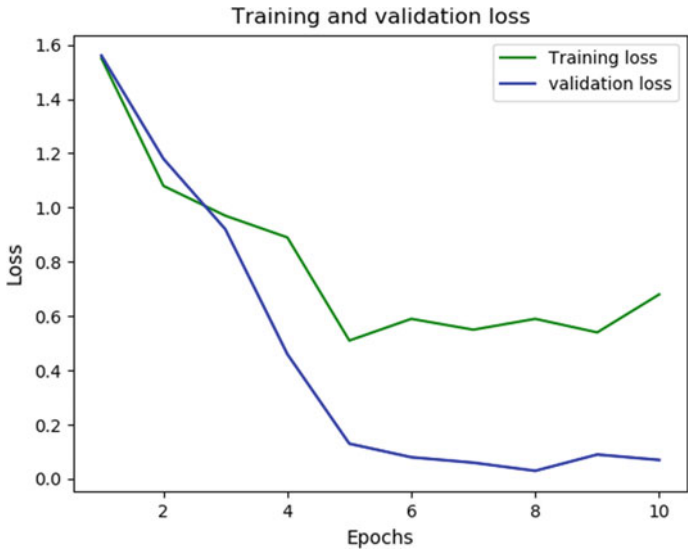


Fig. 16 Loss of R-CNN model

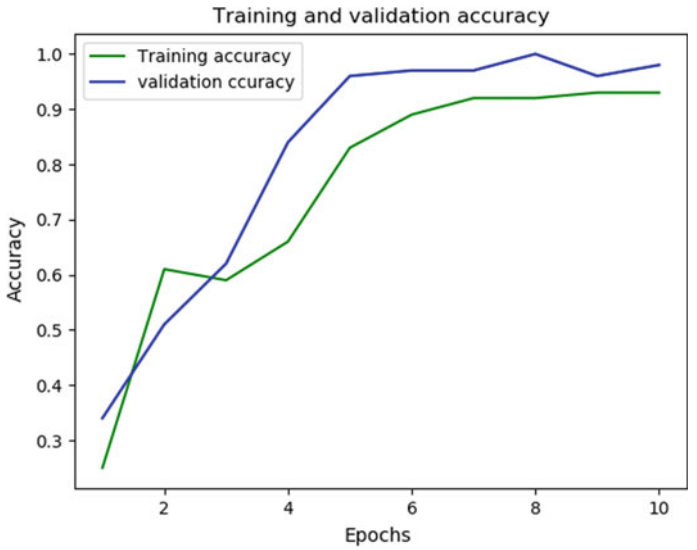


Fig. 17 Accuracy of R-CNN model

Fig. 18 Error rate and accuracy for estimated calories

$$Error = cal.of\ GT - Cal.of\ prediction\ \frac{100}{Cal.of\ GT}$$

$$Accuracy = 100 - Error$$

References

1. V.B. Kasyap, N. Jayapandian, Food calorie estimation using convolutionalneural network, in *2021 3rd International Conference on Signal Processing and Communi cation (ICPSC)* (IEEE, 2021)
2. S. Naritomi, K. Yanai, CalorieCaptorGlass: food calorie estimation based onactual size using HoloLens and deep learning, in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (IEEE, 2020)
3. P. Poply, J. Angel Arul Jothi, Refined image segmentation for calorie estimationof multiple-dish food items, in *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (IEEE, 2021)
4. P. Poply, An instance segmentation approach to food calorie estimation using maskR-CNN, in *Proceedings of the 2020 3rd International Conference on Signal Processing and Machine Learning* (2020)
5. E.S. Muralidhar, et al., Development of health monitoring application using machine learning on android platform, in *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (IEEE, 2020)
6. J.R. Balbin, et al., Determination of calorie content in different type of foods using image processing, in *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)* (IEEE, 2019)
7. J. Sun, K. Radecka, Z. Zilic, FoodTracker: a real-time food detectionmobile application by deep convolutional neural networks (2019). arXiv preprint arXiv:1909.05994
8. R.D. Yogaswara, E.M. Yuniarno, A.D. Wibawa, Instance—aware semantic segmentation for food calorie estimation using mask R-CNN, in *2019 International Seminar on Intelligent Technology and Its Applications (ISITIA)* (IEEE, 2019)
9. R. Rewane, P.M. Chouragade, Food nutritional detection, visualization andrecommendation for health monitoring using image processing, in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (IEEE, 2019)
10. Y. Lu, et al., A multi-task learning approach for meal assessment, in *Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management* (2018)

Adaptive Serious Games to Teach Cybersecurity Concepts Using a Machine Learning Approach

Devottam Gaurav, Yash Kaushik, Santhoshi Supraja, Manav Yadav, Manmohan Prasad Gupta, and Manmohan Chaturvedi

Abstract Use of serious games for teaching various important concepts including cybersecurity has been on rise. However, serious games with fixed trajectories tend to be less effective in learning outcomes as this approach of *one size fits all* fails to hold attention of players with different background knowledge and learning abilities. To address this important gap, we attempt to classify the players, based on initial game play data, in two classes, viz. beginner and expert. The two games, that we describe in this paper, dynamically adapt the teaching and testing phase of the game appropriate to the player type. This feature of the games is expected to suitably adapt the difficulty level of the game to ensure that both beginner and expert players are given appropriate challenge through the game and thus provide the crucial flow towards achieving pedagogical objectives. We also attempt to validate these games using both objective and subjective measures through use of recorded player data and embedded questionnaire on various aspects of user experience (UX). The validation process, leading to a measure of learning outcome, is expected to facilitate suitable improvements to the game design and thus achievement of pedagogical objective while maintaining a sense of play to the potential learner.

Keywords Adaptive serious games · Cybersecurity · Machine learning · Learning outcome

This research work is funded by MHRD, Government of India, as IMPRINT project no. 7804.

D. Gaurav (✉) · Y. Kaushik · S. Supraja
Computer Science and Engineering, Indian Institute of Technology Delhi, Delhi, India
e-mail: gauravpurusho@gmail.com

S. Supraja
e-mail: cs1190362@iitd.ac.in

M. Yadav
Computer Science and Engineering, Delhi Technological University (DTU), Delhi, India
e-mail: manavyadav_co20b4_67@dtu.ac.in

M. P. Gupta · M. Chaturvedi
Department of Management Studies, Indian Institute of Technology Delhi, Delhi, India
e-mail: mpgupta@iitd.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_35

1 Introduction

Malicious players in online world are using sophisticated tools to penetrate protected systems and depend on mistakes by a human to get a way in using social engineering as method of choice. With a focus to have more trained professionals to counter these cyberattacks, we notice a gap in the education domain and industry. Even though there are traditional methods for training professionals using classroom environment supported by YouTube, distance learning and various security competitions [1, 2] which appear to be quite effective, these techniques require special efforts by a domain expert and are found to be costly to implement. Therefore, there is a need to set up an online; adaptive and responsive medium that can train/educate professionals about cyberattacks. The web-based serious games (SGs) can be a suitable candidate for cybersecurity training at a lower cost while providing necessary sense of flow and fun to the player. In literature, a couple of studies have detailed positive outcomes demonstrating that SGs are powerful when utilized for training/creating awareness among professionals related to cybersecurity [3–5]. Consequently, our games may motivate people and organizations to use the SGs-based training as a viable method.

SGs are computer-based games intended for purposes other than amusement [6]. SGs possess the benefits of gaming innovation to make a fun filled, persuasive and intelligent virtual learning environment that enhances user engagement. The utilization of SGs in educational sector for educating and preparing is acquiring momentum and is expected to become mainstream in near future [7]. However, the SGs possess the demerit of not adapting the gaming environment based on the learner's capacity to absorb the content. To overcome this demerit of SGs, the SGs must be designed in such a way that they can adapt to the changing behaviour of players as per the needs [8]. This kind of adaption in SGs helps the players to compete with the gaming challenges/difficulties. Hence, such SGs can be termed as 'Adaptive SGs' [9]. Therefore, to provide the adaptive feature in SGs, there is a need of an intelligent tool, i.e., Machine Learning (ML) [10] that can act as an agent which provides instantaneous suggestions/hints to the players based on the players' ability. In past, computer games have used Artificial Intelligence (AI) to facilitate game of chess besides many other applications to mimic human intelligence [11]. The challenge to defeat human expert players in rule-based strategy games such as chess has greatly advanced the domain of AI research [12]. New AI methods have been used to generate game levels, scenarios and story-lines [13, 14]. Use of Machine Learning (ML) which is a subset of AI by using historical player data has evolved in recent time. In this paper, we describe a very basic use of historical player data to make two of the cybersecurity games adaptable to player profile. In literature, a few studies have reported attempts to make the SGs adaptive [15–20]. Some of the papers in literature of adaptive SGs, prediction models were also developed by the authors for predicting the performance of players. In our case, we have focused only on classifying the players into two classes, i.e. beginner and expert using ML approach.

The major contributions of our paper are:

1. We describe two online playable adaptive SGs to teach about Firewalls and SQL injection attack concepts.
2. We evaluate the effectiveness of the two web-based games using both objective measure; by capturing the player data using Unity analytics and subjective measure;by embedding a questionnaire to elicit user experience (UX).

The organization of our paper is described below. Section 2 discusses the methodology used to make the SGs adaptable as per the player needs. Section 3 presents the design considerations of the two games. Section 4 addresses the limitations of our research. Finally, the concluding remarks and future directions are placed in Sect. 5.

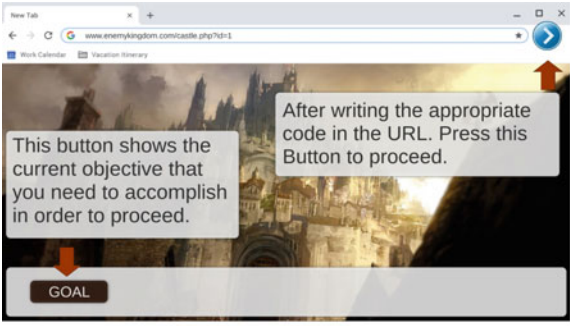
2 Methodology

To make the SGs adaptive as per the player needs, the ML approach is used. The ML approach uses the player classification strategy in two games, i.e. SQL injection [21] and Firewall. In the player classification strategy, the players are classified into two classes, i.e. beginner and expert based on certain player attribute. In our case,we use median value of these attributes from the historical player data stored in a database using Unity analytics feature. Median is a better indicator of the central tendency of a data set and less affected by outliers and skewed data points.[1] To proceed with the computation of median in our paper, first we need to train the ML model with the training set of players data, i.e. objective data. In the SQL injection game, the objective data is considered to be time taken, number of hints, and number of mistakes and in the Firewall game, only time taken is considered. With the trained model, the median is computed for the most recent 50 players' data. The 50 players data is fetched from the Firebase.[2] Based on the computed median, the players are classified into two classes, i.e. beginner and expert. If the player takes more time to complete a level than the median value of historical data, he is classified as beginner, else he is classified as Expert. Similar approach is used while considering other player attributes like number of hints and number of mistakes. Use of less hints or less mistakes characterize an Expert player. The player using more hints and making more mistakes as compared the current median value needs careful hand holding by way of special tutorials.Thus difference between the two classes are that more suggestions/hints and additional tutorials are provided to the players classified as beginner to ensure that the player finds the learning process suitable for him. The classification of players into two classes for two games are shown in Figs. 1 and 2.

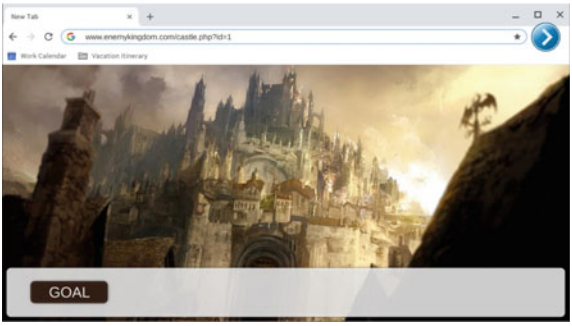
In Fig. 1, for beginner players, a hand-holding tutorial of user interface is presented which helps the players to avoid future mistakes and for expert players, no such hand-

[1] https://www.graduatetutor.com/statistics-tutor/measures-of-central-tendency-mean-median-mode/.

[2] https://firebase.google.com/.



(a) Beginner Classification of players for SQL Injection



(b) Expert Classification of players for SQL Injection

Fig. 1 Classification of players for SQL injection

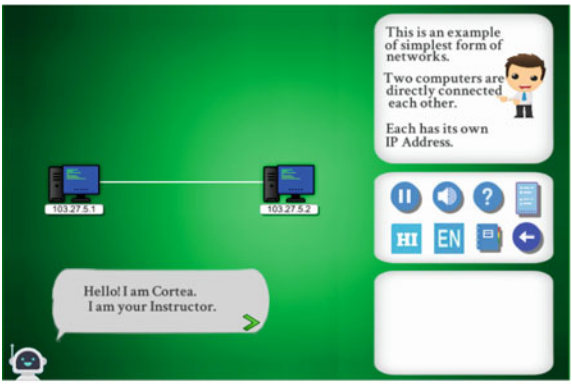
holding tutorial is presented. In Fig. 2, for beginner players, the beginner level, i.e. Level 1 is opened in the Firewall game, where the 'Bot' helps the players in providing the suggestions/hints to complete the level successfully and for expert players, next expert level, i.e. Level 2 is opened. Based on the player classification in the Firewall game, the players are rated with stars accordingly, i.e. 3 stars for expert and 1 star for beginner players.

3 Design Considerations

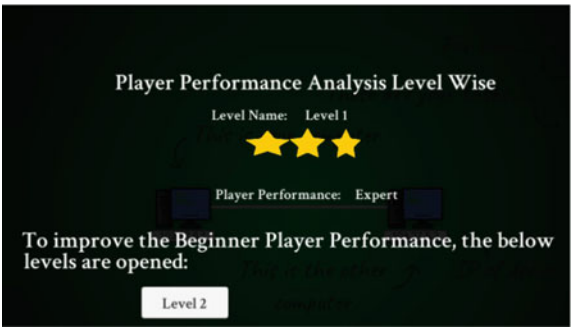
The two web-based games, i.e. SQL injection and Firewall are developed using Unity [22]. Unity is acting as our development platform and also a test bed to elicit objective player data using its 'Unity analytics' feature. Unity is a cross-platform game engine that deploys easily on several platforms.

Fig. 2 Classification of players for firewall

(a) Beginner Classification of players for Firewall



(b) Bot for Beginner Players In Firewall Level



(c) Expert Classification of players for Firewall

The purpose of the two web-based games is to teach the players cybersecurity concepts. While the Firewall game is focused on network security of a host by suitable deployment of Firewall, the SQL injection game describes the vulnerability and protection of an online database to the SQL injection-based attacks. SQL injection is one of the top ten vulnerabilities described by Open Web Application Security Project (OWASP).[3] The target audience for the two web-based games is system administrators/entry-level IT users. The two games are web-based games. These two games are online playable and are accessible on IIT Delhi website.[4] While designing these games, two detailed flow charts for each stage of game play were prepared as a design document. These flow charts were peer reviewed for possible improvements as shown in Figs. 3 and 4. The steps of Figs. 3 and 4 are taken into consideration as per the players' gaming ability.

In these game, we have divided the games in two phases, namely teaching phase and learning phase. The two phases have multiple learning goals. The initial assessment score (IA) relates to the teaching phase and the final assessment score (FA) relates to the testing phase. We have adapted the approach described in [23] to compute these scores. We have been able to collect player data from only 13 players. The objective assessment of the player has been computed by using time taken only to complete a task and other attributes like number of mistakes and number of hints of the captured data have not been used in this study. The results of the objective scores for set of 13 players are clustered in four categories, namely Active Learners, Slow Learners, Masters and Outliers based on the modified values of IA concerning Initial Threshold (IT) and FA regarding Final Threshold (FT).

This classification of players into four categories for two games are done using k-means clustering [24], is shown in Fig. 5.

The Active Learners are one who display good performance during FA as against a Slow Learners who display low performance in both IA and FA. Masters are characterized by higher score in IA as they seem to have background knowledge. Outliers are those players who show poor scores in FA even after getting good scores during IA phase. This clustering would be more meaningful when number of players is large, i.e. greater than 30 and therefore, the clustering shown in Fig. 5 is only for illustration purpose.

The subjective analysis evaluates the user experience (UX) of the two games. The players after finishing these games reflect on these games in a brief anonymous questionnaire embedded within these games. The questionnaire has been conceived with a multi-dimensional aspect of UX and gathers data about various traits as illustrated in [25]. The players rated these games on the five-step Likert scales where (1 belongs to 'Strongly Disagree' and 5 belongs to 'Strongly Agree'). The mean scores of subjective feedback for both games are given in Table 1.

[3] https://owasp.org/www-project-top-ten/.

[4] http://gost.iitd.ac.in/serious_games/pages/ser.html.

Table 1 Subjective feedback mean score for two games

Dimension	Attribute	Survey item	Mean score of SQL injection	Mean score of firewall
Gaming experience	Challenge	The experience was challenging. I found the game stimulating	3	3.3
	Flow	I was able to achieve the goals set in the game	3.3	3.3
	Immersion	I remained focused on the game throughout	2.8	3.4
	Affect	The overall experience was positive	3.4	3.1
Learning experience	Learning goal	The learning goals of the game were clear	2.6	3.5
	Content appropriateness	The game scenario had relevance to the subject	3.1	2.9
	Integration	The game required me to use skills being taught	2.7	3.3
	Feedback	The game provided opportunities to receive feedback	3.4	3.6
Usability	Interface interaction	The user interface was easy to use	2.8	3.4
		It was easy to get started with the game	3.1	3
		I learnt how to play the game quickly	3.2	3
Fidelity	Visual appeal	The playing environment was visually appealing	2.9	3.2
	Identification	I can identify with the components used in the game	3.6	2.7
	Verisimilitude	I can identify with the story/scenario in the game. The experience felt real	2.5	3.4

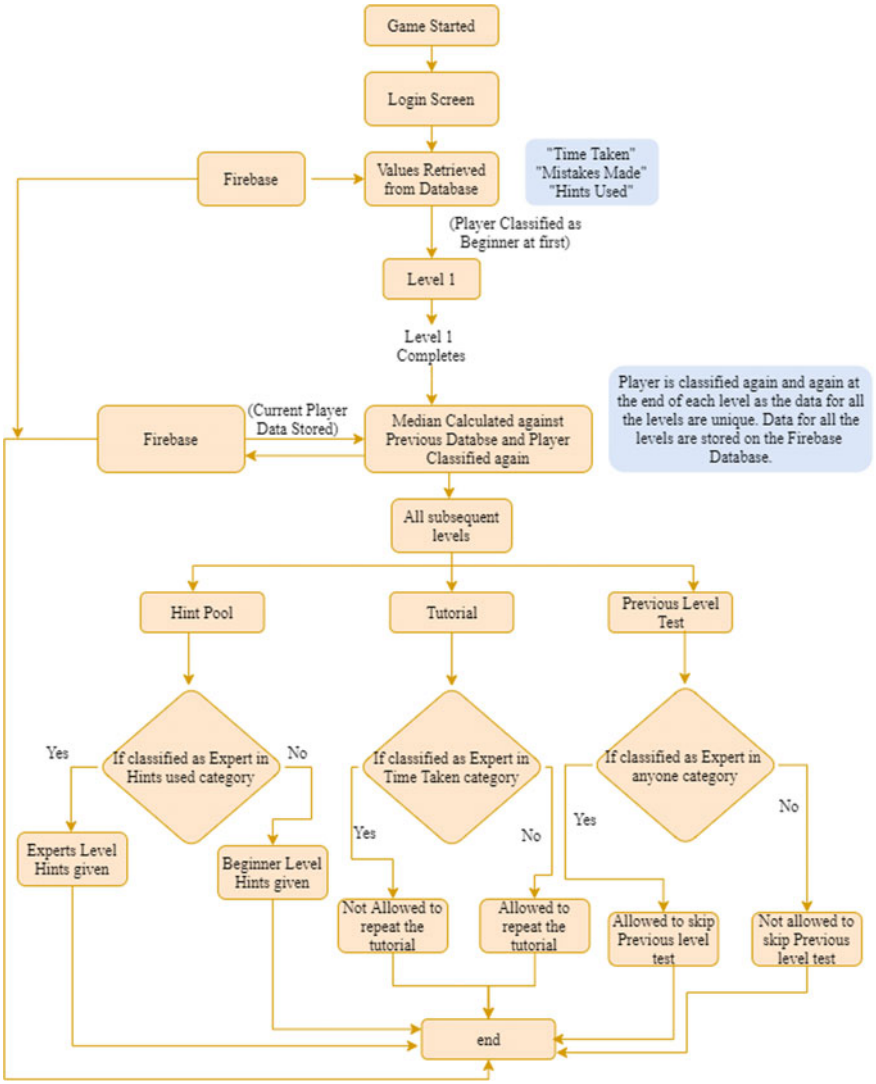


Fig. 3 Illustrative flow chart for adaptive SQL injection

4 Limitations

Our research paper has some limitations.Though our games are able to capture the objective player data using Unity analytics and subjective feedback through a questionnaire given in Table 1, we have been able to measure the learning outcome of these games using these artefacts only for 13 number of players because of non-availability of players due to on-going COVID pandemic. In the two games, among

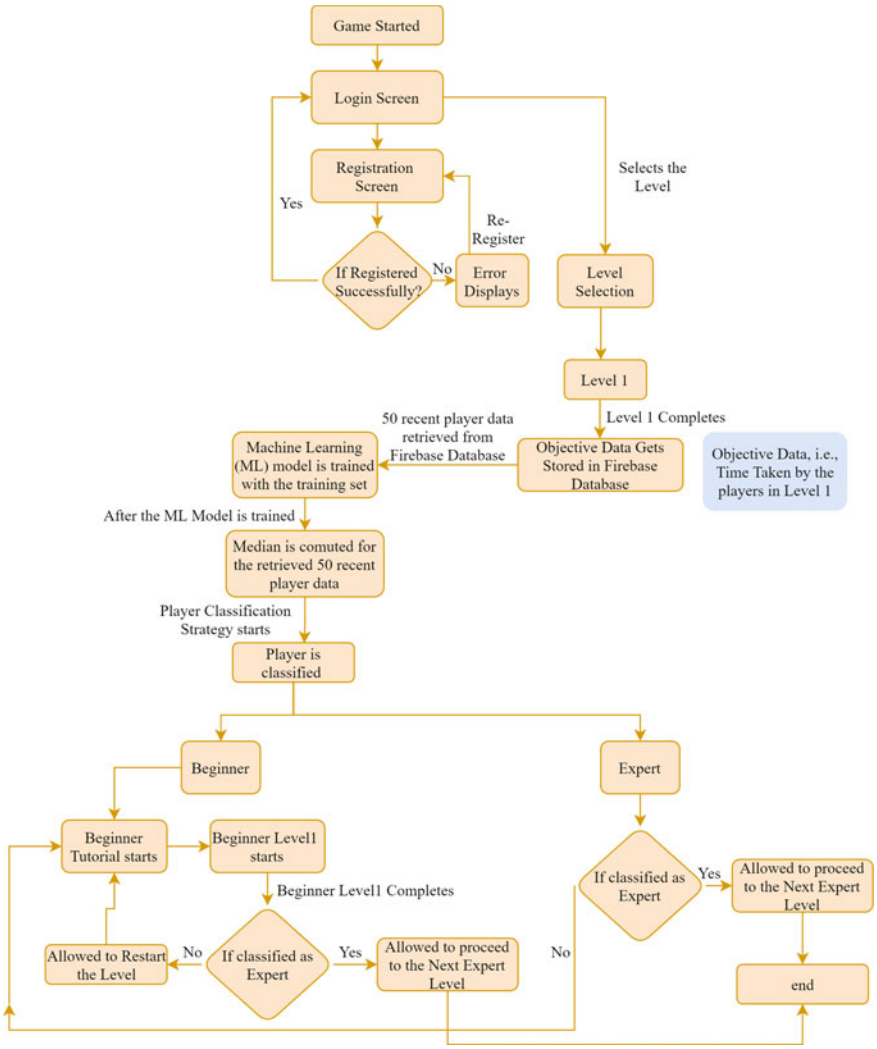
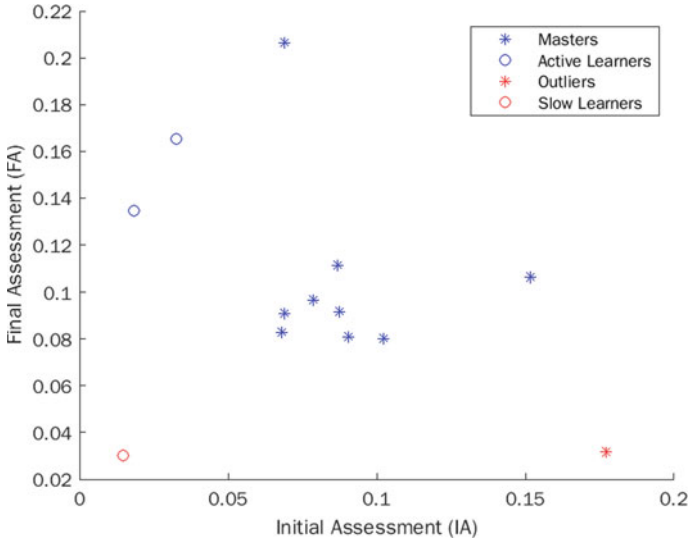
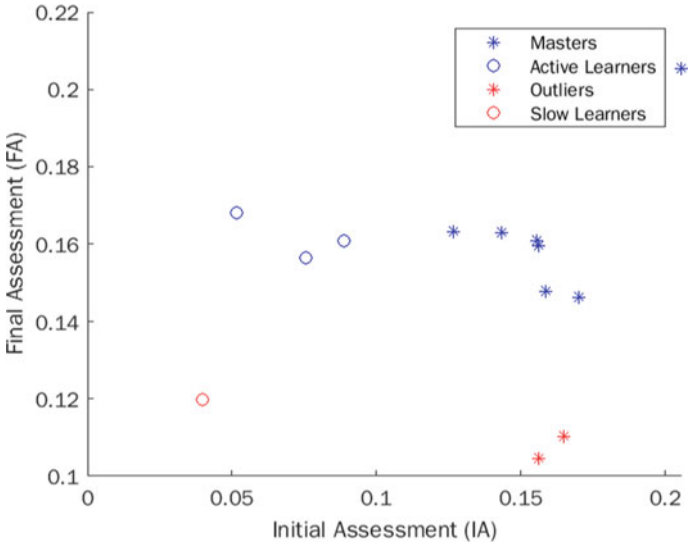


Fig. 4 Illustrative flow chart for adaptive firewall

the different metrics that evaluate the learning outcome, we have restricted the evaluation based on the time taken only. The clustering of players are done using k-means clustering based on the time taken. Hence, only one metric is considered, i.e. time taken. However, we plan to carry out detailed analysis of the efficacy of both games by involving sufficient number of players in near future to help us improve the pedagogical outcome of these games.



(a) Clustering of players for SQL Injection Game



(b) Clustering of players for Firewall Game

Fig. 5 Clustering of players for both games

5 Concluding Remarks and Future Directions

This paper has attempted to describe design of serious games to teach one of the Open Web Application Security Project (OWASP) top 10 vulnerabilities, namely SQL Injection and about network security by using a Firewall.

The focus on UI and UX has been suitably emphasized and the game design process had regular interaction with a student expert in the design aspects. The user interface went through various iterations to ensure that basic tenets of design are manifested in these games. We have attempted to view the subjective feedback from the player besides objective measures using learning analytic.This comprehensive view is expected to help us improve our game design in realistic manner.

We plan to estimate learning outcome by involving adequate number of players, in near future. We hope to muster students' participation under control group settings, when physical participation becomes feasible. Further improvements in terms of content and UI/UX would be attempted based on structured objective analytics and subjective feedback.

References

1. L. McDaniel, E. Talvi, B. Hay, Capture the flag as cyber security introduction, in *2016 49th Hawaii International Conference on System Sciences (HICSS)* (2016), pp. 5479–5486
2. M. Bishop, Teaching computer security, in *SEC*. Citeseer (1993), pp. 65–74
3. M. Hendrix, A. Al-Sherbaz, V. Bloom, Game based cyber security training: are serious games suitable for cyber security training? Int. J. Serious Games **3**(1) (2016). http://journal.seriousgamessociety.org/index.php?journal=IJSG&page=article&op=view&path[]=107
4. J. Jones, X. Yuan, E. Carr, H. Yu, A comparative study of CyberCIEGE game and department of defense information assurance awareness video, in *Conference Proceedings—IEEE SOUTHEASTCON* (2010), pp. 176–180
5. S. Sheng, B. Magnien, Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish, in *Proceedings of SOUPS 2007* (2007), pp. 88–99. http://dl.acm.org/citation.cfm?id=1280692
6. M. Ulicsak, Games in education: serious games. A FutureLab literature review (2010), p. 139. http://www.futurelab.org.uk/projects/games-in-education
7. T.M. Connolly, E.A. Boyle, E. MacArthur, T. Hainey, J.M. Boyle, A systematic literature review of empirical evidence on computer games and serious games. Comput. Educ. **59**(2), 661– 686 (2012). http://www.sciencedirect.com/science/article/pii/S0360131512000619
8. J.L.F. de Almeida, L. dos Santos Machado, Design requirements for educational serious games with focus on player enjoyment. Entertainment Comput. **38**, 100413 (2021)
9. D. Ismailović, J. Haladjian, B. Köhler, D. Pagano, B. Brügge, Adaptive serious game development, in *2012 Second International Workshop on Games and Software Engineering: Realizing User Engagement with Game Engineering Techniques (GAS)*. IEEE (2012), pp. 23–26
10. M. Kuhn, K. Johnson, et al., *Applied Predictive Modeling*, vol. 26 (Springer, Berlin, 2013)
11. W. Westera, R. Prada, S. Mascarenhas, P.A. Santos, J. Dias, M. Guimarães, K. Georgiadis, E. Nyamsuren, K. Bahreini, Z. Yumak et al., Artificial intelligence moving serious gaming: presenting reusable game AI components. Educ. Inf. Technol. **25**(1), 351–380 (2020)
12. H. Fujita, I. Wu, et al., A special issue on artificial intelligence in computer games: AICG (2012)

13. G.N. Yannakakis, J. Togelius, A panorama of artificial and computational intelligence in games. IEEE Trans. Comput. Intell. AI Games **7**(4), 317–335 (2014)
14. G.N. Yannakakis, J. Togelius, *Artificial Intelligence and Games* (Springer, Berlin, 2018)
15. M.C. Gombolay, R.E. Jensen, S.H. Son, Machine learning techniques for analyzing training behavior in serious gaming. IEEE Trans. Games **11**(2), 109–120 (2017)
16. A. Khakpour, R. Colomo-Palacios, Convergence of gamification and machine learning: a systematic literature review. Technol. Knowl. Learning **26**(3), 597–636 (2021)
17. C. Alonso-Fernandez, M. Freire, I. Martinez-Ortiz, B. Fernandez-Manjon, Improving evidence-based assessment of players using serious games. Telematics Inform. **60**, 101–583 (2021)
18. C. Lopez, C. Tucker, Toward personalized adaptive gamification: a machine learning model for predicting performance. IEEE Trans. Games **12**(2), 155–168 (2018)
19. B. Monterrat, E. Lavoue, S. George, Motivation for learning: adaptive gamification for web-based learning environments, in *6th International Conference on Computer Supported Education (CSEDU 2014)* (2014), pp. 117–125
20. B. Monterrat, E. Lavoue, S. George, Toward an adaptive gamification system for learning environments, in *International Conference on Computer Supported Education* (Springer, Berlin, 2014), pp. 115–129
21. K. Patel, A survey on vulnerability assessment and penetration testing for secure communication, in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (IEEE, 2019), pp. 320–325
22. Technologies, Unity game engine. https://unity3d.com/
23. Á. Serrano-Laguna, B. Manero, M. Freire, & B. Fernández-Manjón, A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes. Multimedia Tools Appl. **77**(2), 2849–2871 (2018)
24. D.M.J. Garbade, k-means-clustering (2018). https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1
25. J. Moizer, J. Lean, E. Dell'Aquila, P. Walsh, A. Keary, D. O'Byrne, A. Di Ferdinando, O. Miglino, R. Friedrich, R. Asperges, L. Sica, An approach to evaluating the user experience of serious games. Comput. Educ. **136** (2019)

Remote Accessible Security System with IoT Using LabVIEW

Kalathiripi Rambabu, M. C. Chinnaiah, Sanjay Dubey, and Samatha Yellanki

Abstract In this generation, safety and security have become one of the essential things. So, for having a secured home or industry, we designed a system with IoT using LabVIEW. We are using secured locks and camera, QR code and pass code for more added security for the homes and industrial purposes. We connect the calling bell; power supply, LabVIEW and the remote based IoT system are connected to a microcontroller. In which it is further connected to camera, pass code and QR code to check the person and opens the door if known persons comes using a secured lock which is connected to door and send the alert to owner if an unknown person tries to enter home using IoT. These results were displayed on the LabVIEW. This remote accessible security system is designed for securing the homes and industries with an upgraded system using IoT and interfacing with LabVIEW.

Keywords IoT · LabVIEW · Security · Remote access

1 Introduction

By using Internet of things, we can get an eye on our home from anywhere by just using the Internet. By using IoT, we can connect many devices like phones and TV's everything as the world is in fast pace and getting smart with technology [1]. There are many existing systems for securing homes and industries for gas detecting and some systems for detecting fire catch ups and even for not allowing unknown persons but they are all very expensive and not having much security so that intruders [2] are even escaping from that security systems also. So we designed this system which is

K. Rambabu (✉) · M. C. Chinnaiah · S. Dubey · S. Yellanki
Department of ECE, B V Raju Institute of Technology, Narsapur, Medak, Telangana 502313, India
e-mail: rambabukala@gmail.com

M. C. Chinnaiah
e-mail: chinnaaiah.mc@bvrit.ac.in

S. Dubey
e-mail: sanjay.dubey@bvrit.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_36

more secured and easily detectable system as we are using NodeMCU and IoT [3] and camera, QR code and pass code which will easily detect the robbers and even quickly sends the alert to the owner if any unauthorized person tries to break or enter in to the home. LabVIEW [4] is the software which we can use cost effectively for building this security system. By using this system, we can reduce the thefts and entering of the intruders in to the home or industry.

2 Problem Statement

As the thefts are increasing day by day the new technologies are also emerging and using those innovative technologies to secure the homes and industries in all the best possible ways as the security has become one of the vital thing and a robbery can be done within seconds even if we lock the house also. Every 3 min a robbery is taking place, and we have to reconsider the safety levels as intruders are taking the advantage of lack of security [5] for the home or industries and destroying the properties and even the lives of the people. As per the National Crime Records Bureau 244,119 cases have recorded on robbery. Even the highly secured security cameras also failed to detect the robbers. It has become major problem for the cops to find the thief, so to solve all these problems we have designed this system and can send the alerts to the owner via mobile if any unknown person tries to enter the home.

3 Existing System

There are many existing systems in which they have used IoT remote accessing using GSM module, PIR sensor and some other gas sensors [6] for alerting the people around about the fire and smoke or any harmful gases. They have used motors and sensors which are very expensive. The hardware components they used were DAQ, Raspberry pi are linked to LabVIEW [7]. But DAQ does not have any access to the Wi-Fi [8] and IP address. So we are using a NodeMCU [9] which has access to IP address and can be remotely accessible. Some systems they have used GSM module in which it will directly calls the owner and some security systems for alerting the fire catch ups and detecting gas leakages and alerts the people in the surroundings [10].

4 Proposed System

In this proposed system, we are using NodeMCU. The benefit is with the help of IoT we can access remotely through IP address of NodeMCU [11] from our mobiles. We are using GSM module, alarm buzzer and door unlocking tool which is like as

hydraulic lock, LCD panel and camera will be connected to the NodeMCU. The door open and close will be displayed on the LCD panel according to the pass code entered or QR code scanned. To enter the pass code or start the system, we should click on the calling bell after that the camera, pass code will be turn on and after pressing on the bell if pass code is not entered within 30 s then using IoT [12] it will alert the person in charge by sending the message. So he can monitor the camera which is placed in front of the door. Everything like data will be stored in LabVIEW [13] itself. We used LabVIEW as it is more efficient and gives the results accurately. We used camera, pass code and QR code for more accurate results, and camera takes the picture when unknown person tries to enter and later can be seen by authorized owner.

5 Implementation

5.1 Working Model

The system consists of a NodeMCU ESP32, to which lock tool HCE-12 no electric strike and relay module HW-307 are connected. This system gives the live data to the cloud sent through protocol [14] by ESP32 module. The whole programme is written in the Arduino IDE, and it is dumped into NodeMCU. Once it is connected, it searches for client presence and it will be directed to the LabVIEW through cloud and whole securities are managed in LabVIEW [15]. If the passcode, QR code are correct, then it will again directed to NodeMCU and lock gets open. We can operate the system in mobile or other cloud connected webpages from our house.

5.2 Block Diagram

In the block diagram, NodeMCU is supplied with continuous power supply and that supply is controlled by the relay module HW-307 which is connected to lock HCE-12. From the Wi-Fi module of ESP32, webpage is created and also be controlled from webpage. To communicate with LabVIEW, we are using TCP/IP protocol by that data is transferred to LabVIEW in which entered passcode and QR code are matched with the existed (Fig. 1).

5.3 Flowchart

See Fig. 2.

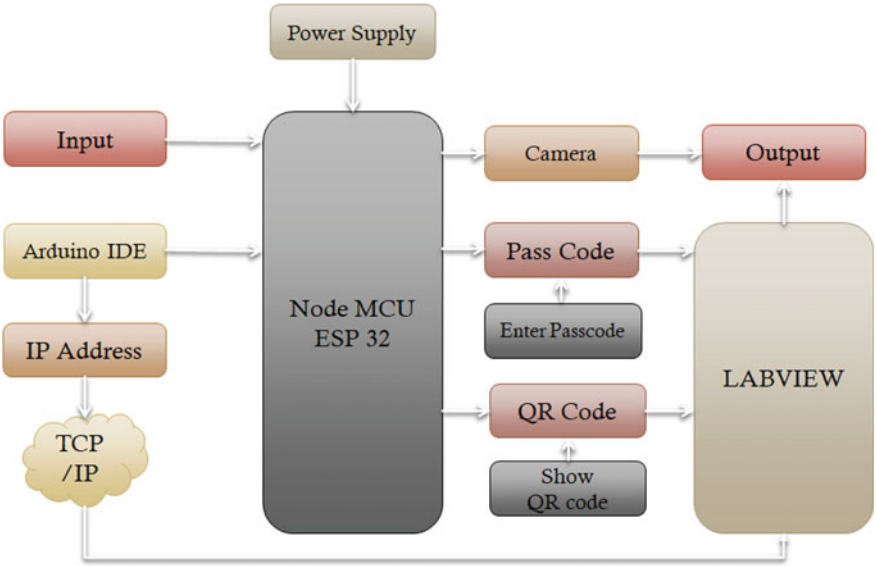


Fig. 1 Block diagram

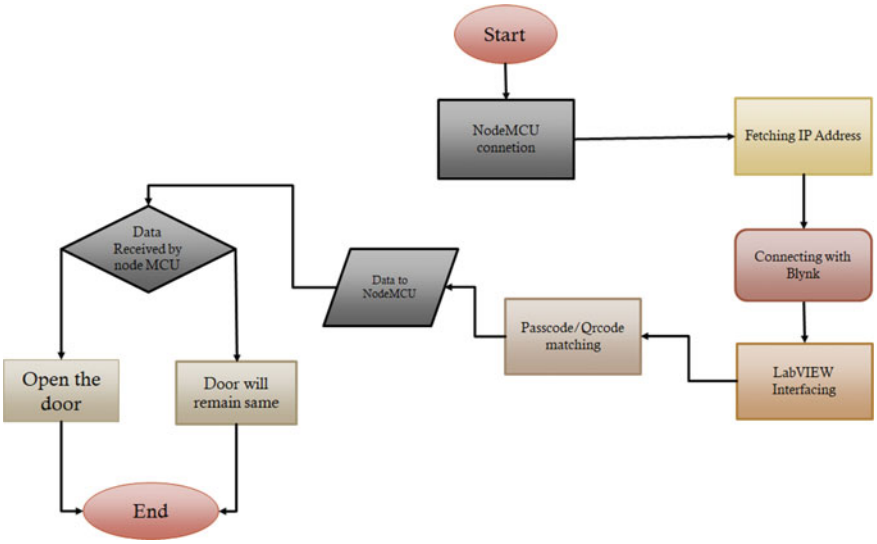


Fig. 2 Flow chart of remote accessible security system

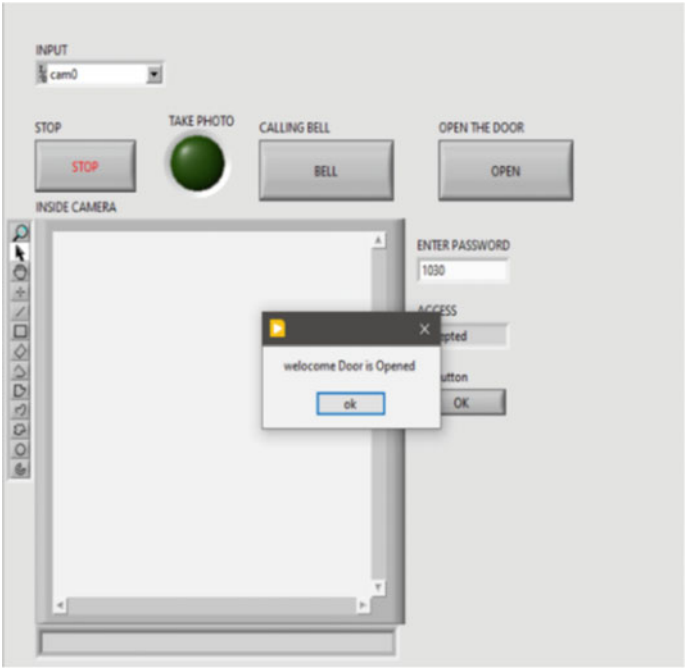


Fig. 3 Passcode output

6 Results

6.1 Software Results

Figures 3 and 4 show the output or results for the passcode and QR code. If they matches with the desired codes, then they will send the signal to the NodeMCU and lock gets unlocked or locked depends on the signal from LabVIEW.

Figures 5 and 6 show the result for the passcode and QR code. Camera gets on when we click on a button or calling bell which acts like a trigger or a starting step for QR code. If QR code and passcode match with the desired code, then it pop-up a window showing "welcome door is opened" and door gets opened. If the desired code not matches with the code, then lock remains locked.

6.2 Hardware Results

NodeMCU is connected to the relay module with 3.3 V output voltage given as supply voltage to the relay module. On other end of relay connected to the 12 V

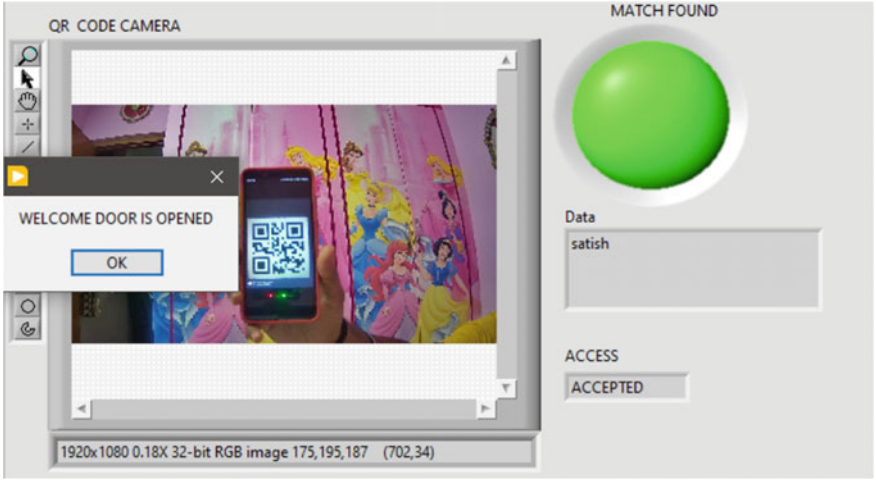
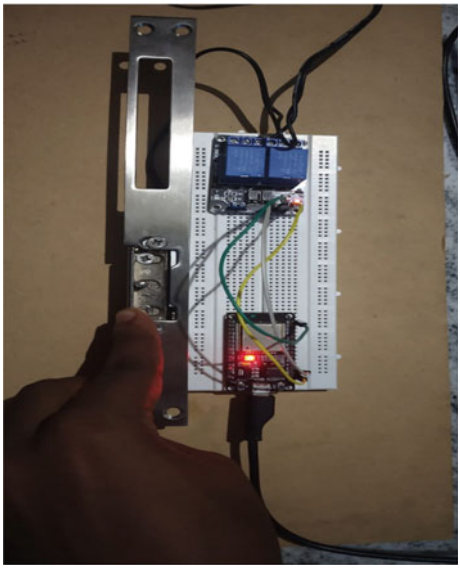


Fig. 4 QR code output

Fig. 5 Lock closed



adapter and connected to the lock. Locking and unlocking operation is decided by signal passed by the NodeMCU.

When the passcode or the QR code entered is wrong, then it remains closed as shown in Fig. 5. The lock opens only when a force is applied towards it. When the passcode or the QR code entered is correct, then it opens as shown in Fig. 6.

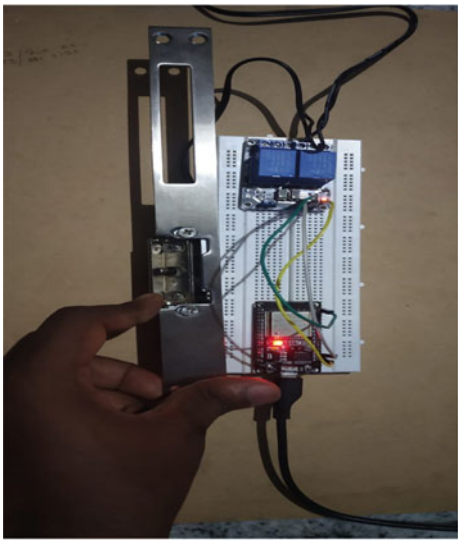


Fig. 6 Lock opened

6.3 Remote Access

In the above figure, two buttons are created, one for the unlocking the lock and other for the alarm alert in surroundings. The third is the video streaming when the camera is connected and power is supplied, then the video capturing in the camera will be visible in this panel (Figs. 7 and 8).

7 Conclusion

There are so many robberies now-a-days by this project the robberies can be controlled or avoidance of robbery will be done. By means of this project, we are focusing on intrusion of unknown persons who may leads to threats for the owners or the persons of the house. We can reduce that intrusion and threats by this particular system using LabVIEW and by connecting through Internet so to access remotely from any corner of the world. We used NodeMCU ESP32 which has access to IP address and connection with IoT helps in securing home or industries more by sending alerts to the owner.

Fig. 7 Mobile application

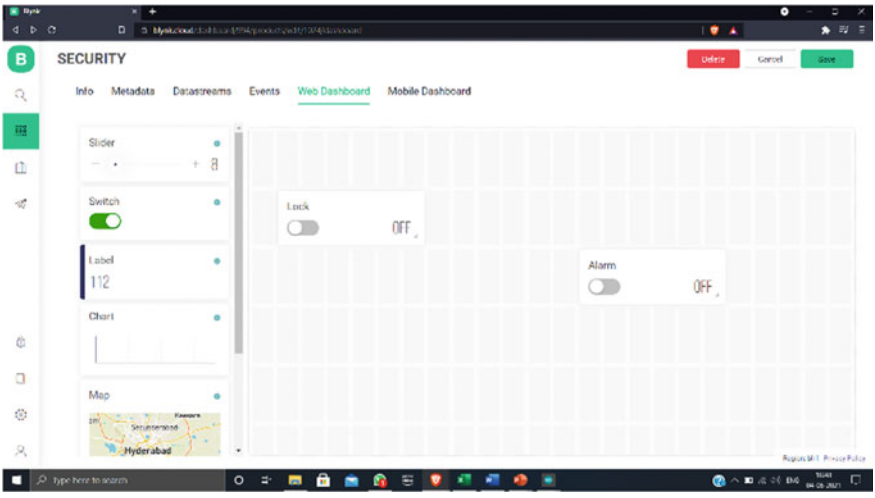


Fig. 8 Web application

References

1. B. Hamid, Design and implementation of smart house control using LabVIEW. IJSCE **1**(6) 98–106 (2012). ISSN: 2231-2307
2. K. Rambabu, N. Venkatram, Contemporary affirmation of security and intrusion handling

strategies of internet of things in recent literature. J. Theor. Appl. Inf. Technol. (JATIT) **96**(9) (2018). ISSN 1992-8645
3. Y. Singh, Internet of things and NodeMCU. JETIR **6**(6) 1085–1088 (2019). ISSN: 2349-5162
4. K. Rambabu, V. Haritha, S. Nikhil Srinivas, P. Sanjana Reddy, IoT based human intrusion detection system using LabVIEW. Int. J. Innov. Technol. Exploring Eng. (IJITEE) **8**(6S4) (2019). ISSN: 2278-3075
5. V. Adat, B.B. Gupta, Security in internet of things: issues, challenge taxonomy, and architecture. Telecommun. Syst. 1–19 (2017)
6. B. Udaya Kumar, D.S. Murty, Ch.R. Phani Kumar, Implementation of low cost ethernet based home security using wireless sensor network. Algorithms Res. **2**(1) 1–7 (2013). https://doi.org/10.5923/j.algorithms.20130201.01.
7. R. Sanudin, W.K. Huei, I.L. Ahmad, Small scale monitoring system in LabVIEW platform. IEEE (2009)
8. I. Baig, C. Mazamil, S. Dalvi, Home automation using Wifi module. Kalsekar Technical Campus (2015)
9. S. Kulakarni, S. Yelpale, O. Pawar, K. Deshpande, R. Singh, P. Shelke, A NodeMCU based home automation system. IRJET **05** (06) 127–129 (2018). e-ISSN: 2395-0056
10. P. Srinivasarao, K. Vamsi Saiteja, K. Prudviraj, N. Prashanth Reddy, R. Tejaswini, Industrial device control using WiFi module. IRE J. **1**(8) 35–39 (2018).
11. K. Ghosh, M. Bhowmick, D. Joddar, Globally controlled multiple relays using NodeMCU (2018). rcciit.org/
12. B.L.R. Stojkoska, K.V. Trivodaliev, A review of internet of things for smart home: challenges and solutions. J. Clean. Prod. **140**, 1454–1464 (2017)
13. J. Meshram, N. Deshmukh, S. Sayed, Home automation using LabVIEW. IJSRD **3**(1) 1104–1106 (2015).
14. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Ayyash, Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun. Surv. Tutorials, **17**(4) 2347–376
15. K. Rambabu, P. Aravind Sai, E. Samuel, P.N.V.S.M. Varama, M. Sumanth, Monitoring and controlling of home security system using IoT and LabVIEW. Int. J. Control Autom. (IJCA) **13**, 596–605 (2020). ISSN: 2005-4297

Human–Machine Interface for Wheelchair Control Using sEMG Signals

M. Gopichand, K. Rajeswari, and E. Deepthi

Abstract Electromyogram (EMG) signal is the electrical activity that is generated by alpha motor neurons in response to the impulse from the brain. EMG is divided into two types: surface EMG and intramuscular EMG. Surface EMG signals provide information about the intensity of muscle activation. EMG signals are used to make myoelectric control system-powered upper-limb prostheses, and electric-powered wheelchairs are the major applications of the myoelectric control system. Tremendous growth has been seen in human–machine interface (HMI) wheelchairs over the last few decades. The manual wheelchair that can be moved by pushing the wheels with hands is replaced by Joystick and a voice-controlled wheelchair. However, even with the advances in technology elderly and paralyzed people have difficulties in intuitive control and navigation of wheelchairs. Therefore, a smart wheelchair based on surface EMG signals and an accelerometer are proposed. The signals from EMG sensors act as input signals which get processed by Arduino Uno, and the output signal based on detected hand gesture is wirelessly transmitted. Proper training is performed to determine the threshold value for each gesture identification. The RF receiver sends the received signal to Arduino mega to process the signal, and Arduino mega sends a command to motor drive to move the wheelchair. The smart wheelchair is controlled in left, right, forward and backward directions. The ultrasonic sensor is being used in the wheelchair to detect obstacles. The hardware design is properly tested and validated; thereby smart wheelchair is cost-effective, easy to use, and safety is ensured.

Keywords EMG · Gesture recognition · Wheel chair control · RF transmitter and receiver · Arduino

M. Gopichand · K. Rajeswari (✉) · E. Deepthi
Department of ECE, Thiagarajar College of Engineering, Madurai, Tamil Nadu 625015, India
e-mail: rajeswari@tce.edu

M. Gopichand
e-mail: gopichand@student.tce.edu

E. Deepthi
e-mail: deepthie@student.tce.edu

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_37

1 Introduction

Human–machine interfaces and brain–machine interfaces are developing technologies with a bright future ahead of them. These devices have gained a lot of importance in recent years because they not only minimize human effort in various industries, but they also provide a great help for disabled people to work normally. Bio-signals produced by our bodies serve as the basis for these technologies. The most important bio-signals to remember are electroencephalography (EEG), electrooculography (EOG), electromyography (EMG) and electrocardiography. Paralyzed or disabled patients need specially designed medical equipment, such as EEG-, EOG- or EMG-controlled robotic hands, prosthetics and smart wheelchairs, that is controlled by these signals. An EEG brain headset is used to monitor a wheelchair, but it could only move in one direction using brain waves, while other directions are chosen using EOG. Advanced bio-sensors with multiple channels are used to achieve 2D regulation of the same. However, these bio-sensors are costly and may not be suitable for real-time applications, with accuracy being the primary concern. So, for this prototype, a smart wheelchair, which will be powered by a low-cost single channel EMG sensor is developed.

Few muscles are regulated actively and willingly by our will, i.e., voluntary muscles for voluntary acts, while others function involuntarily but play an important role in our survival. We normally focus on skeletal muscles in order to support and assist a disabled person. In a few special cases, such as quadriplegia, we must use other muscles such as facial muscles or anterior and posterior side muscles in the leg. Skeletal muscles are linked to the skeleton, which functions under the direction of the person and assists in the movement of various body parts. Muscle fibers in the skeleton are grouped together to form skeletal muscles. The nervous system sends signals to the muscles, which cause them to contract or relax. A muscle can generate enough force to shift the body when its fibers contract in unison, allowing us to move our hand. The EMG records these movement thresholds, which can be used to monitor a variety of devices depending on the application. The use of EMG in HMI is common because EMG recording can be done with surface electrodes and needs very little hardware and software resources, making it a suitable low-cost solution.

A smart wheelchair based on surface EMG signal is designed [1]. Cyber-link sensing device is used to obtain the forehead surface EMG signals generated by facial movements. The autoregressive model is used to compress the data and extract surface EMG features. Then, backpropagation artificial neural network (BPANN) improved by Levenberg–Marquardt algorithm is used to identify different facial movement patterns. The wheelchair makes simple movements like forward, left, right, backward and stop based on facial movements. But, face movements during talking and looking may cause adverse effects.

Innovative driving assistance system of wheel chair, for paralyzed people, is developed using laser dots [2]. A pan-tilt mounted laser is used to track user's head posture for projecting a colored spot on the ground ahead. The wheelchair is provided with depth-camera to model a traversability map. The user can travel to the spot marked by

laser pointer but that spot should also be present in the traversability map; otherwise, the wheel chair will not move and laser dot will glow in red color. If reachable, the laser pointer will turn from red to green color. The wheelchair is then controlled by micro-controller to follow the path to reach the destination.

In [3], the wheel chair is controlled by joystick. Initially, the joystick is in exact middle position so that the motor will not run. When the joystick is moved, the potentiometer sends analog values to Arduino board. Arduino converts it into digital signal. The digital signal is sent to the motor drive for controlling the DC motors. It makes the motors to move forward, backward, left or right based on the signal. This requires a person at least to move his forehand forcefully. Stroke patients' mobility is limited as some of their limbs do not function properly. But the brain signals of stroke patients are observed as normal. Hence, EEG signals are used as input for wheelchair to rehabilitate post-stroke patient (Prajitno et al. [4]). A technology called brain–computer interface (BCI) is used to allow computer to make action based on brain signals. Electroencephalography is used to record brain waves. The Neurosky Mindwave Mobile 2 headset is used as input signal. The obtained signal is processed and classified by MATLAB software. It classifies the signal for moving wheelchair in four directions. The output signal is fed to Arduino board to control the motors of wheelchair.

The wheelchair-assisted system based on voice of a person is developed [5]. The user needs to download a voice recognition application in android phone. It should be connected to a Bluetooth device. Then, the user can speak specific commands which will be processed by the application. The word command is converted into text by google voice service. The text format is processed by micro-controller; it checks for valid input and gives command to motor drive for moving wheelchair forward, left, backward right and stop. As second part, the user can use an Android device with a GUI app to submit commands by pressing a specific button, and the wheelchair will react appropriately. The disadvantage of this system is that it is sensitive to pronunciations accuracy and clarity of voice of the user.

A wearable electromyography device is made to interact with computing systems and associated electrical devices [6]. Sensing, preprocessing, feature extraction and classification are the four main components [7–9]. EMG electrodes are placed on the arm muscles, mainly the flexor carpi radialis and the palmaris longus, for sensing signal. Preprocessing of signal is done for reducing noises. The extraction of features including common statistical features is done, and the patterns that represent gesture movements are matched by using the minimum distance classifier technique. The wristwatch can measure biological impedance body movements, heart beat rate, respiration.

The EMG electrodes are mounted on the upper trapezium muscles for signal acquisition. The trapezium EMG signals are analyzed and converted into control movements like forward and reverse that are fed into a PIC microcontroller. The wheel chair considers two forms of neck muscle rotations when navigating: flexion and lateral rotations. EMG-based wheelchair with robotic manipulator is designed for persons with transradial amputation [10]. EMG-controlled wheelchair for disable patients is available, but they use computationally expensive neural network for

classification [11]. In the literature, even though a number of works are available to control wheelchair signals [4, 12–14], in this work, two simple hand gestures are used to control the movement of wheelchair and its implementation is easier.

2 Materials and Methods

Smart wheelchair consists of two parts: wheelchair transmitter and wheelchair receiver. Myoware muscle sensor and accelerometer are used as input sensors in wheelchair transmitter. Arduino Uno compares the signal from these sensors with threshold value and sends command through RF transmitter. In wheelchair receiver, RF receiver captures the transmitted signal and sends it to Arduino Mega for processing of signal. After processing the signal, Arduino Mega sends command to motor drive to run the motor in desired direction. Ultrasonic sensor is used to avoid obstacle by smart wheelchair.

In EEG-based driving assistance, more than ten electrodes are placed in forehead to sense the EEG signal. The electrode can be used for at least two times, after that it losses adhesive capacity. This increases the maintenance cost. Before placing the electrodes, surface of skin should be cleaned and hair should be removed. This makes the user to feel difficult in using the wheelchair for day-to-day life. In accelerometer-based driving assistance, the hand should not be tilted unwantedly which may trigger the signal of accelerometer. Hence, a combination of Myoware sensor, accelerometer and ultrasonic sensor is proposed as an upgrade to wheelchair.

2.1 Wheelchair Transmitter Module

The wireless control of wheel chair through hand gesture is the methodology used here. In transmitter part, accelerometer and Myoware muscle sensor acts as input, Arduino uno as motherboard, RF transmitter as wireless transmitter. Figures 1 and 2 show the physical and functional block diagram of wheel chair transmitter module. Figure 3 presents the actual hardware implemented for the transmitter. Myoware sensor and accelerometer are used as sensing elements. Myoware sensors are placed in forearm on the brachioradialis muscle for EMG signal acquisition. Contraction of these muscles generates an EMG signal which can be acquired through the Myoware sensors. Index finger movement and hand closure are the two movements used to move the wheelchair in left and right directions. The signal acquired from Myoware sensors is fed to the Arduino Uno which is microcontroller board based on ATmega328P. This board is used as motherboard to process the captured EMG signal. RF transmitter module is connected to Arduino Uno for wireless transmission of the signal.

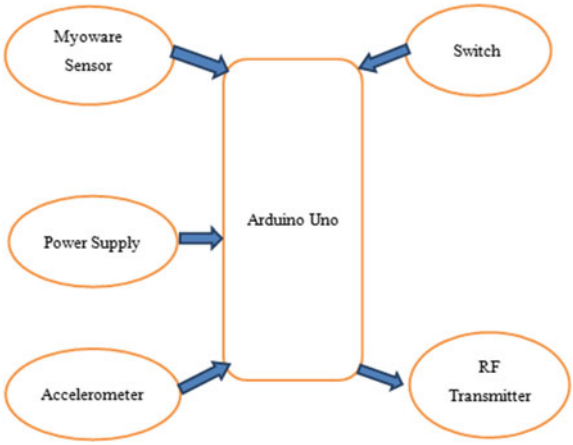


Fig. 1 Physical architecture of transmitter module

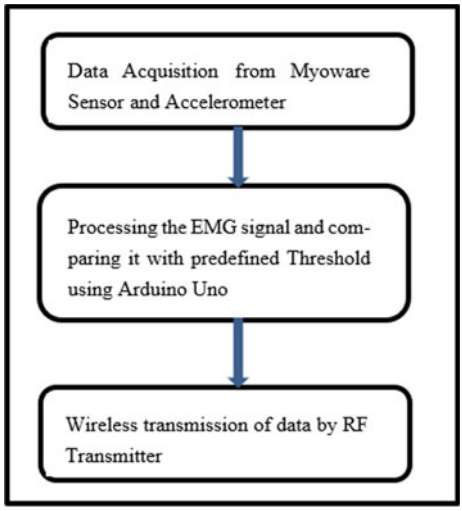


Fig. 2 Functional block diagram wheelchair transmitter module

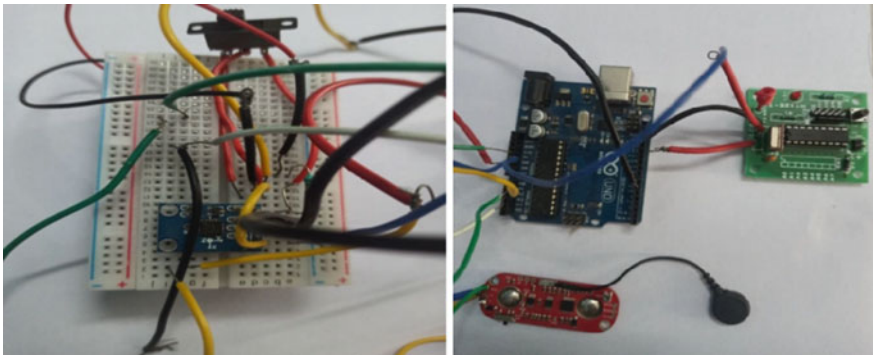


Fig. 3 Actual hardware of wheel chair transmitter module

Arduino Uno is uploaded with predefined threshold for four movements of wheel chair. When the intended movement takes place, the acquired EMG signals are processed and compared with predefined threshold by Arduino Uno. The EMG signal value corresponding to the identified movement is sent wirelessly to the wheelchair through RF transmitter.

2.2 Wheelchair Receiver Module

Figure 4 shows the physical architecture of wheel chair receiver module, and Fig. 5 shows the functional block diagram of the receiver. Figure 6 presents the actual hardware used for implementing the module. The receiver part of wheel chair has RF receiver, Arduino Mega and ultrasonic sensor. The signal transmitted from transmitter module is received by the RF receiver module that consists of RF tuned circuit, a pair of operational amplifiers, PLL and a decoder. The received signal is amplified by the operational amplifiers and fed into PLL, which allows the decoder to lock into a stream of digital bits, resulting in improved decoded output and noise immunity. Arduino Mega is used as motherboard to process the incoming signal obtained through RF receiver module. After processing, to control the movement of wheelchair, signal from Arduino mega is fed to motor drive which is responsible for driving DC motors connected to the wheelchair. L298N motor drive used is able to control the speed and direction of two DC motors simultaneously. The DC motors enable the wheelchair to be moved in all four directions.

Module for Obstacle Avoidance: The ultrasonic sensor monitors the environment and provides distance between wheel chair and surrounding. If there is an obstacle, Arduino Mega sends signal to motor drive to stop the wheel chair and move in reverse direction till the distance increases to threshold value. Now, user will decide

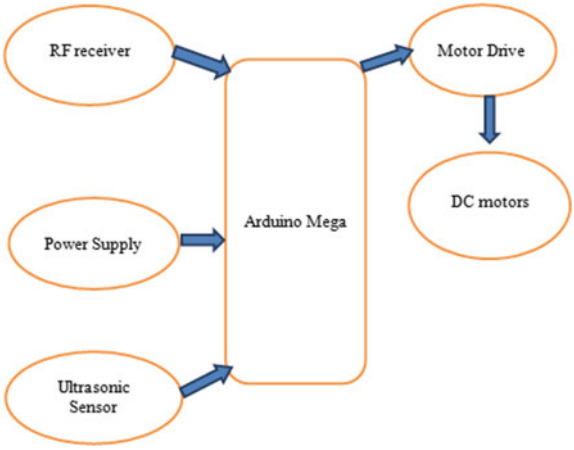


Fig. 4 Physical architecture of wheelchair receiver module

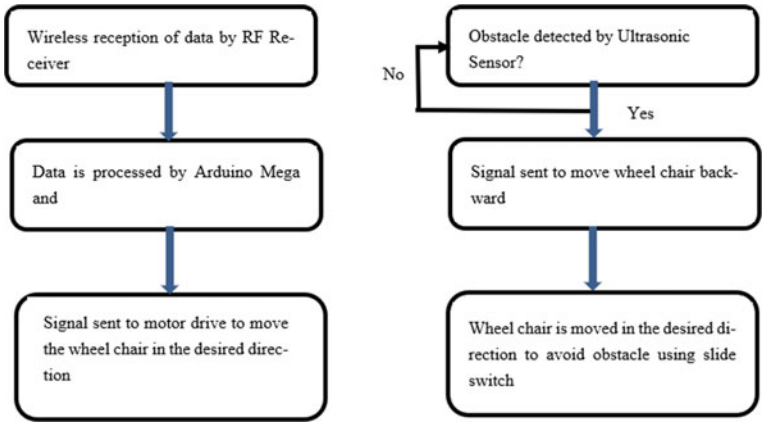


Fig. 5 Functional block diagram of wheelchair receiver module

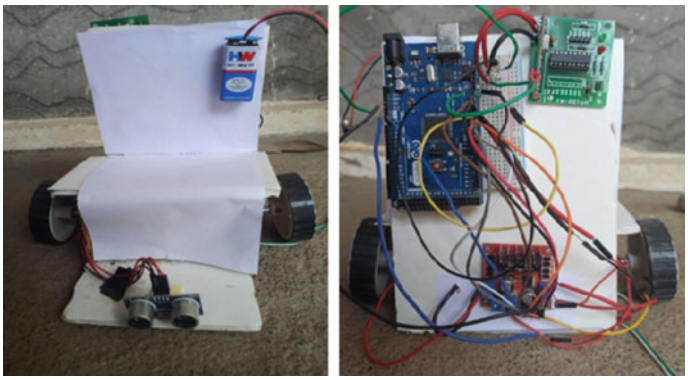


Fig. 6 Actual hardware of wheel chair receiver module (left: frontside of wheelchair, right: backside of wheelchair)

the direction that avoids obstacle. User can also use switch in wheel chair transmitter which makes Arduino Mega send signal to motor drive that moves wheel chair in left, forward, right sequence to avoid the obstacle.

3 Results and Discussion

The Ardunio IDE handles all of the programming, and the application is loaded onto the Arduino board. Forward and reverse movements are carried out using accelerometer, whereas right and left movements are performed using different hand gestures.

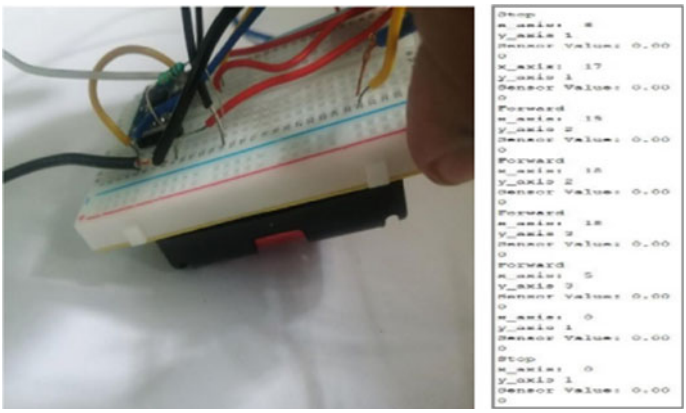


Fig. 7 Accelerometer tilted forward (left) and output (right)

Training is carried out to decide the threshold level for performing each movement using corresponding direction of tilt of the accelerometer and gesture.

3.1 Forward Movement

When the accelerometer is tilted forward at the transmitter side as in Fig. 7, the *X*-axis value gets increased. Arduino Uno compares the obtained value with predefined *X*-axis value. If present value is higher than value 15 (corresponding command output shown in right side of Fig. 7), signal to move wheel chair forward is send through RF transmitter. The threshold values are obtained by proper training with the desired gesture. Arduino Mega processes the received signal and sends signal to motor drive to move the wheel chair forward.

3.2 Backward Movement

When the accelerometer is tilted backward as in Fig. 8, the *X*-axis of the accelerometer value gets increased. Arduino Uno compares the obtained value with predefined *X*-axis value. If present value is lesser than value 15 (corresponding command output shown in right side of Fig. 8), signal to move wheel chair backward is send through RF transmitter. Arduino Mega processes the received signal and sends signal to motor drive to move the wheel chair backward.

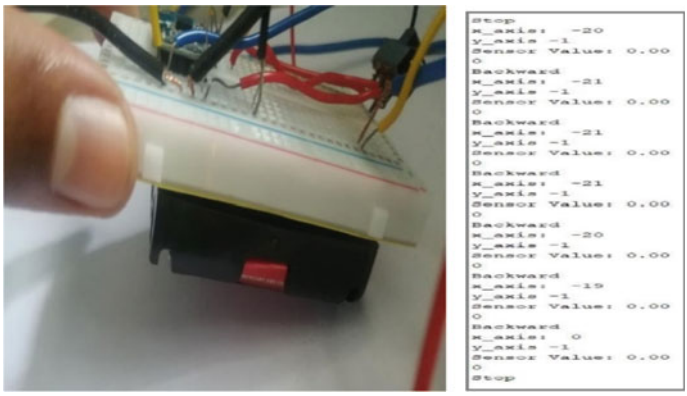


Fig. 8 Accelerometer tilted backward (left) and output (right)

3.3 Left Movement

When hand gesture shown in Fig. 9 is applied, the sensor value of Myoware muscle sensor gets increased. Arduino Uno compares the obtained value with predefined sensor value, and if value is between 60 and 150, it sends signal to move wheel chair left through RF transmitter. Here also, the threshold values by proper training with the desire gesture. RF receiver receives the transmitted signal, and Arduino Mega processes the received signal, thereby sending signal to motor drive to move the wheelchair left.

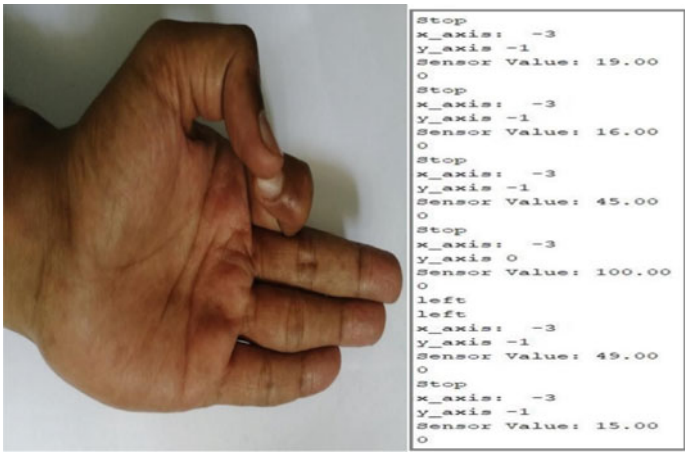


Fig. 9 Hand gesture for left movement (left) and output (right)

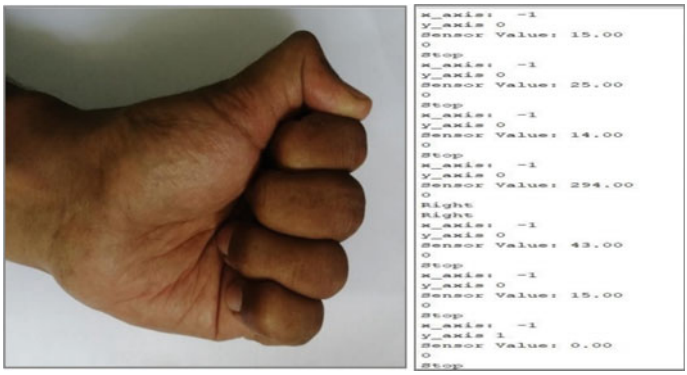


Fig. 10 Hand gesture for right movement (left) and output (right)

3.4 Right Movement

When hand gesture shown in Fig. 10 is applied, the sensor value gets increased. Arduino Uno compares that value with predefined sensor value. If it is between 160 and 300, signal to move wheel chair right through RF transmitter is sent. The signal received by RF receiver is processed by Arduino Mega and signal sent to move the wheel chair right.

3.5 Obstacle Avoidance

Ultrasonic sensor continuously monitors the environment for obstacle and provides distance between wheel chair and surrounding. If the distance value less than 10, it sends signal to Arduino that an obstacle is present in the path. Arduino Mega sends command to motor drive to take reverse direction till the distance value gets 10. Now, user will decide the direction that avoids obstacle for the wheelchair. The user can press switch in transmitter which makes wheelchair to move in left, forward and right direction to avoid obstacle. Figure 11 is used to depict this process.

4 Conclusion and Future Work

Hand gesture recognition has become a very active research subject in recent years, because of its possible use in human–machine interaction. A novel EMG-based HMI device is designed for gesture-based control of an intelligent wheelchair. The benefit of a surface electromyogram is that it is simple to record and non-invasive. The system is used to control the simple movements of the intelligent wheelchair, such

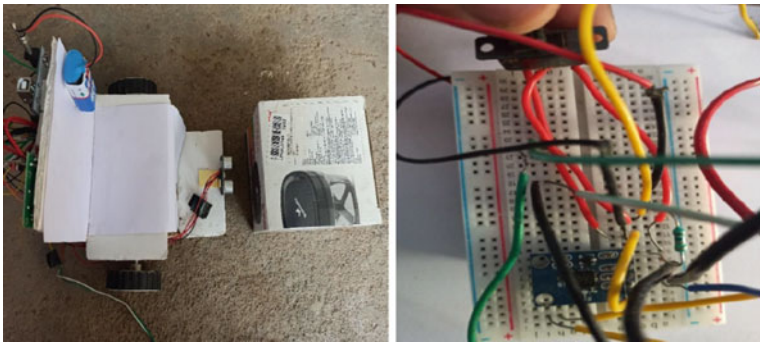


Fig. 11 Detection of obstacle (left), switch pressed in transmitter module (right)

as forward, left, right, backward and stop. The method is real-time and has a high rate of recognition. It has an important research value in many aspects of real-world applications such as assisting the elderly and disabled for their rehabilitation. It is provided with obstacle avoidance mechanism which helps the user to avoid obstacle.

The prototype is cost-effective and the mechanism helps user to control the wheelchair, it has a few drawbacks. The patient must be trained to operate the above-mentioned mechanisms for the wheelchair to work properly, which is a drawback of this model. If these limitations are resolved, these types of assisting wheelchairs can be very useful for patients on a daily basis. The Myoware sensor is used for controlling the wheelchair in left and right directions. In future, Myoware sensor can be used for controlling all the four directions with help of machine learning algorithm.

References

1. Z. Yi, D. Lingling, L. Yuan, H. Hu, Design of a surface EMG based human-machine-interface for an intelligent wheelchair, in *International Conference on Electronic Measurement & Instruments* (2011)
2. E. Rohmer, P. Pinheir, E. Cardozo, M. Bellone, G. Reina, Laser based driving assistance for smart robotic wheelchairs, in *IEEE 20th Conference on Emerging Technologies and Factory Automation*, 05 (2015)
3. T. Saharia, J. Bauri, C. Bhagabati, Joystick controlled wheelchair. Int. Res. J. Eng. Technol. **4** (2017)
4. P. Prajitno, K. Permana, Controlled wheelchair based on brain computer interface using neurosky mindwave mobile 2, in *AIP Conference Proceedings*, 04 (2019)
5. M.I. Malik, T. Bashir, O.F. Khan, Voice controlled wheel chair system. Int. J. Comput. Sci. Mob. Comput. **6** (2017)
6. M.S. Khan, Real time monitoring application using sEMG: wrist watch, in *IEEE International Conferences on Advances in Engineering and Technology Research* (2014)
7. R. Merletti, D. Farina, *Surface Electromyography: Physiology, Engineering and Application* (Wiley, 2016)

8. Y. Omama, C. Hadda, M. Machaalany, A. Hamoudi, M. Hajj-Hassan, M. Abou Ali, L. Hamawy, Surface EMG classification of basic hand movement, in *Fifth International Conference on Advances in Biomedical Engineering* (2019)
9. Z.A. Choudry, S. Aziz, M.A. Imtiaz, S.Z.H. Naqvi, Biometric authentication based on EMG signals of speech, in *2nd International Conference on Electrical, Communication and Computer Engineering* (2020)
10. R.A.M. Abayasiri, A.G.B.P. Jayasekara, R.A.R.C. Gopura, K. Kiguchi, EMG based controller for a wheelchair with robotic manipulator, in *2021 3rd International Conference on Electrical Engineering (EECon)* (2021), pp. 125–130. http://doi.org/10.1109/EECon52960.2021.9580949
11. A. Turnip, D.E. Kusumandari, G.W.G. Arson, D. Setiadikarunia, Electric wheelchair controlled-based EMG with backpropagation neural network classifier, in *Cyber Physical, Computer and Automation System. Advances in Intelligent Systems and Computing*, vol. 1291, ed. by E. Joelianto, A. Turnip, A. Widyotriatmo (Springer, Singapore, 2021). http://doi.org/10.1007/978-981-33-4062-6_13
12. A.G. Dandamudi, D.N. Rao, V.P. Aravilli, R. Sunitha, Single channel electromyography controlled wheelchair implemented in virtual instrumentation, in *3rd International Conference on Computing Methodologies and Communication* (2019)
13. A. Zakaria, H.A. Azeez, N.A. Rahim, M.J.A.B. Safar, An intelligent EMG controlled wheelchair—a review. Int. Acad. Eng. **1**(3) (2016)
14. J. Vigliotta, J. Cipleu, A. Mikell, R. Alba-Flores, EMG controlled electric wheelchair, in *Intelligent Systems and Applications. IntelliSys 2021*. Lecture Notes in Networks and Systems, vol. 296, ed. by K. Arai (Springer, Cham, 2021). http://doi.org/10.1007/978-3-030-82199-9_29

Component Probabilistic Oversampling-Based Classification for Prediction of Dyslexia

M. Shyamala Devi, R. Aruna, Mudragada Ravi Kiran, K. Puneeth, and Tatiparthi Chakradhar Reddy

Abstract Dyslexia is a specific learning medical problem characterized by difficulty reading because of difficulties trying to identify sounds of speech and acquiring knowledge of how they relate to words and letters (decoding). Dyslexia is a reading disability, directly affects areas of the brain that process language. People with dyslexia typically have high intelligence and sense of direction. With mentoring or a specialized educational intervention, most kids with dyslexia can improve learning. With this overview, this project aims to predict the dyslexia kids by using the Dyslexia dataset retrieved from the KAGGLE machine learning repository. The Dyslexia dataset contains 197 features and 3645 student details and it is preprocessed with encoding and missing values. The raw dataset is applied with all classifiers with and without feature scaling to analyze the performance metrics. The raw dataset is subjected to apply with principal component analysis with 40 components with and without feature scaling and then applied with various classification algorithms and performance metrics are compared. The target data distribution is found to have 89.2% of normal people and 10.8% of Dyslexia people, which is imbalanced target. So the 40 component PCA reduced dataset is then applied with oversampling methods like borderline, smote, SVM smote, and ADASYN methods. The PCA oversampled dataset by the above methods are then experienced with various classifiers to predict the dyslexia before and after feature scaling and the performance indices are precision, recall, fscore, and accuracy. Experimental results show that the Random forest classifier shows accuracy of 88% before applying principal component analysis. The random forest classifier shows the accuracy of 96% after applying the principal component analysis along with all the oversampling methods.

Keywords Machine learning · Classifier · Scaling · Classification · Sampling

M. Shyamala Devi (✉) · R. Aruna · M. R. Kiran · K. Puneeth · T. C. Reddy
Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India
e-mail: shyamaladevim@veltech.edu.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_38

1 Introduction

Dyslexia has existed for many years and has already been defined in various ways. The World Federation of Neurologists, for instance, defined dyslexia in 1968 as a problem in children who, contrary to the conventional classroom experience, fail to attain linguistic skills of reading, writing, and spelling comparable with their cognitive capabilities. Dyslexia is the most common teaching disability throughout children and can last a lifetime. Dyslexia can range in severity from mild to severe. The earlier dyslexia is addressed, the better the outcome. Kids with dyslexia, on the other hand, can learn to enhance the learning experience at any age. In the early elementary of school, dyslexia can go unrecognized. The problem of learning to read can frustrate children. The paper is organized in which literature review is discussed in Sect. 2 followed by the paper contributions in Sect. 3. The implementation setup and results are discussed in Sect. 4 and concluded in Sect. 5.

2 Literature Review

This paper [1] proposed a philosophy that will work with the improvement of learning techniques and innovations that are lined up with dyslexia types and manifestations. The paper begins with a conversation of space cosmology and inspects how learning destinations that think about an understudy's abilities and requirements can be coordinated with proper assistive innovation to convey powerful e-learning encounters and instructive assets that can be reliably utilized. The metaphysics utilized inside this review was created utilizing Ontology Web Language, a data handling framework that permits applications to deal with both the substance and the introduction of the data accessible on the web. In this paper [2], the proposed inspiration model for individuals with dyslexia in their utilization of assistive learning devices. To this end, it completes a limited-scale exact review in genuine learning situations. Then, at that point, this paper conduct individual meetings to acquire direct information about the critical factors and elements influencing the learning experience. Utilizing coding and topical investigation techniques, this paper examines principle subjects with respect to persuasive variables and their interrelationships distinguished. In light of these discoveries, custom-made model toward understudies with dyslexia, which will assist with upgrading learning encounters and further develop the learning proficiency.

This study [3] recommends a compensatory job of the cerebellum in perusing for kids with dyslexia. What's more, a propensity for higher cerebellar initiation

in dyslexic people was found in the phonological errand. In addition, the practical network was more grounded for dyslexic kids compared with normally creating per users between the lobule VI of the right cerebellum and the left fusiform gyrus during the orthographic assignment and between the lobule VI of the left cerebellum and the left supramarginal gyrus during the phonological undertaking. This example of results recommends that the cerebellum make up for perusing debilitation through the associations with explicit cerebrum areas liable for the continuous understanding errand. These discoveries improve our comprehension of the cerebellum's association in perusing and understanding weakness. In this paper [4], proposed model is planned and executed to assist with uncovering the basic issues that might influence figuring out how to peruse or compose just as issues that may likewise cause issues with remembers cognizance. This model is carried out to help advocates and guardians comprehend the trouble and get kid in the right way to training achievement. In this paper [5], an expectation model has been recommended that utilizes factual strategies to separate dyslexics from non-dyslexics utilizing their eye development. The eye developments are followed an eye tracker. Eye development has many elements like obsessions, saccades, drifters, and bends. From the crude information of eye tracker, undeniable level highlights are separated utilizing Principal Component Analysis. This paper proposes a Particle Swarm Optimization based Hybrid Kernel SVM-PSO for the expectation of dyslexia in people. The proposed model gives better prescient precision of 95% contrasted with a Linear SVM model. In this study [6], the review with 178 members communicating in various dialects language with and without dyslexia utilizing an online game worked with melodic and visual components that are language autonomous. The review uncovers eighth game measures with huge contrasts for Spanish youngsters with and without dyslexia, which could be utilized in future work as a reason for language autonomous discovery.

In this paper, [7] the measures were registered from those diagrams, and 37 False Discovery Rate rectified elements were chosen as contribution to a Support Vector Machine and a typical K Nearest Neighbors classifier. Cross-approval was utilized to survey the AI execution and irregular rearranging to guarantee the exhibition propriety of the classifier and stay away from highlights overfitting. The best presentation was for the SVM utilizing a polynomial part. Kids were characterized with 95% precision dependent on nearby organization highlights from various recurrence groups. The programmed arrangement methods applied to EEG chart measures were demonstrated to be both strong and solid in recognizing run-of-the-mill. In this paper [8], the numerical and AI based procedures is suggested to naturally find novel complex highlights that consider dependable qualification among Dyslexic and Normal Control Skilled perusers and to approve the supposition that the greater part of the contrasts among Dyslexic and Skilled perusers situated in the left side of the equator. Strangely, these apparatuses additionally highlighted the way that High

Pass signals indeed contain critical important data. At last, the proposed plan can be utilized for examination of any ERP-based investigations. In this paper [9], the proposed approach that uses Bag of Features picture in order to foresee understudy commitment toward the learning content. The commitment expectation was somewhat utilizing front facing face of the 30 understudies. This paper utilizes Speeded-Up Robust Feature central issue descriptor and grouped utilizing k-Means technique for codebook in this BOF model. In this paper [10], it identifies some extraordinary EEG designs relating to dyslexia, which is a learning incapacity with a neurological beginning. In spite of the fact that EEG signals hold significant bits of knowledge of mind practices, uncovering these experiences are not generally direct because of its intricacy. The aqueous solubility nature of the compound is found through machine learning algorithms.

3 Our Contributions

The overall architecture of the work is shown in Fig. 1. The following contributions are provided in this work.

- Firstly, the Dyslexia dataset with 197 features and 3645 student details are preprocessed with encoding and missing values.

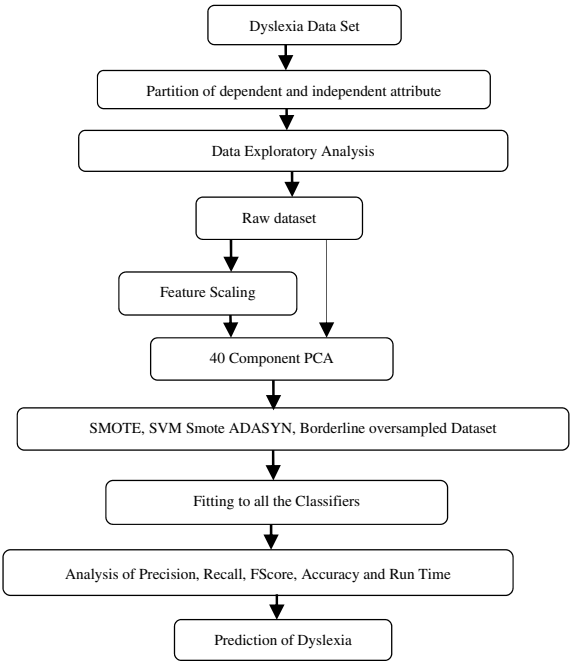


Fig. 1 Architecture system workflow

- Secondly, the raw dataset is applied with all classifiers with and without feature scaling to analyze the performance metrics.
- Thirdly, the raw dataset is subjected to apply with principal component analysis with 40 components with and without feature scaling and then applied with various classification algorithms and the performance metrics are compared.
- Fourth, the target distribution is found to have 89.2% of normal people and 10.8% of Dyslexia people, which is imbalanced target. Therefore, the 40 component PCA reduced dataset is then applied with oversampling methods like borderline, smote, SVM smote, and ADASYN methods.
- Fifth, the PCA oversampled dataset by the above methods are then experienced with various classifiers to predict the dyslexia before and after feature scaling and the performances indices are precision, recall, fscore, and accuracy.

4 Implementation Setup

The Dyslexia dataset with 3645 rows and 197 feature attributes from KAGGLE repository is subjected with the data preprocessing such as refilling the missing values, Encoding of the categorical variables. The target distribution of the dataset is shown in Fig. 2 and the raw dataset is applied with all classifiers and is shown in Table 1.

The dataset is applied with principal component analysis with 40 components and then fitted to all the classifiers to predict the dyslexia with and without feature scaling and performance is analyzed, and shown in Table 2.

The 40-component PCA is then fitted with Borderline oversampling and then fitted to all the classifiers to predict the dyslexia with and without feature scaling and performance is analyzed, and shown in Table 3 and Fig. 3.

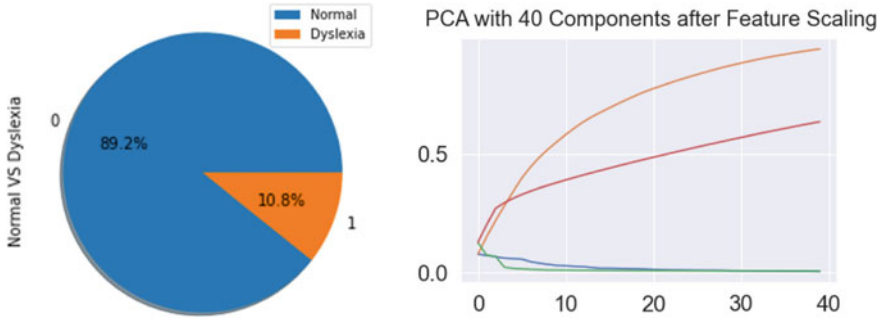


Fig. 2 Target distribution of the dyslexia dataset and 40 component PCA

Table 1 Classifier performance indices of raw dataset before and after feature scaling

Classifier	Before feature scaling					After feature scaling				
	Precision	Rcall	*F*Score	Accu	RunTime	Precision	Recall	*F*Score	Accu	RunTime
LReg	0.85	0.87	0.85	0.87	0.09	0.86	0.88	0.87	0.88	0.11
KNN	0.86	0.88	0.84	0.88	0.77	0.87	0.88	0.83	0.88	0.78
KSVM	0.76	0.87	0.81	0.87	1.02	0.85	0.87	0.82	0.87	0.88
GNB	0.81	0.36	0.42	0.36	0.02	0.81	0.36	0.42	0.36	0.02
Dtree	0.83	0.84	0.83	0.84	0.14	0.83	0.84	0.83	0.84	0.11
Etree	0.82	0.83	0.82	0.83	0.02	0.82	0.83	0.82	0.83	0.02
RFor	0.86	0.88	0.83	0.88	0.17	0.86	0.88	0.83	0.88	0.16
AdaB	0.88	0.89	0.88	0.89	0.82	0.88	0.89	0.88	0.89	0.82
Ridge	0.85	0.88	0.83	0.88	0.08	0.85	0.88	0.83	0.88	0.02
RCV	0.84	0.87	0.83	0.87	0.08	0.84	0.87	0.83	0.87	0.05
SGD	0.86	0.88	0.87	0.88	0.09	0.87	0.87	0.87	0.87	0.10
PAg	0.85	0.87	0.86	0.87	0.06	0.86	0.84	0.85	0.84	0.02
Bagg	0.85	0.88	0.86	0.88	1.04	0.86	0.88	0.85	0.88	1.05

Table 2 Classifier performance of 40 component PCA before and after feature scaling

Classifier	Before feature scaling					After feature scaling				
	Precision	Rcall	*F*Score	Accu	RunTime	Precision	Recall	*F*Score	Accu	RunTime
LReg	0.85	0.87	0.85	0.87	0.12	0.86	0.88	0.87	0.88	0.12
KNN	0.86	0.88	0.84	0.88	0.78	0.87	0.88	0.83	0.88	0.79
KSVM	0.76	0.87	0.81	0.87	1.02	0.85	0.87	0.82	0.87	0.90
GNB	0.81	0.36	0.42	0.36	0.02	0.81	0.36	0.42	0.36	0.02
Dtree	0.83	0.84	0.83	0.84	0.12	0.83	0.84	0.83	0.84	0.11
Etree	0.82	0.83	0.82	0.83	0.02	0.82	0.83	0.82	0.83	0.00
RFor	0.86	0.88	0.83	0.88	0.16	0.86	0.88	0.83	0.88	0.16
AdaB	0.88	0.89	0.88	0.89	0.87	0.88	0.89	0.88	0.89	0.83
Ridge	0.85	0.88	0.83	0.88	0.02	0.85	0.88	0.83	0.88	0.02
RCV	0.84	0.87	0.83	0.87	0.07	0.84	0.87	0.83	0.87	0.05
SGD	0.86	0.87	0.87	0.87	0.09	0.86	0.88	0.86	0.88	0.08
PAg	0.84	0.78	0.80	0.78	0.03	0.86	0.85	0.86	0.85	0.03
Bagg	0.86	0.88	0.86	0.88	1.11	0.87	0.89	0.86	0.89	0.99

Table 3 Classifier performance of borderline oversampling before and after feature scaling

Classifier	Before feature scaling					After feature scaling				
	Precision	Rcall	*F*Score	Accu	RunTime	Precision	Recall	*F*Score	Accu	RunTime
LReg	0.85	0.85	0.85	0.85	0.20	0.88	0.88	0.88	0.88	0.20
KNN	0.82	0.79	0.78	0.79	2.23	0.83	0.77	0.76	0.77	3.67
KSVM	0.79	0.76	0.75	0.76	7.34	0.94	0.94	0.94	0.94	4.31
GNB	0.72	0.67	0.65	0.67	0.03	0.72	0.67	0.65	0.67	0.03
Dtree	0.91	0.91	0.91	0.91	0.51	0.91	0.91	0.91	0.91	0.51
Etree	0.86	0.86	0.86	0.86	0.02	0.86	0.86	0.86	0.86	0.01
RFor	0.96	0.96	0.96	0.96	0.79	0.96	0.96	0.96	0.96	0.78
AdaB	0.93	0.93	0.93	0.93	3.62	0.93	0.93	0.93	0.93	3.34
Ridge	0.87	0.86	0.86	0.86	0.02	0.86	0.86	0.86	0.86	0.02
RCV	0.86	0.86	0.86	0.86	0.10	0.87	0.86	0.86	0.86	0.11
SGD	0.82	0.80	0.80	0.80	0.26	0.81	0.81	0.81	0.81	0.19
PAg	0.81	0.79	0.79	0.79	0.09	0.85	0.85	0.85	0.85	0.05
Bagg	0.95	0.95	0.95	0.95	4.27	0.94	0.94	0.94	0.94	4.51

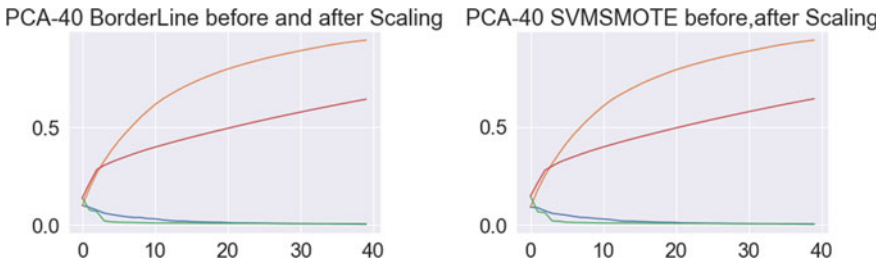


Fig. 3 PCA 40 component of borderline and SVM SMOTE oversampling

The 40-component PCA is then fitted with SMOTE oversampling and then fitted to all the classifiers to predict the dyslexia with and without feature scaling and performance is analyzed, and shown in Table 4.

The 40-component PCA is then fitted with SVM SMOTE oversampling and then fitted to all the classifiers to predict the dyslexia with and without feature scaling and performance is analyzed, and shown in Table 5.

The 40-component PCA is then fitted with ADASYN oversampling and then fitted to all the classifiers to predict the dyslexia with and without feature scaling and performance is analyzed, and shown in Table 6.

The dataset is applied to oversampling and the distribution of the target Dyslexia feature is shown in Fig. 4.

5 Conclusion

This paper explores the performance of classifying the dyslexia based on the target class distribution. The dyslexia dataset is examined to explore the imbalanced target class and the dataset have large number of 197 features. So the dataset is reduced with principal component analysis with 40 features. This 40-component PCA is applied with the oversampling methods like smote, SVM smote, ADASYN and Borderline smote to verify the performance of the classifiers toward predicting the dyslexia people. Experimental results show that the Random forest classifier shows accuracy of 88% before applying principal component analysis. The random forest classifier shows the accuracy of 96% after applying the principal component analysis along with all the oversampling methods.

Table 4 Classifier performance of SMOTE oversampling before and after scaling

Classifier	Before feature scaling					After feature scaling				
	Precision	Rcall	*F*Score	Accu	RunTime	Precision	Recall	*F*Score	Accu	RunTime
LReg	0.83	0.82	0.82	0.82	0.36	0.85	0.85	0.85	0.85	0.38
KNN	0.80	0.77	0.77	0.77	4.70	0.80	0.73	0.71	0.73	7.08
KSVM	0.76	0.74	0.74	0.74	18.74	0.94	0.93	0.93	0.93	7.66
GNB	0.68	0.64	0.61	0.64	0.04	0.68	0.64	0.61	0.64	0.03
Dtree	0.89	0.89	0.89	0.89	1.26	0.89	0.89	0.89	0.89	0.90
Etree	0.86	0.86	0.86	0.86	0.03	0.86	0.86	0.86	0.86	0.02
RFor	0.95	0.95	0.95	0.95	1.89	0.95	0.95	0.96	0.96	1.15
AdaB	0.93	0.93	0.93	0.93	7.54	0.93	0.93	0.93	0.93	6.18
Ridge	0.85	0.85	0.85	0.85	0.04	0.85	0.85	0.85	0.85	0.05
RCV	0.85	0.85	0.85	0.85	0.15	0.85	0.85	0.85	0.85	0.18
SGD	0.78	0.78	0.78	0.78	0.41	0.80	0.80	0.80	0.80	0.62
PAg	0.78	0.76	0.75	0.76	0.33	0.77	0.77	0.77	0.77	0.22
Bagg	0.93	0.93	0.93	0.93	7.77	0.94	0.93	0.93	0.93	10.06

Table 5 Classifier performance of SVM SMOTE oversampling before and after scaling

Classifier	Before feature scaling					After feature scaling				
	Precision	Rcall	*F*Score	Accu	RunTime	Precision	Recall	*F*Score	Accu	RunTime
LReg	0.87	0.87	0.87	0.87	0.31	0.87	0.87	0.87	0.87	0.32
KNN	0.85	0.83	0.83	0.83	5.61	0.84	0.80	0.80	0.80	9.34
KSVM	0.82	0.82	0.82	0.82	15.31	0.94	0.94	0.94	0.94	9.99
GNB	0.74	0.68	0.66	0.68	0.05	0.74	0.68	0.66	0.68	0.07
Dtree	0.90	0.90	0.90	0.90	1.22	0.90	0.90	0.90	0.90	1.98
Etree	0.88	0.88	0.88	0.88	0.05	0.88	0.88	0.88	0.88	0.06
RFor	0.96	0.96	0.96	0.96	1.55	0.96	0.96	0.96	0.96	2.62
AdaB	0.93	0.93	0.93	0.93	7.46	0.93	0.93	0.93	0.93	8.75
Ridge	0.87	0.87	0.87	0.87	0.04	0.87	0.87	0.87	0.87	0.04
RCV	0.87	0.87	0.87	0.87	0.17	0.87	0.87	0.87	0.87	0.19
SGD	0.83	0.83	0.83	0.83	0.26	0.86	0.86	0.86	0.86	0.43
PAg	0.84	0.83	0.83	0.83	0.25	0.84	0.84	0.84	0.84	0.15
Bagg	0.94	0.94	0.94	0.94	10.52	0.93	0.93	0.93	0.93	9.97

Table 6 Classifier performance of ADASYN oversampling before and after scaling

Classifier	Before feature scaling					After feature scaling				
	Precision	Rcall	*F*Score	Accu	RunTime	Precision	Recall	*F*Score	Accu	RunTime
LReg	0.82	0.82	0.82	0.82	0.23	0.86	0.86	0.86	0.86	0.28
KNN	0.82	0.79	0.78	0.79	4.74	0.83	0.74	0.72	0.74	5.51
KSVM	0.78	0.74	0.73	0.74	15.70	0.95	0.94	0.94	0.94	8.40
GNB	0.70	0.65	0.62	0.65	0.04	0.70	0.65	0.62	0.65	0.04
Dtree	0.90	0.90	0.90	0.90	0.88	0.90	0.90	0.90	0.90	0.76
Etree	0.86	0.86	0.86	0.86	0.04	0.86	0.86	0.86	0.86	0.03
RFor	0.96	0.96	0.96	0.96	1.12	0.96	0.96	0.96	0.96	1.16
AdaB	0.93	0.93	0.93	0.93	5.80	0.93	0.93	0.93	0.93	5.68
Ridge	0.85	0.85	0.85	0.85	0.04	0.85	0.85	0.85	0.85	0.03
RCV	0.86	0.85	0.85	0.85	0.13	0.85	0.85	0.85	0.85	0.11
SGD	0.77	0.74	0.73	0.74	0.42	0.80	0.80	0.80	0.80	0.22
PAg	0.77	0.75	0.74	0.75	0.14	0.73	0.73	0.73	0.73	0.12
Bagg	0.95	0.94	0.94	0.94	5.67	0.95	0.94	0.94	0.94	5.68

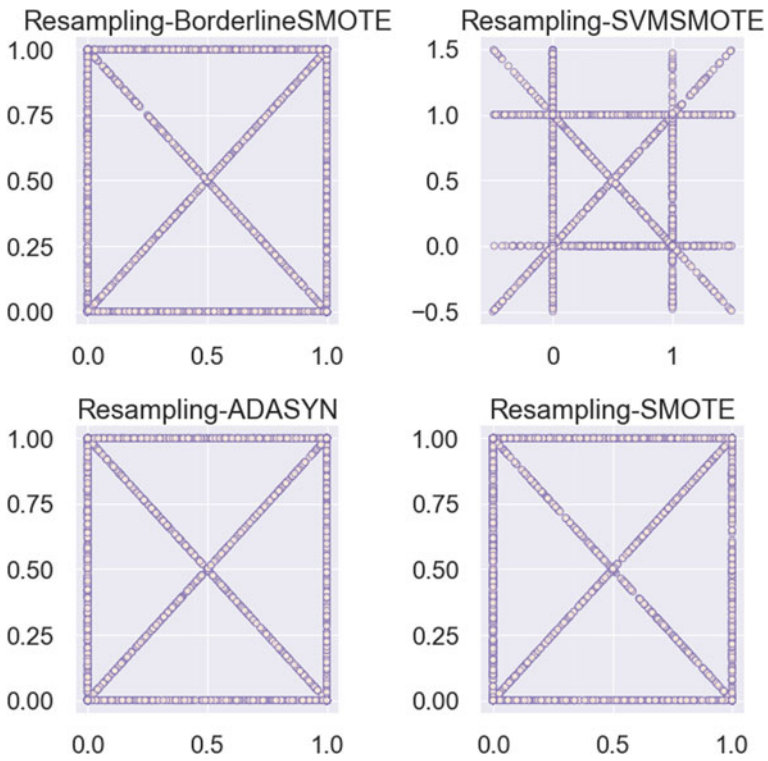


Fig. 4 Target glass type distribution with oversampling methods

References

1. A.Y. Alsobhi, N. Khan, H. Rahanu, Personalised learning materials based on dyslexia types: ontological approach. Proc. Comput. Sci. **60**, 113–121 (2015)
2. R. Wang, L. Chen, I. Solheim, T. Schulz, A. Ayesh, Conceptual motivation modeling for students with dyslexia for enhanced assistive learning, in *Proceedings of the 2017 ACM Workshop on Intelligent Interfaces for Ubiquitous and Smart Learning* (2017), pp. 11–18
3. Z. Feng, T. Yang, D. Xie, Dyslexic children show atypical cerebellar activation and cerebro-cerebellar functional connectivity in orthographic and phonological processing. Cerebellum **16**(2), 496–507 (2017)
4. H.M. Al-Barhamtoshy, D.M. Motaweh, Diagnosis of dyslexia using computation analysis, in *Informatics, Health & Technology (ICIHT).* (IEEE, 2017), pp. 1–7
5. A. Jothi Prabha, R. Bhargavi, Prediction of dyslexia from eye movements using machine learning. IETE J. Res. 2019:1–10
6. M. Rauschenberger, L. Rello, R. Baeza-Yates, J.P. Bigham, Towards language independent detection of dyslexia with a web-based game, in *Proceedings of the Internet of Accessible Things* (2018), pp. 1–10
7. Z. Rezvani, M. Zare, M. Zari, C.M. Bonte, J. Tijms, M. Van der Molen, M.F. Gonzalez, Machine learning classification of dyslexic children based on EEG local network features (2019), pp. 1–23

8. A. Frid, L.M. Manevitz, Features and machine learning for correlating and classifying between brain areas and dyslexia (2018)
9. S.S.A. Hamid, N. Admodisastro, N. Manshor, A. Kamaruddin, A.A.A. Ghani, Dyslexia adaptive learning model: student engagement prediction using machine learning approach, in *International Conference on Soft Computing and Data Mining* (Springer, Berlin, 2018), pp. 372–384
10. H. Perera, M.F. Shiratuddin, K.W. Wong, K. Fullarton, EEG signal analysis of writing and typing between adults with dyslexia and normal controls. Int. J. Interact. Multimedia Artif. Intell. **5**(1), 62 (2018)

Topic-Wise Speech Summarization with Emotion Classification

Aniket Anand, Harsh Choudhary, Anand Singhania, Aditya Manuraj, and R. Jayashree

Abstract In the era of Internet, there is a tremendous amount of textual and audio data spread all over the place, and it becomes very important to develop a method to fetch the most important information efficiently and quickly. Extracting summary manually is a very redundant and time-consuming process. A good summarizing technique is one where we discern all the important points and topics of a speech or document without leaving out any valuable information. Summarizing a speech without losing the actual context has always been a challenge for programmers for a long time. This paper explores a method to divide a large speech into multiple small speeches to summarize them individually to generate an efficient and precise summary. Each sub-speech is further processed to predict the emotion of the speaker at various points during the speech. These individual emotions are used to classify a generalized emotion of the speaker throughout the speech.

Keywords Speech summarization · Emotion recognition · Textual data · Automatic speech recognition · Speech-to-text · Sentence similarity

1 Introduction

Summarization can be explained as a process to discern most important ideas from a speech or document without losing its actual meaning. We live in an era where data is increasing very rapidly, and hence, it is important to store it in a concise manner. Summarization plays an important role in our daily life. It can be classified based on input type, purpose, and output type. Summarization can be single or multi-document depending on the input type or it can be extractive or abstractive depending on the output type. Extractive summarization, as the word suggests, is a method in which

A. Anand · H. Choudhary (✉) · A. Singhania · A. Manuraj · R. Jayashree
PES University, Bengaluru, Karnataka 560085, India
e-mail: choudhary.harsh6@gmail.com

R. Jayashree
e-mail: jayashree@pes.edu

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_39

the summary is generated by extracting the important sentences from the original source text and creating a summary consisting of the most important sections of the original source text. On the other hand, abstractive summarization fetches the context from the source by building a semantic representation of the source file and generates a summary which is contextually similar. The sentences present in the summary is not picked directly from the source document, instead, new sentences are generated. Modern NLP techniques are used to generate a shorter text which holds almost all the critical information of the original source text. Abstractive summarization is comparatively more advanced as compared to extractive summarization. Even though abstractive summarization has more future and possibilities, traditional methods prove to yield better results at this point of time. This paper aims to explore abstractive summarization technique.

Speech is considered as one of the most commonly used method of communication. Humans usually tend to interact in speech, whether it be over a phone call or a personal meeting. However, quick review of speech documents can be a really tough task if it is simply stored as audio signals. Hence, transcribing speech is one of the most important milestones toward summarization. The transcribed text can be summarized in various styles, be it abstractive or extractive. An ideal summarized speech should hold all the information of the original source and should also maintain a proper flow in the summary. Flow control has been an important challenge for researchers working in the field of summarization. As speech is mostly impromptu in nature, the speaker usually switches from topic to topic as one goes through his/her speech.

In current scenario, human–machine interaction (HMI) plays an important role in technical advancement. Emotion recognition can be done using speech, text, facial expression, or a combination of these [1]. In our regular life, emotion holds an important role and is an interesting challenge for programmers.

Emotion along with summary can provide a better idea about the overall context. Emotion provides a better understanding of the summary as it describes the tone of the speaker. Considering the spontaneity of a speech, emotion might change depending on the topic which is being discussed in a speech.

To achieve our objective of generating an efficient and precise summary holding almost all the important information of the input speech, a tool is developed which divides the input speech into multiple sub-speeches based on the sentence similarity and summarizes each of the sub-speeches individually. Ideally, a sub-speech should only discuss about a single topic. The idea is to maintain the contextual flow in the generated summary. Emotion is classified for each sub-speech, and a generalized emotion is classified based on these individual emotions. In the end, these individual summaries are combined to prepare a final summary.

2 System Design

The overall architecture can be divided into five major modules, namely automatic speech recognition, sub-speech generation, summarization, emotion recognition, final summary generation. An audio input file is considered.

2.1 Automatic Speech Recognition

With regard to the current technological development, we might be some years away from becoming fully autonomous, but we are moving toward that goal in a truly fast pace. Automatic speech recognition (ASR) is one of the major spearheads leading that race. It is a method/process that helps in converting human voice to textual data capturing all the subtle human nuisances in a precise manner. We have taken long strides in this direction, but a truly efficient ASR is still some way ahead.

ASR follows a very systematic process and can be broken down in certain phases. Firstly, a speech is recorded in a soft format. One should make sure that the audio file is free from any major background noise. Cleaning of audio file is done by removing the background noise and by normalizing the volume. The file is broken into pieces (phonemes). We can understand phonemes like basic blocks of sound of the language one is peaking. We can consider a sentence as a long sequential chain consisting of multiple phonemes. ASR technique takes help of statistical probability for analyzing a chain and then deducing words and sentences (Figs. 1 and 2).

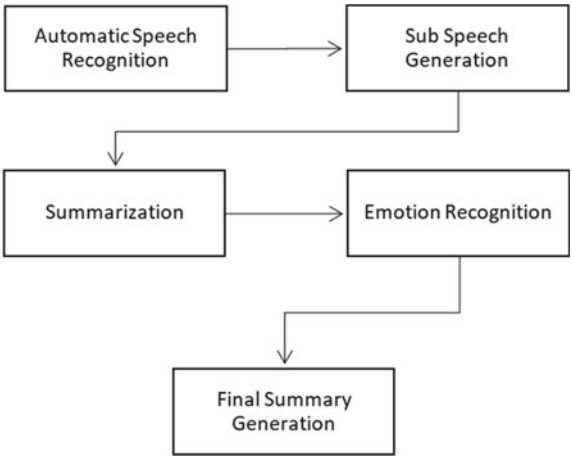


Fig. 1 System design

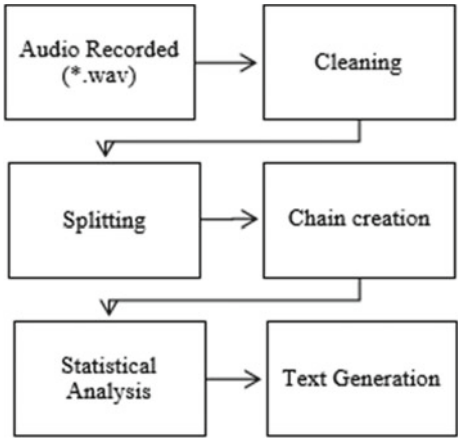


Fig. 2 Flowchart–automatic speech recognition

2.2 Sub-speech Generation

The module mainly focuses on dividing the speech into multiple sub-sections that ideally are different topics covered during the speech. The idea behind this is to summary each topic individually so that no topics are left untouched, and there is no mix-up of information while summarization.

In this module, the transcript of the speech is passed as an array of sentences which is then used to generate a sentence similarity matrix by calculating semantic similarity between each pair of them [4]. This similarity matrix is then used to combine consecutive sentences which are similar to each other. Sentences are considered to be similar if the similarity between them is over a threshold point (i.e., considered to be 0.4). The array generated after combining similar sentences is then iterated again and again until no further combination is possible, and the final array is considered as array of sub-sections (Fig. 3).

The major part of this module is generating similarity matrix [4]. The idea is to determine similarity between sentences. Two sentences that are talking about same topic will have higher similarity than those which discuss different topics. For this calculation, a BERT model [5] (which expands as bidirectional encoder representations from transformers) is used which takes array of sentences/combined sentences and embed them. Then, this embedded sentence is used to calculate cosine similarity between each pair of these embedded sentences.

Initially, single iteration was done to combine the sentences into sub-section, but currently, it is iterated until no more combining of sub-sections is possible. This improvement gave better result. The graph of similarity matrix for a sample date for zeroth iteration and n-th iteration is given Figs. 4 and 5.

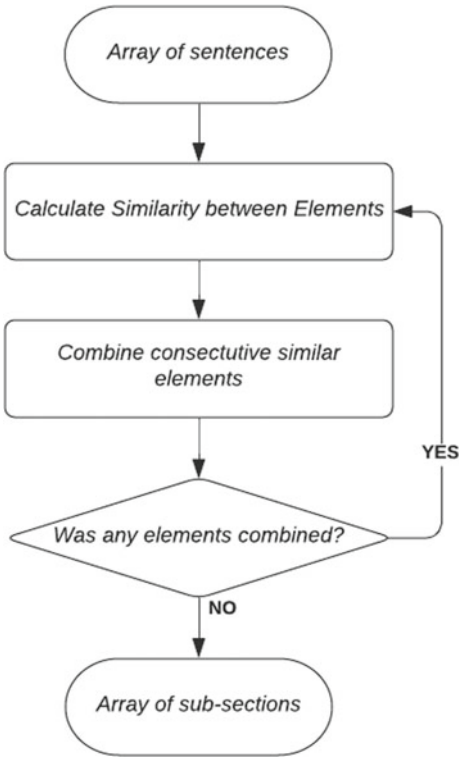


Fig. 3 Flowchart–sub-section generation

2.3 Summarization

Speeches from top leaders were initially summarized using existing tools. These summarized speeches were then compared to manual summaries. It was found that the summary generated by summarizers is not able to maintain a flow whenever a topic is switched during a speech. To overcome this, we divided each speech into multiple sup-speeches and then created summaries for each sub-speech. The dataset being used is called Gigaword, and it is a popular dataset for text summarization. It consists of about 4 million small articles along with their summaries.

Python library 'NLTK' [6] (natural language toolkit) was used as it is a leading library for Python programmers to work with human readable data. It provides easy user interface for about 50 corpora and lexical resources to work with human readable data such as tokenization, stemming, parsing, and semantic reasoning.

A dictionary was build using work tokenize feature for all the words which exist in our dataset. Dictionary key was the word, and the value was its numbers starting from one in order of the decreasing frequency. We also added two values to our dictionary named '<unk>' and '<padding>'. <unk> for unknown words and <padding> to keep each input length equal. The dictionary was dumped as pickle file for later use and easy availability. This dictionary was then reversed for decoding purpose.

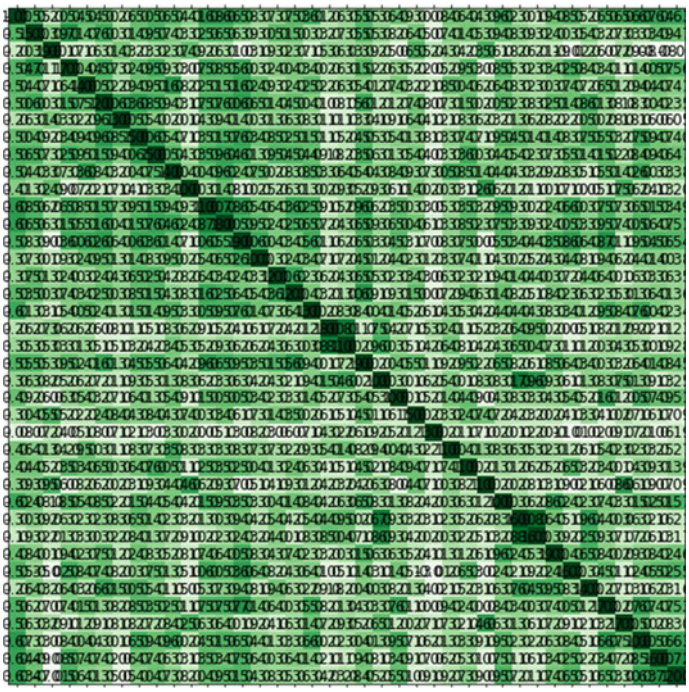


Fig. 4 Sentence similarity graph zeroth iteration

Dataset for model was prepared where we read both the article file and the summary file, and we created an array of array which contains all the tokenized word later converted to their encoded values as per the dictionary, and if word is not found, it was replaced with <unk> value, and each line max length was defined to be 50, whereas for summary, it was defined as 15. For the last line, if length of data was less than defined data length, then paddings were used.

A model was defined which is basically an encoding–decoding model with attention mechanism. A RNN model [7] was created with network size 150 and network depth 2. Glove pre-trained vectors were used to initialize word embedding. The embedding size used was 300. Learning rate was 1e−3 (0.001), and probing was set to 0.8. Batch size of 64 was used, and the model was trained for 10 epochs (Figs. 6, 7 and 8).

For encoding, LSTM cell was used with stack_bidirectional_dynamic_rnn. It gives combined forward and backward layer outputs to provide input to next layer.

For decoder, LSTM BasicDecoder [7] for training and BeamSearchDecoder [8] for inference were used, and for attention mechanism [9], Bahdanau attention with weight normalization was used.

The model was trained for 10 epochs with step size of 1000. After the training was completed, the loss was reported as shown in Fig. 9.

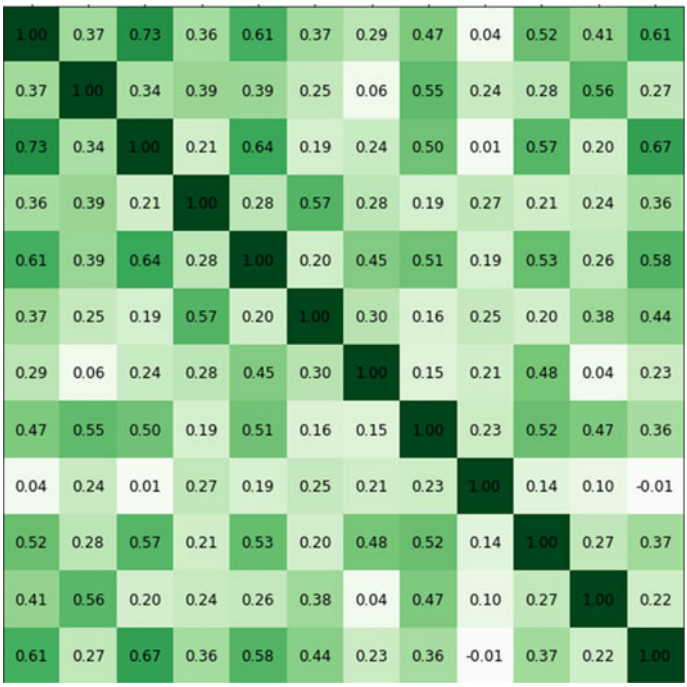


Fig. 5 Sentence similarity graph after *n*-iteration

For predictions, a speech was pre-processed as desired for easy computation which includes sentence tokenization. Each sentence was read one by one and was converted to array which consists of values based on dictionary created earlier. This input was passed to the model, and an output was generated for each line. Now, we use the reversed dictionary to convert the numbers back to human readable data. Later, all the predicted lines were concatenated to form a single paragraph.

After repeating the same step for every sub-speech, sub-summaries were generated. These sub-summaries were also concatenated so that the context flow is maintained.

2.4 Emotion Recognition

An input audio file is passed to a machine learning model to classify the emotion of the speaker. Features are extracted, and a model is trained. Selection of important attributes and features is very important for any classifier, and this process is called feature extraction. Different features can be grouped into three major categories, namely spectral features, source features, and prosodic features. This paper uses models based on spectral features. These are fetched at frame level. Spectral

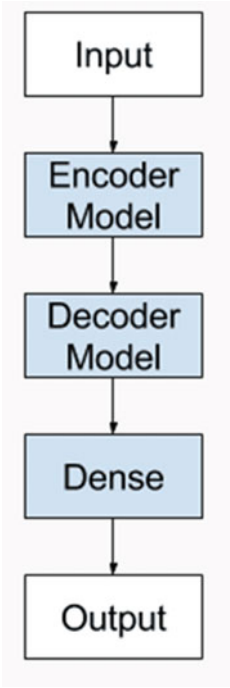


Fig. 6 Encoder–decoder LSTM model architecture

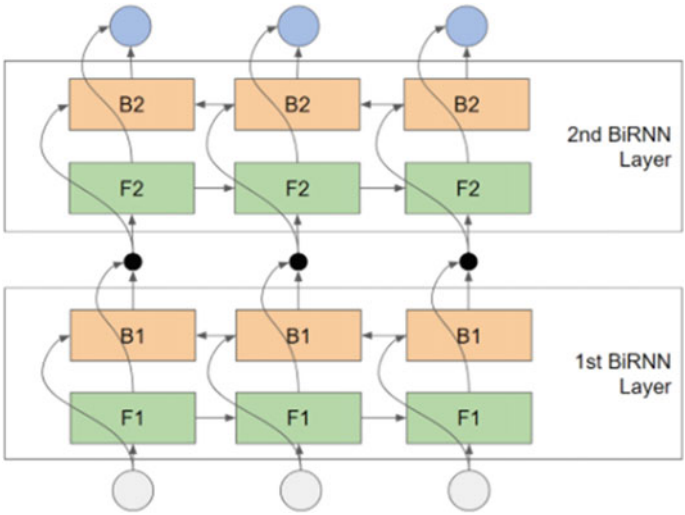


Fig. 7 Stack bidirectional dynamic RNN

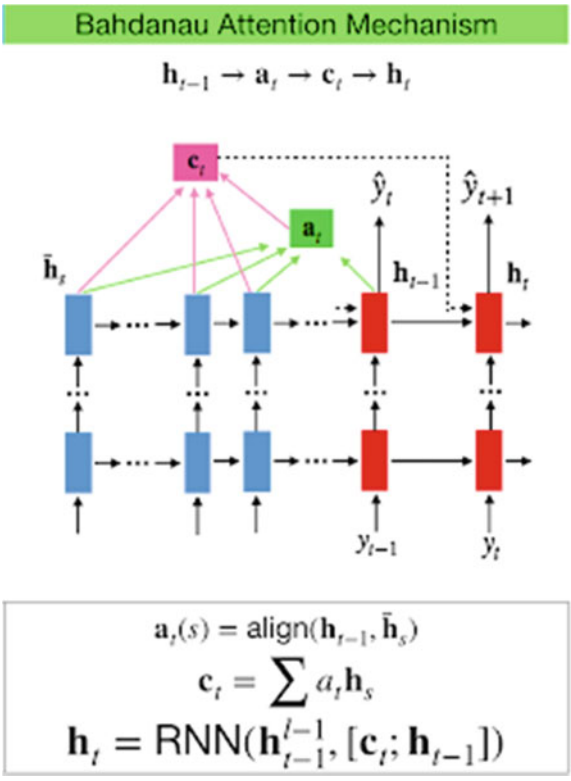


Fig. 8 Bahdanau attention

Fig. 9 Loss after training

features are found in the cepstral domain. The cepstrum, being used in homomorphic signal processing, is a representation to convert signals which are combined by convolution into their cepstra sums, for linear separation. The cepstrum provides us with short period information. Mel frequency cepstral coefficients (MFCCs) [11], linear frequency cepstral coefficients (LFCCs) [12], and linear prediction cepstral coefficients (LPCCs) [13] are a few popularly known spectral features. The accuracy of classifiers can be enhanced by using the combination of several types of features. In this analysis, MFCCs features based classifiers are used.

LibROSA [14] is a package in Python and is used to deal with audio files. In LibROSA, feature extraction can be explained as rhythm feature, spectral feature, and feature inversion and manipulation. MFCCs fall into spectral features list and are a quite popular and commonly used spectral feature. As mentioned earlier, feature extraction is one of the crucial steps in developing an emotion classification model for speech. Feature extraction is a process which can be explained as extracting

important components from an audio file which are crucial for recognizing content and leaving the remaining unnecessary parts, for example, noise. MFCCs are based on mel scale, and it was set up to portray how people perceive musical tones. In technical explanation of mathematics, mel scale is developed after frequency scale is transformed linearly. Mathematically, a reference point can be defined between the two scales, normal frequency and mel, by equating a 1000 Hz tone, 40 dB above the listener's threshold, with a pitch of 1000 mels. On a normal frequency or Hertz scale, a frequency difference of 500 Hz and 900 Hz is perceptible, whereas the frequency difference of 8600 Hz and 9000 Hz is quite difficult to notice. LibROSA helps us to create small bins by partitioning the Hz scale. Each bin is then transformed into an appropriate bin in the mel scale. A spectrogram with mel scale as *y* axis is called a mel spectrogram. The mathematical formula given below is a way to convert or transform a Hz frequency to its respective mel frequency and vice-versa.

$$f(\text{mel}) = \left[2595 * \log 10\left[1 + f(\text{Hz})\right]700\right] \tag{1}$$

MFCCs features were used to create a feature set emotion classification. The complete audio directory was scanned, and each and all audio files were loaded sequentially using LibROSA. MFCCs features were extracted for each clip and stored in a one-dimensional array. Different folders contained different emotions, and each filename was a unique identification. For instance, filename '4.2.angry-10' stands for speaker '4', session '2', class 'angry', and file number '10'. This unique identification helped in ordered loading of files for feature extraction. The computed features and their respective classes were stored in a one-dimensional array. These features were then used as an input for the classifier. A lightweight pipelining was also required and was done using a set of tools in Python called Joblib [15]. Joblib was used for lazy re-evaluation which helped in saving computational time and was used to store feature set and its respective class to avoid reloading of audio files again and again. A sample of the loaded audio file is shown in Fig. 10. The feature set was used to train several traditional and modern classifiers. Among the different classifiers, CNN [16] showed best results with minimum prediction time. The mentioned CNN model was compiled and then trained for up to 100 epochs. When tested using the validation data's feature set, an accuracy of 82% was achieved.

3 Results and Conclusion

Abstractive summarization can vary from individual to individual, hence a quantitative result may not be valid. The trained model was tested on various speeches from different leaders across the world. The model generated summary was then compared to manual written summaries. The comparison showed satisfactory results; however, there is scope for improvement. A few words from the speeches are not present in the dictionary being used. One way to tackle this problem is to increase the scope of the training data. Populating the training dataset with speeches from different regions of

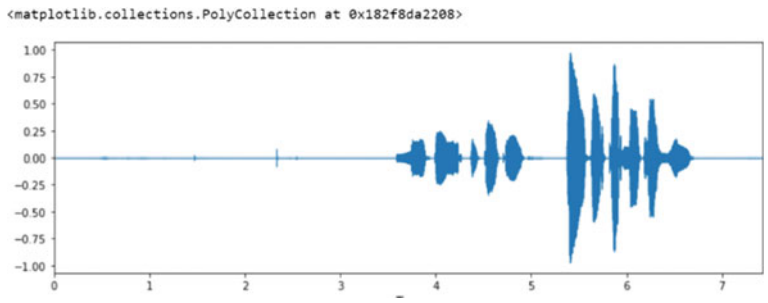


Fig. 10 Waveform of sample audio

the world could help us improve the results. Future advancements of this project can cover all the above-mentioned topics to enhance the results and could help us get a better summary altogether.

As mentioned earlier, CNN was proved to be the most effective emotion classifier for our respective dataset. Different audio files were collected from different sources. Interview clips, news broadcast, online channels, one-to-one conversation, etc., were used to collect audio files. The classifier was tested on these test audio clips, and the classifications were verified manually. The classifier when used with speeches gave satisfactory results. However, in some cases, the classifier was not very efficient as some speeches are properly scripted and are expressed with controlled emotion.

4 Future Scope

The results obtained so far are satisfactory. In future, the scope of the summarization training data can be increased. During analysis, it was found that a lot of words in the test data were not present in the dictionary which was created during training. Training on datasets from different regions across the world can solve this particular problem. If the dictionary is heavily populated, it will result in a word-rich summary. The testing technique can also be improved in future. Different techniques can be used to evaluate computer generated summaries with human generated summaries. In conclusion, the above-mentioned points can result in a better version of the overall system.

References

1. C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, S. Narayanan, Analysis of emotion recognition using facial expressions, speech and multimodal information, in *Proceedings of the 6th International Conference on Multimodal Interfaces* (2004), pp. 205–211

2. C. Zhang, F.L. Kreyssig, Q. Li, P.C. Woodland, PyHTK: python library and ASR pipelines for HTK, in *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2019), pp. 6470–6474
3. T. Giannakopoulos, Pyaudioanalysis: an open-source python library for audio signal analysis. PLoS ONE **10**(12), e0144610 (2015)
4. P. Achananuparp, X. Hu, X. Shen, The evaluation of sentence similarity measures, in *International Conference on Data Warehousing and Knowledge Discovery* (Springer, Berlin, 2008), pp. 305–316
5. N. Reimers, I. Gurevych, Sentence-Bert: sentence embeddings using Siamese Bert-networks (2019). arXiv preprint arXiv:1908.10084. J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd edn., vol. 2 (Clarendon, Oxford, 1892), pp. 68–73
6. E. Loper, S. Bird, Nltk: the natural language toolkit (2002). arXiv preprint cs/0205028
7. K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation (2014). arXiv preprint arXiv:1406.1078
8. P. Koehn, Pharaoh: a beam search decoder for phrase-based statistical machine translation models, in *Conference of the Association for Machine Translation in the Americas* (Springer, Berlin, 2004), pp. 115–124
9. R. Sahba, N. Ebadi, M. Jamshidi, P. Rad, Automatic text summarization using customizable fuzzy features and attention on the context and vocabulary, in *2018 World Automation Congress (WAC)* (IEEE, 2018), pp. 1–5
10. A. Milton, S.S. Roy, S.T. Selvi, SVM scheme for speech emotion recognition using MFCC feature. Int. J. Comput. Appl. **69**(9) (2013)
11. B. Logan, Mel frequency cepstral coefficients for music modeling, in *Ismir*, vol. 270 (2000), pp. 1–11
12. X. Fan, J.H. Hansen, Speaker identification with whispered speech based on modified LFCC parameters and feature mapping, in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, 2009), pp. 4553–4556
13. O.C. Ai, M. Hariharan, S. Yaacob, L.S. Chee, Classification of speech dysfluencies with MFCC and LPCC features. Expert Syst. Appl. **39**(2), 2157–2165 (2012)
14. B. McFee, C. Raffel, D. Liang, D.P. Ellis, M. McVicar, E. Battenberg, O. Nieto, Librosa: audio and music signal analysis in python, in *Proceedings of the 14th Python in Science Conference*, vol. 8 (2015), pp. 18–25
15. G. Varoquaux, O. Grisel, Joblib: running python function as pipeline jobs (2009). packages.python.org/joblib
16. R. Yamashita, M. Nishio, R.K.G. Do et al., Convolutional neural networks: an overview and application in radiology. Insights Imaging **9**, 611–629 (2018). https://doi.org/10.1007/s13244-018-0639-9

Analysis of Existing Algorithms for Verifying Gurmukhi Scripts and the Shortfall

Urvashi Sharma Mishra and Jagdeep Kaur

Abstract Handwriting in context of writer identification and verification are important areas of Forensic Science. In English language, a considerable amount of technological support has been achieved but in Gurmukhi script a lot needs to be done. Some researchers have worked and achieved success in writer identification, but in the area of writer verification, the expert opinions before the courts of law are manual-based which can be named as conjectures. The courts do not feel convinced with such opinions and decide the case otherwise. This author has made a humble beginning to use computer technology for handwriting verification. For the purpose, ten unique characteristics of Gurmukhi script have been identified, and underscoring the need for computer methods. In implementing these methods, this author has suggested some characteristics like measurements of height and width of characters in the word, their angularity like slant measurement, proportional size of characters, etc. have been found applicable in computer methodology to make verification of Gurmukhi handwriting technology-based and capable of convincing to the judges handling cases involving document verification (of Gurmukhi script) as a crucial piece of evidence. The findings of the paper based on 30 samples are a beginning that needs to be carried forward to develop a full-fledged branch of forensics.

Keywords Writer verification · Writer identification · Handwriting · Relative size · Gurmukhi script

U. S. Mishra (✉)
C.Sc. & IT, Hans Raj Mahila Maha Vidyalaya, Jalandhar, Punjab, India
e-mail: urvimishra923@gmail.com

J. Kaur
CSE, UIET, Sant Baba Bhag Singh University, Padhiana, Punjab, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_40

1 Introduction

The identification of writer with the help of handwriting plays an important role in forensic science [1]. Typically, the parameters considered for writer identification for any language includes the comprehensiveness, uniqueness, aging, accessibility, etc. but these parameters differ when script varies [2]. There are 12 characteristics common to all handwritings for its verification viz.—line quality, pen-lifts, size consistency, pen pressure, cursive letters, connecting strokes, word and letter spacing, slant, embellishments, diacritic placements and baseline habits. In addition to these common characteristics of handwriting, characteristics particular to script under question should also be considered for the efficient verification and identification. To explain: there is a lot of difference between Indic and non-Indic scripts and even within Indic scripts, detailed differences can be found, especially in case of formation of characters and baseline habits (i.e., writing on the baseline, above line or below line), etc.

2 Handwriting Identification of Gurmukhi Script

The researcher has scanned and analytically studied standard papers on handwriting verification/identification of authors' writing in Gurmukhi script. The following summary of such available literature arises for presentation:

In [1], the authors have explored the different features and various classification methods for the purpose of writer identification for Gurmukhi handwriting. The authors have worked upon various stages starting from the preprocessing of the image, then extracting the features, training and finally classification. The problem of identification of the writer differs with the change in language. Much work has been done in non-Indic scripts like English, Greek, Arabic, Farsi, etc. The problem of writer identification can be categorized in two parts—text dependent and text independent. In case of text dependent, two documents of the same text are compared. That is why the writer is supposed to write the same text—same words and same characters from two different documents—one questioned and one written by the alleged author are compared. On the other hand, in case of text independent, the identification of the writer is independent of the content of the text. Very few studies in Indian languages have documented it so far. The authors have worked on the text dependent identification model. Various features explored by the authors are vertical and horizontal peak extent, diagonal, zoning, intersection and open-end points, transition, centroid, parabola and power curve fitting-based features. The features are explored using the hierarchical technique. To classify, the writers have relied on above mentioned features, Decision Tree, Naïve Bayes, AdaBoostM1, and Random Forest classification techniques are used by the authors. Each leaf node in the decision tree represents the class of the writer. Naïve Bayes represents probabilistic approach based on the assumption that predictive attributes in a given class are independent. Averaging is

used by random forest to aid in controlling the over fitting problem of decision tree and in getting better predictive accuracy. AdaBoostM1 algorithm have been used in conjunction with random forest classifier to improve the accuracy of the proposed writer identification system. The authors have achieved the maximum accuracy with the combination of centroid features and AdaBoostM1 classifier, i.e., 81.75% and 81.70% in centroid features using random forest classifier. The AdaBoostM1 classifier has overall achieved maximum accuracy in all the features as compared to other classifier methods, i.e., Naïve Bayes, Random forest and decision tree. The authors claim that this accuracy can be improved with the use of different techniques used for optimal feature selection like correlation feature set, PCA, etc. or by increasing the span of the training dataset.

In [3], the authors have proposed the system for writer identification from the offline handwritten characters in Gurmukhi using feature extraction approach and the classification algorithm in WEKA tool. They have implemented the extraction of zoning and open-end point intersection features and classified the writers using k-NN and MLP classification approach and then evaluated the performance parameters. In the proposed methodology, the accuracy of the writer identification largely depends on recognition and normalization. The authors have shown the classification accuracies of 30 writers based on features extraction in terms of k-NN and MLP based on zoning and open-end intersection features and proved that the Multilayer Perceptron classification approach perform better feature extraction process than k-nearest neighborhood with the accuracy of 55% recognition rate than k-NN having 53% recognition rate accuracy. The system is able to recognize characters in varied styles of handwriting. According to authors, the feature extraction step should be carefully diagnosed before the recognition since the recognition accuracy highly depends on it. The authors' approach is based on some complexities like skewed text, overlapped text, and connected components that make the recognition a difficult task. But their approach considers only the word spotting within the different writing styles. According to authors, learning mechanism should be introduced to cope with more different writing styles and dataset should be increased to make it more practical.

According to the authors of [4], every person has adequate individual handwriting characteristics that can be utilized as an identification strategy in spite of high variation. The authors have discussed a novel approach for text dependent identification of the writer based on pre-segmented characters of Gurmukhi script following various stages like digitization, preprocessing of the image, feature extraction and classification. To extract the properties of handwriting, the authors have used transition, diagonal, zoning and peak extent-based feature extraction techniques and to evaluate the prominence of these features, k-NN and SVM classification techniques have been used. The various challenges in writer identification based on handwriting are variation in size, shape of letters, angles of letters, styles of human writing, low accuracy rate in recognition, poor input quality, constrained and unconstrained handwriting, etc. Additional challenges faced by Indic scripts are lack of standard databases, large character set due to modifiers, etc. The methodology followed by authors is—after the digitization of the image, its normalization into 100×100 pixels is done using

k-NN interpolation scheme followed by converting it into bitmap image. They used the parallel thinning scheme to reduce the width of the character to single pixel. Peak extent-based features, diagonal, zoning, and transition features have been considered in the feature extraction phase. The authors have used K-nearest neighborhood classification method for classifying the samples using Euclidean method to find the distance between stored and candidate featured vectors. They have also used the supervised learning classifier—support vector machine with a linear kernel. By dividing the data 70% as training set and remaining 30% as testing set, the authors have achieved the accuracy of 89.85% by combining transition, peak extent-based features, and zoning with linear-SVM classifier and tenfold cross validation accuracy of 94.76% using the same combination of features and classifiers.

The authors of [5] claim that handwriting is the behavioral biometric modality and this skill of personality cannot be imitated. According to them even the twins sharing the hereditary characteristics do not share the same handwriting. They argue that whenever a person writes a letter, each person has his allograph that makes his identification possible based on his writing behavior. They have discussed the difference between handwriting identification versus handwriting recognition; text dependent versus text independent; writer identification versus writer verification and script dependent versus script independent. Authors' work is based on the framework of text dependent writer identification system based on biometric modality and through documents that contain calligraphy. Discussing the issues and challenges during the development of writer identification model, they observe that to achieve the higher accuracy rates in writer identification process, efficient preprocessing of document, and proper data collection in respect to paper quality, dust on the paper, ink-color, presence of oil, overlapping of characters and any kind of dullness—all such factors play a vital role. The authors have used the diagonal, transition, zoning and peak extent feature extraction algorithms and to classify, based on the values of features obtained, Random forest, Artificial neural network and multi-layered perceptron model have been used. The algorithm evaluation is done based on three parameters which are accuracy, true positive and false positive values. To determine the best identification accuracy rate, they have computed and compared these parameters by using individual and hybridization of feature extraction methods with classifiers and have achieved the highest accuracy rate of 93.6% using all the four features with random forest classifier and with 0.39% false positive rate.

The authors of [6] have discussed the water reservoir-based technique for the purpose of identification and segmentation of handwritten Gurmukhi words with the problem of touching character. They have used the statistical analysis-based projection profile techniques for the segmentation of touching characters. Authors have quoted the problem of touching and overlapping characters in case of segmentation process. Both horizontal and vertical projection profile techniques combined with water reservoir technique have been applied on the Gurmukhi handwritten documents under three categories viz. same document written by different writers achieving average accuracy of 92.53%; different document written by same writer with average accuracy of 93.97% and different documents written by different writers

with 94.04% accuracy rate. The authors have achieved the overall average accuracy of 93.51% for segmenting the touching and isolated Gurmukhi characters.

Kumar et al. in paper [6] have presented multiple feature extraction techniques along with the hybridization of the classification algorithms. The system is based not only on the identification of the writer based on handwriting but on gender classification too. The gender classification system relies on both behavioral and physiological modalities. Handwriting is always done with the conscious force that is controlled by the writer and the unconscious force that is controlled by the individual's own spontaneity. The authors have used curve fitting based technique and the intersection open-endpoint based techniques for the purpose of feature extraction and k-nearest neighborhood, SVM, MLP, HMM, Bayesian classifiers. They have achieved the gender classification accuracy rate of 90.57% with TPR 89.47% and FPR 0.36%.

Kaur et al. in paper [7] have applied various principles for handwriting examination like alignment of the writing, slant, size, style of handwriting, beginning and ending strokes and their angularity. According to them sometimes there is possibility that the questioned and specimen writings are not available in the same script. In such cases, the authors suggest the interscript comparison to be done. They claim that the same writer will write the particular allograph or the part of the allograph, i.e., stroke in the same manner irrespective of the script since many strokes in different scripts like English, Devnagari and Gurmukhi are common and are put in the same manner by the same writer, e.g., the similarity in 'k' in English, 'kakka' in Gurmukhi and 'r' in Hindi.

Singh and Sharma in paper [8] have discussed about two levels of features in Gurmukhi handwriting—low level and high level. According to them, the curliness of the characters, the linearity, their slope and aspect ratio are described as the low level features whereas features like crossing of strokes, straight line, headline, loops, and dots are considered as high level features. But both features are useful in verifying the recognized stroke. According to the authors of [9], Gurmukhi script is cursive in nature and it faces various problems like difficulty in segmenting the words into characters since some writers write more than one characters in single stroke. That is why the word-based approach yields more accurate results than stroke-based approach in handwriting recognition systems.

In all the above papers related to the writer identification of Gurmukhi handwriting, the authors have discussed various feature extraction techniques—zoning, transition, diagonal, intersection, open-end point features, peak extent-based features, parabola and power curve fitting features, and classification methods—decision tree, k-nearest neighborhood, naïve Bayes, AdaBoost, Multilayer Perceptron (MLP), Random forest (RF), Support Vector Machine (SVM) and Artificial Neural Networks (ANN)—and claim to have achieved good accuracies. But they all have focused on the techniques and methods for feature extraction and classification phase of writer identification. No doubt, Dargan and Kumar [5] in her paper has named various attributes of handwriting skill: like size, shape and orientation of letters; slope of the letters, pen pressure, average size of letters, regular or irregular gaps between letters, rhythmic repetition of the elements or not, and thickness and width of letters but these authors also do not go beyond writer identification.

3 Need for Writer Verification

Looking from the demand of court cases, there arises a need for verification of writing in Gurmukhi script for ascertaining its writer. In such cases, a writing is before the court undertaking trial of a case during the stage of evidence-giving by both the parties. Both contest on the writing being examined, one claiming to be genuinely written and the other calling it the forged one. To make a comparison, both agree on another writing which is by the author not available to come before the court. The forensic experts apply non-scientific and non-computer-based techniques to arrive at their findings. Many times, the court does not feel convinced with the expert opinion for the reason of its non-scientific basis. Here lies the importance of handwriting verification—may be text dependent or text independent. The available research on Gurmukhi handwriting has not reached to the level of such computer-based verification. The present paper attempts in a humble way to enter into such advancement from the present research so as to fulfill the need of court cases, which is undoubtedly an urgent and important social need.

This paper focuses on the attributes required to be considered for the handwriting verification in context of Gurmukhi script. Marking a unique characteristic of Punjabi handwriting, the topline of each character is normally put before writing the character but, like Devanagari script, some writers write the characters of the word without putting the topline and after completing the word, they put the topline on the entire word; while some other writers put the headline of each character separately and then join them with their top edges (which is the most common method). There are some other writers also who do not have the habit of connecting the characters in the word. They just reduce the gap within the characters in the word keeping all the characters isolated. So, this becomes an important characteristic of Gurmukhi handwriting.

Some writers have the inbuilt characteristic of writing overlapping and touching characters and some keep the characters isolated (as shown in Fig. 1). Similarly, some writers have the habit of putting the diacritics and matras properly whether it is above the character touching its headline, before or after the character touching its edge or below the character touching the foot of the character and the baseline, while some keep them isolated from the character and do not touch them with character or with headline or baseline at all (Figs. 2, 3, 4, 5 and 6).

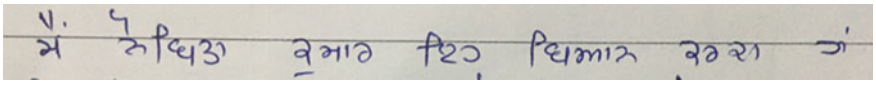


Fig. 1 Isolated characters in the word

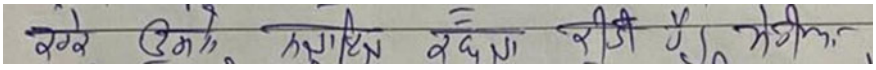


Fig. 2 Single headline on whole word

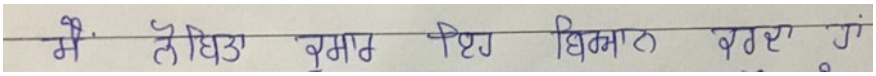


Fig. 3 Characters connected with edges having individual headline

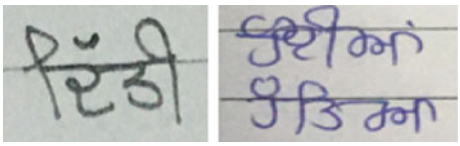


Fig. 4 Matras at proper place and touching headline

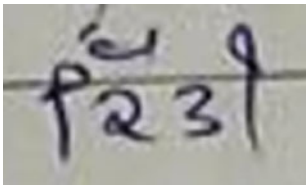


Fig. 5 Sihari and Bihari combined and not combined with character

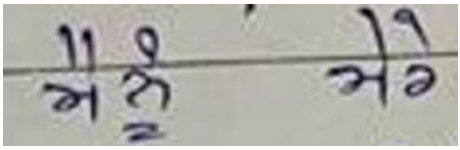
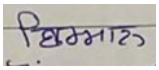


Fig. 6 Isolated characters and diacritics

All these characteristics of Gurmukhi script play a vital role in the identification or verification of the writer. The presence of such characteristics in the script is shown in the following samples.

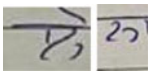
From 30 different samples of the same script writing by different authors, the researcher has identified the following important characteristics of variance between different writers which can lead to the satisfactory application for handwriting verification mentioned as below:

1. Retouching the headline of the character.



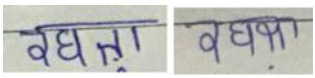
In the above figure, the character 'babba' is constructed by retouching the headline—once while writing the character and once by putting the headline over it.

2. Broken curved characters like 'nanna', 'lalla'



Some writers have the habit of writing the curved characters as the broken curved character as shown in the above figures of characters ‘lalla’ and ‘nanna’. In both the characters, one part of the curve is isolated from the whole character and such habit is inbuilt in the writers which they imply for all curved characters.

3. Joining all the characters in the word with single headline or putting individual headline of each character.



Some writers have the habit of putting the single headline over the word as written in Devanagari script (shown in first figure) where as some put individual headline of each character and then combine it with other character to form the word as shown in second above figure.

4. Joining the character ‘ooda’ with its diacritic at the base like ‘unkad’.

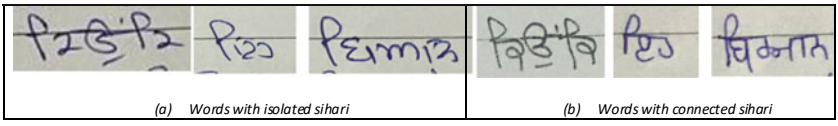
This is one of the very special characteristics that is inbuilt in some writers that they draw the matra with the character in the single stroke treating it as the part of the character like in the above figures, two different writers have combined the ‘unkad’ with the top of the character ‘ooda’ overlapping the character.

5. Slant in the matras on the top of the character like ‘laan’, ‘dulaan’.

(a) (b) (c) (d) (e) (f)

Normally the matra ‘laan’ and ‘dulaan’ are slanted in Gurmukhi script. But as we can see in the above figures, some writers put them straight (a, f) without any slant, some put it in two different slants (c) while some have the habit to put with same slant (d, b, e).

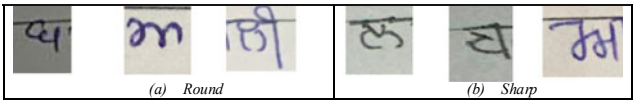
6. Isolated matra at the start of the word like ‘sihari’.



(a) Words with isolated sihari (b) Words with connected sihari

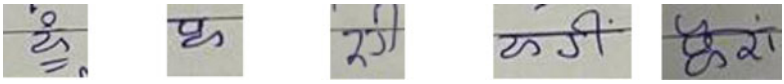
The matra ‘sihari’ plays a very important role in writing the Gurmukhi script. Every writer has his own way of putting this matra either with the open loop or with the closed loop and some writers have a unique characteristic of always putting ‘sihari’ isolated from the character whenever they use this matra at the start of the word.

7. Roundness or sharpness in the characters like ‘lalla’, ‘babba’



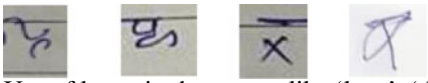
In writing the curved characters specially ‘lalla’, ‘babba’, and ‘adaa’ in Gurmukhi, some writers have the habit of putting the upper part of ‘lalla’ as sharp triangle while others put it in round way. In the similar way, some writers put ‘babba’ or ‘adaa’ with sharp edges while some others have the habit of writing it in rounded form.

8. No pen-lift in the curved characters like ‘nanna’, ‘lalla’.

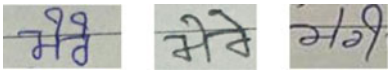


Mostly the characters ‘lalla’ and ‘nanna’ in Gurmukhi are written with 2–3 pen-lifts, but some writers have this unique habit of writing these characters in the single go without any pen-lift creating a different form of these characters.

9. Style of writing some particular characters like in case of ‘lalla’ some first write the upper part and then connect the both sides of lower part with it while some cross both the lines of it. Same is the case with character ‘jhajha’, ‘naanaa’.



10. Use of loops in the matras like ‘laan’, ‘dulaan’.



Some writers make use of loops in putting the matras ‘laan’ and ‘dulaan’ while some do not create any loop while writing such matras. Even we can see a variety of loops in such matras like big round loops in first figure or small loops in the second one.

Thus, the unique formation of the characters and specially the matras play a significant role in the verification of the Gurmukhi handwriting.

4 Making Use of Computer Methods for Implementing Such Methods

The above mentioned characteristics are already considered manually by the handwriting experts, but it will be of use to the courts only if such comparisons can be scientifically/technologically shown. For the purpose, some characteristics like measurement of the height and width of characters, angularity of characters—slant

measurement, the proportional size of various characters and parts of characters, ratios of characters; in addition to taking few from above mentioned characteristics particular to Gurmukhi script can prove the verification task more reliable to the courts. To make it computer-based, use of Machine learning or Deep Learning or in particular Convolutional Neural Networks (CNN or ConvNet)—the most popular algorithm of deep learning, using Python or MATLAB tools can be successfully applied. Deep learning is a very powerful technique as it automatically identifies deep powerful features and it can be used in practical situations due to its high prediction accuracy [10]. The layers of CNN perform three important functions—convolution, pooling and rectified linear unit (ReLU). Convolution activates certain features of the images; pooling simplifies the output by reducing the number of parameters to learn and ReLU allows faster and effective training. No doubt, deep learning learns features and classifies automatically with unlimited accuracy but it requires very large datasets. Thus, the algorithm for handwriting verification depends on the amount of data [11–13]. Classification models like Support Vector Machine (SVM) and Random Forest (RF) can be used for classification purpose with good accuracies (97.5% and 97.7%, respectively) [14].

To make the verification more efficient in Gurmukhi script, besides size and slant of writing, ratios of letters play an important part [15]. Slant or slope of writing is the inclination of letters or their strokes with the actual or imaginary base line and size of writing vary from person to person and play a significant role in determining the genuineness. In most of the cases of disguised writings, the writer has the tendency to change the size of writing which is mostly increased or decreased. Although the overall size of writing may increase or decrease due to the internal and external factors, but the habit with regard to the relative size of one letter to the other letter(s) is almost a fixed habit and it remains almost constant. According to the famous handwriting expert Harrison, the ratios of letters are maintained to a great extent, despite changes in size, speed or intent of the writing (disguised or original). Thus, to make verification process more authentic and acceptable to the courts, if we automate the process of finding the relative size and relative slant of characters in some most commonly used words in the original document and then comparing these relativities with the simulated document and demonstrating it through histograms and tabular form that will prove to be more authentic and beneficial.

5 Conclusion and Future Scope

The verification of Gurmukhi handwriting is still in its infancy stage leaving a large scope for its further growth and development. The excuses, this author experienced that a scientific and computer-based methodology is impossible to be applied in Gurmukhi script has been belied. The author picking threads from the works on identification of Gurmukhi handwriting, has found the characteristic differences in writing by different authors preparing the same document in Gurmukhi script. Such characteristics are many and all of them have a scope for being adopted in verification

by computer-based technology. This author has given a beginning and the end is unending. Even use of a few characteristics can lead to satisfactory results and every addition has a scope for improvement, thereby reaching to the stage of perfection. It is rising step-by-step and development of a full-fledged branch of computer forensics where lies a vast scope of further research in the area of much socio-legal significance.

Acknowledgements We wish to extend our special thanks to 30 different writers for providing the samples without which the study does not have been possible.

References

1. N.G. Sakhi, M. Kumar, Writer identification system for handwritten Gurmukhi characters: study of different feature-classifier combinations, in *Proceedings of International Conference on Computational Intelligence and Data Engineering* (Springer Nature, 2017), pp. 125–131. http://doi.org/10.1007/978-981-10-6319-0_11
2. Lucy, What are the 12 characteristics of handwriting? Available at: https://www.thepencompany.com/blog/handwriting/what-are-the-12-characteristics-of-handwriting/ (2019)
3. K. Kalra, et al., Writer identification from offline isolated handwritten Gurmukhi characters, in *Advances in Computational Sciences and Technology*, vol. 10, no. 5 (Research India Publications, 2017), pp. 903–914. ISSN 0973-6107
4. M. Kumar, et al., A novel framework for writer identification based on pre-segmented Gurmukhi characters. Sadhana **43**, 197 (2018). http://doi.org/10.1007/s12046-018-0966-z
5. S. Dargan, M. Kumar, Writer identification system based on offline handwritten text in Gurumukhi script, in *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (IEEE, 2020), pp. 544–549. http://doi.org/10.1109/PDGC50313.2020.9315842
6. M. Kumar, et al., Segmentation of isolated and touching characters in offline handwritten Gurmukhi script recognition. Int. J. Inf. Technol. Comput. Sci. **6**(2), 58–63 (2014)
7. A. Kaur et al., Interscript comparison of handwriting features leading to their identification and authorship. Nowa Kodyfikacja Prawa Karnego, Tom 45, Wroclaw (2017). http://doi.org/10.19195/2084-5065.45.3
8. S. Singh, A. Sharma, Recognition of online handwritten Gurmukhi characters through neural networks, in *Advances in Communication and Computational Technology*. Lecture Notes in Electrical Engineering, vol. 668, ed. by G. Hura, A. Singh, L. Siong Hoe (Springer, Singapore, 2021). http://doi.org/10.1007/978-981-15-5341-7_18
9. S. Singh, A. Sharma, V.K. Chauhan, Online handwritten Gurmukhi word recognition using fine-tuned deep convolutional neural network on offline features, in *Machine Learning with Applications*, vol. 5 (2021)
10. e-book, Introducing deep learning with MATLAB. Available at: www.in.mathworks.com
11. A. Castelblanco, et al., Machine learning techniques for identity document verification in uncontrolled environments: a case study, in *MCPR 2020: Pattern Recognition* (2020), pp. 271–281
12. S.N.M. Khosroshahi, S.N. Razavi, A.B. Sangar, et al., Deep neural networks-based offline writer identification using heterogeneous handwriting data: an evaluation via a novel standard dataset. J. Ambient Intell. Human Comput. http://doi.org/10.1007/s12652-021-03253-2
13. P.K. Sarangi, A.K. Sahoo, S.R. Nayak, A. Agarwal, A. Sethy, Recognition of isolated handwritten Gurumukhi numerals using Hopfield neural network, in *Computational Intelligence in Pattern Recognition. Advances in Intelligent Systems and Computing*, vol. 1349, ed. by A.K. Das, J. Nayak, B. Naik, S. Dutta, D. Pelusi (Springer, Singapore, 2021). http://doi.org/10.1007/978-981-16-2543-5_51

14. A. Sethy, P.K. Patra, S.R. Nayak, Offline handwritten numeral recognition using convolution neural network, in *Machine Vision Inspection System: Image Processing, Methodologies and Applications*, vol. 1 (2020), pp. 197–212
15. V. Aubin, M. Mora, M.S. Penas, Off-line writer verification based on simple graphemes. Pattern Recogn. **79**, 414–426 (2018)

Efficient Designing of Mockups Using Convolutional Neural Network

Suman De and Debjyoti Das

Abstract Software development methodologies have seen major changes in terms of functional programming and open-source libraries. User Interface design takes a major proportion of development effort, and the creation of mockups is still a manual process. It not only involves time but also requires continuous feedback from the end user, making human behavior a major component for prototype designing. The emergence of an artificial neural network accounts for simplifying human behavior and replicates the actions systematically taken by a human. This paper discusses the usage of the convolutional neural network model of artificial neural networks to generate mockups as per user responses and suggesting User Interface mockups that best suits a business scenario. The auto-generation of mockups leads to a significant reduction of time and cost to the company for developing the initial mockups of a system. The reduction in time can be counted in not Person Days (PDs) but in months because it reduces the time taken in the initial research and leads to direct validation from the customer. The paper proposes the possibility of using end user data and predicting a relevant mockup by eliminating the initial human error regarding the templates that are stored in the system.

Keywords Science · Convolutional neural network · Software prototyping · Usability engineering · User-centered design · User interface design

1 Introduction

Business applications require continuous upgradation and simplification of the features required to address end-to-end business scenarios. Customer needs are increasing with the advancements of technology resulting in the faster and efficient delivery of products. The goal is to empathize with the customer and identify the actual problems before the initiation of software development. Deep learning has impacted industries like health care, supply chain, e-commerce, science, and other

S. De (✉) · D. Das
SAP Labs India Pvt. Ltd., Bangalore, India
e-mail: suman.de@sap.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_41

research fields in recent times [5]. Artificial neural network facilitates the prediction of the right choice among multiple possibilities using deep learning for a given scenario. Supervised and unsupervised learning is used to classify a group of objects for known and unknown clusters, respectively, that hints at the existence of relevant patterns [8]. Pattern matching is then used for categorizing business scenarios that are mapped to a suitable template [12].

This paper proposes an algorithm that optimizes mockup development using a convolutional neural network (CNN) model [19] that suggests relevant mockups for business scenarios by processing end user behavior from an existing application without disturbing the existing behavior. The model analyzes end user behavior and identifies the closest accurate mockup concerning end user behavior, stored templates, and usability principles. Mockups are then generated by comparison with conditional biases which are used to redesign the existing product. The paper looks at how artificial neural network [2] gets a set of inputs and produces possible outputs after passing through multiple hidden nodes which decides a path for the final optimized output. The paper identifies the required mockup by having inputs related to human behavior, templates, and standard guidelines which after processing with conditional biases, gives outputs of mockups that is used to simplify the existing applications.

2 Proposed Idea

The User Interface design space has seen major progress involving the abovementioned tools and literature surveys, but the possibility of reducing human intervention for generating mockups using artificial intelligence or machine learning hasn't been explored yet. This paper proposes an idea that uses a convolutional neural network model to analyze human behavior and suggests relevant mockups as per the required business scenario [7]. So, a set of input variables (x) and output variables (y) are used in an algorithm to learn through a mapping function (1) and get the output from the input [15].

$$y = f(x) \tag{1}$$

Here, the objective focuses on the mapping function and predicted to contain the new input data (x) and estimate the right output variables (y) for that data. A CNN model is used as a training model. The two neural networks work simultaneously to predict the required fields and then map an already stored template, respectively, and involves:

- Data Accumulation—*Customer Behavior, Standard, and Customized Templates.*
- Output and relation to User Experience.
- Control flow and steps involved.

The identified fields are classified for their weights and estimate specific order based on their usage. Next, the most suitable template is mapped to the fields identified post the first neural network. This neural network selects the appropriate template for the given scenario and recreates it. The selected template is used for the development of the application. The idea looks at alternate mockups in case the suggested mockup does not satisfy the business requirements. This is achieved using backpropagation. It is used in artificial neural networks to find a gradient descent of a network that supports calculating the accurate weights and bias. Backpropagation takes place multiple times for a single node in a network until a template is selected. The vital activity in the neural network is how the network of nodes is trained.

(1) Convolutional Phase

First, initialize all the weights and biases of the CNN to a small value
Set learning Ω rate such that $0 < \Omega < 1$
$n = 1$
repeat
For $m = 1$ to M do
propagate pattern Xm through the network
For $k = 1$ to the number of neurons in the output layer
Get error
End For
For layers $L - 1$ to 1 do
For maps, $j = 1$ to J *do*
Find the error factor which needs to be backpropagated
End For
End For
For $i = 1$ to L do
For $j = 1$ to J *do*
For all weights of the map, j do
Find $d(w)$
Update the weights and biases
$w(new) = w(old) + d(w)$
End For
End For
End For
$n = n + 1$
Find mean square error (MSEI)
This continues until MSEI $< \varepsilon$ or $n >$ maximum bounds. The forward pass of the algorithm can also be described through (2). The output of the neuron k, column y in the lth convolution layer, and kth feature pattern:

$$O_{x,y}^{(l,k)} = \tanh\left(\sum_{i=0}^{f-1} \cdot \sum_{r=0}^{Kh} \cdot \sum_{c=0}^{Kw} W_{(r,c)}^{(k,l)} O_{(x+r,x+c)}^{(l-1,i)} + \text{Bias}^{(l,k)}\right) \tag{2}$$

where f is the number of convolution cores in a feature pattern. The output of neuron of row x, column y in the lth sub-sample layer, and kth feature pattern:

$$O_{x,y}^{(l,k)} = \tanh\left(W^{(k)} \sum_{r=0}^{Sh} \cdot \sum_{c=0}^{Sw} \cdot O_{(x*Sh+r,y*Sw+c)}^{(l-1,k)} + \text{Bias}^{(l,k)}\right) \tag{3}$$

The output of the jth neuron in lth hide layer H:

$$O_{(l,j)}^{\cdot} = \tanh\left(\sum_{k=0}^{s-1} \cdot \sum_{x=0}^{Sh} \cdot \sum_{y=0}^{Sw} W_{(x,y)}^{(j,k)} O_{(x,y)}^{(l-1,k)} + \text{Bias}^{(l,j)}\right) \tag{4}$$

Among the given symbols, the feature pattern in the sample layer is s. The output of the ith neuron lth output layer F

$$O_{(l,i)}^{\cdot} = \tanh\left(\sum_{j=0}^{H} \cdot O_{(l-1,j)} W_{(i,j)}^{l} + \text{Bias}^{(l,i)}\right) \tag{5}$$

(2) Knowledge Transfer Phase

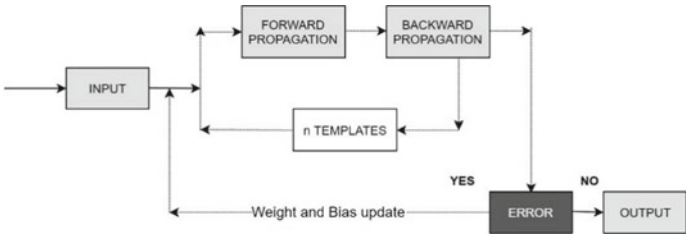
Repeat
  For $tk = 1$ to TK (number of training samples)
    Propagate pattern x_{tk} through the network
    For $z = 1$ to the number of neurons in the last convolutional
      layer (Z)
      Find output $Oz = (O1, O2, O3, \ldots, Oz)$
      Find O_{ztk} using the TSL framework
    End For
  End For

(3) Weight Update and Learning Phase

This acts as the basis for the identification of the required set of fields that are generally used by the end user. It trains the feedforward layers (last convolutional layer) using O_{ztk} of phase II. A gradient decent algorithm is used here. This goes until MSE2 < ε or $n >$ maximum bounds.

3 Results and Discussion

Time for serial execution is $(T1 + T2) \times N + T3$.



For parallel execution for the n nodes in the hidden layers, the time complexity comes

$$\sum_{i=1}^{i=n} .((T1 + T2) \times N) + T3 \tag{6}$$

where $T1$ is the time taken for forwarding propagation, $T2$ is the time taken for backpropagation of a node, $T3$ is time take to update the weight and bias, N is the number of templates, and n is the number of nodes. Similar iterations take place for other n nodes too. In the serial realization method, the total execution time is N times of $T1$ and $T2$.

As observed in Table 1 [16], there is a significant reduction of effort in requirements and specifications gathering, design, and overall ongoing lifecycle activities of the project. There is an expected reduction in efforts of around 10.4% possible with the auto-generation of mockups for existing applications. The design phase that involves mockup design gets an approx. 33.33% reduction of efforts with faster iteration cycles.

Table 1 Comparison with a previous study [16] and the new expected effort

Software phase	Effort	Effort percentage	New estimated effort
Requirements gathering	18	7.5	14
Specifications	18	7.5	12
Design	24	10	16
Implementation	24	10	24
Integration testing	18	7.5	18
Acceptance and deployment	18	7.5	16
Ongoing lifecycle activities	120	50	115

4 Conclusion

The topic of this research was motivated by the field study of Usability Engineering. The development of user-friendly User Interfaces forms the cornerstone for modern-day software. Customers demand high performant and consistent applications on the cloud and manual creation of mockups for both time and irregularities. There is a higher Cost-To-Company in such scenarios resulting in delayed delivery and compromise on quality. The proposed algorithm eliminates manual intervention for generating User Interface mockups and automates the process by considering standard usability engineering principles. The comparison with an existing effort model showed that the automation of this process in a software development lifecycle fastens the development process by a good percentage as well as ensuring user-friendliness of the delivered solution. This idea solves a critical challenge for product-based software companies in rebuilding existing applications and developing new solutions.

Acknowledgements We would like to highlight the continuous support of Deepak Padmanabha and guidance from Atreya Biswas, who helped us understand the basics of artificial neural networks, which was vital for clarifying queries about the topic. We also would like to thank IGDTUW and AIST-2019 for publishing our initial work on the topic which has been extended in this paper.

References

1. A. Ali, M. Alrasheedi, A. Ouda, L.F. Capretz, A study of the interface usability issues of mobile learning applications for smart phones from the user's perspectives. Int. J. Integr. Technol. Educ. (IJITE) **3**(4) (2014). https://doi.org/10.5121/ijite.2014.3401
2. A. Anuse, V. Vyas, A novel training algorithm for a convolutional neural network. Complex Intell. Syst. **2**, 221 (2016). https://doi.org/10.1007/s40747-016-0024-6
3. M. Bakaev, V. Khvorostov, Component-based engineering of web user interface designs for evolutionary optimization, in *2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Busan, 2018, pp. 335–340. https://doi.org/10.1109/SNPD.2018.8441135
4. R. Balabhadrapathruni, S. De, A study on analysing the impact of feature selection on predictive machine learning algorithms, in *6th IEEE PDGC 2020*, Shimla, India, Nov 2020, pp. 10–15. https://doi.org/10.1109/PDGC50313.2020.9315801
5. M. Bano, D. Zowghi, F. da Rimini, User involvement in software development: the good, the bad, and the ugly. IEEE Softw. **35**(6), 8–11 (2018). https://doi.org/10.1109/MS.2018.4321252
6. D. Das, S. De, An algorithm for auto-generation of mockups of legacy applications using behavioral analysis and artificial neural network, in *International Conference on Artificial Intelligence & Speech Technology (AIST-2019)*, Nov 2019, vol. 1, no. 1, pp. 1–7
7. S. De, D. Agarwal, A novel model of supervised clustering using sentiment and contextual analysis for fake news detection, in *IEEE MPCIT 2020*, Karnataka, India, Dec 2020, pp. 112–117. https://doi.org/10.1109/MPCIT51588.2020.9350457
8. N. Li, Q. Hua, S. Wang, K. Yu, L. Wang, Research on a pattern-based user interface development method, in *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Hangzhou, 2015, pp. 443–447. https://doi.org/10.1109/IHMSC.2015.203
9. T. Liu, S. Fang, Y. Zhao, P. Wang, J. Zhang, Implementation of Training Convolutional Neural Networks. https://arxiv.org/abs/1506.01195. Accessed 4 Nov 2019

10. N. Mezhoudi, User interface adaptation based on user feedback and machine learning, in *IUI'13 Companion*, Santa Monica, CA, USA, 19–22 March 2013
11. M. Mirović, M. Miličević, I. Obradović,A framework for dynamic data-driven user interfaces, in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija, 2018, pp. 1421–1426. https://doi.org/10.23919/MIPRO.2018.8400256
12. A. Mordvintsev, C. Olah, M. Tyka, Inceptionism: Going Deeper into Neural Networks. https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html. Accessed 17 June 2015
13. K. Moran, C. Bernal-Cardenas, M. Curcio, R. Bonett, D. Poshyvanyk, Machine learning-based prototyping of graphical user interfaces for mobile apps. IEEE Trans. Softw. Eng. **46**(2), 196–221 (2020). https://doi.org/10.1109/TSE.2018.2844788
14. S. Nam, G. Ko, K. Suh, J. Kwon, User experience- and design-oriented virtual product prototyping system, in *2019 11th International Conference on Knowledge and Smart Technology (KST)*, Phuket, Thailand, 2019, pp. 116–120. https://doi.org/10.1109/KST.2019.8687418
15. P.L. Primandari, A. Sholiq, Effort distribution to estimate cost in small to medium software development project with use case points. Procedia Comput. Sci. **72**, 78–85 (2015). ISSN: 1877-0509. https://doi.org/10.1016/j.procs.2015.12.107
16. C.R. Rubi, A review: speech recognition with deep learning methods. Int. J. Comput. Sci. Mob. Comput. **4**(5), 1017–1024 (2015). ISSN: 2320-088X
17. B.K. Triwijoyo, W. Budiharto, E, Abdurachman, The classification of hypertensive retinopathy using convolutional neural network. Procedia Comput. Sci. **116**, 166 (2017). https://doi.org/10.1016/j.procs.2017.10.066

Fake News Detection in Social Networks Using Attention Mechanism

Om Prakash and Rajeev Kumar

Abstract Online social networks (OSNs) play a vital role in society by connecting people, but their credibility and reliability are questionable due to unverified content. Some people misuse these platforms for cybercrime, e.g., spreading fake contents, which damages social integrity and harmony. Hence, maintaining social networks' integrity is essential by identifying fake contents. In this paper, we have considered fake news as a serious and alarming anomaly threat in OSNs. We have proposed a deep learning-based multimodal architecture that uses *Attention Mechanism* for fake news detection. We have used the benchmark *Weibo* dataset collected from the Weibo micro-blogging Website, containing textual and visual data. LSTM is used to deal with textual data and different pre-trained deep learning models to deal with visual data. The attention mechanism binds textual and visual features, and finally, the softmax layer is used for classification. We achieved 5% more accuracy than the existing base model. Results show that the proposed model can effectively detect fake news.

Keywords Social networks · Fake news · Multimodal · Deep learning · Convolutional neural network · LSTM · Attention mechanism

1 Introduction

Online social networks (OSNs) play a vital role in society by connecting people, but their credibility and reliability are questionable due to unverified contents. Anyone can write or share on social media, which has no credibility. Even Websites or pages are created on popular social media to spread fake news. Most of the fake news on social media focus on trending events to get more attention from users because it is difficult to verify due to its real-time nature. According to Standford University:

O. Prakash (✉) · R. Kumar
School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India
e-mail: omprak16_scs@jnu.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_42

"The news article that is intentionally created and false verifiable could mislead people." It misleads and confuses social media users and changes the opinion and the way people interact in real life, Even though it threatens or damages our social integrity and harmony. It is challenging to tell when and how far-reaching consequences of fake news are.

Hundreds of thousands of fake contents are uploaded every hour; it is hard to keep track of each piece of contents manually. Recently, fake viral messages on social media in India have led to death, killing, and lynching. One of the best examples of misleading video or news share on social media and mainstream media is that a Rs. 2000 currency note contains a GPS tracker. Social media is a platform where fake news and hoaxes thrive. Doctored photos and videos are shared on social media, mainly in India, to set political or social agendas and reshape public opinion. Fake news has low-quality content with poor grammar content with more emotional content.

In this paper, we explore fake news detection techniques from various aspects. First, we describe the proposed model based on multimodal features using deep neural networks. LSTM is used to deal with textual data, and six different pre-trained deep learning models deal with visual data. Then, we combined textual and visual features using attention mechanism to building multimodal features. The attention mechanism helps in focusing keywords of textual data with associated visual information.

The rest of the paper is organized as follows. Section 2 includes the literature survey. Section 3 demonstrates the proposed model, whereas Sect. 4 deals with dataset and preprocessing steps. In Sect. 5, we exhibited results and analysis. Finally, we conclude the paper and include directions for future work in Sect. 6.

2 Related Work

Any news or article intentionally created using manipulated content may not be cross-verified using authentic sources or proof is considered fake news. At first glance, these stories may appear authentic. Such news is mainly created to mislead the people for personal, political, and commercial goals. Online social media platforms are flooded with false information. The information available on social media for fake news is in the form of textual, visual, and mixed data. In this section, we have explored the contribution of researchers to deal with fake news.

Smith and Bharath [1] have used count vectors, TF-IDE, and word embedding to generate feature vectors and used seven different classifiers for a comparative study between fake or real news. Similarly, Meel et al. [2] have used five popular machine learning algorithms: SVM, KNN, LR, RF, Bernoulli Naïve Bayes, and clubbed them to create an ensemble learning model to detect or debunk fake news. In addition, the authors extracted only textual features and implemented max-voting techniques for classification after preprocessing. Baarir et al. [3] have proposed a dataset for fake news detection and used TF-IDF, n-gram for features extraction, and finally, an SVM classifier to classify news. Jin et al. [4] have proposed statistical and visual

characteristics of an image to detect fake news on micro-blogging sites automatically. They considered visual clarity score, visual clustering score, visual coherence score, visual diversity score, and visual similarity distribution histogram as visual features of an image used in micro-blogging sites, image count, image ratio, and popularity of image as a statistical feature of the image. They also extracted textual elements from the publisher profile. They used a publicly available benchmark dataset, "Weibo" and various machine learning algorithms such as SVM, LR, and random forest for the implementation purpose. Qi et al. [5] have used a multi-domain visual network (MVNN) to detect fake news on social networks using Visual contents in frequency and pixel levels. To identify tempered or misleading images, the author used frequency domain sub-network at a physical level and pixel domain sub-networks at pixel level of the image to know the semantic of an image and finally, fusion sub-networks to fuse these images extracted features. A multimodal benchmarked dataset is used provided by "Webio" and tested on a real-world dataset. For better visualization of the results, different pre-trained deep CNN techniques are used as VGG19.

Similarly, Lago et al. [6] described that images used in online social news do not represent the event accurately, as mentioned in the news article. Identifying tempered images composes forensic image algorithms, whereas textual information is a complementary source of information supporting tempered images. In this paper, image forensic tool based on statistical features and convolutional neural network (CNN) is used, including analysis of textual information associated with the text. In traditional, image forensic techniques, date, time, and location were considered; now, in statistical forensic, different statistical methods are used mainly in JPEG digital fingerprints. For textual analysis, different NLP techniques as sentiment analysis are used, and three different datasets and classification techniques are considered to identify fake news. Shu et al. [7] have proposed a repository for fake news detection called FakeNewsNet, which contains a comprehensive dataset with diverse features in news content including social contest and spatiotemporal information. It is a multimodal dataset contain textual as well as visual information which can be used to detect fake news, fake news evolution, fake news mitigation as well as it can be used to detect malicious accounts. The authors have used one hot-vector encoding vector and then applied some standard supervised machine learning models including SVM, LR, NB, and CNN. Saleh et al. [8] have used grid search and hyperopt optimization for optimizing CNN mode for fake news detection. Four different benchmark datasets have been used, and N-gram and TF-IDF have been used to extract features from the dataset. The model's performance compares with RNN and LSTM, including standard ML techniques such as DT, LR, KNN, RF, SVM, and NB. Umer et al. [9] have proposed a hybrid model by combining CNN and LSTM's capability and reducing the feature vector using PCA and Chi-Square. Shrivastava et al. [10] described the propagation of misinformation and its impacts on other misinformation. A different equation is used to get the reproduction number (Ro) by which they can predict message spreading over OSNs and control using reproduction number. Tan et al. [11] have proposed a hybrid model using CNN and LSTM to detect fake news in which CNN extracts high-level sequence features and LSTM long-term dependencies of sequence features.

3 The Proposed Model for Multimodal News

Micro-blogging sites like *Facebook* and *Twitter* provide facilities for tweets or posts in different modularities such as audio, video, text, and images. These tweets contain comprehensive textual and visual information, including social context. The proposed hybrid deep learning model will use such comprehensive information to identify fake news or post being shared on social media.

The objective of our model is to distinguish between fake and non-fake contents. The proposed model has three sub-modules, as demonstrated in Fig. 1: (1) textual features extractor, (2) visual features extractors, and (3) classifier. The model first learns transferable feature representation from user's posts. It extracts different types of features from textual and visual information associated with tweets. Once the textual and visual representation are learned, we concatenate it to make the final multimodal feature. Finally, the multimodal features are input to the classifier for final detection. Here, the instance of tweets can be represented as $I = \{T, S, V\}$ as a tuple containing three modularity as tweets of users T, social context of tweet S, and V as visual information of an attached image. The proposed model extracts features (F) from the given instance and aggregates the final features representation after concatenating them $F_I = \{F_T, F_S, F_V\}$.

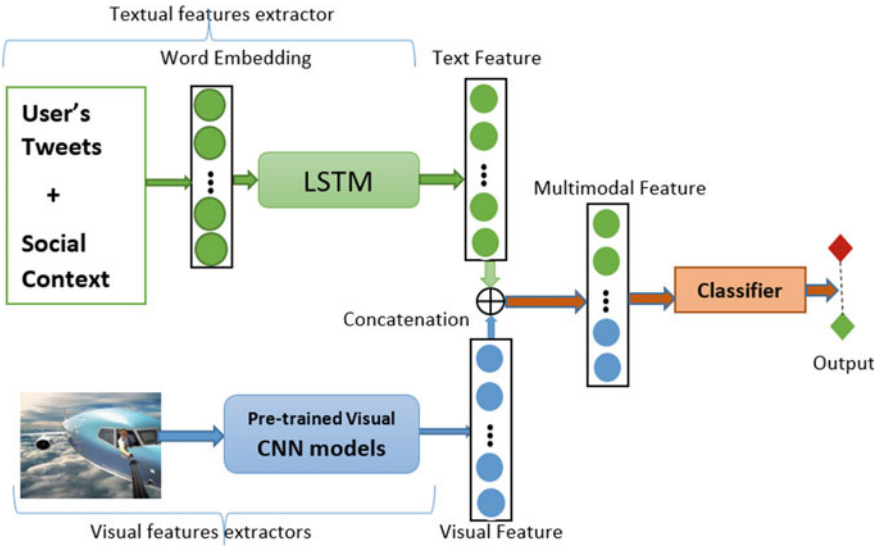


Fig. 1 Architecture of multimodal fake news detector

3.1 Textual Representation

In the first phase of textual feature representation, F_{Text} is formed using user tweet feature F_T and social context of tweets F_S, so the textual features representation $F_{\text{Text}} = [F_T; F_S]$. We employ an RNN with long short-term memory (LSTM) units to learn the joint representation F_{Text} of text and social context in the proposed model. It is the default behavior of LSTM to remember information for a long period. Whereas simple recurrent networks (SRNs) struggles to remember long-term dependencies and optimization hurdles [12].

3.2 Visual Representation

Images of tweets are given to this pre-trained sub-modal to extract visual features denoted as V. We use four different pre-trained models as ResNet-101, GoogleNet, MobileNet-V2, and DenseNet121 in place of the visual sub-model and further added two fully connected layers: Vis-Fc1 and Vis-FC2 to adjust the dimension. The output of the first layer, Vis-Fc1 layer with 512 input size and ReLU activation function, goes as input into the second layer Vis-Fc2 with a softmax function, as $V = \{v1, v2, \ldots, V512\}^T$ for each image. F_V shows the visual feature that can be represented as an Eq. 1.

$$F_V = W_{vf2} \cdot \sigma(W_{vf1} \cdot R_v) \tag{1}$$

In Eq. 1, W_{vf1} and W_{vf2} are weights of fully connected layers. W_{vf1} is weights in the first fully connected layer with ReLU activation function, and W_{vf2} is the weights in the second fully connected layer with softmax function(σ). Obtained features are kept static with concatenation of textual features that are output from LSTM. R_V is the pre-trained CNN model's output representing image features and acting as an input to the first visual fully connected layer with the ReLU activation function.

3.3 Attention Mechanism for Visual Representation

Here, we assume that tweets' images correlate with the textual context of tweets, whereas images play an essential role in highlighting textual features, as we know images contain thousands of information. Visual components are valuable attributes component of fake news. The attention mechanism is similar to the human processing model; they focused on the high-resolution part of an image as explained by [13] and [14]. Textual semantics are somehow related to the semantics of images. Our primary goal of the attention mechanism is to determine semantics' connections. In this multimodal approach, neurons with similarities will assign more weights according to neurons' contributions for different words. Such an attention mechanism for visual

and textual data, having semantic similarities, has been successfully implemented in a computer vision or language vision task.

We take input of LSTM hidden output H_n and pass to non-linear ReLU function and fully connected layer with softmax function to get attention vector Am as shown in Eq. 2 where W_{f1} and W_{f2} are weights of fully connected layers.

$$A_m = W_{f1}\psi(W_{f2} \cdot H_n) \tag{2}$$

A_m is the attention vector obtained from the fully connected layer with softmax function, which has the same dimension as the visual neurons R_V that is output of the pre-trained CNN model and represents image features, and ψ represents softmax function.

Classifier: This sub-model built is on the top of multimodal feature extractor with fully connected layer and softmax activation function. It takes multimodal features as input and predicts an output that particular post is fake or not fake.

4 Dataset and Preprocessing

This "Multimodal fake news detection" dataset used in paper [15] is collected from Weibo (A Chinese micro-blogging Website). This Weibo dataset contains textual as well as visual information and labeled as rumor and non-rumor. Images have unique names corresponding to tweets. Tweets hold three types of information into separate lines; the first line contains meta-information of particular tweets, separated by "|." The second line is the list of uniform resource locators (URLs) of the attached images of a particular tweet. These URLs are also separated by "|" if more than one URLs are available, and finally, a NULL placeholder at the end of URLs. The third line contains the text of tweets, which may be empty. In the form of metadata, tweets contain fifteen different types of information, including publisher and tweets details.

The end-to-end training part is done in two steps. The first step is transforming and preprocessing visual as well as textual data. We manually filter out all images in the ".gif" file format, then resize or transform the remaining images into RGB with 224 * 224 pixels. For textual data, we performed a Regex operation to clean all punctuation marks, then the tokenixation operation was performed to segment sentences into individual linguistic meaning; for this operation, we used "Jieba" Chinese text segmentation library because our textual data is in the Chinese language. Then, the removal of stop words operation is performed to remove all commonly used words, which do not bring any valuable information on text analysis. Finally, we store these processed data into a pickle file format. We also created a separate word vectors for unknown words having the least min_dif document, and 0.25 is chosen from them so that unknown vectors have approximately the same variance as per-trained.

Table 1 Comparative results of the proposed model

Method	Accuracy	Rumor			Non-rumor		
		Precision	Recall	F1 score	Precision	Recall	F1 score
att-RNN [15]	0.788	**0.862**	0.686	0.764	0.738	0.80	0.807
DenseNet-161	0.7829	0.79	0.79	0.79	0.77	0.78	0.78
MobileNet-V2	0.7904	0.76	**0.86**	0.81	**0.83**	0.72	0.77
GoogLeNet	0.80	0.82	0.81	0.81	0.80	0.81	0.80
ResNet-50	0.8157	0.84	0.80	0.82	0.79	0.83	0.81
ResNet-152	0.7945	0.80	0.80	0.80	0.79	0.79	0.79
ResNet-101	**0.8307**	0.85	0.82	**0.83**	0.82	**0.84**	**0.83**

We split a processed dataset into a 7:1:2 ratio for training, validation, and testing purposes. Social semantic features are added with textual features using RNN with LSTM; here, we set $k = 32$ for word dimension, extracted from textual features, and 32 for hidden size to fully connected layer to extract visual. Textual and visual features are concatenated using an attention mechanism, and then, it is passed to a fake news detector. Here, we used optimal parameters for the model while training, dropout probability is 0.4, and the learning rate is 0.001. We used adaptive moment estimation (adam) optimizer and cross-entropy for loss function. For all different pre-trained networks for this model, we set the same parameters. We used 32 batch sizes of an instance for local machines and trained up to 100 epochs.

5 Results and Analysis

We conduct a series of experiments with various hyper-parameters tuning of each pre-trained neural network. Table 1 demonstrates the performance of the above model in terms of accuracy, precision, recall, and F1 score for fake and non-fake. The first row of the table is the highest accuracy of an existing model, where att-RNN indicates the attention mechanism with RNN. The remaining rows demonstrate the performance of the proposed model with different pre-trained CNN models. The highlighted with bold indicates the maximum performance gain by models. As our problem was a binary classification problem (news is fake or not), we plot the confusion matrix of the best model, where one (1) indicates fake, or rumor news and zero (0) indicates non-rumor or non-fake news. Finally, we plot the AUC-ROC curve and loss for training and validation.

We observed from Table 1 that use of DensNet-161 in place of VGG-19 for visual features extraction increased only 3% F1 score and 10% recall than the existing model. Here, accuracy is almost the same. The major difference between the existing model and our model is that we used batch size 32 in place of 128 and decreased dropout probability from 0.5 to 0.4%. In another experiment, when we replaced DenseNet-161 with the MobileNet-V2 CNN model for visual features extraction

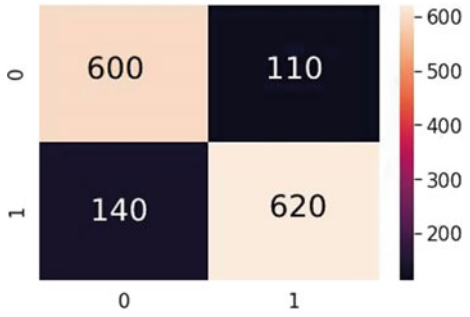


Fig. 2 Confusion matrix of ResNet-101

by keeping the same parameter as DensNet-161, we get 1% better accuracy than the existing base model. We also get 20% better recall, 5% better F1 score for rumor, and 10% better precision for non-rumor. Recall of rumor and precision of non-rumor is the highest prediction score by MobileNet-V2 among all experiments with different CNN models. By replacing MobileNet-V2 with GoogLeNet with the same parameter as used in MobileNet-V2, we get 2% more accuracy than the existing att-RNN model. We also get 13% better recall and 5% better F1 score for rumor, whereas 7% precision and 1% recall for non-rumor. The accuracy of this CNN model is 1% higher than the MobileNet-V2 model.

Finally, we used three variants ResNet: ResNet-50, ResNet-101, and ResNet-152 for visual feature extraction. We use the same parameter tuning as used for the previous model. Using the ResNet-50, we get 3% more accuracy than the existing model, including 12% in recall, 6% F1 score for fake news, 6% precision, and 1% F1 score for non-rumor news, respectively, than the base model. In contrast, we decrease the accuracy by 2% by using ResNet-152 than ResNet-50. Hence, we decided to use a lower version of ResNet than ResNet-152 and succeeded in accuracy by using ResNet-101. It increased the accuracy by 5%, F1 score for fake by seven percentage, recall and F1 score by 4% and 3%, respectively, for non-fake than att-RNN base mode, which is highest among all models. Confusion matrix ResNet-101 is shown in Fig. 2.

Figure 3 demonstrates the training and validation AUC-ROC curves of proposed model with ResNet-101. The learning patterns of various models are similar so we have only shown the AUC-ROC curve of ResNet-101 since it has the highest test accuracy. We trained and validated this model with 100 epochs, where the model has learnt under 20 epochs thereafter, it stabilizes the training; hence, there is no improvement in training and validation accuracy. We also tested the model for longer epochs up to 100 and different dropout rates, but there is no improvement in training. However, the maximum accuracy gained by the model is 99% in the training and around 80% in validation.

Training and validation losses are also almost similar for all different CNN models. The initial label cost of training and validation is decreasing. As it can be observed from Fig. 4, the cost of training and validation is decreasing up to 20 epochs after it keeps on fluctuation, and there is no further improvement. However, the training

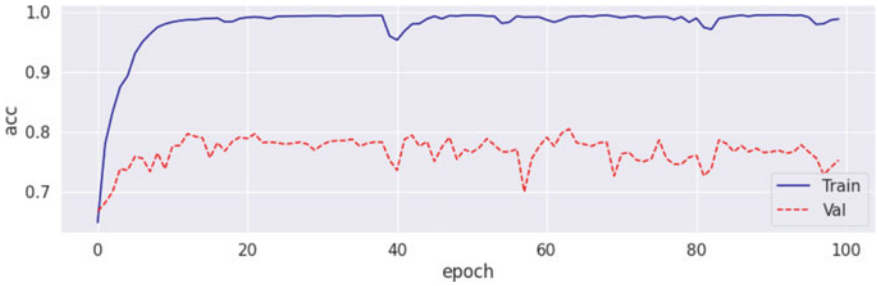


Fig. 3 Training and validation AUC-ROC curves

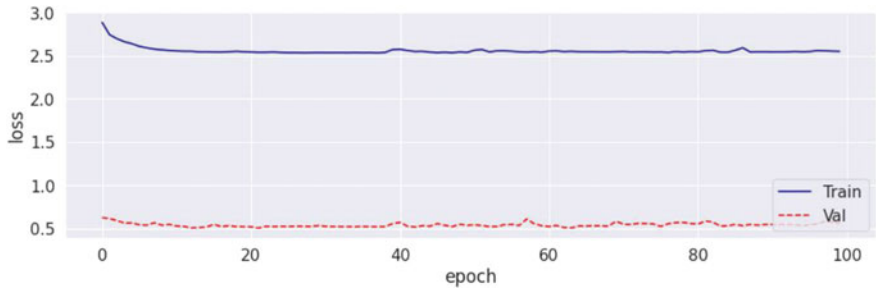


Fig. 4 Training and validation loss

cost has reduced from 2.85 to 2.55, and validation cost has decreased from 0.60 to 0.52.

6 Conclusion

With the rise of digitization and social media, it is challenging to deal with fake news because it seems to be real. Hence, to deal with fake news, we have proposed a multimodal fake news detector that classifies news as fake or not fake. In the proposed model, LSTM is used to extract textual features from the text, whereas the pre-trained CNN model is used to extract visual features from the image. Here, we used four most popular pre-trained CNN models in VGG-19 as used in the base model. The Weibo dataset is used for training and testing the model. The proposed model with ResNet-101 works well to identify fake news with 83% accuracy that is 5% more than the base model. As a future work, we would like to explore further user-independent features related to fake news and expand the research work by identifying patterns of social media posts. In addition, we would like further improve the performance of the proposed model by collecting larger datasets and stacking techniques.

References

1. N. Smitha, R. Bharath, Performance comparison of machine learning classifiers for fake news detection, in *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications*, 2020, pp. 696–700
2. P. Meel, P. Chawla, S. Jain, U. Rai, Web text content credibility analysis using max voting and stacking ensemble classifiers, in *Proceedings of the Advanced Computing & Communication Technologies for High Performance Applications*, 2020, pp. 157–161
3. N.F. Baarir, A. Djeffal, Fake news detection using machine learning, in *Proceedings of the International Workshop on Human-Centric Smart Environments for Health and Well-being*, 2021, pp. 125–130
4. Z. Jin, J. Cao, Y. Zhang, J. Zhou, Q. Tian, Novel visual and statistical image features for microblogs news verification. IEEE Trans. Multimed. **19**(3), 598–608 (2016)
5. P. Qi, J. Cao, T. Yang, J. Guo, J. Li, Exploiting multi-domain visual information for fake news detection, in *IEEE International Conference on Data Mining (ICDM)*, 2019, pp. 518–527
6. F. Lago, Q.-T. Phan, G. Boato, Visual and textual analysis for image trustworthiness assessment within online news. Secur. Commun. Netw. (2019)
7. K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, Fakenewsnet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data **8**(3), 171–188 (2020)
8. H. Saleh, A. Alharbi, S.H. Alsamhi, OPCNN-FAKE: Optimized convolutional neural network for fake news detection. IEEE Access **9**, 129471–129489 (2021)
9. M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G.S. Choi, B.-W. On, Fake news stance detection using deep learning architecture CNN-LSTM. IEEE Access **8**, 156695–156706 (2020)
10. G. Shrivastava, P. Kumar, R.P. Ojha, P.K. Srivastava, S. Mohan, G. Srivastava, Defensive modeling of fake news through online social networks. IEEE Trans. Comput. Soc. Syst. **7**(5), 1159–1167 (2020)
11. K.L. Tan, C.P. Lee, K.M. Lim, Fake news detection with hybrid CNN-LSTM, in *Proceedings of the International Conference on Information and Communication Technology (ICoICT)*, 2021, pp. 606–610
12. K. Greff, R.K. Srivastava, J. Koutník, B.R. Steunebrink, J. Schmidhuber, LSTM: a search space odyssey. IEEE Trans. Neural Netw. Learn. Syst. **28**(10), 2222–2232 (2016)
13. S. Li, T. Xiao, H. Li, B. Zhou, D. Yue, X. Wang, Person search with natural language description, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1970–1979 (2017)
14. Q. You, L. Cao, H. Jin, J. Luo, Robust visual-textual sentiment analysis: when attention meets tree-structured recursive neural networks, in *Proceedings of the 24th ACM International Conference on Multimedia*, 2016, pp. 1008–1017
15. Z. Jin, J. Cao, H. Guo, Y. Zhang, J. Luo, Multimodal fusion with recurrent neural networks for rumor detection on microblogs, in *Proceedings of the 25th ACM International Conference on Multimedia*, 2017, pp. 795–816

Comparative Analysis of Process Mining Algorithms in Industrial Applications

M. Shanmuga Sundari and Rudra Kalyan Nayak

Abstract In businesses, people have to manage the processes efficiently to build products and services to meet customer requirements. The process generally consists of a large number of activities that users need to manage efficiently. Process mining plays a vital role in all business domains. It helps organizations to organize activities and improve productivity. These processes in business generate activity data and produce overwhelming process data. Process mining helps connect these activities, applies business intelligence to the process, and finds hidden patterns in the business process. Process mining provides different process models and algorithms. The models use event log data of activities and find patterns in the process data. Process mining algorithms work on different scenarios in business and usage of domains in the application. This paper explains the various algorithms and determines the use of process mining applications in business domains. This research gives the solution to find the most suitable model for business applications.

Keywords Alpha miner · Fuzzy miner · Genetic miner · Heuristic miner · Process mining

1 Introduction

Process mining [1] is the intermediate bridge between process-based analysis and data-oriented analysis techniques. The process-based analysis provides the knowledge of activities in the process. On the other hand, data-oriented analysis uses machine learning techniques to analyze data, evaluate the accuracy of the model, and find insights from data. Process mining can bridge the gap between process-based analysis and data-centric analysis [2]. Process mining is a tool to seek the confrontation between the event logs and models used in mining. For example, business administrators define the processes in the organization and their sequence. In

M. Shanmuga Sundari (✉) · R. K. Nayak
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
e-mail: mshanmugasundari@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_43

supply chain management, the logistics route from source to destination is considered a process. In the online delivery business, the product delivery to the end consumer without delay is considered a process. The number of activities will be different for each trace. The information which is the backbone of the process is mined with a proper model. Process mining is a relatively good discipline to achieve the analysis of the activities. The process mining tool ingests real-time information or event logs [3] based on the extracted information. Process mining uses the real-time event log data and applies the models. It can picture the real flow of the process in the organization that shows the efficacy [4] of the resources.

2 Process Mining Algorithm

There are various algorithms widely used to evaluate the business process. Each algorithm will execute various scenarios and fit into the particular domain.

2.1 Alpha Miner

Process mining has been implemented in three sequence steps. The steps will help to prepare the dataset [5] and make it ready for the process analysis: Pre-processing, Processing, and Post-processing. The α-algorithm [6] is the basic algorithm that works in the sequence of activities and builds the Petri net diagram.

Let T be a set of activities. W be the event log in T. *first* (σ) is the first trace in the event log. I define the first trace in the event. O is the last trace in an event which is calculated using the *last* (σ). X is the pair between the events in activities. X, Y are the two different event flows. Y is the set of minimal pairs of activities. P represents the place of traces. F determines the connectivity matrix in between the activities. Let W consider as a workflow log over U. The (W) is defined as follows.

1. $U_W = \{u \in U | \exists_{\sigma \in W} u \in \sigma\}$
2. $U_I = \{u \in U | \exists_{\sigma \in W} u = \text{first}(\sigma)\}$
3. $U_W = \{u \in U | \exists_{\sigma \in W} t = \text{last}(\sigma)\}$
4. $V_W = \{(X, Y) | X \subseteq U_W \wedge Y \subseteq U_W$
   $\wedge \forall_{\sigma \in X} \forall_{y \in Y} x \rightarrow Wy \wedge \forall_{x1,x2 \in X} x1 \#_W x2 \wedge \forall_{y1} \#_W y2\}$
5. $W_w = \{(X, Y) \in X_W | \forall_{(X,Y) \in X W} X \subseteq X^{'} \wedge Y \subseteq Y^{'} \Rightarrow (X, Y) = (X^{'}, Y^{'})\}$
6. $P_W = \{P(X, Y) | (X, Y) \in Y_W\} \cup \{i_w, o_w\}$
7. $F_W = \{(a, P(X, Y)) | (X, Y) \in Y_w \wedge x \in X\} \cup \{(P(X, Y), y) | (X, Y)$
   $\in Y_W \wedge y \in Y \cup \{i_w, u\} | u \in U I\} \cup \{(u, o_w) | u \in U_o\}$
8. $\alpha(W) = (P_W, U_W, F_W)$

2.2 Heuristic Miner

Heuristics miner [6] works with frequencies of events and activities and builds the frequency matrix [7]. It can deal with noisy events and find the behavior of the process. It can identify the exceptions of the event log which is not identified using the alpha algorithm. Consider there is N number of activities. A is the sequence of activity. E is the event log. In the event log, the activities will have happened multiple times. The following steps of activities show the relationship used in the algorithm. Let E consider as event log over N, Let x,y are the activities in $x,y \in N$:

1. $x\ E\ y$ if any traces are noted in $N = t_1 \ldots t_n$, t is the time stamp for the activity
2. $x \rightarrow E\ y$ whenever the activity meets $x > E\ y$ and $y > E\ x$
3. $x \neq E\ y$ and $x \parallel E\ y$ whenever the activity meets $x > E\ y$ and $y > E\ x$.
4. $x > > E\ y$ and $x > > > E\ y$ whenever the activity meets a trace of $N = ts_1, ts2, ts3 \ldots . t_n$ and $i\ \mathfrak{E}\ \{1;:::;\ n\ -\ 2\}$ that based on $N \in E$ and $ts_i = x$ and $ts_i + 1 = y$ and $ts_i + 2 = x$.

The dependency graph shows the value of the dependency relation (1) between x and y:

$$|x =>_L y| = \left\{ \begin{array}{ll} \frac{x=>_L y|-|y+>_L x|}{|x=>_L y|+|y+>_L x|+1} & \text{if } x \neq y \\ \frac{|x=>_L y|}{|x=>_L y|+1} & \text{if } x = y \end{array} \right\} \tag{1}$$

This will produce the dependency graph from the traces [8] of activities. We can identify the dependency in between the activities based on the out product of the formula. The join and split process will find the proper flow of the activities. Based on the flow it can find the deviation [9] in the process in real-time application.

2.3 Genetic Miner

The genetic algorithm is started with an initial population [10] and crossover [11] which are created using dependency measures [12]. The initial population is derived from the dependency measures between the activities. For example $x1$, $x2$ are the activities. The dependency measure $D(x1, x2)$ is high than the causal high probability then we have it true.

$$\begin{aligned} \text{Fitness}_S(\text{PM}, L) = 0.20 \times & \frac{\text{parsed\_activities}_S(\text{PM}, L)}{\text{num\_activities\_log}(L)} + .0.30 \\ & \times \frac{\text{completed\_log\_traces}_S(\text{PM}, L)}{\text{num\_traces\_log}(L)} + 0.50 \\ & \times \frac{\text{completed\_log\_traces}_S(\text{Pm}, L)}{\text{num\_traces\_log}(L)} \end{aligned} \tag{2}$$

After the initial population [13] is derived, then the fitness of the events is calculated using the below formula. The stop semantics is calculated using the activity log traces. Let L be an event log and PM be a process model. The fitness (2) and (3) is calculated based on the completed traces with a successful time frame. The continuous semantics [14] are calculated using the traces of the event log and shown below. The fitness of the continuous semantics is:

$$\text{Fitness}_C(\text{PM}, L) = 0.40 \times \frac{\text{Parsed\_activities}_C(\text{PM}, L)}{\text{all\_actiivies\_log}_C(L)} + 0.60 \times \frac{\text{completed\_traces}_C(\text{PM}, L)}{\text{num\_traces\_log}(L)} \tag{3}$$

Then crossover and mutation are evaluated with the fitness values [15]. Crossover generates the new individuals from the fittest [16] in the current population.

2.4 Fuzzy Miner

The Fuzzy Miner [17] is applicable for semi-structured processes in a large database. The complex and uncertain activities [18] will happen in a real-time process. The unstructured process [19] is very difficult to find the flow. Fuzzy miner has been obtained using the flow of sub-logs in the particular event. It will execute up to the low-level abstraction in the sub-log also. A fuzzy net is generated as a Simple Precedence Diagram (SPD) [20]. An SPD consists of nodes and edges called a directed graph. The nodes are considered as activities, and edges are considered as transactions. A Fuzzy Performance Diagram (FPD) is the visual output diagram of an SPD. It also shows the performance analysis information. It can predict the control flow of a process in a visual manner that can help the process mining analysis. This will be captured in all nodes in the SPD in terms of clusters.

3 Comparative Analysis of Algorithms

The comparative analysis of process mining algorithms is displayed in Table 1. The resultant table depicts the various aspects of criteria and explained the differences of the algorithms.

4 Conclusion

The importance and usage of process mining increase in real-world dynamic applications. Many of the domains find this process mining useful. In this paper, we have

Table 1 Analysis of various process mining algorithms in industry aspect

	Alpha	Heuristic	Genetic	Fuzzy
Input	Basic event and traces	Simple and less events	Random population of process	Unstructured log data
Output	Petri net graph	Heuristic Net	Petri net graph	Fuzzy model
Operation	Using split and join do replay of process	Dependency matrix, split, and join operation	Control flow operation	Zoom in and control the process
Behavior	Flow of process through petrinet	Frequency of process	Natural evolution	Dependency graph
Technology	Local strategy technique	Local strategy	Global strategy technique	Global strategy
Log	Unstructured log	Unstructured log	Structured and unstructured log	Less structured log
Data	Incomplete data	Incomplete data	Incomplete data	Incomplete data
Problems	Noise and incompleteness	Noise log	Hidden task cannot mine	Cannot create petri net
Domains	Reconstruct causality events	Behavior events	Comparison native event	Dynamic process behavior

carried a comparative analysis of various process mining algorithms and domain usage. At the end of this research based on many papers, we can conclude that:

Alpha miner: This can be used for concrete process log and using replay it can find the deviation in the process. This is the basic model in process mining, which can show the flow of traces in the process.
Heuristic miner: This can be used for fewer noise data and having a simple sequence of activities in the process. This can produce the frequency dependency between the activities.
Genetic miner: This can be used with the noisy processes, multiple repetitions of activities in process with any kind of choices inflow. This can populate the evolution and analysis with the native events.
Fuzzy miner: This can be used with complex and irrelevant processes. This can depict the traits and eliminate irrelevant activities.

References

1. M.S. Sundari, R.K. Nayak, Process mining in healthcare systems: a critical review and its future. Int. J. Emerg. Trends Eng. Res. **8**(9), 5197–5208 (2020). https://doi.org/10.30534/ijeter/2020/50892020
2. W.M.P. Van Der Aalst et al., Business process mining: an industrial application. Inf. Syst. **32**(5), 713–732 (2007)

3. J. Geyer-klingeberg, J. Nakladal, F. Baldauf, F. Veit, Process mining and robotic process automation : a perfect match process mining as enabler for RPA ımplementation, in *16th Int. Conf. Bus. Process Manag.*, vol. i, no. July, 2018
4. R.R. Reddy, Y. Ramadevi, K.V.N. Sunitha, Robust Data Model for Enhanced Anomaly Detection," in *Proceedings of the International Congress on Information and Communication Technology*, 2016, pp. 439–446
5. Q. Cai, H. Wang, Z. Li, X. Liu, A survey on multimodal data-driven smart healthcare systems: approaches and applications. IEEE Access (2019). https://doi.org/10.1109/ACCESS.2019.2941419
6. R. Tripathy, R. K. Nayak, P. Das, and D. Mishra, Cellular cholesterol prediction of mammalian ATP-binding cassette (ABC) proteins based on fuzzy c-means with support vector machine algorithms, *J. Intell. Fuzzy Syst.*, vol. 39, no. 2, 2020, doi: https://doi.org/10.3233/JIFS-179934.
7. S. Chen, S. Yang, M. Zhou, R. Burd, I. Marsic, Process-Oriented Iterative Multiple Alignment for Medical Process Mining, 2017, doi: https://doi.org/10.1109/ICDMW.2017.63
8. M. Prodel, V. Augusto, B. Jouaneton, L. Lamarsalle, X. Xie, Optimal process mining for large and complex event logs. IEEE Trans. Autom. Sci. Eng. (2018). https://doi.org/10.1109/TASE.2017.2784436
9. W. M. P. Van Der Aalst, A. K. A. De Medeiros, and A. J. M. M. Weijters, Genetic process mining, 2005, doi: https://doi.org/10.1007/11494744_5
10. S.A. Shinde, P.R. Rajeswari, Intelligent health risk prediction systems using machine learning: a review. Int. J. Eng. Technol. **7**(3), 1019–1023 (2018). https://doi.org/10.14419/ijet.v7i3.12654
11. M. D. N. Salman, P. T. Rao, U. R. Rahman, M. D. Zıa, Adaptıve noıse cancellers for cardıac sıgnal enhancement for ıot based health care systems. *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 10, 2017
12. A. Mubarakali, M. Ashwin, D. Mavaluru, A.D. Kumar, Design an attribute based health record protection algorithm for healthcare services in cloud environment. Multimed. Tools Appl. **79**(5), 3943–3956 (2020)
13. P. Weber, B. Bordbar, P. Tino, A framework for the analysis of process mining algorithms, *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 43, no. 2, pp. 303–317, 2012
14. M. Weidlich, A. Polyvyanyy, N. Desai, J. Mendling, M. Weske, Process compliance analysis based on behavioural profiles. Inf. Syst. **36**(7), 1009–1025 (2011)
15. R. Gatta *et al.*, A framework for event log generation and knowledge representation for process mining in healthcare, 2018, doi: https://doi.org/10.1109/ICTAI.2018.00103
16. D. Van, Process mining discovery, conformance and enhancement of business processes, *Springer Heidelb. Dordr. London New York. ISBN*, 2011
17. A. K. A. de Medeiros, B. F. van Dongen, W. M. P. van der Aalst, and A. J. M. M. Weijters, Process mining: Extending the a algorithm to mine short loops, *BETA Work. Pap. Ser.*, 2004
18. S. Chandra Sekaran, V. Saravanan, R. K. Nayak, R. Tripathy, and S. Siva Shankar, Human health and velocity aware network selection scheme for WLAN/WiMAX integrated networks with QoS, *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 12, pp. 198–205, 2019, doi: https://doi.org/10.35940/ijitee.L3571.1081219
19. S. S. Mirza and M. Z. Ur Rahman, Efficient adaptive filtering techniques for thoracic electrical bio-impedance analysis in health care systems, *J. Med. Imaging Heal. Informatics*, vol. 7, no. 6, pp. 1126–1138, 2017
20. B. Tibeme, H. Shahriar, and C. Zhang, Process Mining Algorithms for Clinical Workflow Analysis, 2018, doi: https://doi.org/10.1109/SECON.2018.8479176

Ridge Regression for PSNR of Restored Images by Recursive Median Filter

Shweta Aggarwal and Himanshu Agarwal

Abstract The study and development of image noise removal algorithm (INRA) are very important research area, where the goal is to remove unwanted noise from an image such that maximum significant information can be observed in noise removed image. Peak-signal-to-noise-ratio (PSNR) and execution time (ET) are the main criteria for INRA evaluation. In this paper, ridge regression models of degree three have been developed for prediction of PSNR of recursive median filter-based INRA without its implementation. Prediction of PSNR of INRA by regression model is a novel contribution which helps to save ET. Accurate prediction is a challenging problem due to randomness in noise, images, and sampled data. *Theory based on Bias Variance Decomposition Theorem (BVDT) and minimum testing error* has been used to select 'good' regression models. Experiments ensure computation time advantage by a factor of 10^{-5} at the cost of 1% and 10% error in PSNR values of the database with one and ten images, respectively.

Keywords Image noise removal algorithm (INRA) · Ridge regression · Bias variance decomposition theorem (BVDT)

1 Introduction

Digital images are very important data that provide visual information of facts in science, engineering, and technology. Display of image on the screen (image rendering), image capturing, and image transmission over the Internet is the most frequent phenomenon associated with digital images. During each phenomenon, the quality of the image can be degraded. For example, noise due to electronic and photometric

S. Aggarwal (✉) · H. Agarwal
Department of Mathematics, Jaypee Institute of Information Technology, Noida 62, Uttar Pradesh, India
e-mail: aggarwal.shweta15@gmail.com

H. Agarwal
e-mail: himanshu.agarwal@jiit.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_44

effect in image sensors, atmospheric turbulence, relative motion between the camera and the original scene, etc. [1].

The quality of degraded images must be improved by removing the unwanted noise from the image to observe the maximum possible significant information available in an image [2]. This process is known as image denoising (known as image restoration if noise model is known), and it is often the first step in many image processing applications such as photography, publishing, astronomy, forensic science, medical imaging, computed tomography, magnetic resonance imaging, recognition (face, license plate, fingerprint, signature), and various fields of image processing (registration, segmentation, classification, etc.).

It is a general observation that the peak-signal-to-noise-ratio (PSNR) of a restored image depends on the image under study, parameters of noise model, image noise removal algorithm (INRA), and randomness due to noise. Mafi et al. [3] used execution time, correlation, PSNR, structural similarity index measure (SSIM) as the assessment parameters (AP) of INRA. They considered the randomness factor of noise and provided the interval for AP. Moreover, the study of AP with respect to (w.r.t.) noise density and different images open the way to fit regression models for the prediction of AP.

In this paper, prediction of PSNR of INRA by regression model [4] is a novel contribution. This work has applications to predict the PSNR value without actual implementation of the INRA and thus has the potential to reduce the computation time of prediction. Selection of regression model for accurate prediction of PSNR values is a challenging problem due to randomness in noise, images, and sampled data. The main contributions are as follows:

- Ten standard images [5] have been used to generate two databases. In these images, random generated salt and pepper noise are added and then removed by recursive median filter (RMF) [6] based INRA.
- Three experiments have been performed to study the distribution of coefficients and prediction error of ridge regression models. Randomness due to noise, images, and sampled data has been considered in the experiments 1, 2, and 3, respectively.
- *Theory based on bias variance decomposition theorem (BVDT) and minimum testing error* have been used for the selection of 'good' regression models.

Experiments ensure computation time advantage by a factor of 10^{-5} at the cost of 1%, 13%, and 9% average error in PSNR values for experiment 1, 2, and 3, respectively.

Rest of the paper is organized as follows. Related work is explained in Sect. 2. The algorithm for PSNR computation of the restored image followed by data generation is discussed in Sect. 3. Ridge regression model is explained in Sect. 4. Three experiments have been performed in Sect. 5 for results and analysis of coefficients, PSNR prediction, and computation time. Conclusions and future work are highlighted in Sect. 6.

2 Literature Survey

Gaussian noise, salt and pepper noise (impulse noise) [7], and rain streak [8] are the popular noise models. Development and evaluation of INRA are an active research area. Evaluation of PSNR [9] on a set of images is the widely used strategy for the assessment of INRA.

In this paper, the standard median filter (SMF) [1] is the first choice for salt and pepper noise removal due to its root and deterministic properties [10, 11]. RMF [6] improves image quality repeatedly before its convergence by repeated application of the median filter. Several other filters based on median approach are weighted median filter, center-weighted median filter, and adaptive length median filter [9]. Mafi et al. [7] provided a very good survey on the impulse noise INRA. Few recent impulse noise INRA are improvement boundary discriminative noise detection (IBDND) [9], decision-based unsymmetrical trimmed modified winsorized mean filter (DBUTMWMF) [12], switching adaptive median and fixed weighted mean filter with shrinkage window (SAMFWMF) [3], unbiased-based weighted mean filter (UWMF) [13], two-stage algorithm by combination of multilevel weighted graphs model and induced generalized ordered weighted average operator (MWGIGOWA) [2], an adaptive window of size $n \times n$ [14], multilayer decision-based iterative filter [15].

3 Algorithm for PSNR of Restored Image and Data Generation

Algorithm for computation of PSNR of restored image and details of data generation are described.

3.1 Algorithm

Inputs are noise density ρ of salt and pepper noise and image A of size $m \times n$ pixels and output recorded is PSNR value (PSE) of a restored image. Stepwise details are as follows:

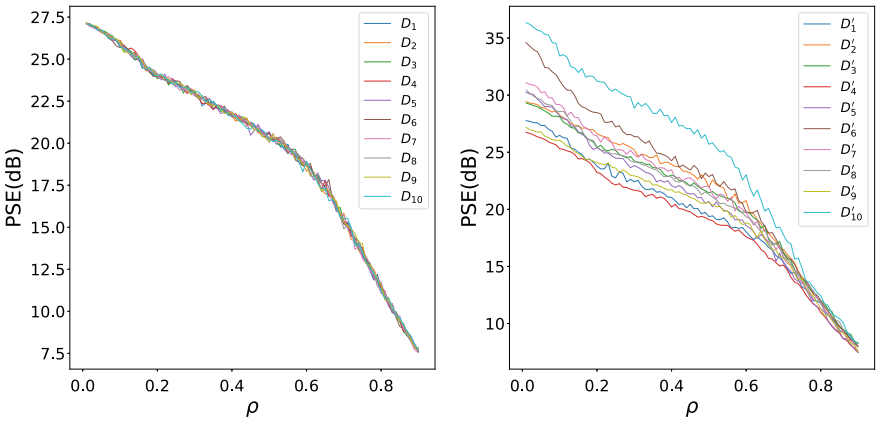
Step 1: Salt and pepper noise of density ρ is added to image A to obtain image A'. Note that noise is a random quantity. Every time distribution of noise is different for a fixed value of ρ.

Step 2: Median filter of size 3×3 is applied recursively on A' for 30 iterations. Store the obtained images as B_1, B_2,…, B_{30}.

Step 3: Compute PSNR P_i between A and B_i where, $i = 1, 2, \ldots, 30$. Store the results in P_i.

Step 4: PSE = $\max_{i=1}^{30} P_i$ is the final value of PSNR.

Note that step 2 to step 4 are equivalent to noise removal by RMF [6, 16].



(a) DB1 (b) DB2

Fig. 1 Noise density ρ versus PSE

3.2 Data Generation

Two databases (say DB1 and DB2) have been generated. DB1 and DB2 consist of ten datasets (say, D_1, D_2, …, D_{10} and D'_1, D'_2, …, D'_{10}, respectively). Each dataset has two attributes (ρ, PSE), where ρ is the density of salt and pepper noise, and PSE is PSNR value of enhanced image as described in step 4 of Sect. 3.1. Datasets of DB1 are generated by applying the algorithm of Sect. 3.1 on 'cameraman' image [5] for $\rho \in \{0.01, 0.02, \ldots, 0.9\}$. Random salt and pepper noise is generated 10 times for fixed ρ. The suffixes in the datasets represent randomness due to noise generation. These ten datasets are visualized in Fig. 1a. Datasets of DB2 are generated by applying the algorithm of Sect. 3.1 on a set of 10 images (Lena, Barbara, Boat, Cat, Airplane, Peppers, Fruits, Monarch, Cameraman, Zelda) [5] for $\rho \in \{0.01, 0.02, \ldots, 0.9\}$. The suffixes in the datasets represent an image. Database DB2 is presented in Fig. 1b.

4 Ridge Regression Model

Theory and design of regression models are discussed to predict the PSE values for a given ρ. Available data is divided into training (TR) and testing (TS) sets for model development and validation, respectively. Sum of the square of the error between predicted PSE (Y) and actual PSE of the training set is a common loss function for the design of regression models [4]. If data is limited, then overfitting is an issue with sum of the square of the error. Ridge regression model (proposed by Hoerl et al. [17, 18]) utilizes a regulation parameter in loss function to solve the problem of overfitting. The prediction formula by a polynomial of degree three is

$$Y(\rho\ , \boldsymbol{w}) = \sum_{i=0}^{3} w_i \rho^i \tag{1}$$

where $\boldsymbol{w} = (w_0, w_1, w_2, w_3)$ are the coefficients. The loss function is

$$E = \sum_{j=1}^{N_1} \left(\text{PSE}_j - \sum_{i=0}^{3} w_i \rho_j^i \right)^2 + \frac{\lambda}{2} \sum_{i=0}^{3} w_i^2 \tag{2}$$

where N_1 is the number of instances in training dataset, j is the index for data points in the training set, $\lambda > 0$ is the regularization parameter. If $\lambda = 0$, then loss function is sum of the square of the error. The coefficients are computed by minimization of E. Therefore,

$$\frac{\partial E}{\partial w_i} = \sum_{j=1}^{N_1} \left(w_0 + w_1 \rho_j + w_2 \rho_j^2 + w_3 \rho_j^3 - \text{PSE}_j \right) \rho_j^i + w_i \lambda = 0, i = 0, 1, 2, 3 \tag{3}$$

The solution of Eq. (3) provides the coefficients. It is worthy to note that if $\lambda > 0$, then a unique solution of Eq. (3) always exists. The existence of a unique solution was the main motivation for Hoerl et al. [17, 18] to develop the ridge regression model. Training and testing errors defined in Eqs. (4) and (5) are used for model validation.

$$\text{Training error} = \sqrt{\frac{\sum_{j=1}^{N_1} \left(Y(\rho_j\ , w) - \text{PSE}_j \right)^2}{N_1}} \tag{4}$$

$$\text{Testing error} = \sqrt{\frac{\sum_{k=1}^{N_2} \left(Y(\rho_k\ , w) - \text{PSE}_k \right)^2}{N_2}} \tag{5}$$

where N_2 and k are number of instances and index for data points, respectively, in testing dataset.

5 Experiments, Results, and Analysis

Ridge regression models are constructed for $\lambda \in \{0.0, 0.01, 0.02, \ldots, 0.1\}$, data DB1 and DB2 to predict the PSE values of the INRA (as discussed in Sect. 3.1). Three experiments have been performed to study the distributions of coefficients (w_0, w_1, w_2, w_3), training and testing errors of constructed models w.r.t. randomness due to noise (Experiment 1), various images (Experiment 2), and data selection (Experiment 3). Predicted PSE values and error histograms have been described at a

good estimate of coefficients and λ in Sect. 5.1. A comparison of training and prediction time is provided in Sect. 5.3 followed by PSNR performance of state-of-the-art INRA.

5.1 Distribution of Coefficients and Error

The purpose of this section is (i) to study the distribution of coefficients w.r.t. randomness due to noise, various images, and data selection and (ii) to find a good estimate of coefficients and λ.

The coefficients are estimated for $\lambda \in \{0.0, 0.01, 0.02, \ldots, 0.1\}$ by solving Eq. (3). Quality of estimator is decided by (i) error in prediction of training and testing data (ii) mean squared error of estimator which is the sum of bias square, variance, and irreducible noise according to *bias variance decomposition theorem* (BVDT) [19]. Choice of a good estimate is a challenging task due to the trade-off among various errors.

5.1.1 Experiment 1

Database DB1 is used to study the distribution of coefficients w.r.t. randomness due to noise. For each dataset of the DB1, 70% data (63 instances) is utilized for the training purpose, and the rest 30% (27 instances) data is used for testing. Ten ridge regression models have been constructed, one corresponding to each dataset. The average and standard deviation (SD) of training and testing errors of these ten models are shown in Fig. 2a and b, respectively. The average and SD of coefficients of the models are provided in Table 1. Important observations are as follows:

1. Average training and testing errors increase w.r.t. λ, and variation in coefficients decreases w.r.t. λ.
2. λ-w tuple $(\lambda, w_0, w_1, w_2, w_3)$ for the minimum training error (0.269 dB) is $(0, 27.51, -20.85, 30.42, -36.09)$ and $(0, 27.59, -22.00, 34.42, -39.33)$ for the minimum testing error (0.211 dB).
3. λ-w tuple for the maximum training error (0.546 dB) is $(0.1, 26.28, -6.77, -6.58, -9.47)$ and $(0.1, 26.32, -6.86, -6.71, -9.40)$ for the maximum testing error (0.632 dB).
4. Small values of SD (Fig. 2b and Table 1) ensure that coefficients, training, and testing errors are almost independent of randomness due to noise.

5.1.2 Experiment 2

Second experiment is performed on the database DB2 to study the distribution of coefficients w.r.t. randomness due to images, and its procedure is similar to as dis-

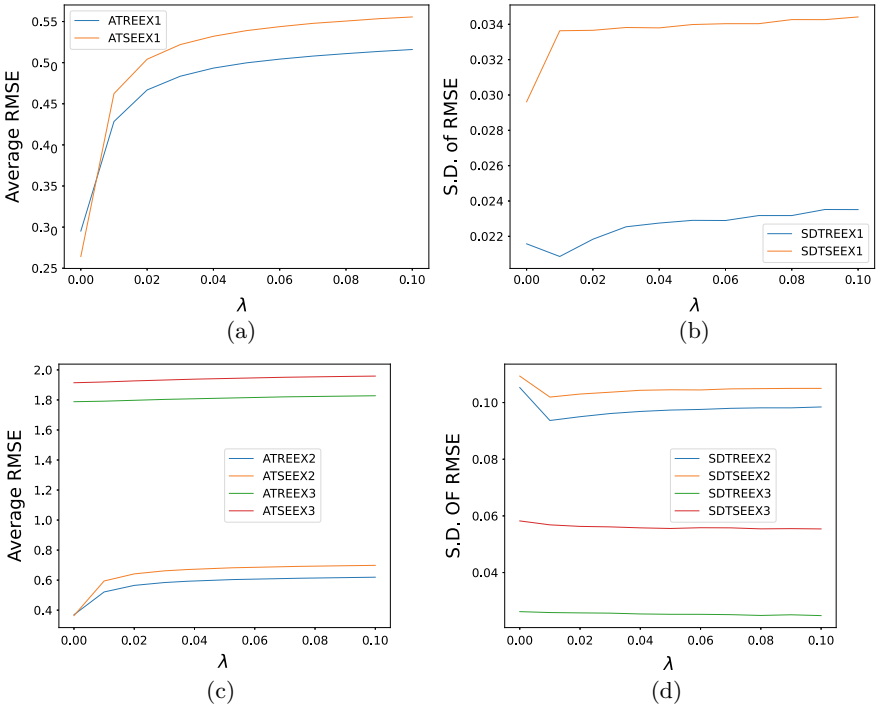


Fig. 2 Results of experiment 1 (EX1), experiment 2 (EX2), and experiment 3 (EX3). **a** Average of RMSE versus λ of EX1. **b** Standard deviation (SD) of RMSE versus λ of EX1. ATREEX1: average of training errors (TRE) of EX1, ATSEEX1: average of testing errors (TSE) of EX1, SDTREEX1: SD of TRE of EX1, SDTSEEX1: SD of TSE of EX1. **c** Average of RMSE versus λ of EX2 and EX3. **d** SD of RMSE versus λ of EX2 and EX3. ATREEX2: average of TRE of EX2, ATREEX3: average of TRE of EX3, ATSEEX2: average of TSE of EX2, ATSEEX3: average of TSE of EX3, SDTREEX2: SD of TRE of EX2, SDTSEEX2: SD of TSE of EX2, SDTREEX3: SD of TRE of EX3, SDTSEEX3: SD of TSE of EX3

cussed above for the database DB1. The average and SD of training and testing errors of these ten models are shown in Fig. 2c and d, respectively. The average and SD of coefficients of the models are provided in Table 1. Important observations are as follows:

1. As expected from experiment 1, average training and testing errors increase w.r.t. λ, and variation in coefficients decreases w.r.t. λ. The SD of the training and testing errors is minimum at $\lambda = 0.01$.
2. λ-w tuple for the minimum training error (0.271 dB) is (0, 27.59, −22.21, 35.12, −39.86), and the image is 'cameraman'. (0, 28.70, −30.89, 46.74, −42.97) for the minimum testing error (0.231 dB) and image is 'lena'.
3. λ-w tuple for the maximum training error (0.759 dB) is (0.1, 34.91, −11.90, −10.09, −11.39) and (0.1, 32.38, −14.66, −7.00, −7.50) for the maximum testing error (0.908 dB).

Table 1 Average and SD of coefficients of models in experiments 1, 2, and 3

λ	Average				SD			
	w_0	w_1	w_2	w_3	w_0	w_1	w_2	w_3
Experiment 1								
0	27.56	− 21.79	33.97	− 39.16	0.05	0.64	1.89	1.44
0.02	26.44	− 8.41	− 1.44	− 13.77	0.02	0.17	0.31	0.34
0.04	26.35	− 7.32	− 4.57	− 11.36	0.02	0.16	0.17	0.25
0.06	26.33	− 6.98	− 5.72	− 10.39	0.02	0.16	0.13	0.22
0.08	26.32	− 6.84	− 6.30	− 9.83	0.02	0.15	0.10	0.19
0.1	26.31	− 6.78	− 6.65	− 9.45	0.02	0.15	0.09	0.18
Experiment 2								
0	30.78	− 28.8	40.48	− 41.97	2.83	6.03	9.76	4.51
0.02	29.43	− 13.05	− 0.64	− 12.75	2.78	3.2	2.36	1.33
0.04	29.29	− 11.58	− 4.33	− 10.15	2.77	2.85	1.78	1.59
0.06	29.22	− 11	− 5.71	− 9.19	2.77	2.66	1.57	1.65
0.08	29.18	− 10.68	− 6.43	− 8.69	2.76	2.52	1.47	1.64
0.1	29.14	− 10.46	− 6.87	− 8.39	2.76	2.41	1.41	1.62
Experiment 3								
0	30.75	− 29.37	43.33	− 44.62	0.22	1.70	4.19	2.94
0.02	30.19	− 22.52	25.20	− 31.61	0.17	1.13	2.72	1.92
0.04	29.92	− 19.21	16.43	− 25.31	0.15	0.86	2.03	1.44
0.06	29.75	− 17.25	11.25	− 21.60	0.15	0.71	1.62	1.15
0.08	29.65	− 15.96	7.84	− 19.15	0.14	0.62	1.35	0.97
0.1	29.57	− 15.04	5.42	− 17.41	0.14	0.55	1.16	0.84

4. Larger values of SD (Fig. 2d and Table 1) ensure that coefficients, training, and testing errors depend on randomness due to image.

5.1.3 Experiment 3

Third experiment is performed on the database DB2 to study the distribution of coefficients w.r.t. randomness due to data selection. In this experiment, all the datasets of the DB2 are mixed to construct a single dataset. 70% data (630 instances) is randomly selected for the training purpose, and the rest 30% (270 instances) data is used for testing. Random selection has been done ten times to construct ten ridge regression models. The average and SD of training and testing errors of ten models are shown in Fig. 2c and d, respectively. The average and SD of coefficients of the models are provided in Table 1. Important observations are as follows:

1. Average training and testing errors increase w.r.t. λ, and variation in coefficients decreases w.r.t. λ.

2. λ-w tuple for the minimum training error (1.746 dB) is (0, 30.51, −28.10, 40.40, −42.52) and (0, 30.85, −29.79, 45.19, −46.52) for the minimum testing error (1.847 dB).
3. λ-w tuple for the maximum training error (1.859 dB) is (0.1, 29.60, −14.69, 5.40, −18.04) and (0.1, 29.45, −14.81, 4.76, −16.68) for the maximum testing error (2.065 dB).
4. Small values of SD (Fig. 2d and Table 1) ensure weak relation of coefficients, training, and testing errors with randomness due to random selection of data.

5.2 PSNR Prediction and Error Analysis

Results of experiments 1, 2, and 3 have been used to select 'good' λ-w tuples according to the principle of BVDT (step 1: select λ such that range of testing errors is the minimum, step 2: select w corresponding to median testing error, where λ is obtained from step 1) or minimum testing error. According to the principle of minimum testing error, T1: (0, 27.59, −22.00, 34.42, −39.33),
T3: (0, 28.70, −30.89, 46.74, −42.97) and T5: (0, 30.85, −29.79, 45.19, −46.52) have been selected from experiments 1, 2, and 3, respectively. According to the principle of BVDT, T2: (0, 27.56, −22.21, 35.17, −39.99),
T4: (0.01, 29.68, −17.23, 7.92, −17.05) and T6: (0, 30.98, −31.36, 47.78, −47.44) have been selected from experiments 1, 2, and 3 respectively. Prediction of PSE values of DB1 by T1 and T2 is presented in Fig. 3a and b presents the prediction of PSE values of DB2 by T3, T4, T5, and T6.

PSE predictions by T1 and T2 are almost equal. The histogram of the difference between predicted PSE values by T1 and PSE values of the database DB1 (error

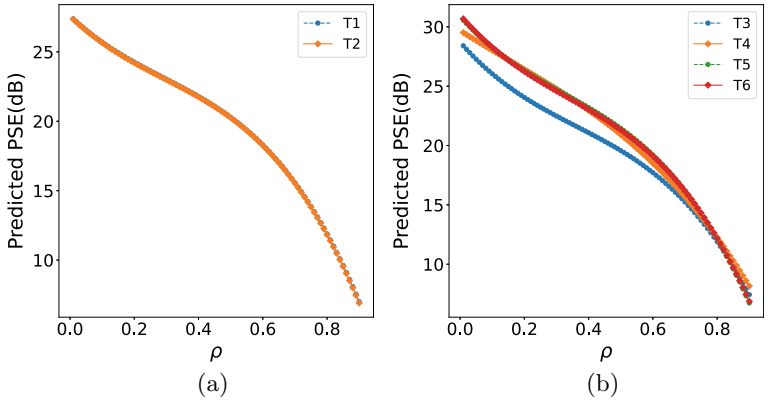


Fig. 3 Predicted PSE (Y in dB) versus noise density (ρ). **a** Experiment on database DB1. **b** Experiment on database DB2

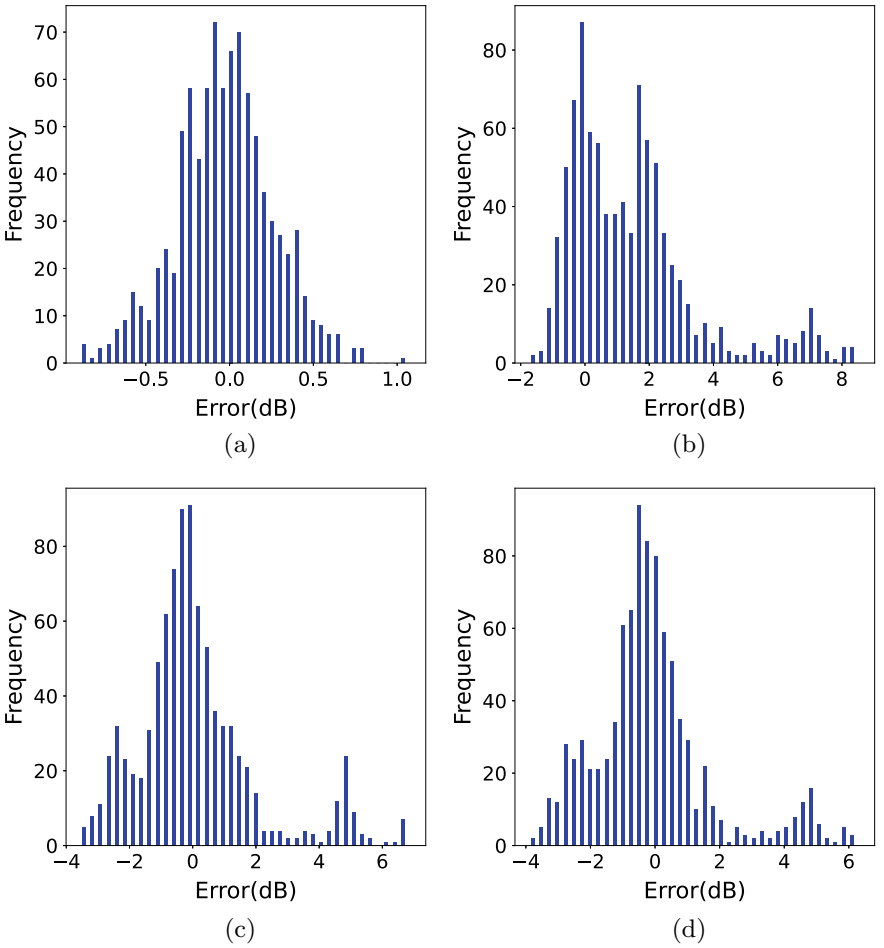


Fig. 4 Histogram of error between predicted PSE(Y) and PSNR obtained by INRA of Sect. 3.1 (PSE). **a** Model T1, **b** model T3, **c** model T4, **d** model T5

Table 2 RMSE and SD of data of error histogram

DRMSE				DSDE			
T1	T3	T4	T5	T1	T3	T4	T5
0.289	2.508	1.880	1.82	0.289	1.998	1.874	1.826

Unit is dB

histogram) is shown in Fig. 4a. The error histogram of T3, T4, T5 is shown in Fig. 4b, c, d, respectively. Table 2 provides the root mean square error (RMSE) and SD of the histogram of error data (say, DRMSE and DSDE). It is observed (Table 2, histograms Fig. 4b–d) that the tuple T5 is better than the tuples T3, T4, T6. Thus, experiment 3 is better than experiment 2.

Table 3 Comparison of computation time (unit is second)

Training time			Prediction time			Generation time	
TT1	TT3	TT5	TP1	TP3	TP5	TDB1	TDB2
1.011	0.734	2.19	0.0001	0.00008	0.0001	8.68	8.77

Each activity has been repeated five times to compute average time. TT1, TT3, TT5: average training time of models T1, T3, and T5, respectively. TP1, TP3, TP5: average prediction time by models T1, T3 and T5 respectively. TDB1, TDB2: average time to compute ninety PSE values by INRA of Sect. 3.1, $\rho \in \{0.01, 0.02, \ldots, 0.9\}$. Image is cameraman for TDB1, and an arbitrary image of DB2 is taken for TDB2

Table 4 PSNR (dB) comparison of state-of-the-art INRA

ρ	IBDNDF [9]		UWMF [13]		SAMFWMF [3]		IGOWA [2]		**RMF** [6]		
	Lena	Ca	Lena	Ca	Lena	Ca	Lena	Boat	Lena	Ca	Boat
0.1	26.88	22.69	26.98	22.88	29.496	26.592	42.86	39.68	26.47	25.68–26.13	28.10
0.3	26.38	21.82	26.68	22.40	28.89	25.22	36.75	34.36	22.40	22.82–23.17	24.22
0.5	23.31	18.75	23.64	19.33	24.41–24.55	20.15–20.44	32.73	31.02	19.36	20.03–20.68	21.44
0.7	22.84	18.19	23.18	18.67	23.81–24.00	19.41–19.67	29.54	28.2	15.41	14.90–15.83	16.23
0.9	19.1	19.81	19.94	15.65	20.09–20.28	15.84–16.01	25.32	24.12	7.96	7.5–7.81	8.29

Ca Cameraman

5.3 Study of Computation Time and PSNR Comparison

The average execution time for training of models, prediction by trained models (prediction has been done at $\rho \in \{0.01, 0.02, \ldots, 0.9\}$), and data generation have been provided in Table 3. It is observed that predictions time (TP1, TP3, TP5) are significantly lower than executions time (TDB1, TDB2) based on RMF.

According to Table 4, state-of-the-art INRA is better than RMF for salt and pepper noise. Study of executions time, training time, and prediction time for state-of-the-art INRA will be interesting future work. The PSNR comparison of RMF [6] with state-of-the-art is provided in Table 4. It is observed that INRAs better than RMF are available for salt and pepper noise. However, regression models to predict PSNR need to be developed.

6 Conclusions

Prediction of PSNR of INRA by third degree ridge regression model is the main aim. SD is used to study the relation of models coefficients w.r.t randomness due to noise, images, and data selection. SD values by experiments ensure that coefficients are

weakly related with randomness of noise and data selection while strong relation is observed with images. *Theory of BVDT and minimum testing error* is used to identify the following good ridge regression models.

1. T1: λ-w tuple is (0, 27.59, −22.00, 34.42, −39.33), DRMSE is 0.28 dB.
2. T3: λ-w tuple is (0, 28.70, −30.89, 46.74, −42.97), DRMSE is 2.50 dB.
3. T5: λ-w tuple is (0, 30.85, −29.79, 45.19, −46.52), DRMSE is 1.82 dB.

DRMSE of T1 is less than T3 and T5 as T1 considers data of one image, while T3 and T5 consider data of ten images. The training time of T1, T3, and T5 is approximately 1.0 s, 0.7 s, 2.1 s respectively, and respective training data size is 63, 63, and 630 instances. Prediction accuracy of T5 is better than T3 at the cost of training time and training data size. The prediction time of all the models for 90 PSE values is 0.0001 s, while the time to generate 90 PSE values by INRA section (3.1) is approximately eight seconds. Thus, time advantage by a factor of 10^{-5} is ensured by all the models at the cost of prediction error (measured by DRMSE).

In future, we want to develop regression models (i) with less DRMSE, (ii) for state-of-the-art INRA, for example, (IBDNDF [9], UWMF [13], SAMFWMF [3], IGOWA [2], RMF [6]), (iii) for the database with more images.

References

1. R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 4th ed. (Pearson education Inc., 2018)
2. Q. Xu, Q. Zhang, D. Hu, J. Liu, Removal of salt and pepper noise in corrupted image based on multilevel weighted graphs and IGOWA operator. Hindwai J. Math. Probl. Eng. **2018**, 1–11 (2018). https://doi.org/10.1155/2018/7975248
3. M. Mafi, H. Rajaei, M. Cabrerizo, M. Adjouadi, A robust edge detection approach in the presence of high impulse noise intensity through switching adaptive median and fixed weighted mean filtering. IEEE Trans. Image Process. **27**(11), 5475–5490 (2018)
4. M. Svensén, C.M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer, Berlin, Heidelberg, 2007)
5. https://homepages.cae.wisc.edu/~ece533/images/. Accessed 10 Sept 2021
6. T.S. Huang, G.J. Yang, *Median Filter and Their Applications to Image Processing* (School of Elec. Engineering, Purdue University, West Lafayette, IN, TR-EE, 80-1, 1980)
7. M. Mafi, H. Martin, M. Cabrerizo, J. Andrian, A. Barreto, M. Adjouadi, A comprehensive survey on impulse and Gaussian denoising filters for digital images. Signal Process. **157**(4), 236–260 (2019). https://doi.org/10.1016/j.sigpro.2018.12.006
8. Y. Ding, M. Li, T. Yan, F. Zhang, Y. Liu, R.W.H. Lau, Rain streak removal from light field images. IEEE Trans. Circ. Syst. Video Technol. **32**, 467–482 (2021). https://doi.org/10.1109/TCSVT.2021.3063853
9. I.F. Jafar, R.A. AlNa'mneh, K.A. Darabkh, Efficient improvements on the BDND filtering algorithm for the removal of high-density impulse noise. IEEE Trans. Image Process. **22**(3), 1223–1232 (2013)
10. J. Fitch, E. Coyle, N. Gallagher, Root properties and convergence rates of median filters. IEEE Trans. Acoust. Speech Signal Process. **33**(1), 230–240 (1985)
11. M. McLoughlin, G. Arce, Deterministic properties of the recursive separable median filter. IEEE Trans. Acoust. Speech Signal Process. **35**(1), 98–106 (1987)

12. K. Vasanth, T.G. Manjunath, S. Raj, A decision based unsymmetrical trimmed modified winsorized mean filter for the removal of high density salt and pepper noise in images and videos. Procedia Comput. Sci. **54**, 595–604 (2015). https://doi.org/10.1016/j.procs.2015.06.069
13. C. Kandemir, C. Kalyoncu, O. Toygar, A weighted mean filter with spatial-bias elimination for impulse noise removal. Digit. Signal Process. **46**(11), 164–174 (2015). https://doi.org/10.1016/j.dsp.2015.08.012
14. J. Ebrahimnejad, A. Naghsh, Adaptive removal of high-density salt-and-pepper noise (ARSPN) for robust ROI detection used in watermarking of MRI images of the brain. Comput. Biol. Med. **137**(10), 104831 (2021). https://doi.org/10.1016/j.compbiomed.2021.104831
15. N. Sharma, P.J.S. Sohi, B. Garg et al., A novel multilayer decision based iterative filter for removal of salt and pepper noise. Multimed. Tools Appl. **80**(5), 26531–26545 (2021). https://doi.org/10.1007/s11042-021-10958-1
16. A. MacAndrew, *An Introduction to Digital Image Processing with Matlab*. Notes for SCM2511 Image Processing, vol. 1 (Brooks/Cole, School of Computer Science and Mathematics, Victoria University of Technology, Pacific Grove, CA, 2004)
17. A.E. Hoerl, R.W. Kennard, Ridge regression: applications to nonorthogonal problems. Technometrics **12**(1), 69–82 (1970)
18. A.E. Hoerl, R.W. Kennard, Ridge regression: biased estimation for nonorthogonal problems. Technometrics **12**(1), 55–67 (1970)
19. J. Friedman, T. Hastie, R. Tibshirani, *The Elements of Statistical Learning*, 2nd ed. Springer Series in Statistics (2001)

Load Balancing Algorithms in Cloud Computing

K. Krishna Sowjanya and S. K. Mouleeswaran

Abstract The evolving nature of cloud computing in recent years has increased the demand for accessing resources online. This has resulted in a need to manage more traffic and a greater number of requests for cloud servers. This continuing growth of requests necessitates a more effective load balancer which distributes the workload across various instances. It reduces the risk of experiencing performance issues, the response time of a task can be minimized, and the resource utilization can also be improved. Many cloud-based companies are using both static and dynamic load balancing algorithms. This paper critically reviews various types of load balancing algorithms based on performance metrics.

Keywords Cloud computing · Load balancer · Load balancing algorithms · Virtual machines · QoS

1 Introduction

Cloud computing is a way of utilizing resources over the Internet instead of a computer's hard drive. To avoid the overhead of maintaining the infrastructure, many companies are moving to cloud service providers. These resources are provided by cloud servers/data centers which are located all over the world. With the exponential growth of the Internet, requests to cloud servers are piling up from hundreds to millions. The impending growth of cloud usage results in generating a huge number of user requests. It generates about 20% of traffic in the data centers [1]. These requests must be addressed efficiently without any latency. Thus, there is a need for a load balancer that distributes the load equally to all the servers to avoid performance issues, maintain the Quality of Services and also improve resource utilization.

K. Krishna Sowjanya (✉) · S. K. Mouleeswaran
Dayananda Sagar University, Bangalore, Karnataka, India
e-mail: ksowjanya-cse@dsu.edu.in; krishnasowjanyak.res-soe-cse@dsu.edu.in

S. K. Mouleeswaran
e-mail: mouleeswaran-cse@dsu.edu.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_45

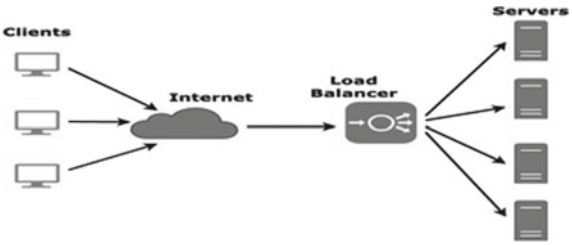


Fig. 1 Overall functionality of load balancer in cloud [6]

Load balancing refers to the fair distribution of workload across numerous computing servers. It mainly focuses on improving resource usage, maximum throughput, less response time, and avoiding overutilization/underutilization of the resources [2, 3]. A load balancer is a software that makes sure that all the servers are provided with an equal amount of load. It also improves the overall performance by decreasing the burden on servers. "Fig. 1" depicts the overall functionality of the load balancer in the cloud.

1.1 Need for Load Balancing

Load balancing has become an essential component of cloud computing. It defines how the incoming traffic can be distributed to the virtual machines at the data center by maintaining a proper balance between the user's request and the resource allocation [4]. The most important aim of load balancing is that no single node should be overwhelmed with requests [5]. Distributing the load equally among the virtual machines is the main concern of the load balancer. It reduces the possibility of overprovisioning and under-provisioning of the resources, thereby handling many requests concurrently. The incoming load is transferred dynamically to various resources using the load balancer and is processed to meet the satisfaction of the user. The load balancer also helps in managing a sudden spike of the requests from the user. Doing so, the throughput will be more and response time will be less which enhances the overall performance of the cloud and also meets various QoS metrics given in Table 1.

2 Literature Survey

In [2], the authors have presented different metrics that affect load balancing. Along with the metrics, they have discussed necessary policies that facilitate the dynamic transfer of the load from one server to another. In addition to this, the authors pointed out the metrics that need to be focused on by load balancing algorithms in the future. Similar work has been carried out by Sidhu and Kinger [3]. A comparative analysis of load balancing algorithms is conducted, where the authors divided the algorithms

Table 1 QoS metrics

Metric	Description
Scalability	The ability to add/delete the resources based on the incoming requests
Resource utilization	It is the percentage of devices/resources used from the set of resources provided. The more value of resource utilization the more efficiency of the cloud
Performance	The overall time taken for a request to be processed. If the time taken for execution is more, then the performance is less and vice versa
Response time	The time taken for giving a response to a request
Overhead	The number of requests that a single virtual machine is handling at the given point of time
Migration time	Time taken to migrate the virtual machines for the ongoing request
Throughput	The services delivered to the client within the specified time
Energy consumption	The energy consumed to process a request

into two types: static and dynamic load balancing algorithms. To emphasize the importance of load balancing, Raghava and Singh [4] performed a comparative study on various load balancing algorithms. In this study, the authors have compared a few algorithms based on QoS metrics.

3 Load Balancing Algorithms

The load balancing algorithms are prominently categorized into two types: static algorithms and dynamic algorithms.

3.1 Static Load Balancing Algorithms

Static load balancing algorithms follow the procedure of allocating the required resources before the execution of the requests. This requires ahead information on the prerequisites. It is not adaptable for any dynamic change in the requests at that particular time. Various static load balancing algorithms are discussed as follows:

Shortest Job First (SJF): SJF algorithm executes the job which is having the shortest time. Mondal et al. [10] have given a basic introduction about the shortest job first algorithm and also explained the working of the algorithm in detail. This algorithm reduced the waiting time and turn-around time. Seth and Singh [7], minimized the CPU time, and overall execution time by considering dynamic requests along with the SJF algorithm. The proposed algorithm is implemented in both homogenous and heterogeneous clouds. However, the resource utilization is less. Ru and Keung [8] grouped the incoming requests based on the file size and prioritized based on the

Table 2 Analysis of various SJF load balancing algorithms

S. No.	Technique	Pros	Cons
1	[7]	Minimizes the CPU time and overall execution time	Resource utilization is less
2	[8]	Waiting time is lessened and makespan time is reduced Resource utilization is maximized	Not suitable for heterogeneous resources
3	[9]	Enhanced resource utilization Avoids the starvation time	Overall performance is degraded if the number of requests is more
4	[10]	Waiting time and turn-around time are reduced	Could not predict the burst time of a process results in the overall degradation of the performance

bandwidth requirement, and applies SJF to select a request from the pool of the requests. The task waiting time and makespan time are considerably reduced, and the resource utilization is also maximized. Singh et al. [9] implemented the SJF load balancing algorithm in the private cloud to enhance resource utilization. To avoid starvation, the authors assigned each task a time of arrival such that if the arrival time is more, it reduces the time based on the incoming requests and also prioritized the users based on system-level having more priority and user-level having less priority. The resource utilization and the processing time are improved only for a smaller number of requests. Processing time and waiting time are increased by the increase in the requests. The overall analysis of various variants of the SJF algorithm is given in Table 2.

Round-Robin Load Balancing Algorithms: Round-Robin is an easy and old algorithm that reduces the starvation time of a process by providing a time slice and gives every process in the request queue a chance to execute in a cyclic manner. Balharith and Alhaidari [11] provided a detailed literature survey on various types of Round-Robin load balancing algorithms and categorized them based on quantum time, i.e., either static quantum or dynamic quantum. Sangwan et al. [12] proposed an improved Round-Robin algorithm by providing the quantum that changes with time and the incoming requests. That change is predicted or calculated based on the mean of the burst time of all the incoming requests. There is a considerable amount of reduction in waiting time and turn-around time. However, other factors such as resource utilization, the performance of the system have not been addressed. Fataniya and Patel [13] tried to lessen the average waiting time and turn-around time by choosing the dynamic time slice. The slice is calculated by taking the mean and median of the burst time of all the incoming requests and taking an average of it. This method was unable to address the issues like resource utilization and overall performance. For few processes having less burst time, the waiting time will be more because of the time slice. However, Rajput and Gupta [14] proposed an advanced algorithm that assigns priority for a process in addition to the time quantum. If a process is having less burst time, then it is given the highest priority, thereby reducing waiting time.

Table 3 Analysis of various Round-Robin load balancing algorithms

S. No.	Technique	Pros	Cons
1	[12]	Reduced waiting time and turn-around time	Resource utilization and overall performance have not been addressed
2	[13]	Reduced waiting time and turn-around time	Resource utilization and overall performance have not been addressed
3	[14]	Reduced waiting and turn-around time	–

The overall analysis of various variants of the Round-Robin algorithm is given in Table 3.

Min–Min and Max–Min Load Balancing Algorithms: Min–Min algorithm chooses a task that takes minimum time to complete the process and based on that the resources are allocated to it. Max–Min algorithm chooses a task with a maximum time of completion and assigns it to the relevant resources. These algorithms focus more on execution time but not on resource utilization. To avoid that, Chen et al. [15] combined two algorithms that solely focus on the load of the resource. Based on the load, it will allocate the tasks to the resource. Another algorithm focused on the priority of the tasks. So that higher priority tasks can have less waiting time. However, both the algorithms cannot address the performance when it comes to the heterogeneity of the resources and also other QoS parameters. In addition to this algorithm, Suntharam [16] modified the existing algorithm to balance the load in the cloud equally based on Max–Min algorithm. Here, before the starting up of the load balancer, all the resources like VMs and storage give the status like available storage space and executing time. So, when the requests come, the average mean-time of all the requests is calculated and related resources are provided to the task to avoid over-usage and under usage of the resources. This algorithm is not feasible for heterogeneous resources and more requests. To balance the load, the task having the maximum execution time is allocated to the resources with minimum completion time. In this case, the load on a resource is not checked. To improve that [17], the resource that takes minimum time to complete is assigned only with the task that takes maximum time to complete. This algorithm only focused on the resource allocation but not on the quality metrics like waiting time, turn-around time, etc. The overall analysis of various variants of Min-Min and Max–Min load balancing algorithms is given in Table 4.

Table 4 Analysis of various Min–Min and Max–Min load balancing algorithms

S. No.	Technique	Pros	Cons
1	[15]	Focused mainly on load distributing and in the reduction of waiting time	Overall performance of heterogeneous resources and tasks has not been addressed
2	[16]	Focused on load distribution mainly	Not focused on waiting time, performance, and various metrics
3	[17]	Focused on resource allocation	Not focused on waiting time, performance, and various metrics

3.2 Dynamic Load Balancing Algorithms

Dynamic load balancing algorithms distribute the resources to the requests at the runtime. The distribution is done by checking the status (overloaded/underloaded) of the resources. The request in real time will be dynamic and will be changing from time to time. So, the dynamic load balancing algorithms are most commonly used. Various dynamic load balancing algorithms are discussed as follows:

Honey Bee Foraging Load Balancing Algorithms: This algorithm is based on the natural foraging behavior of the honey bees to collect the honey and notify the other scout bees about the availability of the resource. Hashem et al. [18] attempted to reduce the execution time, response time, and makespan time by implementing the honey bee algorithm for load balancing. The results have shown a considerable amount of improvement in the performance when compared to the other algorithms like ABC, Throttled, and Round-Robin algorithms. However, the overall performance is degraded because of the overhead of the migration of VM during the task allocation. Performance plays an important role to evaluate an algorithm as the mentioned above algorithm cannot improve the performance level, and Dinesh Babu and Krishna [19] tried to improve the overall performance by taking into priority as a consideration along with the honey bee foraging method. This method applies to heterogeneous environments also. But the cost and energy consumption is more. Similar to the above approach, Gupta and Sahu [20] proposed a honey bee algorithm to lessen the turn-around time, makespan time, and waiting time. But the overall cost and energy consumption are more. The overall analysis of various variants of honey bee load balancing algorithms is given in Table 5.

Ant Colony Optimization Load Balancing Algorithms: This algorithm is based on the natural behavior of ants for searching food. To search the food, the ant searches randomly and places markers through the path, if it finds the food. The same path will be followed by other ants which also leave the markers if the resources are plenty. Stronger the marker and stronger the resources. The same concept has been implemented by Mishra and Jaiswal [21], to balance the load by allocating the VMs based on the availability. This procedure reduced the average waiting time and response

Table 5 Analysis of various honey bee foraging load balancing algorithms

S. No.	Technique	Pros	Cons
1	[18]	Execution time, response time, and makespan time are reduced	Overall performance is degraded because of the overhead of the VM migration
2	[19]	Overall performance is increased by considering the priority of the tasks	Energy consumption is more
3	[20]	Throughput, response time, and makespan time are reduced	Cost and energy consumption are more

time but it is not fault-tolerant. Similar to this algorithm [22], proposed a hybrid algorithm to increases resource allocation. They tried to achieve that with ant generation of forward ant and backward ant. If the resources are full the ant will go backward and search for new resources. If the resources are not full, then a forward ant is generated and is forwarded to the resources. This algorithm improved resource allocation, but the time taken to search the resources is more which degraded the performance and efficiency. To reduce the searching time, Xu et al. [23] included the combination of VMs and PMs (physical machines). The requests are directed to the VMs using the ant colony optimization and that VM, in turn, is connected to the PMs to increase the overall performance. However, this algorithm has not addressed the cost parameter. The overall analysis of various variants of ant colony optimization load balancing algorithms is given in Table 6.

Throttled Load Balancing Algorithms: This algorithm maintains a table that maintains the availability of VMs. Based on the requests and the data in the availability table, the resources are allocated to the user requests. The main disadvantage of this method is, for every new request, it starts to search the availability status from the beginning and only one request can be given to one VM. To overcome these disadvantages, Ahmad et al. [24] enhanced the traditional Throttled load balancing algorithm by assigning more than one task to VMs by maintaining a queue and priority and searching is done from the last recent searched VM. This method maintains a single node to store the table and is not fault-tolerant. Similarly, Phi et al. [25] maintained two index tables, one for availability status and one for busy status. If a

Table 6 Analysis of various ant colony optimization load balancing algorithms

S. No.	Technique	Pros	Cons
1	[21]	Response time and waiting time are reduced, and resource allocation is increased	Not fault-tolerant
2	[22]	Focused more on resource allocation	Overall performance and efficiency are degraded as it is taking more time to search the resources
3	[23]	Overall performance is increased	Cost is more

Table 7 Analysis of various throttled load balancing algorithms

S. No.	Technique	Pros	Cons
1	[24]	Reduced response time and waiting time	Has a single point of failure
2	[25]	Enhanced resource allocation	Performance degradation with a greater number of requests
3	[26]	Load is equally distributed to the VMs	Overall performance is degraded with a greater number of requests

request comes, the VMs availability table is checked and its status is changed to busy using the modified algorithm. But the time taken for the response and time taken for processing is more if the number of requests is more. To reduce the response time and waiting time, Bagwaiya and Raghuwanshi [26] used a hybrid approach. First, the availability of VMs is checked in the index table and the load is equally distributed using the equal spread algorithm. But the overall performance is degraded when the requests are more. The overall analysis of various variants of Throttled load balancing algorithms is given in Table 7.

Table 8 Various papers on load balancing algorithms have been analyzed according to QoS parameters mentioned in Table 1, and the comparative analysis of those algorithms is given in Table 8.

4 Conclusion

Load balancing ensures the equal distribution of the resources to the user's request and also enhances the functionality of the cloud to an optimal level. Various load balancing algorithms have been analyzed and based on those algorithms, the pros and cons are represented. In addition to that, QoS metrics addressed by these algorithms are also analyzed comparatively. Static algorithms need a heads-up of all the required resources before starting processing the requests which is not much feasible in the present scenario due to the explosive growth of the Internet. To balance the rising request rates with proper processing time, dynamic algorithms are more suitable. Both static and dynamic algorithms focus on various performance metrics depending on the methodology. Few algorithms like ant colony, Min–Min, and SJF solely focused on the resource allocation which affected the overall performance, and few algorithms like honey bee and Round-Robin focused on the overall performance which will be expensive and consumes a lot of energy. The above-mentioned algorithms focus only on a part of the metrics or partial metrics, but not the complete set of metrics. So, keeping in context of dynamic requests from the users, there is a need for an efficient dynamic load balancing algorithm that addresses the maximum metrics which in turn involves enhancing the overall performance of the cloud as well as user satisfaction in terms of cost and Quality of Service.

Table 8 Comparative analysis of various load balancing algorithms

Technique	Resource utilization	Performance	Response time	Waiting time	Cost	Migration time	Throughput	Energy consumption
[7]	No	Yes	Yes	Yes	No	No	Yes	No
[8]	Yes	Yes	Yes	Yes	No	No	Yes	No
[9]	Yes	No	No	Yes	No	No	No	No
[10]	No	No	Yes	Yes	No	No	Yes	No
[12]	No	No	Yes	Yes	No	No	Yes	No
[13]	No	No	Yes	Yes	No	No	No	No
[14]	No	No	Yes	Yes	No	No	No	No
[15]	Yes	No	Yes	Yes	No	No	Yes	No
[16]	Yes	No	No	No	No	No	No	No
[17]	Yes	No	No	No	No	No	No	No
[18]	Yes	No	Yes	Yes	No	No	Yes	No
[19]	Yes	Yes	Yes	Yes	No	Yes	Yes	No
[20]	Yes	Yes	Yes	Yes	No	Yes	Yes	No
[21]	Yes	Yes	Yes	Yes	No	No	Yes	No
[22]	Yes	No	No	Yes	No	No	No	No
[23]	Yes	Yes	Yes	Yes	No	No	Yes	No
[24]	Yes	Yes	Yes	Yes	No	Yes	Yes	No
[25]	Yes	No	Yes	No	Yes	Yes	No	No
[26]	Yes	No	Yes	No	No	Yes	Yes	No

References

1. Global cloud index projects cloud traffic to represent 95 percent of total data center traffic by 2021. CISCO (2018, February 05). https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=1908858
2. E.J. Ghomi, A.M. Rahmani, N.N. Qader, Load-balancing algorithms in cloud computing: a survey. J. Netw. Comput. Appl. **88**, 50–71 (2017)
3. A.K. Sidhu, S. Kinger, Analysis of load balancing techniques in cloud computing. Int. J. Comput. Technol. **4**(2), 737–741 (2013)
4. N.S. Raghava, D. Singh, Comparative study on load balancing techniques in cloud computing. Open J. Mob. Comput. Cloud Comput. **1**(1), 18–25 (2014)
5. N.J. Kansal, I. Chana, Cloud load balancing techniques: a step towards green computing. IJCSI Int. J. Comput. Sci. Issues **9**(1), 238–246 (2012)
6. V. (2020, June 12). Load balancers, an analogy—codeburst. Medium. https://codeburst.io/load-balancers-an-analogy-cc64d9430db0
7. S. Seth, N. Singh, Dynamic heterogeneous shortest job first (DHSJF): a task scheduling approach for heterogeneous cloud computing systems. Int. J. Inf. Technol. **11**(4), 653–657 (2019)
8. J. Ru, J. Keung, An empirical investigation on the simulation of priority and shortest-job-first scheduling for cloud-based software systems, in *2013 22nd Australian Software Engineering Conference* (IEEE, 2013), pp. 78–87
9. A.K. Singh, S. Sahu, K.K. Gautam, M.N. Tiwari, Private cloud scheduling with SJF, bound waiting, priority and load balancing. Int. J. **4**(1) (2014)
10. R.K. Mondal, E. Nandi, D. Sarddar, Load balancing scheduling with shortest load first. Int. J. Grid Distrib. Comput. **8**(4), 171–178 (2015)
11. T. Balharith, F. Alhaidari, Round robin scheduling algorithm in CPU and cloud computing: a review, in *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, May (IEEE, 2019), pp. 1–7
12. P. Sangwan, M. Sharma, A. Kumar, Improved round robin scheduling in cloud computing. Adv. Comput. Sci. Technol. **10**(4), 639–644 (2017)
13. B. Fataniya, M. Patel, Dynamic time quantum approach to improve round robin scheduling algorithm in cloud environment. IJSRSET **4**(4), 963–969 (2018)
14. I.S. Rajput, D. Gupta, A priority based round robin CPU scheduling algorithm for real time systems. Int. J. Innov. Eng. Technol. **1**(3), 1–11 (2012)
15. H. Chen, F. Wang, N. Helian, G. Akanmu, User-priority guided Min–Min scheduling algorithm for load balancing in cloud computing, in *2013 National Conference on Parallel Computing Technologies (PARCOMPTECH)* (IEEE, 2013, February), pp. 1–8
16. S.M.S. Suntharam, Load balancing by max-min algorithm in private cloud environment. Int. J. Sci. Res. (IJSR) (2013). ISSN (Online): 2319-7064. Index Copernicus Value (2013): 6.14 Impact Factor, 4, 438
17. R. Kaur, P. Luthra, Load balancing in cloud system using max min and min min algorithm. Int. J. Comput. Appl. **975**, 8887 (2014)
18. W. Hashem, H. Nashaat, R. Rizk, Honey bee based load balancing in cloud computing. KSII Trans. Internet Inf. Syst. **11**(12) (2017)
19. L.D. Dinesh Babu, P.V. Krishna, Honey bee behavior inspired load balancing of tasks in cloud computing environments. Appl. Soft Comput. **13**(5), 2292–2303 (2013)
20. H. Gupta, K. Sahu, Honey bee behavior based load balancing of tasks in cloud computing. Int. J. Sci. Res. **3**(6) (2014)
21. R. Mishra, A. Jaiswal, Ant colony optimization: a solution of load balancing in cloud. Int. J. Web Semant. Technol. **3**(2), 33 (2012)
22. P. Verma, S. Shrivastava, R.K. Pateriya, Enhancing load balancing in cloud computing by ant colony optimization method. Int. J. Comput. Eng. Res. Trends **4**(6), 277–284 (2017)

23. P. Xu, G. He, Z. Li, Z. Zhang, An efficient load balancing algorithm for virtual machine allocation based on ant colony optimization. Int. J. Distrib. Sens. Netw. **14**(12), 1550147718793799 (2018)
24. E.I. Ahmad, S. Ahmad, E.S. Mirdha, An enhanced throttled load balancing approach for cloud environment. Algorithms **4**, 5 (2017)
25. N.X. Phi, C.T. Tin, L.N.K. Thu, T.C. Hung, Proposed load balancing algorithm to reduce response time and processing time on cloud computing. Int. J. Comput. Netw. Commun. **10**(3), 87–98 (2018)
26. V. Bagwaiya, S.K. Raghuwanshi, Hybrid approach using throttled and ESCE load balancing algorithms in cloud computing, in *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, March (IEEE, 2014), pp. 1–6

Using Marketing Cloud Applications for Better Analysis of Consumer Insights

Ivy Baroi and Suman De

Abstract Digitization has led to the evolution of technology over various domains and has provided better tools to understand the end consumer for any business. Marketers have leveraged technology solutions from leading vendors to better understand their customers and also predict future trends and segment their users based on various parameters. Enterprise resource planning has developed over the last decade and made a complete transition to the cloud. Currently, there exists a huge list of software providers from big organizations to startups who provide solutions to companies to achieve marketing goals. It is a major challenge to understand the technology side of the offerings and better evaluate the needs that can be solved by a Marketing Cloud application. In this paper, we explore Marketing Cloud applications and the features that can be used by a marketer. We discuss various benefits and analytical insights that a marketer will be able to consume for better understanding a customer. In this paper, we also look at adoption strategies of such solutions and how it helps in overall marketing of a product.

Keywords Marketing cloud · Cloud · ERP · Analytics · Consumer insights

1 Introduction

Technology has evolved with the introduction of cloud and analytics imbibed with the flavor of machine learning and artificial intelligence [1]. The introduction of new applications and solutions has seen a major upgradation in terms of their deployment strategies from being on-premises to the cloud where the customer is not responsible and not subjected to their infrastructure. Cloud solutions provide the option for a subscription-based pay-per-use model that is more attractive to

I. Baroi
Symbiosis Institute of Business Management, Nagpur, India

S. De (✉)
SAP Labs India Pvt. Ltd., Bangalore, India
e-mail: suman.de@sap.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_46

customers. Marketers have also leveraged these new options to better understand their consumers. This paper talks about the usage of tools through a solution offering of Marketing Cloud by various software vendors and how the adoption of the same helps the marketers. Marketing Cloud is a Customer Relationship Management platform that helps marketers create and manage excellent marketing relationships with their customers. It is a unique platform that bestows detailed customer information in real-time. It gives a relevant and personalized approach across all channels and devices throughout the customer journey and also for omnichannel customer communication. It includes fused solutions for customer journey management, email, mobile, social, web personalization, advertising, content creation and management and data analysis [2]. Every company, big and small, that identified the worth of developing and nurturing authentic relationships to make their customer successful, bring revenue, and increase their ROI is using the Marketing Cloud platform [3]. Companies like Nestle, Cisco, Amazon, Audi, Philips, Dior, Vodafone, and countless companies use this platform to bestow their customers the most pleasing experience with their products and increase brand loyalty. Companies can understand their customers and provide a personalized experience by having a 360-degree view of every customer with the help of Marketing Cloud.

Industry 4.0 automation solution's primary facilitator is in Cloud. This paper studies the basic concept of Marketing Cloud, which has become a boon to marketers in today's world [4]. Cloud computing has been playing a vital role in the field of information technology in recent years. This next study focuses on the factors that increase or decrease the probability of adopting Marketing Cloud by SMEs and proposes a three-dimensional model that describes the crucial factors affecting SME decisions relating to the adoption of the Marketing Cloud [5]. Finally, we look at the works of Uroš Arsenijevic and Marija Jovic for their paper titled, 'Artificial Intelligence Marketing: Chatbots.' In total, 60 survey respondents were polled on their behaviors, habits, and expectations when employing multiple communication channels, with a focus on chatbots and their advantages and drawbacks in contrast to other communication channels. The findings revealed that the greatest benefit of employing chatbots in marketing services was when offering simple, quick information, but they also revealed respondents' worry of chatbots giving them incorrect information [6].

2 Marketing Cloud and Source Tools or Channels

Marketing Cloud gives a single location for all cross-channel content, allowing brands to have their single voice and provide a consistent experience in real-time. It is a channel to connect marketing, sales, and service with the following features [7–12].

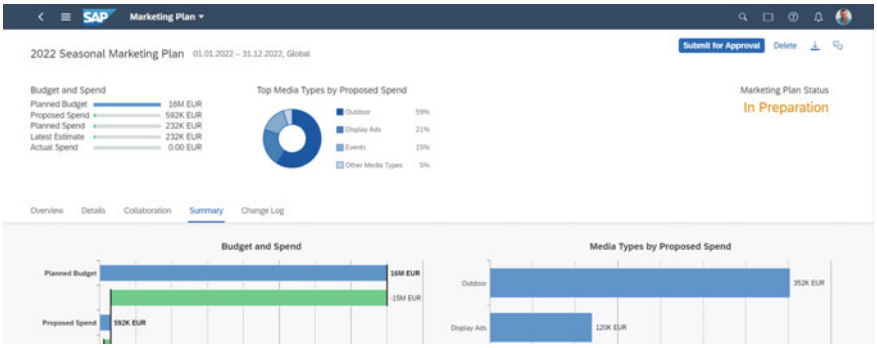


Fig. 1 Marketing Cloud solution providing insight on seasonal marketing plans. *Source* https://www.g2.com/products/sap-marketing-cloud

2.1 Content Management Tool

It consists of tools to manage a company's assets and support workflows, approvals, even versioning, assign tags, campaign associations, and owners to the content and then use the quick search functionality to track it easily.

2.2 Customer Journeys Through Emails

Email marketing is a marketing strategy in which email is used to promote a company's products to potential customers with dynamic content blocks for personalization, incorporated and automated prescient choices, and predefined occasion triggers.

2.3 Managing Mobile Marketing Strategy

Ecommerce email marketing is a marketing strategy with the Marketing Cloud's mobile marketing features; a marketer can drive instant customer engagement (Fig. 1).

2.4 Social Media Engagement Tool—Social Studio

The social media marketing tools are a user-friendly single interface that allows a company to operate its marketing strategies across all social media channels.

2.5 Lead Generation and Account Management

Marketing Cloud solutions offer support to a marketer and make it easy to contact, track and process necessary action items to manage leads and specific accounts.

2.6 Market Segmentation

It facilitates to better understand their customers by preparing reports and forming clusters of end-users with the right segments marked with correct distinctions.

2.7 Recommendation Service

It provides recommendations that help marketers in acting toward a change and take actions according to the suggestions received from the recommendation services.

2.8 Providing Predictive Intelligence

Predictive intelligence is a process that starts with the collection of consumer data and potential consumer behavior data to get insights on the consumer characteristics.

2.9 Campaign Management

It provides a seamless journey of managing campaigns specific to a product, brand, organization, etc., and acts as a common platform to promote an object.

2.10 Personalized Experience on Websites

Marketing Cloud consists of web marketing tools that help marketers easily create and measure the effectiveness of dynamic web pages and can develop and deploy consistent branded content over landing pages, microsites, Facebook tabs, etc.

3 Industry Solutions and Current Scenario

Marketing Cloud is highly used by business-to-business companies, business-to-consumer companies, and even nonprofit organizations, that identified the worth of developing and nurturing authentic relationships to make their customer successful, bring revenue, and increase their ROI are using the Marketing Cloud platform. It is always linked with all major social networking sights like Facebook, Twitter, Linked In, YouTube, Pinterest, and Google+. This solution has access to the Pinterest Business Insights API, allowing marketers to get data better to understand their company's brand value on Pinterest. It will enable marketers to easily manage various campaigns, target the best audiences with LinkedIn's unique demographic data, schedule updates, and track and optimize promotion campaigns. The customer journey can be guided by a marketer with the help of cross-channel marketing strategies, which compiles digital and real-world experiences with the company. In Fig. 2, we observe an integration scenario for a such solutions with other relevant modules in the marketing scenario.

During every stage of the customer lifecycle, the Marketing Cloud creates a visual plan and maps out the company's interactions with its various customers individually. This software plans and maps out the interactions during any specific moment or situation and also does this throughout the customer lifecycle. This helps the company create coordination between every series of events and interactions between every customer's behavior and the company's objective. It enables the company to use automation to build customer journeys throughout the customer lifecycle across email, mobile phone, advertisement, website, and IoT to provide a flawless experience through marketing, sales, and service. It keeps customer data in sync with marketing, sales, and service interactions to provide a seamless experience to the customer.

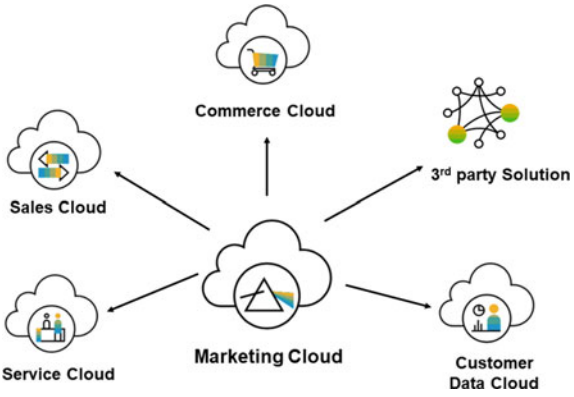


Fig. 2 Integration of Marketing Cloud with other modules [13]

4 Proposed Feature Upgradations

Marketing Cloud vendors are constantly upgrading their feature list and helping marketers to better gauge their customers and identify behavior patterns. We propose the features of Neuromarketing that involves multiple ways to get a clear understanding of an end-user. Considering the amount of money required to implement and the complexity of managing devices for fMRI, EEG, etc., we mention the most common yet effective measure in Neuromarketing that can be implemented, eye-tracking [14]. It allows a marketer to see from the perspective of the consumer and how they engage with the product. It gives unbiased insights into how consumers truly interact with advertisements, Web sites, interfaces, or products across digital platforms and the real world. It reveals deep insights into what captures the attention of the consumer and drives decision-making in the purchase process. Some tools can use this method and give consumer insights as companies are using such tools to drive better consumer behavior. If this feature is added to existing Marketing Cloud solutions, companies will get a complete package of tools in one platform that would give better consumer understanding. The amalgamation of Neuromarketing and Marketing Cloud would lead to a complete solution for marketers to get detailed consumer insights.

5 Advantages and Limitations

Marketing Cloud conveys data and application functionality through the web, instead of having them within a user's laptop/desktop or mobile device, hardware drive, or in any system known as on-premises solutions. Any user with web access can log on to the application to immediately access, share, and collaborate on relevant data and business processes. The integration of the proposed methodology of Neuromarketing with the existing features of a Marketing Cloud provides a strong suit for marketers to perform tasks with ease like lead management, content and campaign management, market prediction, segmentation, get recommendations, and so on. Communicating with the customer at the right time with the right message helps the company make a more robust and reliable relationship with its customer. When a company can fill the gap between itself and its customers, it becomes a 'customer company,' and Marketing Cloud helps a company become a 'customer company.' There are very few limitations of using these applications but the focus on the human element is necessary to carry out tasks using these tools. The usage of these applications may lead to limited attention to detail, and a marketer needs to focus on performing tasks with empathy and not completely depend on the recommendations. Machine learning algorithms and implementations of chatbots may make the processes smarter, but the right training of data with marketing heuristics will only make the systems respond accurately.

6 Conclusion

Consumers are transforming and evolving with time and require better tools to understand them. Marketing Cloud applications are providing various advantages to a marketer in identifying gray areas that can be leveraged to better understand an end-user. Additionally, it is important to have the right report and analytics available to gauge what the customer base is looking for and to address them with the right product or feature. Campaign management in a larger journey for a customer is of prime importance and overall helps the product gain a larger focus in the customer's subconscious. Marketing cloud solutions offer a marketer to keep track of leads and their accounts, run campaigns, get predictions, achieve reports generation in real-time for their consumers, segment markets, and also get help in running the right social media campaigns. Cloud solutions are adding flavor to it just as we observe for Marketing Cloud offerings from the likes of SAP, HubSpot, Salesforce, etc., vendors.

References

1. R. Balabhadrapathruni, S. De, A study on analysing the impact of feature selection on predictive machine learning algorithms, in *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2020, pp. 10–15. https://doi.org/10.1109/PDGC50313.2020.9315801
2. S. De, D. Agarwal, A novel model of supervised clustering using sentiment and contextual analysis for fake news detection, in *2020 Third International Conference on Multimedia Processing, Communication & Information Technology (MPCIT)*, 2020, pp. 112–117. https://doi.org/10.1109/MPCIT51588.2020.9350457
3. V.A. Titova, Y.V. Tomilina, T.V. Titova, Information technology in the system of purchasing decision-making based on models of long-term and short-term marketing actions, in *2014 9th International Forum on Strategic Technology (IFOST)*, 2014, pp. 234–237. https://doi.org/10.1109/IFOST.2014.6991111
4. Z. Gavrilović, M. Maksimović, The concept of cloud marketing. New Econ./Novi Ekonomist **26**, 49–55 (2019). https://doi.org/10.7251/NOE1926049G
5. S. Fazli, H. Shirdastian, M. Laroche, Effective factors of successful cloud marketing adoption by SMEs: the case of Iran. Int. J. Bus. Environ. **7**(4), 415–434 (2015). https://doi.org/10.1504/IJBE.2015.073186
6. U. Arsenijevic, M. Jovic, Artificial intelligence marketing: chatbots, in *2019 International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI)*, 2019, pp. 19–193. https://doi.org/10.1109/IC-AIAI48757.2019.00010
7. What's new in SAP Marketing Cloud 2005? Anodius. Available at: https://www.anodius.com/en/whats-new-sap-marketing-cloud-2005. Accessed 19 Sept 2021
8. SAP Marketing Cloud. Available at: https://www.sapstore.com/solutions/40081/SAP-Marketing-Cloud---standard-edition. Accessed 20 Sept 2021
9. V.P. Semenov, E.V. Budrina, I.K. Soldatov, A.G. Budrin, A.V. Soldatova, E.A. Eniushkina,Factor analysis of the results of digital technology applications in the company's marketing activities, in *2017 XX IEEE International Conference on Soft Computing and Measurements (SCM)*, 2017, pp. 879–882. https://doi.org/10.1109/SCM.2017.7970753
10. Salesforce Marketing Cloud. Available at: https://www.salesforce.com/in/products/marketing-cloud/faq/. Accessed 20 Sept 2021

11. A. Bismo, S. Putra, Melysa, Application of digital marketing (social media and email marketing) and its impact on customer engagement in purchase intention: a case study at PT. Soltius Indonesia, in *2019 International Conference on Information Management and Technology (ICIMTech)*, 2019, pp. 109–114. https://doi.org/10.1109/ICIMTech.2019.8843763
12. Maihiro, SAP Marketing Cloud. Available at: https://www.maihiro.com/en/solutions/sap-marketing-cloud. Accessed 20 Sept 2021
13. J. Ehbauer, Extended Capabilities of SAP Marketing Cloud using Business Event Handling, SAP Community, 16th April 2020. Available: https://blogs.sap.com/2020/04/16/extended-capabilities-of-sap-marketing-cloud-using-business-event-handling/. Accessed 20 Sept 2021
14. I. Baroi, S. De, A Novel Application of neuromarketing for designing user interface mockups to enhance user experience in software development, in *2021 IEEE 2nd International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship, and Talent (TEMSMET)*, 2021, pp. 1–6. https://doi.org/10.1109/TEMSMET53515.2021.9768683

Confiscate Boisterous from Color-Based Images Using Rule-Based Technique

D. Saravanan and Shubhangi V. Urkude

Abstract This paper explains how noise removed from any motion videos using the new sorting technique. Removing an error or a corrupted pixel in a motion video is not an easy process. Identify the error indicator is not easy for the researchers. Using the sorting technique, finding the disturbance is one of the important steps, and this is achieved through a step-by-step process. Most of the existing techniques create problems because while removing the corrupted pixel, it also removes the originality of the content. The proposed technique finds the errors using a step-by-step process using a blurry ruling technique. Using this, noisy pixel is identified and removed, with a disturbance of the original. The experiments show the proposed method suitable for any type of input videos.

Keywords Motion image · Image processing · Corrupted pixel · Image noise removal · Image edge detection

1 Introduction

An image places an important role today. Most of the research and day-to-day functions are carried through image-based input. This image-based input gives better results than text-based extraction. Selecting the appropriate input image is one of the challenging tasks. Motion images are the other sources of important information transporter present technology fields like traffic control, medical image analysis, education, weather and for costing and more. The help of the Internet or any media makes the creation of this image one of the simple tasks for the researchers. Among this group, finding the quality videos is really tough for many users. Technology

D. Saravanan (✉) · S. V. Urkude
Faculty of Operations & IT, ICFAI Business School (IBS), The ICFAI Foundation for Higher Education (IFHE) (Deemed to be University u/s 3 of the UGC Act 1956), Hyderabad, India
e-mail: sa_roing@yahoo.com

S. V. Urkude
e-mail: shubhangini@ibsindia.org

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_47

allows the user to upload many videos without any proper training or proper knowledge on the domain. Due to this, an enormous number of video files are damaged due to various reasons. Images are captured through unprofessional cameras, poor lighting, poor lens focal and more. Due to these, recorded image are get damaged and the originality of the images get disturbed. This noisy image gives more trouble for many image processing application and research. For that, many sorting techniques are necessary to remove the errors from the image before it is processed. This process is carried step by step because of the existing technique which not only removes the errors, it also removes the quality and content of the images [1]. Various techniques are currently available for removing the error from the image file. One of the techniques is blurry ruling techniques. This technique works as a map technology association method. Using this distinguishing occupation, a map with original is created. These functions are compared with blurry policy such as good, bad or average. Using image, threshold values help to find the above criteria. These threshold value helps to find the corrupted pixel value based on the blurry ruling operation. Based on the ruling option, the filtering functions are working.

Many image processing techniques work with appropriate tools. Difference between the images and any noisy factors is identified through these tools only. Because human eyes are not so powerful to find these errors. For that reason, there are number of image analysis tools are available. It helps to find the difference and corrupted pixel easily [2]. The proposed technique works in two phases, first, find the affected phase followed by removal phase. These techniques works well first, find the threshold values of each frame, it helps to find the amount of degradation. Difference between the original frame threshold value and damaged image frame threshold values is compared and finds the error rate. Difference is higher than proper techniques which are used to remove or reduce the error values.

The proposed technique works for static image frames. Due to that, any video files are converted to frame sequence. Then this process is repeated to the rest of the frames in each frame image pixel threshold values are calculated. This value helps with further processing either need to be corrected or need to be modified. These pixel threshold values are stored in the computer memory for any further operations. Using blurry ruling techniques, works well for indecision information. This technique gathering of many techniques it not only analyzes but also gives the solution. The experiments show the proposed method suitable for any type of input videos. Image frame rates play an important role for image analysis, and these frame rates may vary. Example, news video frame rate vary to song video file, and song video files frames are vary to cartoon or animated video files [3]. Animated or sports video files rates are different from any other video files. Based on the video input, these frame presents and frame pixel threshold values are getting changed. Here image threshold values are calculated based on the important three pixels such as red, green and blue values present in the frame. It helps to find the difference between the frames.

2 Existing System

Removing of noisy pixel from image frames is one of the challenging tasks. Number of existing techniques are currently available and those techniques works well but it not only remove the affected pixels it also removes the unaffected pixel from the image frames because of this image quality get affected. Aswathy Mohan et al. [1] described image processing is one of the important filed today, and noise on images gives bad results. Removing noise from the image frames using two phase detection phase and filtering phase using absolute filtering technique. Patidar et al. [3] on his paper removing noise from digital image using wavelet transformation tool also using mean filter and median filter used for removal of noise. Hambal et al. [8] proposed filtering technique for noise remove. Noise can occurred in any digital images any point of time at the time of creation, transmission, or reusing. Noise removal also helps the user to identify what type of noise present in the system. Kaur et al. [9] proposed different noise removing filtering technique performance. From his experimental result, he finds bilateral filter best of removing gauss ion noise, if the image affected with mixed gauss ion noise switching bilateral filter works well.

2.1 Drawback of Existing Technique

- Image extraction required more calculation process.
- Technique based on filtering concept.
- Methods focused on boisterous only.
- Techniques are too complex.
- Extraction process required many steps.

3 Proposed System

Removing noise from the image frames is a challenging task for many researchers. In the proposed technique, blurry ruling method image frame threshold values are calculated [4]. These values are stored in memory for future processing. Image frame values are compared with existing stored value difference when more than two phases of operation are conducted. The first phase called the revealing phase, in this phase, in which part of the image frame which gets disturbed is identified. It is shown in Fig. 1. This affected value only helps us to know how much the originality gets affected. Followed by a filter technique used to remove the noisy pixels [5].

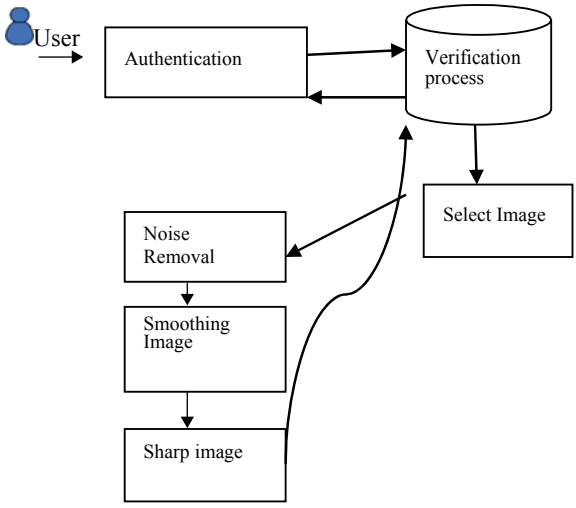


Fig. 1 Proposed architecture

3.1 Advantages in Proposed System

- It helps to identify the affected pixel rate.
- A rule-based technique, identifying ration also gives the solution.
- Two phase approach helps the work more effectively.

4 Experimental Setup

4.1 Validation

Validation is the process to check the credential of the registered user through some validation mechanism. This process also allows the new entry through a proper validation mechanism. Here the system collects the user's information, security confirmation and more [6]. Based on the input attribute, it allows the user to continue their operation after proper validation. One of the most important tasks in validation is to permit only authorized or registered users only. Unregistered users entries are excluded, and they never get any data or information's. Another advantage in this process is if any user submitted irrelevant or mislead information, the system will suspend their accounts.

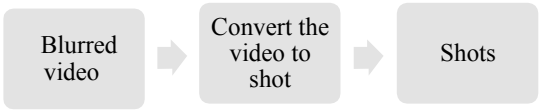


Fig. 2 Converting video process impulse noise

Fig. 3 Impulse noise process

4.2 Convert Motion Frame to Static Frame

Any noise removal can be done with the help of static frames. For that user, first convert the motion frame into static frames [7]. It depends on many image attribute's, one of the most used attributes is the time interval between the frames. The user needs to specify the time delay between each form. Based on that, the motion frames are converted into static frames. This process is shown in Fig. 2. Based on the time interval, frames are stored in the file for further process. The user can't directly implement this concept on video.

4.3 Image Post

This process helps to remove the noise. Any blurred or noise frame is placed. After the noise is removed, user will get new frame without any noise. This process shown in Fig. 3.

4.4 De Noise

Removing noise from the image frame is a challenging task. This phase helps to perform the operation more effectively. It also helps to retain the originality of the image frame. This phase removes only the affected pixel from the image frame due to this originality of the image maintained. Blurry rulings help the process more effectively.

Fig. 4 Smoothing process

4.5 Detecting Edges

This phase helps the user finds the affected image pixel in the image frames. It helps to find at which part of the image pixels are getting changed. It is calculated using intensity of image pixel.

4.6 Smoothing Image

Image smooth can be identified by value of avg pitch in the area. For every avg pitch calculation, preprocessing step of image is required. This process shown in Fig. 4. Calculating of this required first need to find region type based on this values is calculated.

4.6.1 Pseudo for Image Smoothing

```
        Flexiblesoftdesign =          new Flexible soft desing( );
        Design.picture = picture;if ( design. Show text( )
== textoutput.ok  )
        { Placeexude( Desing.exude );} }edgeExdesign form = new edgeexdesing(
);
        Desing. picture = picture;  if ( design. Show text( )
== textoutput.OK )
        {Applyexude( design.exude );}
```

4.7 Sharpening Images

Image sharpening is the process which performs reshape or fine-tuning the image. It also modify the image based on the user required with help of image processing tool.

Fig. 5 Select input image file

4.7.1 Pseudo for Image Sharpening

```
{ EdgedExForm form = innovative EdgeExForm( ); procedure. Picture = Picture;
        if ( Procedure. Display Text( ) == Testoutput
        .OK )
{       Apply exude( procedureexude );
} Bitmap newpicture = exudecover( picture );
            if (enduserbuildnewblog)
            endusernewblog( newpicture );
  } elseif ( enduser.callupshift ) {
if ( storage != null ) storage. Position ( ); storage =
picutre;}else { picture. end( ); } picture = newpicture;
```

5 Experimental Outcomes

First image input is selected, then given input is converted into frames. Converted frames are undergone the image preprocessing technique. Unwanted, repeated and noise frames are removed. After preprocessing, frames are stored separately. After the color-based image technique is applied to the rest of the frames, and desired outputs are obtained. The experimental process shown in Figs. 5, 6, 7, 8, 9, 10, 11, 12 and 13.

6 Conclusion and Future Enhancement

Proposed work based on blurry ruling technique for despoiled color motion images. Image get affected with various reasons such as bad footage, light and more.

Fig. 6 Image frame convection process

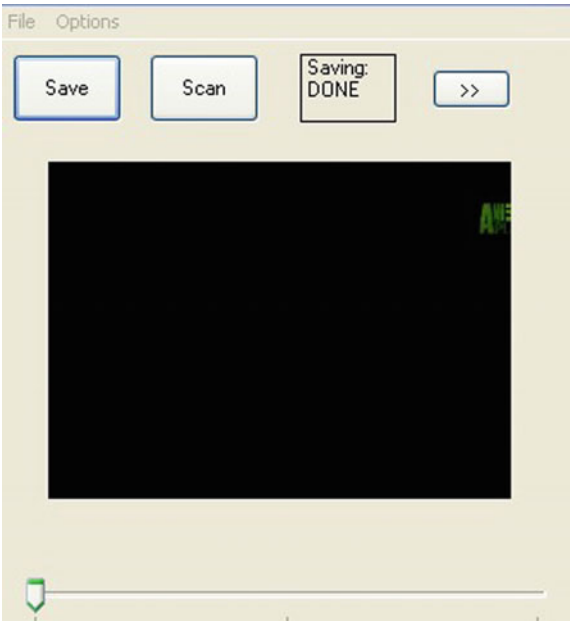


Fig. 7 Frames stored process

Removing the error from the affected frames are done step-by-step process. Technique is based on color threshold values and based on blurry ruling method, and it helps to predict the values with various parameters. This parameter values used to compare with original values.

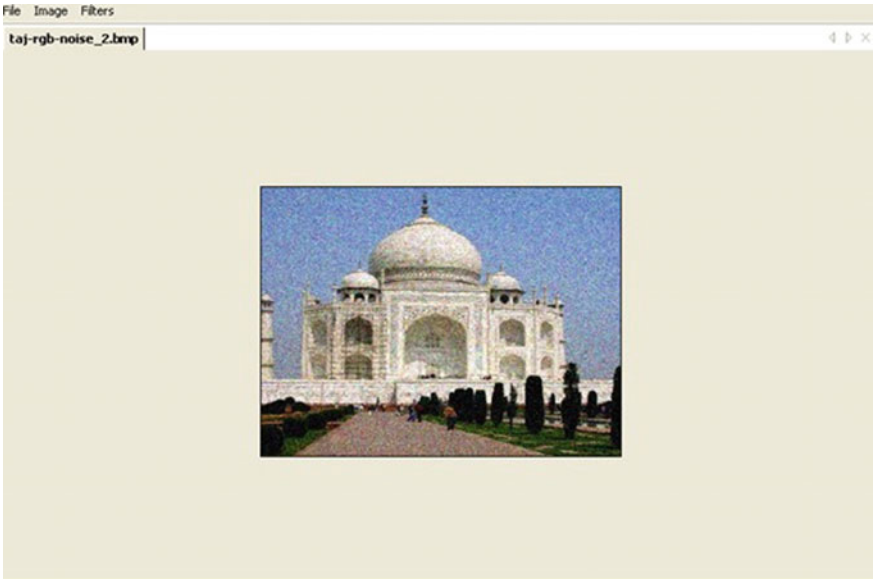


Fig. 8 Filtering impulse noise

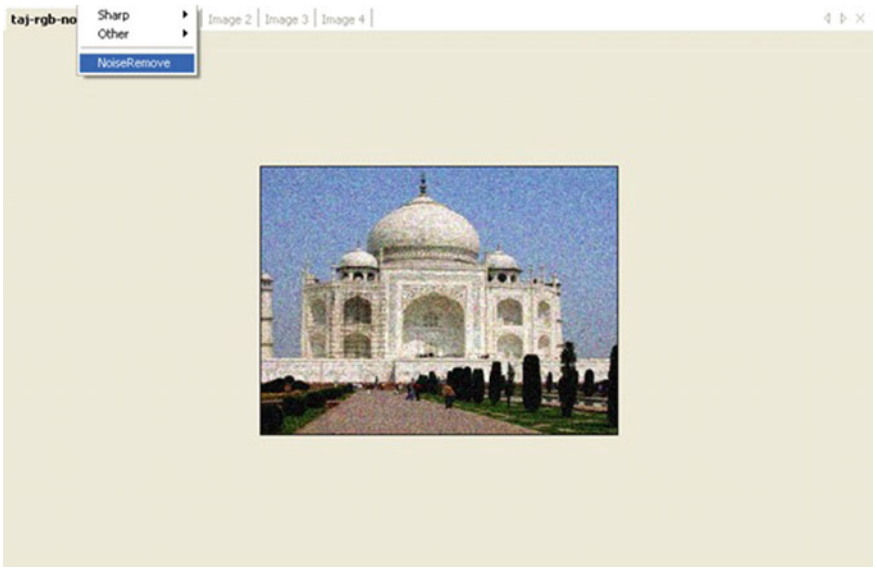


Fig. 9 Noise removal process

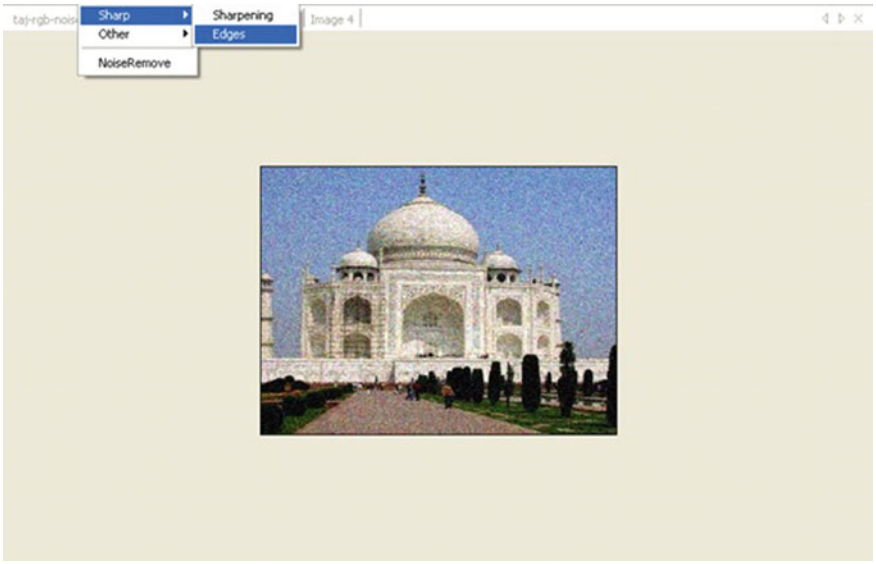


Fig. 10 Image edge detection process

Fig. 11 Output of edge detection

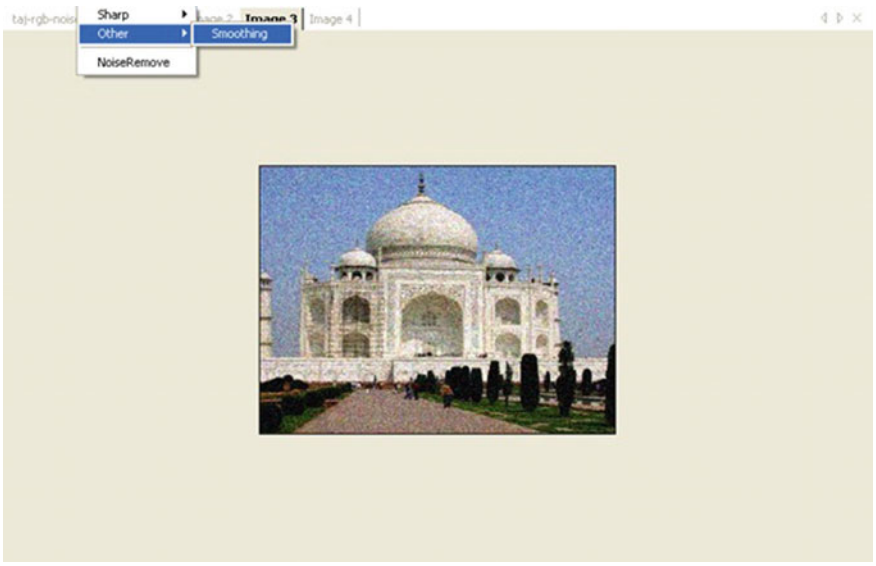


Fig. 12 Image smoothing process

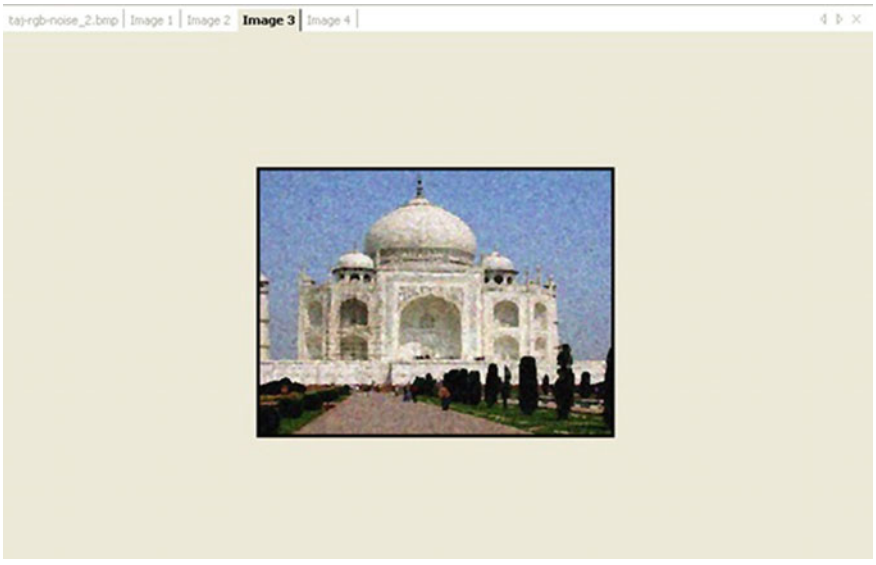


Fig. 13 Output of smoothing

References

1. T.C. Aswathy Mohan, B. Radhakrishnan, R. BijiKumar, An efficient technique for the removal of random valued impulse noise in digital images. Int. J. Eng. Res. Gen. Sci. **3**(3), 353–358 (2015)
2. D. Saravanan, Effective video data retrieval using image key frame selection. Adv. Intell. Syst. Comput. 145–155 (2017)
3. P. Patidar, S. Srivastava, Image de-noising by various filters for different noise. Int. J. Comput. Appl. **9**(4), 45–50 (2010)
4. D. Saravanan, Multimedia data retrieval data mining image pixel comparison technique. *Lecture Notes on Data Engineering and Communications Technology*, vol. 31 (2019), pp. 483–489. ISBN: 978-3-030-24642-6 (Chapter 57)
5. C.W. Ngo, T.C. Pong, H.J. Zhang, On clustering and retrieval of video shots, in *Proceeding of 9th ACM International Conference on Multimedia* (2001), pp. 51–60
6. M.R. Naphade, J.R. Smith, On the detection of semantic concepts at trecvid, in *Proceedings of the 12th Annual ACM International Conference on Multimedia (Multimedia '04)*, Oct 2004, pp. 660–667
7. D. Saravanan, Efficient video indexing and retrieval using hierarchical clustering techniques. *Advances in Intelligence Systems and Computing*, vol. 712 (2018), pp. 1–8. ISBN: 978-981-10-8227-6
8. A.M. Hambal, Z. Pei, F.L. Ishabailu, Image noise reduction and filtering techniques, Int. J. Sci. Res. (IJSR) **6**(3), 2033–2038 (2017)
9. M. Kaur, S. Gupta, Comparison of noise removal techniques using bilateral filter, Int. J. Signal Proc. Image Proc. Pattern Recognit. **9**(2) 433–444 (2016)

A Systematic Literature Review on Pulmonary Disease Detection Using Machine Learning

Biswaranjan Debata, Sudhira Kumar Mohapatra, and Rojalina Priyadarshini

Abstract Pulmonary disease is a category of respiratory tract infection which affects lungs. The objective of automatic pulmonary disease detection is to diagnose the disease with state-of-the-art machine learning techniques. This paper summarizes and analyzes the current knowledge about pulmonary disease detection in a systematic manner using more than hundred research articles. We present a detailed scientific finding, previous work done, and quantitative results from the existing literature.

Keywords Pulmonary diseases · Image processing · Deep learning · Machine learning · CNN · CXR image

1 Introduction

To identifying, evaluating, and interpreting all available research gaps about pulmonary disease detection, we conduct a systematic literature review as a review methodology. There are some reasons why we select a systematic literature review as a review methodology.

- SLR can define search strategies; this ensures the completeness of the primary studies to be assessed in our review.
- It has well-defined primary study inclusion and exclusion criteria which allow us to filter out more relevant documents for our review.

B. Debata (✉)
C.V. Raman Global University, Bhubaneswar, India
e-mail: talkbiswa92@gmail.com

S. K. Mohapatra
Faculty of Emerging Technologies, Sri Sri University, Cuttack, India
e-mail: sudhir.mohapatra@srisriuniversity.edu.in

R. Priyadarshini
Department of CSE, C.V. Raman Global University, Bhubaneswar, India
e-mail: rojalinapriyadarshini@cgu-odisha.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_48

- Since every process of the review is documented, it is easy to understand for the readers.
- It is the well-defined methodology that makes the result of the review less likely biased and reporting the report that supports our review question [1].

The rest of our literature review is organized as follows: Sect. 2 describes the need for conducting a systematic literature review, Sect. 3 describes conducting the review, Sect. 4 describes study selection criteria, Sect. 5 describes primary study selection, Sect. 6 describes result and discussion. Finally, we conclude in Sect. 7.

2 Need for Conducting a Systematic Literature Review

The need for this systematic literature review is to extract some existing information about pulmonary disease detection. We have different reasons to conduct SLR.

- to extract information about pulmonary disease,
- to know how the previous studies identify and detect pulmonary disease,
- to apply different inclusion and exclusion criteria, and
- by applying different search strategies, we can search different documents which are important for our studies.

In addition to this, we did this systematic literature review to identify gaps from the previous pulmonary disease-related studies and to design our methodologies based on the previous studies.

The following review questions are framed to achieve the above objectives:

RQ1: What type of techniques researcher can use to detect pulmonary disease?
RQ2: How other techniques can detect pulmonary disease?
RQ3: Which techniques accurately predict pulmonary disease?

3 Conducting Review

To start the review work, we start existing literature searching. Search keywords are framed, and customization of the search results is done. These things are elaborately discussed in the following subsections.

3.1 Search Keyword

The description of our systematic literature review includes the area of pulmonary disease detection. To find out the primary studies related to pulmonary disease, we

use the following keywords, pulmonary disease, machine learning, CNN, asthma, tuberculosis, chest X-ray, and COVID.

3.2 Search Result

By using the above keyword, we download 160 papers related to pulmonary disease published between 2015 and 2021. Then, we selected the paper specifically on pulmonary disease detection which is published between 2018 and 2021. The primary studies searched for this SLR are in the following database: IEEE Xplore, Google Scholar, Web of Science, Science Direct, Springer Library, ACM Digital Library, and also we use the reference found on the primary studies.

3.3 Documenting the Search Process

Table 1 presents the documentation of the selected article.

4 Study Selection Criteria

In this section, we discussed the inclusion and exclusion criteria that the main studies are related to our studies.

4.1 Inclusion Criteria

The selected papers are expected to answer our defined research questions; the studies that achieve the following criteria are included in SLR:

IC1: Studies must be published between 2018 and 2021
IC2: Studies must be published in the conference proceeding
IC3: Studies must be published in journal issues
IC4: Studies focus on pulmonary disease detection
IC5: Studies discuss pulmonary disease
IC6: Studies can answer one or more of our research questions.

4.2 Exclusion Criteria

The downloaded studies are excluded if it meets one of the following criteria.

Table 1 Documenting the search process

Title	Data source	Documentation
[2]	Conference proceeding	*Database name*: IEEE Xplore
		Search strategy: We use search keywords like lungs cancer, machine learning, and convolution neural network
		Date of search: 4.01.2021
		Year covered by search: 2018–2021
[3]	Conference proceeding	*Database name*: IEEE Xplore
		Search strategy: We use search keywords like lungs cancer prediction, machine learning, and deep learning
		Date of search: 04.01.2021
		Year covered by search: 2018–2021
[4]	Conference proceeding	*Database name*: IEEE Xplore
		Search strategy: We use search keywords like pneumonia detection, machine learning, and CNN
		Date of search: 06.01.2021
		Year covered by search: 2018–2021
[5]	Conference proceeding	*Database name*: IEEE Xplore
		Search strategy: We use search keywords like pneumonia detection, X-ray images, and deep learning
		Date of search: 11.01.2021
		Year covered by search: 2018–2021
[6]	Journal hand searches	*Database name*: IEEE Xplore
		Search strategy: We use search keywords like COVID-19 detection, X-ray images processing, deep learning, and CNN
		Date of search: 15.01.2021
[7]	Journal hand searches	*Database name*: IEEE Xplore
		Search strategy: We use search keywords like COVID-19 detection, chest images, machine learning, and CNN
		Date of search: 22.01.2021
		Year covered by search: 2018–2021
[8]	Conference proceeding	*Database name*: IEEE Xplore
		Search strategy: We use search keywords like automatic detections, COVID-19, chest X-ray (CXR) images, CNN, deep learning, and machine learning
		Date of search: 29.01.2021
		Year covered by search: 2018–2021
[9]	Conference proceeding	*Database name*: IEEE Xplore
		Search strategy: We use search keywords like computer-aided diagnosis, tuberculosis (TB), and support vector machine (SVM)

(continued)

Table 1 (continued)

Title	Data source	Documentation
		Date of search
		Year covered by search: 2018–2021
[10]	Conference proceeding	*Database name*: IEEE Xplore
		Search strategy: We use search keywords like tuberculosis detection, deep learning, and CNN
		Date of search: 08.01.2021
		Year covered by search: 2018–2021
[11]	Conference proceeding	*Database name*: IEEE Xplore
		Search strategy: We use search keywords like image segmentation, CAA, deep learning, CNN, and U-Net segmentation model
		Date of search: 28.01.2021
		Year covered by search: 2018–2021

EC1: Studies do not relate to the pulmonary disease detection
EC2: Studies are not full paper, short paper, and Ph.D. thesis
EC3: Studies not written in English.

5 Primary Study Selection

The primary study selection is presented in Table 2.

6 Result and Discussion

In this section, we explained the research questions described in the above section.

RQ1: What type of techniques researcher can use to detect pulmonary disease?

The researcher used some machine learning techniques and frameworks like convolution neural network (CNN), deep learning, supervised learning, D-CNN, histogram of oriented gradient (HOG), linear binary pattern (LBP), CovFrameNet frame work, computed tomography (CT) scans, gray level co-occurrence matrix (GLCM), support vector machine (SVM), and MR U-Net image segmentation to predict pulmonary diseases [2–11].

Table 2 Selected primary studies

Study no.	Publication year	Study focus	RQ1	RQ2	RQ3
[2]	2019	The study focuses on the classification of multi-nodule, single nodule, and healthy low-dose CT scan images	Yes	Yes	Yes
[3]	2020	The research focuses on a deep neural network system to improve the possibilities of detecting tumors	Yes	Yes	No
[4]	2021	The study focuses on a D-CNN-based method to detect pneumonia	Yes	Yes	Yes
[5]	2020	The study focuses on a system to detect pneumonia with low lung opacity sensitivity	Yes	Yes	No
[6]	2021	The study focuses on the application of image preprocessing and deep learning techniques to automate the process of speeding up diagnoses of the SARS-CoV-2 virus	Yes	No	Yes
[7]	2020	The study focuses on the deep learning method for COVID-19 diagnostics early stages	Yes	No	Yes
[8]	2020	The study focuses on finding a detection method that is not only sensitive and specific but also easily automatable that would solve many of the issues currently faced in the early stages of treatment	Yes	Yes	No
[9]	2019	The study focuses on detecting lung abnormalities and provides an opportunity for doctors to localize abnormal tissue types, both tumors and pulmonary edema	Yes	No	Yes
[10]	2021	The study focuses on diagnosing TB using CXR images obtained from a public dataset (Shenzhen dataset)	Yes	Yes	Yes
[11]	2020	The study focuses on the identification of the pulmonary tuberculosis affected region in the given CT image of the lungs	Yes	Yes	Yes

RQ2: How other techniques can detect pulmonary disease?

The researcher used some traditional and other techniques like image enhancing techniques, automatic B-line detection algorithm, LFM technique, CAD system, bi-spectral analysis method, and noninvasive detection method to predict pulmonary diseases [12–16].

RQ3: Which techniques accurately predict pulmonary disease?

The machine learning techniques [RQ1] detect the pulmonary disease and meet the accuracy rate of 99.5% [2] with F1-score 0.99, whereas the traditional techniques [RQ2] detect the pulmonary disease and meet the accuracy rate 92.5% [11].

7 Conclusion

According to our literature review, different researchers use different techniques to detect pulmonary disease and different techniques and generate different outputs; the output may depend on the input image. According to this literature review, machine learning techniques accuracy is better as compared to traditional technique. In the future, we will add further address to the findings of this systematic literature review.

References

1. B.A. Kitchenham, S. Charters, Guidelines for performing systematic literature reviews in software engineering **2** (2007)
2. S.S. Sanagala et al., A fast and light weight deep convolution neural network model for cancer disease identification in human Lung (s), in *18th IEEE International Conference on Machine Learning and Applications (ICMLA)* (IEEE, 2019)
3. S. Mukherjee, S.U. Bohra, Lung cancer disease diagnosis using machine learning approach, in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)* (IEEE, 2020)
4. S. Jamil et al., A deep convolutional neural network based framework for pneumonia detection, in *2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)* (IEEE, 2021)
5. Z.Y. Yang, Q. Zhao, A multiple deep learner approach for X-ray image-based pneumonia detection, in *2020 International Conference on Machine Learning and Cybernetics (ICMLC)* (IEEE. 2020)
6. O.N. Oyelade, A.E. Ezugwu, H. Chiroma, CovFrameNet: an enhanced deep learning framework for COVID-19 detection, IEEE Access (2021)
7. M.A. Cifci, Deep learning model for diagnosis of corona virus disease from CT images. Int. J. Sci. Eng. Res **11**(4), 273–278 (2020)
8. A.M. Fangoh, S. Selim, Using CNN-XGBoost deep networks for COVID-19 detection in chest X-ray images, in *2020 15th International Conference on Computer Engineering and Systems (ICCES)* (IEEE, 2020)
9. I. Junaedi, E. Yudaningtyas, R. Rahmadwati, Tuberculosis detection in chest X-ray images using optimized gray level co-occurrence matrix features, in *2019 International Conference on Information and Communications Technology (ICOIACT)* (IEEE, 2019)
10. B. Oltu et al., Automated tuberculosis detection using pre-trained CNN and SVM, in *2021 44th International Conference on Telecommunications and Signal Processing (TSP)* (IEEE, 2021)
11. M.O. Ramkumar, D. Jayakumar, R. Yogesh, Multi res U-net based image segmentation of pulmonary tuberculosis using CT images, in *IEEE 7th International Conference on Smart Structures and Systems ICSSS* (2020)
12. D. Abin, S.D. Thepade, S. Dhore, An empirical study of dehazing techniques for chest X-ray in early detection of pneumonia, in *2021 2nd International Conference for Emerging Technology (INCET)* (IEEE, 2021)

13. R. Moshavegh et al., Automatic detection of B-lines in $ In Vivo $ lung ultrasound. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **66**(2), 309–317 (2018)
14. Y. Xiaobo et al., Community detection in TCM network of COPD, in *2017 36th Chinese Control Conference (CCC)* (IEEE, 2017)
15. H. Porieva, K. Bahdasariants, T. Saurova, Detection of Respiratory Diseases Using Bispectral Analysis Method, in *2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO)* (IEEE, 2020)
16. X. Liu et al., Multi-view multi-scale CNNs for lung nodule type classification from CT images. Pattern Recogn. **77**, 262–275 (2018)

Cluster Head Selection Based on Type-II Fuzzy Logic System in Wireless Sensor Networks: A Review

Hetal Panchal and Sachin Gajjar

Abstract Sensor nodes (SNs) in a Wireless sensor network (WSN) sense physical characteristics such as temperature, light, humidity, and forward the data to a base station (BS). Because of the limited power supply of the SNs, WSN becomes an energy constraint network. So, to extend the network's lifetime, it is required to build energy-effectual protocols. Clustering is an important mechanism that offers an effective way to lower the consumption of energy. However, several soft computing techniques are used for cluster head (CH) selection. Due to the dynamic nature of the environment and random deployment of the SNs, the Type-I fuzzy logic system (FLS) is used that deals with uncertainties of the WSNs and provides accurate results compared to the traditional clustering mechanism. Uncertainties of Type-I FLS are improved by Type-II FLS. So, it is widely used in WSNs for cluster head selection which increases the network performance. In this paper, various clustering protocols using Type-II FLS for cluster head selection are reviewed based upon their properties like the energy of node, node density, node centrality and the distance between the nodes, historical contribution as a CH, efficiency, link quality, and moving speed of the nodes, which helps to attain maximum network lifetime, minimum energy consumption, higher packet delivery ratio, secure data transmission, maximum throughput, and scalability. Clustering protocols are described with their simulation results and compared in terms of their fuzzy input parameters, simulation tools used, merits, and demerits.

Keywords Wireless sensor networks · Type-II fuzzy logic system · Cluster head selection

H. Panchal (✉) · S. Gajjar
Institute of Technology, Nirma University, Ahmedabad, Gujarat, India
e-mail: 20ftphde47@nirmauni.ac.in

S. Gajjar
e-mail: sachin.gajjar@nirmauni.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_49

1 Introduction

Recent advancements in the sensor technology based on micro-electro-mechanical systems, less power analog, digital circuits, and less power radio frequency design have validated the process of costly and less power wireless sensor nodes that can be utilized to build up WSN [1]. In recent years, WSNs are grown drastically. WSN has been widely researched due to its use in applications in health monitoring, military applications, disaster management, and many more.

The WSN is formed of hundreds of battery-operated SNs deployed randomly in a physical environment, and a wall-powered base station (BS). The sensor nodes (SNs) collect data, process it, and forward it to the BS via direct or multi-hop communication. The BS gathers the data, analyses it, and, if necessary, takes appropriate action. WSNs are widely used due to their ease of deployment, less deployment cost, and less network complexity compared to wired networks. However, there are many challenges in WSNs. Key challenges are energy efficiency, stability, scalability, and the lifetime of the network. Because of ideal listening, retransmission due to collision, and overhearing, SNs' energy is dissipated. However, it is not feasible to change the batteries of thousands of SNs after deployment. So, designing energy-efficient, secure, scalable, and stable protocols is one of the solutions to deal with these challenges.

Forming clusters and selecting appropriate cluster head is efficient solutions to deal with these challenges. Various traditional clustering algorithms and soft computing techniques are used for the same. In this paper, Type-II FLS is described for CH selection with its CH selection properties that raise the performance of the network by dealing with its input environmental uncertainties.

The remaining paper is described as follows. Section 2 presents the idea of clustering in WSNs. Section 3 presents the properties of a cluster depending on which selection of CH is performed. Section 4 presents the types of FLS that are Type-I FLS and Type-II FLS. Section 5 presents the clustering algorithms using Type-II FLS in WSNs. Section 6 presents the conclusion of the review paper.

2 Clustering Mechanism in Wireless Sensor Networks

Clustering is a mechanism in which SNs in the network are partitioned into groups. A single node in a group is assigned as a leader known as the cluster head, and the leftover nodes of the group are known as cluster members (CMs). The work of a CH is to aggregate the data sensed by its CMs from each cluster and send it to BS for required actions like adding or removing the nodes in case the node dies due to depletion of energy or hardware failure, scheduling time slots for each node, and locality preservation. Clustering offers the following advantages:

- Energy consumption decreases as CH works as a data aggregator node and the responsibility of transmission of data is allocated to CH only and therefore the traffic at BS reduces.
- It excels the network topology management as SNs are grouped in clusters and the responsibility to manage CMs is divided and allocated to respective CH only.
- The bandwidth can be reused for better resource allocation [2].
- It raises the stability, lifetime, and scalability of the WSN as well as provides localization of routing settings [2].

CH can be fixed or variable depending upon the application. The clustering algorithm is split into three stages: (i) CH selection (ii) formation of the cluster, and (iii) data transmission. Various algorithms are implemented for CH selection such as probabilistic algorithms [1], distributive algorithms, centralized algorithms, and fuzzy based algorithms. The cluster head is decided by a randomly generated threshold value, without considering the energy usage of the node in the probabilistic algorithm [1]. As a result, the network's energy usage is not dispersed, and the network's lifetime is reduced. CH is selected based on a small set of input parameters like the available energy of a node, the distance between nodes, etc. in the deterministic algorithm. In centralized algorithms, cluster heads are selected by BS [3]. In the FLS, CH is selected by applying a fuzzy rule base to the fuzzy input descriptor.

3 Cluster Properties

Some of the typical properties and characteristics that help in CH election and cluster formation are as given below:

Residual Energy: The available energy of SNs after each round of communication is known as residual energy. If SN's leftover energy is high, the chances to be CH is high.

Node Density: The amount of surrounding nodes around a cluster's CH is referred to as node density. The likelihood of being CH increases as the nodes around SN grows.

Distance: The distance between CH and BS is used to calculate. The closer SN and BS are, the more likely it is that SN will become CH. The Euclidian difference between adjacent nodes or the received signal strength index can also be used to compute the distance.

Centrality: It is a value that categorizes nodes according to how central they are to the cluster. To compute node centrality, SN adds the squared distances of surrounding nodes to the selected node.

4 Types of Fuzzy Logic System

Lofti A. Zadeh has introduced a fuzzy logic-based mathematical tool that handles the problems related to uncertainty in 1965 [4], it is a mathematical tool. It handles the problem related to uncertainty. The fuzzy logic system is split into two types: (i) The Type-I fuzzy logic system is the standard fuzzy logic system, while the Type-II fuzzy logic system is the extension of the Type-I fuzzy logic system. Figure 1 shows the block diagram of Type-I FLS. The Type-I FLS functions are divided into four stages which are stated below:

Fuzzifier: Fuzzifier is used for the fuzzification process. It maps non-fuzzy inputs into their fuzzy representation. This includes the use of MFs like triangular, trapezoidal, Gaussian, and so on.
Knowledge Base: The If–Then rule is used to define a relationship between input and output.
Fuzzy Inference Engine: To produce a fuzzy output, the inference process maps fuzzified inputs to the rule base. This stage determines the rule's consequence and its membership in the output sets.
Defuzzifier: The defuzzification procedure transforms the fuzzed output into a crisp, non-fuzzy state.

Figure 2 shows the block diagram of the Type-II FLS. Except out processing unit, Type-II FLS is equivalent to a Type-I FLS, which includes a type reducer block and a defuzzifier. The type reducer's function is to reduce a Type-II fuzzy set to a Type-I fuzzy set. A comparison of the Type-I FLS and the Type-II FLS is shown in Table 1 [5].

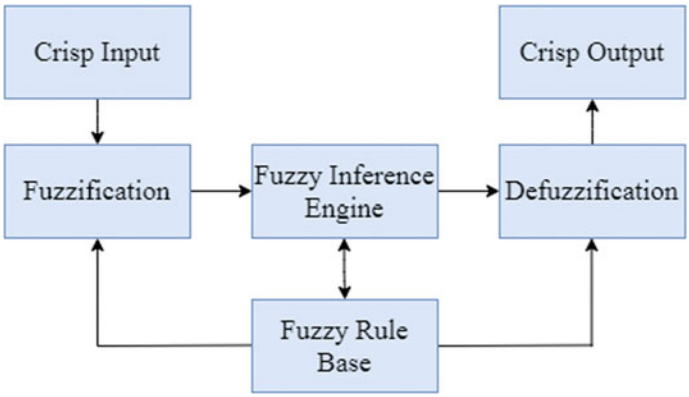


Fig. 1 Block diagram of Type-I FLS

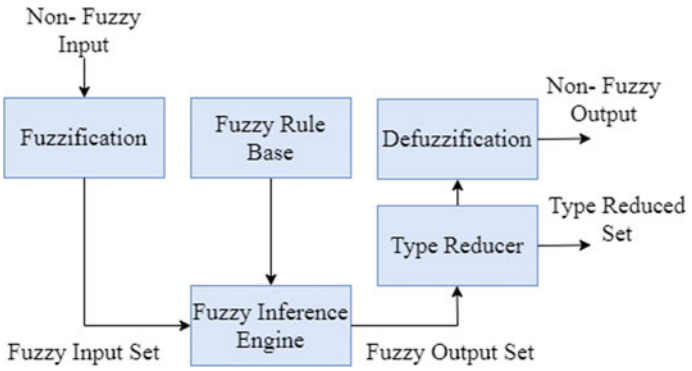


Fig. 2 Block diagram of Type-II FLS

Table 1 Comparison of Type-I FLS and Type-II FLS

Type-I FLS	Type-II FLS
Membership functions (MFs) are crisp	Membership functions are fuzzy
Different sources of uncertainty could not be modeled	Models any source of uncertainties that prevails in the system
The MFs of the type-I FLS are two-dimensional	These FLS have three-dimensional MFs. Additional dimension provides an extra effect for handling the uncertainty
It requires simple calculation and methods	Computational complexity is cumbersome but the modeled output is far better than the Type-I FLS
Determination of exact numeric membership functions is easier	Determination of exact numeric membership function is cumbersome but measurement of uncertainties is done effectively

5 Clustering Algorithms Using Type-II FLS in WSN

Low-energy adaptive clustering hierarchy (LEACH) [1] is a probabilistic clustering algorithm of WSN. It was the first algorithm when clusters were formed and CHs are selected in the setup phase and data transmission is done in the steady-state phase. The number of clusters is determined in advance. Probability is used to select CHs without considering the energy of SNs. CHs are rotated at every round. Work in [1] shows that the energy efficiency of LEACH is more as compared to single-hop data transfer from SNs to sink nodes. LEACH, on the other hand, has a disadvantage in that energy is wasted when there is no CH chosen in some rounds.

To tackle the disadvantages of LEACH [1], Gupta et al. [6] gave a distributed Type-I FLS for CH selection. Three fuzzy variables the available energy of SN, node concentration, and node centrality are applied to select CH. LEACH [1] and LEACH-C [3] are compared with this approach. The performance is evaluated by

the first node dies matrix. Results show that Gupta et al. [6] lengthen the lifetime of a network than compared protocols.

5.1 A New Energy Efficient Cluster-Head and Backup Selection Scheme in WSN

Izadi et al. [7] have proposed the protocol to save the sensed data and decrease the consumption of energy when CH dies during the communication. If the CH fails, data received and data collected can be lost and nodes transmit the data through direct communication. As a result, overhead and energy usage increase. If CH dies and the failure is not noticed, the problem gets worse. Another effect of the problem is when CH dies before the reselection of another CH and the area becomes unmonitored. To prevent the failure of CH, distributed and dynamic Type-II Mamdani FLS-based self-configurable cluster head (SCCH) approach is applied, and in case of failure, provision of the backup node is proposed in this work. The CH is selected using three fuzzy input descriptors: Available energy of node, node centrality, and node density. The fuzzy output of the Type-II FLS is the eligibility of being CH. If the available energy of CH decreases below the threshold value, the backup cluster head (BCH) node is designed to act as a cluster head (CH). SCCH is compared with PDD [8] and LEACH [1]. In PDD [8], three fuzzy input parameters are employed for CH selection, which includes distance from the BS, probability, and the sum of distances between the selected node and the other nodes. Because Type-II FLS presents a sustainable smooth surface that provides better monitoring performance and can handle uncertainties, SCCH has a greater chance of identifying CH than PDD. Average created messages in SCCH for different network sizes are lower than LEACH [1] and PDD [8] because cluster head is selected using available energy of a node after each transmission and distance between nodes. The average created messages in LEACH and PDD is higher because of probabilistic CH selection and due to the absence of BCH when CH fails, respectively. Consumption of the energy is distributed among the SNs in the SCCH is lower as BCH is assigned in case CH fails. Because CHs are selected without considering the energy dissipated of the node and due to direct communication, when CH fails in PDD, energy consumption is higher in LEACH.

5.2 ICT2TSK: An Improved Clustering Algorithm for WSN Using a Type-II Takagi-Sugano-Kang Fuzzy Logic System

Zhang et al. [9] have proposed a method to improve energy efficiency, load balancing of the network, and network longevity. Type-II Takagi-Sugano-Kang (TSK) FLS

(ICT2TSK) is introduced to select the CH in this work, which is better than the Type-I FLS at handling uncertainties. The selection of CH relies on fuzzy input parameters like the remaining energy of the node, neighbor nodes within the fixed radius, and distance to the BS. To balance the network traffic, eliminate CH gathering, and increase energy efficiency, each cluster-head has a fixed competition radius. The probability of becoming CH is an output parameter of the ICT2TSK. Here the selection of CH is done by BS in each round. ICT2TSK is compared with LEACH [1], CHEATS [10], and EEUC [11]. EEUC [11] is a distributive unequal routing protocol. CH selection is done by local competition and routing is performed by multi-hop communication. CHEATS [10] uses Type-I FLS for CH selection using the TSK approach. To evaluate the protocol's performance, three matrices are used: half nodes die (HND), first node dies (FND), and used energy in each turn. Simulation is performed by considering two different scenarios of BS. In scenario 1 when BS is at the center, the FND increase for ICT2TSK is 182.17%, 103.79%, and 11.93%, respectively, when compared to LEACH, CHEATS, and EEUC. The HND increase for ICT2TSK is compared with LEACH, CHEATS, and EEUC is 109.65%, 79.67%, and 78.14%, respectively. In scenario 2 when BS is away from the network, the FND increase for ICT2TSK to LEACH, CHEATS, and EEUC is 56.00%, 37.08%, and 30.56%, respectively, and the HND increase ICT2TSK to LEACH, CHEATS, and EEUC is 307.20%, 266.6%, and 346.39%, respectively.

5.3 A Clustering Routing Protocol for Wireless Sensor Networks Based on Type-II Fuzzy Logic and ACO (CRT2FLACO)

Zhang et al. [12] have proposed CRT2FLACO aiming toward the problems of load balancing, lifetime prolonging, and uncertainties of WSN. Type-II Mamdani FLS is used for CH selection that deals with WSN uncertainties and load balancing. Three fuzzy input parameters: Residual energy, number of neighbor nodes, and distance between nodes and BS are considered for the selection of CH and its competition radius. The selection of CHs is done by BS in every round. An unequal competition radius method is used to prevent CHs from gathering at the same place. Selected CHs are connected in a chain using ACO to pass on the data from CHs to BS, enhancing the network's lifespan. Along with the link, one leader node is chosen depending on the nearest SN to BS. This node integrates information from all CHs and sends it to the BS. The CRT2FLACO is compared with LEACH [1], EEUC [11], PEGASIS [13], and CRT1FLACO. In PEGASIS, chain-based routing is used, in which an SN pass on the data to the nearest node, fuse the data with its data, and continue transmitting till it reaches to leader node. Leader node finally transfers data to BS. CRT1FLACO is the same as CRT2FLACO but Type-I FLS is used for CH selection. Metrics FND, HNA, LND for varying values of node density and position of BS as well as the number of alive nodes per round are considered to evaluate the lifetime of WSNs.

Three scenarios of BS are considered: BS at the center; BS at outside; the BS far from the WSN. There is no variation in FND, HNA, and LND in CRT2FLACO due to the proper selection of CHs using Type-II FLS and unequal clustering mechanism. The network lifetime suffers if the number of SNs in the network is few and BS is far from WSN. This is due to the fewer nodes and larger distance between nodes and BS. Energy consumption per round is very less in CRT2FLACO than LEACH [1], EEUC [11], and PEGASIS [13] in every situation. This is because PEGASIS forms a very long chain for transferring data from one node to BS, LEACH transmits data using single-hop form node to BS, multi-hop routing is used by EEUC to pass on the data and CRT1FLACO uses ACO but CH selection is using Type-I FLS. In CRT2FLACO chain formed by ACO is short as ACO converges in a short time and Type-II FLS selects CH with more accuracy than Type-I. The network lifetime of ICT2TSK [7] is higher than CRT2FLACO as ICT2TSK [7] uses Type-II TSK FLS for CH selection.

5.4 An Alternative Clustering Scheme in WSN

Izadi et al. [14] provide the solution of the problems that is the failure of the CH and CM node, data collected by CH being lost and communication stops due to power instability. Early detection of a node's failure can help to prevent data loss and provide enough time to restore the data. In this work, a self-configurable clustering mechanism (SCCM) is proposed to identify the failure of a node and replace it with another node. The BCH node is used to take over the responsibility of the failure node. CH with a single BCH is not sufficient to solve the problem because the functionality of BCH cannot be predicted accurately every time. BCH may lose energy quickly if it is used more than other nodes. To overcome this problem, a Type-II FLS is used for CH selection in WSN. CHs and BCHs are selected based on Node centrality, Residual energy of the nodes, and local distance in between node deployed and neighbor. The node with the highest capability is designated as CH, while the remaining nodes are designated as BCHs based on their successive capability. Eligibility of BCHs is also monitored as if most eligible BCH fails due to heavy load then another BCH is selected to replace CH. CMs are also monitored and removed from CMs list if they fail in between. TDMA scheduling is used to transfer the data from nodes to CHs. In this section, the SCCM is compared with PDD [8] and DBCH-LEACH-C [15]. In DBCH-LEACH-C [15], two BCH nodes are assigned in case of failure of CH, and CHs are selected by BS. It is calculated how much energy is used to deliver a particular number of packets. The result shows that the SCCM is better than both PDD and DBCH-LEACH-C as it transfers a similar amount of data packets with low energy dissipation, with less data loss rate, and with less traffic because PDD does not consider the CH and CM failure and DBCH is a centralized algorithm and does not take the failure of CMs in consideration. SCCM does not transfer data to the fail nodes. Because the SNs are supposed to stay connected to the BS directly in the

event of CH failure, the created messages in PDD are greater than DBCH-LEACH-C and SCCM.

5.5 *Cluster Head Enhanced Election Type-II Fuzzy Algorithm for Wireless Sensor Networks (CHEETAH)*

Cuevas-Martinez et al. [16] aim to increase network lifetime via a fully distributed approach for clustering in WSNs. Type-II FLS is applied at each node of the WSN for the selection of CH. The input variables are relative distance of SN to BS, residual energy of a node, efficiency, and historical contribution as a CH. Output interval of the FLC represents the minimum as well as the maximum chance for an SN of getting elected as a CH. Here, distance and efficiency involve more uncertainty as they cannot be measured directly by considering parameters of SNs. The protocol runs in three phases: (i) CH selection, (ii) CH advertisement, and (iii) data transmission. Due to sampling precision can be reduced as inputs are uncertain. The performance test is done by evaluating three matrices FND, HND, and LND. Simulations are performed using two scenarios. In test scenario 1 (TS-1), the CHEETAH is compared with LEACH [1] and CRT2FLACO [17]. The rounds until FND of LEACH, CRT2FLACO, and CHEETAH are 689, 1012 and 1099, respectively. The rounds until HNA of LEACH, CRT2FLACO, and CHEETAH are 613, 1074 and 1502, respectively. Rounds until LND of LEACH, CRT2FLACO, and CHEETAH are 881, 1113, and 2055, respectively. In test scenario 2 (TS-2), the CHEETAH is compared with LEACH [1] and DUCF [18]. DUCF is a load-balancing distributed clustering system that employs FLS for CH selection. The fuzzy input parameters are the distance of a node to BS, energy remaining in a node, and node density. The round until FND of LEACH, DUCF, and CHEETAH is 416,715 and 1564, respectively. The rounds until HNA of LEACH, DUCF, and CHEETAH are 671, 767 and 2943, respectively. The rounds until LND of LEACH, DUCF, and CHEETAH are 791, 988 and 3758, respectively. The improvement is due to the proper selection of CH and reduction in control packets.

5.6 *Energy Efficient Clustering Algorithm for Multi-hop Wireless Sensor Network Using Type-II Fuzzy Logic*

Nayak and Vathasavai [2] provide the solution for overburdening the CH during the formation of the cluster. To solve this issue Type-II FLS is used for decision-making. The network is split into several layers. The number of levels is determined by the distance between the nodes and the BS. At each level, CH is decided by applying Type-II FLS. The proposed protocol approaches a multi-hop communication to pass the data collected by CH to the BS. If the final CH fails, one SN with the highest

energy is selected as a standby node (SB-CH) near the BS to resume communication (because if the CH is near to the BS, it consumes more energy). Tentative CH is elected by considering, three fuzzy input variables: Remaining battery power, distance to BS, and node concentration. Output fuzzy parameter is the confidence factor that is utilized to decide CH. This proposed protocol is compared with multi-hop LEACH (LEACH-MH), single-hop LEACH (LEACH-SH), and LEACH Type-I FLS protocols (LEACH-T1FL). The proposed protocol is implemented by considering Footprint of Uncertainty (FOU) values 0.2 and 0.7. FOU is the distance between the primary MF and the secondary MF. The total amount of packets sent in a single round, the total amount of data signals transferred to BS per second, the total alive nodes per round, and the throughput per second of the proposed protocol are higher than LEACH-SH, LEACH-MH, and LEACH-T1FL. The overall energy used per round of the proposed protocol is lesser than LEACH-SH, LEACH-MH, and LEACH-T1FL. Time when FND is higher of the proposed protocol than LEACH-SH, LEACH-MH, and LEACH-T1FL. This is because Type-II FLS deals with the input uncertainties of the environment efficiently. If the FOU of the Type-II FLS is larger than it covers more uncertainties of the input signals.

5.7 EUDFC—Enhanced Unequal Distributed Type-II Fuzzy Clustering Algorithm

Delgado et al. [19] proposed a new set of fuzzy input parameters for a new unequal distributed clustering technique for WSNs. To limit the exchange of data, interference, and energy consumed in the transmission processes, the algorithm depends on local information from the SNs. Because the available data is unknown, the suggested clustering algorithm uses a Type-II FLS as its core, with the knowledge base sampled to make it effective to implement in a node. Unequal clustering decreases energy use. The cluster size varies in the unequal clustering strategy, with clusters closer to the BS being smaller. Because nodes near the BS are utilized more frequently than others, turn is being used by EUDFC. Each turn is separated into two parts. In the first step, CHs are selected using a Type-II FLS. The functionality of the nodes is influenced by non-working or dead CH is readjusted in the second phase without the need for an FLS. The EUDFC is designed by using four fuzzy input variables: Relative distance to the BS, residual energy, contributing nodes' proximity to CH, and factor based on how many rounds have passed since a CH was elected the previous time it was a CH. The output variable is the chance to be CH. To check the sustainability of EUDFC, simulation results of EUDFC are compared with LEACH [1], Gupta [6], CHEETAH [16], CRT2FLACO [12], FUCA [20], and FBUC [21]. Except for LEACH, all the protocols are compared by implementing them in Type-I FLS as well as Type-II FLS. FUCA [20] is implemented to overcome the problem of a hotspot in the WSN with Type-I Fuzzy input parameters like the density of nodes around CH, distance to BS, and residual energy for CH selection. In FBUC [21], CMs connect with the CHs

based on the competition range of CHs and distance to the CH. For the CMs to join with the CH in FBUC, fuzzy logic is used. In EUDFC simulation is done by considering three different positions of the BS: BS at the Centre, outside, and the corner of the network. Three metrics are considered to analyze the lifetime of WSNs are provided from each simulation: FND, HNA, and the round when 10% of the total SNs are alive, or LND. Due to CH selection parameters, Contributing nodes' proximity to CH, and factor based on how many rounds have passed since a CH was elected the previous time it was a CH., EUDFC performs better than LEACH, CHEETAH, CRT2FLACO, FBUC, Gupta, et al., and FUCA in consideration of FND, HNA, and LND for the mentioned scenarios with Type-I FLS and Type-II FLS. Type-II FLC outperforms the Type-I FLS as Type-II FLS deals with input uncertainties effectively.

5.8 An Unequal Clustering Algorithm for Wireless Sensor Networks Based on Interval Type-II TSK Fuzzy Logic Theory (UCT2TSK)

Tao et al. [22] propose an interval Type-II TSK FLS to select CH using an unequal clustering algorithm in WSN. This method reduces the hot spot problem that is typical in clustered WSNs since CH near to BS has severe traffic problems due to inter-cluster communication, and it lengthens the network's lifetime. The residual energy, relative distance to the BS, and node density are considered as the inputs of an interval Type-II FLS. The result of the FLS reshapes the cluster radius and chooses the CHs. Data sensed by CMs is transmitted to CH by single-hop communication and CHs transmits to BS by multi-hop communication. The UCT2TSK is analyzed by comparing it with LEACH [1], EEUC [11], and TTDFP-Tier1 [23] (the TTDFP clustering stage). TTDFP is divided into two parts. Type-I FLS is used for first-tier clustering, and it uses fuzzy inputs such as relative connectivity, remaining node energy, and distance to the BS to adapt the competition radius for provisional CHs. Routing is done using fuzzy logic in the second tier. The UCT2TSK is also compared to the Mamdani Type-I equivalent (UCT1MD). In UCT2TSK, the radius of a cluster depends upon the node density and remaining energy of the node. If the node density is lower and the node's residual energy is higher, UCT2TSK has larger clusters near BS. When looking at the overall residual energy, UCT2TSK is about 49.6%, 20.3%, 10.7%, and 10.7% efficient than LEACH, EEUC, TTDFP-Tier1, and UCT1MD, which suggests that UCT2TSK consumes low energy in scenario 1 with BS out of network and 200 nodes. For the HND metric, UCT2TSK is about 40.5%, 12%, 7.4%, and 7.1% better than LEACH, EEUC, TTDFP-Tier1, and UCT1MD. When looking at the overall residual energy, UCT2TSK is 100%, 13.1%, 10%, and 6.4% efficient than LEACH, EEUC, TTDFP-Tier1, and UCT1MD, indicating that UCT2TSK consumes low energy in scenario 2, where BS is located outside of the network and the nodes in the network is 1000. For the HND metric, the HND in the UCT2TSK scenario

performs roughly 98.7%, 9.1%, 5.5%, and 2.9% better than LEACH, EEUC, TTDFP-Tier1, and UCT1MD. In both instances, UCT2TSK has a faster network performance and fewer nodes die every round than other protocols.

5.9 *A New Centralized Clustering Algorithm for Wireless Sensor Networks*

Cuevas-Martinez et al. [24] aim to lengthen the lifetime, reliability, the durability of the network, and provide energy-efficient data transmission. The proposed approach is a centralized unequal clustering algorithm that includes the CH selection and skip setting selections. CH is elected by Type-II Mamdani FLS. Fuzzy input parameters are residual energy, normalized distance, degree of deviation of energy, and a normalized number of rounds without being CH and the fuzzy output parameter is score value. The highest 5% of sensor nodes are chosen as CH based on their score value. The election of consecutive CHs is postponed for a predetermined round, which is known as the BS's skip parameter. Skip value is adjusted by the value of FND, HNA, and LND. If the round value is the same as the skip value, then new CHs are selected otherwise same set of CHs is used to collect and pass the data to the BS. LEACH [1], CTR2FLACO [17], CHEF [25], EEDCF [26], and EUDCF [19] are compared to the proposed protocol. CHEF [25] is a centralized algorithm. It uses Type-I FLS to choose the CH. Three inputs are energy, centrality, and concentration and the output is the appropriate selection of an SN to become a CH. EEDCF [26] is a distributed algorithm that identifies the CHs from the chances estimated by each node using Type-I FLS and the close nodes coordinate. Because control message exchange is minimized by setting the skip value and uncertainties in the network are managed by Type-II FLS as well as due to unequal clustering, the centralized skip based algorithm (CSBA) has a lengthen the network lifetime than LEACH, CTR2FLACO, CHEF, EEDCF, and EUDCF. However, network performance is reduced in terms of durability and reliability if similar nodes transmit the information to the CH for a long time and if the skip value is not properly selected. FND's skip value is inversely proportional to HNA and LND's. This is because being a CH consumes significantly more energy than being a contributing node. Depending upon the priority in between reliability and network lifetime skip value is selected high and low, respectively.

5.10 *Trust-Aware Clustering Approach Using Type-II Fuzzy Logic for Mobile Wireless Sensor Networks (TAT2F-M)*

Kousar et al. [27], aims to lengthen the network lifetime and to provide security in the transference of data in the network. Because of the cryptographic technique utilized for security, there is a lot of overhead, network complexity, and poor connectivity.

Therefore, trust aware approach is developed here using Type-II Mamdani FLS. Type-II FLS is applied to decide secure and energy-efficient CH based on fuzzy input parameters like remaining battery power, trust value, distance to BS, node concentration, and moving speed of the node and output fuzzy parameter is CH trust factor. Trust value is a combination of indirect trust and direct trust. The self-monitoring of SNs is used to calculate the direct trust value. The indirect trust value is a value that is dependent on the opinions of neighbor nodes. For the CH selection process, CH has decided in a way that it detects and eliminates malicious nodes. The threshold value is used to determine the trust value. A malicious SN is defined as one whose trust value decreases below a predetermined threshold. The random waypoint model is used to decide the mobility (direction, position, and moving speed of node for time) of nodes. Some random malicious nodes are placed in the network initially. The accuracy of malicious node identification is initially determined by varying threshold values. The observed best value of threshold (0.3) is then used to examine the average sending ratio, average consistency ratio, and average packet delivery ratio for the round. LEACH-M (LEACH-Mobile) [28], LEACH-TM (LEACH-Trust) [29], and EESTB [30] are all compared to the TAT2F-M. A variation in LEACH by considering the mobility of SNs is done in LEACH-M. In LEACH-TM, CHs are selected using trust values. To deal with security selection while utilizing the least amount of energy, the secure CH selection method in EESTB is implemented by determining the weight of each node. In the TAT2F-M, all trust factors undergo degradation initially as all the malicious nodes are there in the network are not detected. As the number of rounds increases malicious nodes are detected and three trust factors are increased. TAT2F-M uses less energy and has a higher number of alive nodes than LEACH-M (LEACH-Mobile), LEACH-TM (LEACH-Trust), and EESTB since CH reelection is quick and reliable while utilizing Type-II FLS.

5.11 An Improved Hierarchical Clustering Method for Mobile Wireless Sensor Network Using Type-II Fuzzy Logic (LEACH-MT2FL)

Kousar et al. [31] aims to optimize the energy usage, scalability, and lifetime of the WSN. Because of the node mobility, packets are lost in the transmission is another challenge to deal with in WSN. In this work, Type-II FLS is used to select CH in the network which deals with the CH overload, packet loss, and energy consumption. Fuzzy input parameters are the moving speed of the node, node concentration, distance to BS, remaining battery power, and pause time. The output parameter is a selection of CH, namely, chance. The random waypoint model is utilized to describe the velocity of the SN, location of mobile SNs, and acceleration of the node that vary over a specific time. The deployed SNs are categorized and numbered according to their distance from the BS. Selection of CH is performed at each level by Type-II FLS. Performance is analyzed by comparing the LEACH-MT2FL with LEACH-MF

(LEACH-Mobile using Fuzzy) [32], LEACH-M [33], and mobile-based clustering (MCB) [34]. LEACH-M selects CH without considering the movement of SNs. Mobile SNs leave the cluster and issue a join request to another CH so that the observed data can be sent. This results in packet loss and waste of energy. MBC algorithm selects CH according to the node mobility and available energy of SN. LEACH-MF selects CH using Type-I FLS. For varying amounts of node mobility, LEACH-MT2FL having a 0.7 value of FOU has a better packet delivery ratio than LEACH-MT2FL having 0.2 value of FOU, LEACH-MF, MBC, and LEACH-M. LEACH-MT2FL having a 0.7 value of FOU has a higher number of alive nodes than LEACH-MT2FL having 0.2 value of FOU, LEACH-MF, MBC, and LEACH-M to the round number. LEACH-MT2FL having a 0.7 value of FOU consumes less energy than LEACH-MT2FL having a 0.2 value of FOU, LEACH-MF, MBC, and LEACH-M. A higher value of FOU covers a large distance between the primary MF and secondary MF. So, it provides enough space to deal with input uncertainties for the selection of CH. Table 2 shows the protocols used for CH selection using Type-II FLS in WSN.

6 Conclusion

The clustering approach is integrated with Type-II FLS to solve the major issues in WSNs that are data loss, data transmission security, energy consumption, network lifetime, hotspot problems, the failure of CH in between communication, delay in transmission, and packet drop due to collision. The adaption of unequal clustering algorithms, selecting back up a node in case of CH or CM failure, use of cryptographic technics for security, splitting the network into layers avoids a problem of the hotspot in the network, distributes the energy in WSN, maximize lifetime, scalability, and stability of the network. Type-II FLS with its input parameters like node density, node degree, residual energy of nodes, and distance between nodes results in efficient CH selection, cluster formation, and cluster range computation. This is because Type-II FLS MFs are fuzzy and represented in three dimensions which can model any source of input uncertainties in the WSN.

Another computation intelligence techniques like particle swarm optimization, ant colony optimization, can be used to select CH. By expanding the list of metrics for CH selection, such as packet jitter, latency, reliability, efficiency, and throughput, the efficiency of CH selection can be enhanced.

Table 2 Cluster head selection using Type-II FLS in WSN

References	Fuzzy input parameters	Simulation tool	Merits	Demerits
[7]	(i) Available energy (ii) Node density (iii) Node centrality	MATLAB	(i) Reduces the communication overhead (ii) Lengthen the network lifetime (iii) Reduces packet loss	(i) If BCH fails network becomes unmonitored and direct communication increases energy consumption
[9]	(i) Remaining energy (ii) Node density (iii) Distance from SNs to BS	MATLAB	(i) Increase in network stability (ii) Increase the network lifetime	(i) CHs are selected by BS (ii) GPS receivers are used to locate the nodes that increase the cost
[12]	(i) Remaining energy (ii) Number of neighbor nodes (iii) Distance from node to BS	MATLAB	(i) Equal load balancing (ii) Increase in network scalability, network lifetime, and network stability	(i) CHs are selected by BS (ii) Nodes die quickly after FND
[14]	(i) Residual energy (ii) Node density (iii) Local distance	MATLAB	(i) Increase the network lifetime and throughput (ii) Reduces message overhead	(i) If consecutive BCHs are found to fail in case of CH failure, data loss, and end-to-end delay increase
[16]	(i) Residual energy (ii) Relative distance (iii) Historical contribution as a CH (iv) Efficiency	MATLAB	(i) Increase the network lifetime (ii) Reduction in control packets	(i) Transmission of data is done by single-hop communication
[2]	(i) Remaining battery power (ii) Distance to BS (iii) Node concentration	MATLAB and NETSIM	(i) Increase in network scalability (ii) Increase the network lifetime	(i) Nodes near the BS depletes soon as stand-by CH or CH transmit the data to BS due to chain of CH is formed to transmit the data

(continued)

Table 2 (continued)

References	Fuzzy input parameters	Simulation tool	Merits	Demerits
[19]	(i) Residual energy (ii) Relative distance to the BS (iii) Relative distance of SNs to CH (iv) Number of rounds that node has not been selected as CHs	MATLAB	(i) The communication protocol has been built so that there is no major overhead (ii) Turned-based CH selection reduces control overhead and so the energy consumption	(i) If CH dies in the first turn, then in the second turn CMs transmit the data directly to BS. So, energy consumption increases (ii) Complexity increases due to turn-based scheduling to the entire process
[22]	(i) Relative distance to BS (ii) Residual energy (iii) Node density	MATLAB	(i) TSK method has precise output. So, the reasoning process and output process are simpler (ii) Radius depends upon node density and not on the distance between BS and SNs	(i) The rule base applied for the small-scale and large-scale networks is similar
[24]	(i) Residual energy (ii) Normalized distance (iii) Degree of deviation of energy (iv) Normalized number of rounds without being CH	MATLAB	(i) Increases network lifetime (ii) Reduces energy consumption (iii) Increases the network performance by durability and reliability	(i) CHs are selected by BS (ii) Reduction in durability and reliability if skip value is not properly selected (iii) If any node either CH or CM dies in the skip period data sensed by CM and data gathered by CH is lost
[15]	(i) Remaining battery power (ii) Trust value (iii) Distance to BS (iv) Node concentration (v) Moving speed of the node	MATLAB	(i) Reduces energy consumption and packet loss (ii) Increases network lifetime and security	(i) Because trust value is dependent on both direct and indirect trust value, the failure of a neighbor node can affect the identification of the malicious node

(continued)

Table 2 (continued)

References	Fuzzy input parameters	Simulation tool	Merits	Demerits
[16]	(i) Remaining battery power (ii) Distance to BS (iii) Moving speed (iv) Concentration (v) Pause time	MATLAB	(i) Increased network reliability, scalability, and lifespan	(i) Computation complexity increases due to a greater number of input parameters

References

1. W.R. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in *Proceedings of the Annual Hawaii International Conference on System Sciences 2000*, Jan 2000, pp. 1–10
2. P. Nayak, B. Vathasavai, Energy efficient clustering algorithm for multi-hop wireless sensor network using Type-II fuzzy logic. IEEE Sens. J. **17**, 4492–4499 (2017)
3. W.B. Heinzelman, A.P. Chandrakasan, H. Balakrishnan, An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wirel. Commun. **1**, 660–670 (2002)
4. L. Zadeh, Fuzzy sets. Inf. Control **8**, 338–353 (1965)
5. S.N. Sivanandam, S.N. Deepa, *Principles of Soft Computing* (2005)
6. I. Gupta, D. Riordan, S. Sampalli, Cluster-head election using fuzzy logic for wireless sensor networks, in *Proceedings of the 3rd Annual Communication Networks and Services Research Conference* (2005), pp. 255–260
7. D. Izadi, J. Abawajy, S. Ghanavati, A new energy-efficient cluster-head and backup selection scheme in WSN, in *Proceedings of the 2013 IEEE 14th International Conference on Information Reuse and Integration (IRI)* (2013), pp. 408–415
8. Y. Shen, H. Ju, Energy-efficient cluster-head selection based on a fuzzy expert system in wireless sensor networks, in *Proceedings—2011 IEEE/ACM International Conference on Green Computing and Communications (GreenCom 2011)* (2011), pp. 110–113
9. F. Zhang, Q.Y. Zhang, Z.M. Sun, ICT2TSK: an improved clustering algorithm for WSN using a type-2 Takagi-Sugeno-Kang Fuzzy Logic System, in *IEEE Symposium on Wireless Technology & Applications (ISWTA)* (2013), pp. 153–158
10. A. Pires, C. Silva, E. Cerqueira, D. Monteiro, R. Viegas, CHEATS: a cluster-head election algorithm for WSN using a Takagi-Sugeno fuzzy system, in *2011 IEEE Latin-American Conference on Communications (LATINCOM 2011)—Conference Proceedings* (2011)
11. C. Li, M. Ye, G. Chen, J. Wu, An energy-efficient unequal clustering mechanism for wireless sensor networks, in *2nd IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, 2005 (MASS 2005)* (2005), pp. 597–604
12. W.X. Xie, Q.Y. Zhang, Z.M. Sun, F. Zhang, A clustering routing protocol for WSN based on Type-II fuzzy logic and ant colony optimization. Wirel. Pers. Commun. **84**, 1165–1196 (2015)
13. S. Lindsey, C.S. Raghavendra, PEGASIS: power-efficient gathering in sensor information systems, in *IEEE Aerospace Conference—Proceedings*, vol. 3 (2002), pp. 1125–1130
14. D. Izadi, J. Abawajy, S. Ghanavati, An alternative clustering scheme in WSN. IEEE Sens. J. **15**, 4148–4155 (2015)
15. S.U. Hashmi, H.T. Mouftah, N.D. Georganas, Achieving reliability over cluster-based wireless sensor networks using backup cluster heads, in *GLOBECOM—IEEE Global Telecommunication Conference* (2007), pp. 1149–1153
16. J.C. Cuevas-Martinez, A.J. Yuste-Delgado, A. Trivino-Cabrera, Cluster head enhanced election Type-II fuzzy algorithm for wireless sensor networks. IEEE Commun. Lett. **21**(9), 2069–2072 (2017)

17. Q.Y. Zhang, Z.M. Sun, F. Zhang, A clustering routing protocol for wireless sensor networks based on Type-II fuzzy logic and ACO, in *IEEE International Conference on Fuzzy Systems* (2014)
18. B. Baranidharan, B. Santhi, DUCF: distributed load balancing unequal clustering in wireless sensor networks using fuzzy approach. Appl. Soft Comput. J. **40**, 495–506 (2016)
19. A.J. Yuste-Delgado, J.C. Cuevas-Martinez, A. Trivino-Cabrera, EUDFC—enhanced unequal distributed Type-II fuzzy clustering algorithm. IEEE Sens. J. **19**, 4705–4716 (2019)
20. D. Agrawal, S. Pandey, FUCA: fuzzy-based unequal clustering algorithm to prolong the lifetime of wireless sensor networks. Int. J. Commun. Syst. **31**, 1–18 (2018)
21. R. Logambigai, A. Kannan, Fuzzy logic based unequal clustering for wireless sensor networks. Wirel. Netw. **22**, 945–957 (2016)
22. Y. Tao, J. Zhang, L. Yang, An unequal clustering algorithm for wireless sensor networks based on interval Type-II TSK fuzzy logic theory. IEEE Access **8**, 197173–197183 (2020)
23. S.A. Sert, A. Alchihabi, A. Yazici, A two-tier distributed fuzzy logic based protocol for efficient data aggregation in multihop wireless sensor networks. IEEE Trans. Fuzzy Syst. **26**, 3615–3629 (2018)
24. J.C. Cuevas-Martinez, A.J. Yuste-Delgado, A.J. Leon-Sanchez, A.J. Saez-Castillo, A. Triviño-Cabrera, A new centralized clustering algorithm for wireless sensor networks. Sensors (Switzerland) **19**, 1–19 (2019)
25. J. Kim, S. Park, Y. Han, T. Chung, CHEF: cluster head election mechanism using fuzzy logic in wireless sensor networks, in *International Conference on Advanced Communication Technology*, vol. 3 (2008), pp. 654–659
26. Y. Zhang, J. Wang, D. Han, H. Wu, R. Zhou, Fuzzy-logic based distributed energy-efficient clustering algorithm for wireless sensor networks. Sensors (Switzerland) **17** (2017)
27. A. Kousar, N. Mittal, P. Singhm, Trust-aware clustering approach using Type-II fuzzy logic for mobile wireless sensor networks, in *Advanced Informatics for Computing Research (ICAICR 2019)*. Communications in Computer and Information Science, vol. 1076 (2019), pp. 300–310
28. G.S. Kumar, P.M. Vinu, K.P. Jacob, Mobility metric based leach-mobile protocol, in *2008 16th International Conference on Advanced Computing and Communications* (2008), pp. 248–253
29. W. Wang, F. Du, Q. Xu, An improvement of LEACH routing protocol based on trust for wireless sensor networks, in *5th International Conference on Wireless Communications, Networking and Mobile Computing* (2009), pp. 1–4
30. E. Rehman, M. Sher, S.H.A. Naqvi, K. Badar Khan, K. Ullah, Energy-efficient secure trust-based clustering algorithm for mobile wireless sensor network. J. Comput. Netw. Commun. **8**, Article ID 1630673 (2017)
31. A. Kousar, N. Mittal, P. Singh, An improved hierarchical clustering method for mobile wireless sensor network using Type-II fuzzy logic, in *Proceedings of the ICETIT 2019*. Lecture Notes in Electrical Engineering, vol. 605 (2019), pp. 128–140
32. J.S. Lee, C.L. Teng, An enhanced hierarchical clustering approach for mobile sensor networks using fuzzy inference systems. IEEE Internet Things J. **4**(4), 1095–1103 (2017)
33. D.S. Kim, Y.J. Chung, Self-organization routing protocol supporting mobile nodes for the wireless sensor network, in *2006 First International Multi-symposiums on Computer and Computational Sciences (IMSCCS 2006)*, vol. 2 (2006), pp. 622–626
34. S. Deng, J. Li, L. Shen, Mobility-based clustering protocol for wireless sensor networks with mobile nodes. IET Wirel. Sens. Syst. **1**(1), 39–47 (2011)

PLRec: An Efficient Approach Towards E-Learning Recommendation Using LSTM-CNN Technique

N. Vedavathi and K. M. Anil Kumar

Abstract With the rapid increase in Social Web applications, a significant amount of research has been dedicated to the analysis and development of a personalized recommendation using social tags. The social tags will allow the users to select their preferences and object of interest. Different machine learning and neural networks are applied widely to address the problem of tag ambiguity. However, conventional machine learning algorithms require additional training models to train recommendation models for processing large scale data. Hence they are considered to be inappropriate for generating personalized recommendations in learning systems. This research presents an LSTM-CNN-based approach for generating appropriate learning recommendations. The paper presents an ontology-based similarity metric to overcome the issue of tag ambiguity without requiring any additional training model. The proposed approach semantically examines and quantifies the similarities between user and item profiles by matching the concepts of tags. The effectiveness of this approach was validated using different performance metrics and results validate the performance of the proposed approach.

Keywords E-Learning · Personalized recommendation · Tag disambiguation · Ontological similarity · Concept similarity · LSTM-CNN

1 Introduction

There is a substantial rise in the demand for Social Web-based applications in recent years. A Social Web is a collective form of various social media platforms and has attracted a large number of users because of its effective interaction and ability to provide user-requested information [1]. With the continuous growth of the Social

N. Vedavathi (✉)
Department of Computer Science and Engineering, NIE Institute of Technology, Mysuru, India
e-mail: vedavathiresearch@gmail.com

K. M. Anil Kumar
Department of Computer Science and Engineering, SJCE, Mysuru, India
e-mail: anilkm@sjce.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_50

Web, it has become difficult for users to browse through all the data available online due to increased network traffic and data overload. Hence, personalized recommender (PLRec) systems are developed on the Social Web which provides personalized recommendations for the users by filtering out redundant and irrelevant information. However, it is challenging to provide relevant recommendations to the users with high accuracy. In this context, various researches have discussed how to develop personal recommendation with high accuracy [2–5].

Personalized recommendation has widely applied in various domains such as e-commerce websites, social media platforms, blogging sites, and online learning systems. This research mainly emphasizes the application of a personal recommendation for online learning systems such as e-learning platforms. Online learners generate enormous data with respect to their learning patterns, which allows the learning systems to discover the user requirements. It can be inferred from existing literature works that the current learning systems have undergone a significant transformation with the help of advanced technologies such as personalized recommendation. The personalized recommendation is highly important in online learning systems since it enhances the data acquisition process for different users and enhances their learning experience [6].

Conventional PLRec techniques face several issues in terms of recommendation accuracy, cold-start of recommendation and quantification of the learning effect. Recommendation approaches used in the learning systems are basically designed by leveraging the personal preferences and interests of the learners to build learners' profiles. Correspondingly, these profiles are used for ranking the preference of the learners. The preferences are compared with the information stored in the database and preferences that reflect the learner's personal interest are ranked higher [7]. Basically, the preference or interest of the learners can be obtained by using the information which is available distinctly. Hence, the development of personalized recommender systems requires the learner's data which is publicly available and easy to access. Fortunately, learners can easily provide various social annotations to online resources such as different learning websites, blogs, videos etc. on online platforms with the help of tags and bookmarks. Social annotations are very much suitable for personalized recommendation systems on the Social Web and are available publicly [8]. Since these annotations are directly provided by the learners, it will be helpful for the personalized recommendation system to recommend relevant content to the learner's.

Conventional approaches for personalized learning recommendation are designed focusing only on a single subject or a structured learning process to obtain learning full-path recommendation [9]. It must be noted that the structured learning process recommendation is a step-by-step approach, which makes it cumbersome and complex. Also, it increases the length and duration of the learning process thereby making this recommendation process quite ineffective. To overcome these limitations of conventional approaches, machine learning and neural networks are used widely to provide accurate recommendations to the learners based on their preferences and interests [10, 11]. Because of the diversified nature of learning data, it is difficult for traditional neural networks such as Support Vector Machine (SVM)

[12], Multilayer Perceptron (MLP) [13], and Radial basis function (RBF) to perform effectively for diverse and large scale data systems. In this context, advanced learning methods such as Convolutional Neural Networks (CNN) [14], Autoencoders (AE) [15], and Long-Short Term Memory (LSTM) [16] have been successfully employed for pattern recognition and in the analysis of large scale data. Various existing works have shown the effectiveness of LSTM in providing accurate personalized learning recommendations [17–19].

In this research an integrated LSTM-CNN model is proposed for providing a PLRec for the users in an online learning environment. The key contributions of the current work are listed in the below points:

(i) A novel LSTM-CNN framework is proposed for understanding learners' preferences, interests, and their needs to recommend appropriate learning contents.
(ii) An ontology-based personalized recommender system is proposed which incorporates a similarity measuring metric, known as ontological similarity (OS), to overcome the issue of tag ambiguity.
(iii) A two-step disambiguation algorithm is proposed which includes mapping, and disambiguation using a top-down traversal operation.

The rest of the paper is organized as follows: Sect. 2 discusses the existing literary works done on personalized recommendation models, Sect. 3 provides the architecture of the proposed Personalized Recommendation System. Section 3.4.1 presents a brief analysis of the proposed LSTM-CNN approach. Section 4 details the results of the simulation analysis and Sect. 5 concludes the paper with prominent research observations and future work.

2 Literature Review

Recommender systems are one of the most discussed topics in online learning systems and were investigated for many years with several successful deployments [20–22]. A recommender system provides recommendations for multiple different domains such as books, movies, products, and helps the users in selecting the services based on their object of interest. In learning systems, they personalize the learning preferences and generate recommendations by analyzing the learner's data. The design of recommender systems and their applications to solve various learning related problems has been an active area of research [23, 24]. PLRec systems are characterized according to their domain and features of the learning data wherein they make use of the data sources to provide recommendations. There are four main methods in the design of PLRec systems namely collaborative, content-based, demographic recommender, and Knowledge-based PLRec systems. Collaborative recommender systems collect the feedback of the users in terms of rating and evaluate the similarities of ratings given by different users to generate recommendations. These systems are further

categorized into memory-based and model-based techniques. Content-based recommender systems make use of two different data sources and recommend relevant content to the user based on the ratings and data features. These systems attempt to match the users to the content which are similar to the content liked by the same user previously based on the features of the content liked by the user [25]. While demographic recommender systems generate PLRec based on the demographic details of the user, Knowledge-based recommender systems give recommendations based on the analysis of the user's interest and preferences. The knowledge about the likings of the users can be either in the form of explicit or implicit forms. Collaborative filtering is used widely [26]. However, collaborative filtering does not consider the user attributes while recommending the content and also there is a problem of data sparsity in these systems. To overcome this problem various researchers have suggested the implementation of ontology-based approaches in recommender systems [27]. A personalized ontology-based e-learning approach was proposed in [28]. The learner's profile was categorized and based on that personalized content recommendation was generated. A mathematical model was developed for categorizing the learner requirements using machine learning and neural networks. Grivokostopoulou et al. [29] proposed an ontology-based approach for personalization in e-learning systems. The study addresses the concerns related to personalization and the implementation of the learning techniques to evaluate the characteristics of the learners. This was achieved by modeling the student's characteristics and improving the learning experience. Various existing works have validated the efficacy of machine learning approaches in enhancing the performance of personalized recommendation systems [30, 31].

3 Proposed Research Methodology

The preliminary aim of the proposed research is to understand learners' preferences, interests and their needs to recommend appropriate learning content. The proposed recommendation approach recommends the course tags to the students based on the social media text search using machine learning with text mining techniques. To achieve this, a hybrid LSTM-CNN-based personal recommender approach is implemented. To overcome the need for an advanced computing approach to train recommendation systems, this research proposes an ontology-based personal recommendation model which applies an ontological similarity (OS) metrics to overcome the issue of tag ambiguity without actually training the model [32]. The ontological similarity makes use of an external ontology specific to one domain to disambiguate tag data and to further evaluate the similarities between the user and object details based on the semantic similarity of the matching concepts of tags in the corresponding user profiles. OS has high accuracy compared to other similarity metrics and hence can be employed for enhancing the performance of the PLRec on the online learning platforms. The steps involved in the proposed approach is discussed in the below sub sections.

3.1 Data Collection

Tags enable better personalized recommendations. Appropriate dataset containing tags relevant to e-learning is collected from the relevant datasets for experimental analysis. In this research, the data for experimentation is collected from the Kaggle website wherein the data about the past information about the courses in text format was collected. Here, two datasets are used; a question dataset and a tag dataset. From the question's dataset the data of 1000 users were collected. The first 1000 rows from the datasets are considered for preprocessing and were trained with 1000 row's data with tags. From the tag dataset, only two columns, i.e., Id and Tag were considered with 1000 tags from the dataset. Here, two types of tags namely NLP and Web were considered with each tag containing 500 records.

3.2 Implementation of Tag Disambiguation Algorithm

The implementation of the tag disambiguation algorithm involves two steps [33]: Mapping and Disambiguation. Initially, a mapping algorithm is applied which maps the tags to all relevant and similar information in the ontology. For a specific user profile, the tags are mapped to all possible concepts in the ontology. For tag disambiguation, a top-down traversal operation is performed wherein statistical information is used to disambiguate tags with multiple similar information. This is done to ensure that each tag in the data will have at least one similar concept in the ontology.

Mapping

For a given user profile, the complete string containing information is traversed to recognize all similar concepts. The matching is done based on the similarity of the string values. If a similar concept is not found for a tag, then the tag is regarded as an unmatched tag. On the other hand, if a single concept is identified as similar, then it is considered as a matching concept directly. If a concept is identified as similar for multiple times, then this problem is regarded as a multiple occurrence problem and in this case, all the concepts are marked as matching concepts for the tag and the ambiguity of such tags is eliminated in the disambiguation stage.

To perform disambiguation, the weights of the concepts are increased which is denoted as (c_w) of the matching concept corresponding to all the previous tag weights 'w'.

Tag Disambiguation

Once all the tags are mapped, top-down disambiguation is performed for addressing the issue of multiple occurrences, i.e., a single tag having multiple similarity concepts. The top-down process involves various steps: Initially, the concept variable c_s is assigned with the root of the taxonomy and secondly, the concept weights are compared with the child concepts of c_s which have a minimum of one child concept

as its descendant. Third, a child concept that has the highest weight is selected and all other concepts that are not the descendants of the new c_s are unmarked. These steps are repeated till a single candidate concept is left and that concept is considered as the matching concept for the tag. It must be noted that if more than one child concepts possess similar weights with highest values, then the first concept selected during the process will be identified and selected as the nearest and matching concept.

3.3 Computation of Similarity Measures

Each tag will have a minimum of one similar concept at the end of the tag disambiguation process. A similar concept will allow the users to explore the semantic similarities of the concepts in ontology, and this process is known as concept similarity (CS). The CS process is employed for estimating the semantic similarity of the tags. Correspondingly, the OS is evaluated by determining the weighted sum of all CS values. As discussed previously, the ontological similarity used in this research addresses the problem of tag ambiguity by mapping homonyms to different concepts in the ontology based on user-related information. The evaluation of the ontological similarity is done using two main steps:

- For each and every concept that is matched, i.e., c_u of the user's profile tag t_u, a semantic similarity with the closest concept c_d of the tag t_d is estimated.
- Furthermore, the concept similarities are integrated with the cosine similarities to obtain ontological similarity.

Concept Similarity

This research uses two qualitative metrics such as concept similarity metric and ontological similarity metric for evaluating the similarities between two concepts. These two metrics are based on two prominent aspects of the taxonomies such as the concept of relative depth and the minimum distance between two concepts. The potential capability of these two concepts have been validated in existing literary works [34]. The relative depth holds a greater preference in the taxonomy since it effectively defines the concept specifications; it states that (i) the concepts at the higher levels are less similar compared to the concepts in the lower levels; and (ii) the concepts incorporated through a common ancestor in the upper levels are highly different compared to the ancestors in the lower levels. A traditional hierarchy-based metric [32, 33] can be employed for defining the similarity between two concepts c_u, and c_d as shown in Eq. 1.

$$\mathrm{ConSim}_{1(\mathrm{cu},\,\mathrm{cd})} = \frac{2.l(\mathrm{lca}(\mathrm{cu},\,\mathrm{cd}))}{l(\mathrm{cu}) + l(\mathrm{cd})} \tag{1}$$

In Eq. 1, lca (c_u, c_d) represents the lowest common ancestor of the concepts c_u, and c_d and $l(c)$ represents the concept level of c (with $l(\mathrm{root}) = 0$). Considering

the minimum distance between c_u, c_d to be very less, another metric is used for computing the similarity measures by determining the reciprocal of the length of their minimum distance as shown in Eq. 2.

$$\text{ConSim}_{2(\text{cu, cd})} = \frac{1}{\text{MD}_{(\text{cu, cd})} + 1} \tag{2}$$

In the above equation, MD (c_u, c_d) is defined as the minimum distance between c_u, and c_d which is determined as: $\text{MD}(c_u, c_d) = l(c_u) + l(c_d) - 2 * l(\text{lca}(c_u, c_d))$.

Ontological Similarity

Lastly, the OS of two user profiles x_u and x_d can be evaluated by calculating the weighted sum of all CS values which integrates both the concept and cosine similarity. The cosine-based OS is illustrated in Eq. 3.

$$\text{ontSim}_{\text{Cosine(xu, xd)}} = \frac{\sum_{i=1}^{|\text{Tu}|}(\text{wui} * \text{wdi} * \text{consim(cui, cdi)})}{\sqrt{\sum_{i=1}^{|\text{Tu}|}(\text{wui})^2} * \sqrt{\sum_{j=1}^{|\text{Td}|} k_j * (\text{wdj})^2}} \tag{3}$$

Here, $|T_u|$, $|T_d|$ represents the number of tags in the profiles x_u and $x_{d,}$ respectively. c_{ui} and c_{di} are the matching concepts for tags tag t_{ui} in x_u and tag t_{di} in profile $x_{d,}$ whereas w_{ui}, w_{di}, and w_{dj} are defined as the weights of the tags t_u, t_{di}, and $t_{dj,}$ respectively.

3.4 Implementation of PLRec System

The PLRec system toward learning recommendation uses the ontological similarity between the profile of the user's and the document in the taxonomy for generating personal recommendations on the Social Web. In general, the computing cost of ontological similarity is too high; hence, this research aims to achieve a practical personalized recommendation that assumes that the user's preference is stable. Hence, for a given user profile, initially, offline preprocessing is performed for determining the OS between the users and document profiles. The obtained OS values are stored for further analysis of online PLRec. This research uses an LSTM-CNN technique for the development of an online personalized recommendation system.

3.4.1 LSTM-CNN-Based Personalized Recommendation System

Application of machine learning and neural networks to the personalized recommendation systems which requires data mining and extraction of relevant features from large scale data will not only improve the accuracy of recommendation, but also enhance the performance of the recommendation approach and thereby improve the

user experience. In this research, the proposed algorithm takes both user and document profiles and an external ontology based on a particular domain as input. The processing is performed in three stages: in the initial stage, each concept will be allocated with tags. Then, for each document, the tag will be allocated to the concepts and calculates the OS between the user and document profiles, a list which is finally returned is used for personalized recommendation.

Convolutional Neural Network is applied hugely in various classification and natural language processing (NLP) applications. Various approaches using CNN are discussed for studying sentence analysis and for generating recommendations based on the analysis. While LSTM is a type of recurrent neural network which holds a greater prominence in predictive analysis. LSTM stores the memory of the previous iteration and predicts the outcome based on the previously stored value. For every concept stored in the library, the LSTM assigns probability based on the previously identified concept; it identifies the concept that holds a greater value of probability and stores that concept in the memory unit. The superior memory of the LSTM makes it adaptable for generation of recommendations as it remembers the background of the concepts stored in the library. LSTM has the potential to overcome the disadvantage of fundamental RNN techniques of saving the long length word strings.

The CNN in the hybrid model of LSTM-CNN, extracts and stores the relevant features of the conceptual data which is composed in the form of a matrix. The feature extracting attribute of the CNN has a significant role in extracting the data from the user's profile. The convolution is performed by creating a window of equal length for the conceptual data. This will increase the efficiency of the CNN working by adjusting the convolution variables along with the size and number of convolutional windows. LSTM is used for encoding a complete sentence which concatenates the output of the CNN and provides the resultant output via a fully connected layer and the softmax layer.

The process of the LSTM-CNN is illustrated in Fig. 1.

The matrix placed at the bottom level incorporates the concepts embedded in a vector. Every single matrix will convolute into a window of a specific dimension and the corresponding feature maps of the convolutional kernels are generated. The convolution operation performed by the CNN layers will enable the proposed model to extract and store the local characteristics of the conceptual data. The generated feature map is further subjected to max pooling which in turn generates $m * 2$ feature maps. Lastly, it concatenates and the output is sent to LSTM as shown in the equation below:

$$Z = \oplus\left(P_{\mathrm{x}},\ P_{\mathrm{y}}\right),\ Z \in R^{\left(\frac{l-w+1}{2}\right)*(m*2)} \tag{4}$$

The LSTM generates the result which is similar to the encoding of the whole sentence. The LSTM further creates a penultimate layer as the output via the fully connected layer and the softmax layer. Correspondingly, the probability distribution score is evaluated by performing the softmax operation as illustrated in Eq. 5.

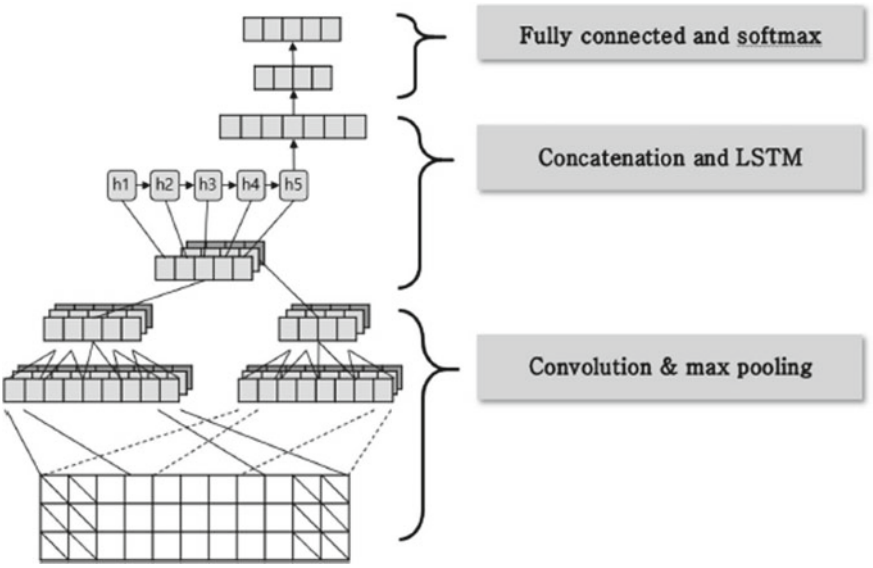


Fig. 1 Process flow of the LSTM-CNN framework

$$\widehat{P_i} = \frac{\exp(o_i)}{\sum_{j=1}^{c} \exp(o_j)} \tag{5}$$

where P_i is the user profile for '*i*th' function, that is the probability that the given user profile belongs to a particular category.

Once the learning process is completed, the LSTM-CNN will provide the final output which can be used for modeling the users' profiles as vectors over tags. The LSTM-CNN discovers the latent features and the extracted features are aggregated to generate recommendations.

4 Experimental Results

This section describes the outcome of the experimental analysis. The efficacy of the proposed approach is determined using different performance metrics such as precision, recall, F1-score, and support. The training and validation accuracy and training and validation loss of the proposed approach is also evaluated.

4.1 Performance Evaluation

The performance metrics are evaluated using four different classification elements namely: True positives (TP), True negatives (TN), False positives (FP), False negatives (FN). These terms collectively define the term confusion matrix which is mainly used for solving the problems related to classification accuracy where the output can be two or more classes. Here, the performance metrics are used to evaluate the recommendation accuracy and the performance of the proposed approach.

Precision score defines the ability of the LSTM-CNN model to accurately predict the true positive out of all the positive predictions. It is calculated as;

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \tag{6}$$

Recall for a function is determined as the ratio of the total generated recommendations that are accurately classified and is given as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

Similarly, F1-score is used for measuring the accuracy of the proposed approach which is given as:

$$\text{F1} - \text{score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

And Support is the count of actual occurrences of the recommendations in the specified dataset.

The experimentation was carried out using two datasets namely, NLP and Web. Correspondingly, the result of the different performance metrics is tabulated in Table 1.

Here, two labels are used from the dataset; one is NLP and the other one is WEB. Each dataset consists of 500 entries. The data obtained from the data was used to determine the performance of the proposed model. The dataset is split into NLP and WEB using a Keras module before building the model. Once the dataset is split, the Keras module will classify the data samples as 0 and 1 (binary format), wherein 0 and 1 represent NLP and Web labels, respectively. In other words, if the features of

Table 1 Confusion matrix of the proposed model

Sample	Precision	Recall	F1-score	Support
0	0.89	0.97	0.93	355
1	1.00	0.28	0.44	359

the data are mapped with the NLP then the classification report shows 0 else it shows 1. The obtained binary data is represented as [0, 1]. As inferred from Table 1, the proposed approach achieves a higher precision of 0.89 and 1.00 for 0 and 1 samples, respectively.

Correspondingly, the simulation results of the training and validation accuracy and training and validation loss of the proposed approach are illustrated in Fig. 2 and 3, respectively.

Fig. 2 Training and validation accuracy

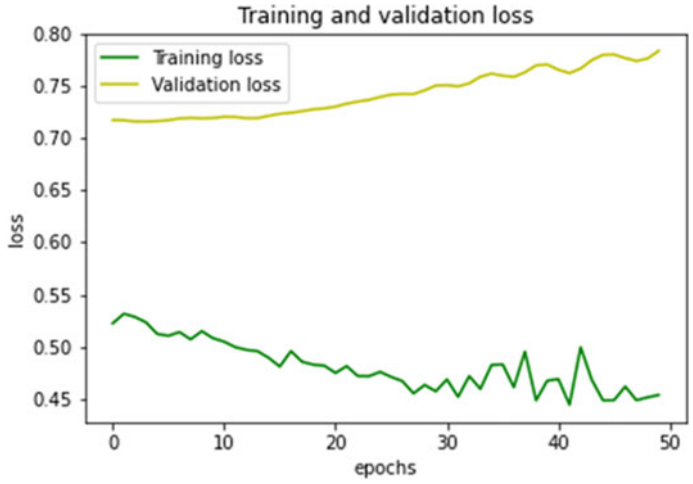


Fig. 3 Training and validation loss

The training and validation loss determines the data loss of the proposed training and validation. The training and validation loss was evaluated based on the performance loss. During experimentation, when the model was trained using the training data, the result was moderate, whereas upon training with validation data, the performance was improved. The accuracy and losses were compared and as observed from Figs. 2 and 3, the training loss shows better performance and is very low when compared with validation loss.

4.2 Comparative Analysis

The performance of the proposed approach was validated with the existing model. The existing model consists of only one LSTM layer with some hidden layers. In the proposed approach, the model was built by modifying the architecture of the neural network by combining the LSTM layer with CNN layers along with some additional hidden layers. This arrangement was done to improve the accuracy of the model. The training and validation accuracy and training and validation loss of the existing model is illustrated in Figs. 4 and 5, respectively.

In comparison with the accuracy and loss curves of the existing model, the proposed model achieves better performance as shown in Figs. 2 and 3. This validates the fact that the addition of neural network layers to the existing model will improve the accuracy. The results of the comparative analysis are presented in Table 2.

As inferred from the results, the proposed model outperforms the existing model in terms of accuracy and precision. Though there is not much difference in the performance of these two models in terms of recall and F1-score, it can be said that the

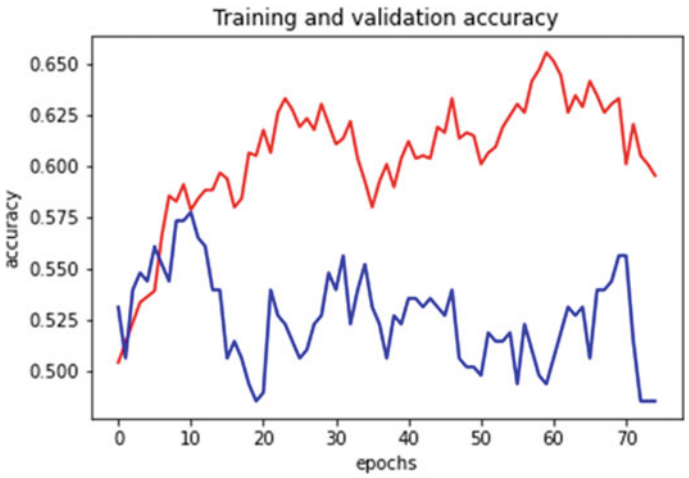


Fig. 4 Training and validation accuracy of the existing model

Fig. 5 Training and validation loss of the existing model

Table 2 Results of the comparative analysis

Parameters	Existing model	Proposed model
Accuracy	60%	87%
Precision	0.50	0.75
Recall	1.00	1.00
F1-Score	0.67	0.86
Mean reciprocal rank	0.25	0.50

addition of LSTM layers to the CNN layers has a significant impact on the accuracy, which is validated through simulation results. In addition, the existing model and the proposed model were compared with respect to their classification performance. The classification results of the existing and proposed model are tabulated in Tables 3 and 4, respectively.

Table 3 Classification performance of the existing model

Sample	Precision	Recall	F1-score	Support
0	0.50	1.00	0.66	119
1	0.00	0.50	0.00	120

Table 4 Classification performance of the proposed model

Sample	Precision	Recall	F1-score	Support
0	0.50	1.00	0.66	355
1	0.50	1.00	0.67	359

The results of the comparative analysis clearly indicate the superior classification performance of the proposed model. As observed, in the existing model, the inaccurate classification results in zero precision and F1-score values as shown in Table 3., which happens for various reasons like sparse and high dimensional data. On the other hand, the proposed model achieves desired classification performance.

5 Conclusion

This research proposes an effective personalized recommendation approach for learning systems using an LSTM-CNN model. The proposed approach evaluated the generation of recommendations based on ontological similarities. The proposed approach also addresses the issue of tag disambiguation problem by using a concept similar approach integrated with the ontological similarity process to overcome the issue of multiple occurrences. The study evaluates how the proposed approach matches the semantically similar matching concepts. The proposed LSTM-CNN model is based on the process which extracts relevant features from the datasets using the CNN model and recommendations are generated using LSTM based on the learner's preference and interests. It generates recommendations based on the user profile and provides useful guidance in the searching process on a Social Web. The performance of the proposed approach was validated with respect to different performance metrics such as precision, recall, F1-score and support. In addition, the performance of the model was also compared with existing models to validate the potential of the LSTM-CNN model. Simulation results validate the effectiveness of the proposed approach. It can be observed that the proposed approach outperforms the existing model in terms of accuracy, and precision. Also, the proposed model achieves superior classification performance which validates its potential of being a better classifier. For future work, the study intends to verify the performance of the proposed approach for different Social Web applications.

References

1. M. Ghenname, M. Abik, R. Ajhoun, J. Subercaze, C. Gravier, F. Laforest, Personalized recommendation-based hashtags on e-learning systems, in *2013 3rd International Symposium ISKO-Maghreb* (IEEE, 2013), pp. 1–8
2. H.H. Kim, A personalized recommendation method using a tagging ontology for a social e-learning system, in *Asian Conference on Intelligent Information and Database Systems* (Springer, Berlin, Heidelberg, 2011), pp. 357–366
3. M.R. Bouadjenek, H. Hacid, M. Bouzeghoub, Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms. Inf. Syst. **56**, 1–18 (2016)
4. A. Klašnja-Milićević, M. Ivanović, B. Vesin, Z. Budimac, Enhancing e-learning systems with personalized recommendation based on collaborative tagging techniques. Appl. Intell. **48**(6), 1519–1535 (2018)

5. H. Li, H. Li, S. Zhang, Z. Zhong, J. Cheng, Intelligent learning system based on personalized recommendation technology. Neural Comput. Appl. **31**(9), 4455–4462 (2019)
6. R. Kim, L. Olfman, T. Ryan, E. Eryilmaz, Leveraging a personalized system to improve self-directed learning in online educational environments. Comput. Educ. **70**, 150–160 (2014)
7. X. Shen, B. Tan, C. Zhai, Implicit user modeling for personalized search, in *Proceedings of the 14th ACM International Conference on Information and knowledge management* (2005), pp. 824–831
8. Z.K. Zhang, T. Zhou, Y.C. Zhang, Tag-aware recommender systems: a state-of-the-art survey. J. Comput. Sci. Technol. **26**(5), 767–777 (2011)
9. Y. Zhou, C. Huang, Q. Hu, J. Zhu, Y. Tang, Personalized learning full-path recommendation model based on LSTM neural networks. Inf. Sci. **444**, 135–152 (2018)
10. H.K. Han, E.H. Suh, Personalized recommendation system for instructing students using multiple machine learning. Int. Inf. Inst. (Tokyo). Inf. **21**(1), 67–74 (2018)
11. L. Tian, B. Yang, X. Yin, Y. Su, A survey of personalized recommendation based on machine learning algorithms, in *Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering* (2020), pp. 602–610
12. X. Wang, J. Wen, F. Luo, W. Zhou, H. Ren, Personalized recommendation system based on support vector machine and particle swarm optimization, in *The* International *Conference on Knowledge Science, Engineering and Management* (Springer, Cham, 2015), pp. 489–495
13. X. Sha, Z. Sun, J. Zhang, Attentive knowledge graph embedding for personalized recommendation. arXiv preprint (2019). arXiv:1910.08288
14. M.J. Er, Y. Zhang, N. Wang, M. Pratama, Attention pooling-based convolutional neural network for sentence modelling. Inf. Sci. **373**, 388–403 (2016)
15. C. Zhang, T. Li, Z. Ren, Z. Hu, Y. Ji, Taxonomy-aware collaborative denoising autoencoder for personalized recommendation. Appl. Intell. **49**(6), 2101–2118 (2019)
16. L. Yang, Y. Zheng, X. Cai, H. Dai, D. Mu, L. Guo, T. Dai, A LSTM based model for personalized context-aware citation recommendation. IEEE Access **6**, 59618–59627 (2018)
17. Z. Yu, J. Lian, A. Mahmoody, G. Liu, X. Xie, Adaptive user modeling with long and short-term preferences for personalized recommendation, in *IJCAI* (2019), pp. 4213–4219
18. T. Dai, L. Zhu, Y. Wang, K.M. Carley, Attentive stacked denoising autoencoder with bi-lstm for personalized context-aware citation recommendation. IEEE/ACM Trans. Audio Speech Lang. Process. **28**, 553–568 (2019)
19. H. Wang, N. Lou, Z. Chao, A personalized movie recommendation system based on LSTM-CNN, in *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)* (IEEE, 2020), pp. 485–490
20. K. Souali, A. El Afia, R. Faizi, R. Chiheb, A new recommender system for e-learning environments, in *2011 International Conference on Multimedia Computing and Systems* (IEEE, 2011), pp. 1–4
21. A. Klašnja-Milićević, M. Ivanović, A. Nanopoulos, Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions. Artif. Intell. Rev. **44**(4), 571–604 (2015)
22. S. Wan, Z. Niu, A hybrid e-learning recommendation approach based on learners' influence propagation. IEEE Trans. Knowl. Data Eng. **32**(5), 827–840 (2019)
23. M. Erdt, A. Fernandez, C. Rensing, Evaluating recommender systems for technology enhanced learning: a quantitative survey. IEEE Trans. Learn. Technol. **8**(4), 326–344 (2015)
24. M. Hassan, M. Hamada, Smart media-based context-aware recommender systems for learning: a conceptual framework, in *2017 16th International Conference on Information Technology Based Higher Education and Training (ITHET)* (IEEE, 2017), pp. 1–4
25. C.C. Aggarwal, Content-based recommender systems, in *Recommender systems* (Springer, Cham, 2016), pp. 139–166
26. Z. Tan, L. He, An efficient similarity measure for user-based collaborative filtering recommender systems inspired by the physical resonance principle. IEEE Access **5**, 27211–27228 (2017)

27. T. Di Noia, C. Magarelli, A. Maurino, M. Palmonari, A. Rula, Using ontology-based data summarization to develop semantics-aware recommender systems, in *The European Semantic Web Conference* (Springer, Cham, 2018), pp. 128–144
28. S. Sarwar, Z.U. Qayyum, R. García-Castro, M. Safyan, R.F. Munir, Ontology based e-learning framework: a personalized, adaptive and context aware model. Multimedia Tools Appl. **78**(24), 34745–34771 (2019)
29. F. Grivokostopoulou, I. Perikos, M. Paraskevas, I. Hatzilygeroudis, An ontology-based approach for user modelling and personalization in e-learning systems, in *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)* (IEEE, 2019), pp. 1–6
30. Z. Batmaz, A. Yurekli, A. Bilge, C. Kaleli, A review on deep learning for recommender systems: challenges and remedies. Artif. Intell. Rev. **52**(1), 1–37 (2019)
31. G. Zhang, Y. Liu, X. Jin, A survey of autoencoder-based recommender systems. Front. Comp. Sci. **14**(2), 430–450 (2020)
32. Z. Xu, O. Tifrea-Marciuska, T. Lukasiewicz, M.V. Martinez, G.I. Simari, C. Chen, Lightweight tag-aware personalized recommendation on the social web using ontological similarity. IEEE (2018). https://doi.org/10.1109/ACCESS.2018
33. P. Ganesan, H. Garcia-Molina, J. Widom, Exploiting hierarchical domain structure to compute similarity. ACM Trans. Inf. Syst. (TOIS) **21**(1), 64–93 (2003)
34. Y. Li, Z.A. Bandar, D. McLean, An approach for measuring semantic similarity between words using multiple information sources. IEEE Trans. Knowl. Data Eng. **15**(4), 871–882 (2003)

Prediction of Swine Flu (H1N1) Virus Using Data Mining and Convolutional Neural Network Techniques

Pilla Srinivas, Debnath Bhattacharyya, and Divya MidhunChakkaravarthy

Abstract Swine flu is a transmitting virus which causes risk to health of humans, and it damages all functioning systems and gives rise to bare high-risk symptoms that which are harmful to community where effected person lives and works. Data mining plays a significant role in predicting diseases. The database report of medical patient is not more efficient; currently, we made an endeavor to detect the most widely spread disease all over the world named swine flu. Swine flu is a respiratory disease which has numeral number of tests and must be requisite from the patient for detecting a disease. Advanced data mining techniques give us help to remedy this situation. This paper describes about a prototype using data mining techniques, namely Naive Bayesian classifier. The data mining is an emerging research trend which helps in finding accurate solutions in many fields. This paper highlights the various data mining techniques and convolutional neural network techniques used for predicting swine flu diseases.

Keywords Data mining · Swine flu disease · Naive Bayes classifier · Convolutional neural network

P. Srinivas (✉) · D. MidhunChakkaravarthy
Department of Computer Science and Multimedia, Lincoln University College, Kuala Lumpur, Malaysia
e-mail: ao@diet.edu.in

D. Bhattacharyya
Department of Computer Science and Engineering, K L Deemed to be University, KLEF, Guntur 522502, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_51

1 Introduction

Swine flu is one of the most infectious diseases which contains the type of virus that causes thousands of deaths per year. Hence, we have used a method of data mining to predict this disease using Naive Bayesian classifier; thus, we can decrease the laboratory test cost as well as time also. "How can we turn useful data into information that can enable healthcare practitioners to make intelligent clinical decisions?" This is the main purpose for this paper.

Swine flu is a respiratory disease caused by influenza viruses that infect the respiratory tract of pigs and result in a barking cough, decreased appetite, nasal secretions, and listless behavior; the virus can be transmitted to humans. Swine flu viruses may mutate or change so that they are easily transmissible among humans. Swine influenza is also called pig influenza or swine flu, hog flu, and pig flu, which is an infection caused by any one of the several types of swine influenza viruses. Swine influenza virus (SIV) or swine origin influenza virus (S-OIV) is any strain of the influenza family of viruses that is endemic in pigs. In 2009, the known influenza is included in strain, and there are some subtypes of influenza A known as H1N1, H1N2, H2N1, H3N1, H3N2, and H2N3. Swine influenza virus is common throughout and comes from pig populations.

The virus transmitted from pigs to humans is not common and does not always tend toward human flu, often resulting only in the productivity of antibodies in the blood of human. If transmission does cause human flu, it is known as zoonotic swine flu. People who regularly come in contact with pigs are at increased risk of swine flu infection. Around the mid-twentieth century, the checking of influenza subtypes became possible, allowing exactly right diagnosis of transmission to humans. Since then, only 50 such transmissions have been confirmed. Swine flu is transmitted from person to person by inhalation or ingestion of droplets containing virus from people sneezing or coughing; it is not transmitted by eating cooked pork products. The newest swine flu virus that has caused swine flu is influenza A H3N2v.

Some common symptoms of swine flu are as follows:

- Cough
- Fever
- Sore throat
- Runny nose
- Headache
- Chill
- Fatigue
- Nausea.

Statistics of swine flu

The cause of the 2009 swine flu was an influenza A virus type designated as H1N1. In 2011, a new swine flu virus detected with new strain was named influenza A (H3N2)v. Only a few people (mainly children) were first infected, but officials from the US

Centers for Disease Control and Prevention (CDC) reported increased numbers of people infected in the 2012–2013 flu season. Currently, the swine flu virus kills more than 1500 people in India, and it will be rapidly spreading all over the world; so recently, there are not noticed large number of people infected with H3N2v. Unfortunately, another virus termed H3N2 (note no "v" in its name) has been found and caused flu, but this strain is different from H3N2v. In general, all of the influenza A viruses have a structure similar to the H1N1 virus; each type has a somewhat different H and/or N structure.

2 Literature Review

Thakkar, Hasan, and Desai were inspired to do this research by the study of mortality rate of swine flu. The paper focuses on the aspect of medical diagnosis by learning predictions through the data collected for swine flu. The authors proposed a method to identify swine flu by studying 110 symptoms to decrease the cost incurred in the test of the disease. The author developed prototype intelligent swine flu prediction software; they used Naive Bayes classifier technique for classifying the patients of swine flu. Based on the possibility of the disease and guaranteed the accuracy of almost 63.3%. In August 2010, the World Health Organization declared the swine flu pandemic officially over. Cases of swine flu have been reported in India, with over 25,000 positive test cases and 1370 deaths till March 2015.

3 Techniques Used

Concepts used are as follows:

A. Data mining
B. Convolutional neural networks

A. Data Mining

- The cause of the 2009 swine flu was an influenza A virus type designated as H1N1. In 2011, a new swine flu virus detected with new strain was named influenza A (H3N2)v. Only a few people (mainly children) were first infected, but officials from the US Centers for Disease Control and Prevention (CDC) reported increased numbers of people infected in the 2012–2013 flu season. Currently, the swine flu virus kills more than 1500 people in India, and it will be rapidly spreading all over the world; so recently, there are not notice large numbers of people infected with H3N2v. Unfortunately, another virus termed H3N2 (note no "v" in its name) has been found and caused flu, but this strain is different from H3N2v. In general, all of the influenza A viruses have

a structure similar to the H1N1 virus; each type has a somewhat different H and/or N structure.

- Technically, data mining is the process of finding correlations or patterns among different fields in large relational databases. The step of the "Knowledge Discovery in Databases or KDD" process, an interdisciplinary subfield of computer science, is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use (Fig. 1).

B. Convolutional Neural Networks

- The convolutional neural network (CNN) is a class of deep learning neural networks. CNNs represent a huge breakthrough in image recognition. They are most commonly used to analyze visual imagery and are frequently working behind the scenes in image classification. They can be found at the core of everything from Facebook's photo tagging to self-driving cars.

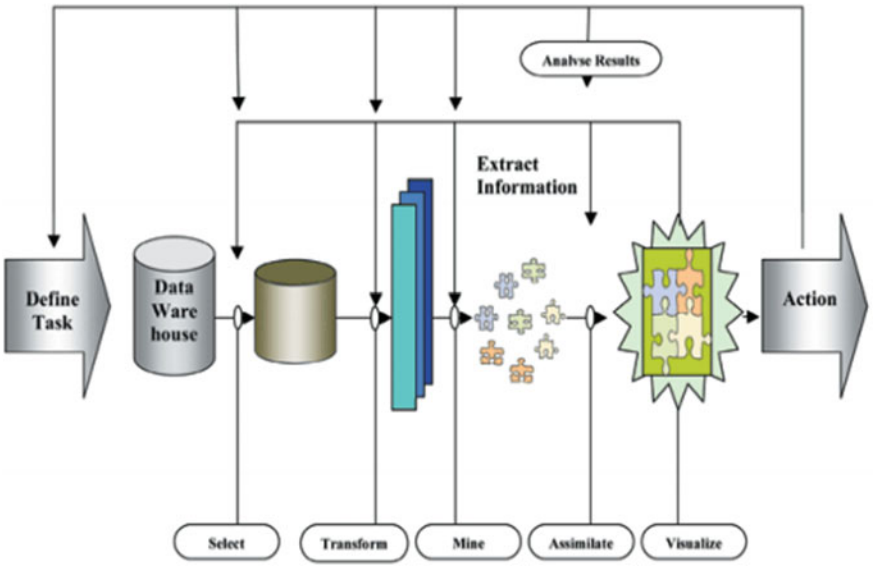


Fig. 1 Working of data mining

They are working hard behind the scenes in everything from health care to security.

- A CNN has

1 Convolutional layers

2 ReLU layers

3 Pooling layers

4 A fully connected layer.

- A classic CNN architecture would look something like this:

  Input ⇒ Convolution ⇒ ReLU ⇒ Convolution ⇒ ReLU ⇒ Pooling ⇒ ReLU ⇒ Convolution ⇒ ReLU ⇒ Pooling ⇒ Fully Connected
- A CNN convolves (not convolutes…) learned features with input data and uses 2D convolutional layers. This means that this type of network is ideal for processing 2D images. Compared to other image classification algorithms, CNNs actually use very little preprocessing. This means that they can learn the filters that have to be handmade in other algorithms. CNNs can be used in tons of applications from image and video recognition, image classification, and recommender systems to natural language processing and medical image analysis.
- CNNs are inspired by biological processes. They are based on some cool research done by Hubel and Wiesel in the 1960s regarding vision in cats and monkeys. The pattern of connectivity in a CNN comes from their research regarding the organization of the visual cortex. In a mammal's eye, individual neurons respond to visual stimuli only in the receptive field, which is a restricted region. The receptive fields of different regions partially overlap so that the entire field of vision is covered (Fig. 2).

CNNs have an input layer, output layer, and hidden layers. The hidden layers usually consist of convolutional layers, ReLU layers, pooling layers, and fully connected layers.

- Convolutional layers apply a convolution operation to the input. This passes the information on to the next layer.
- Pooling combines the outputs of clusters of neurons into a single neuron in the next layer.
- Fully connected layers connect every neuron in one layer to every neuron in the next layer.

In a convolutional layer, neurons only receive input from a subarea of the previous layer. In a fully connected layer, each neuron receives input from every element of the previous layer.

A CNN works by extracting features from images. This eliminates the need for manual feature extraction. The features are not trained. They are learned while the

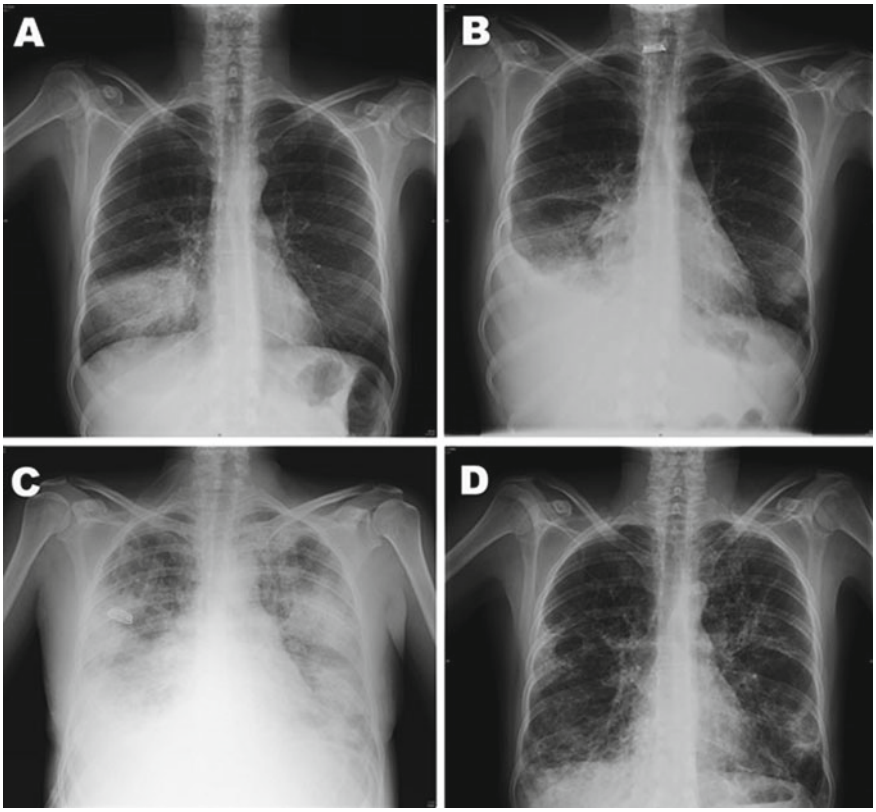


Fig. 2 X-ray image recognition

network trains on a set of images. This makes deep learning models extremely accurate for computer vision tasks. CNNs learn feature detection through tens or hundreds of hidden layers. Each layer increases the complexity of the learned features.

GIF via GIPHY

A CNN

- Starts with an input image
- Applies many different filters to it to create a feature map
- Applies a ReLU function to increase nonlinearity
- Applies a pooling layer to each feature map
- Flattens the pooled images into one long vector
- Inputs the vector into a fully connected artificial neural network
- Processes the features through the network. The final fully connected layer provides the "voting" of the classes that we are after

- Trains through forward propagation and back propagation for many and many epochs respectively. This repeats until we have a well-defined neural network with trained weights and feature detectors.

a. **Very Deep Convolutional Networks for Large-Scale Image Recognition (VGG-16)**

The VGG-16 is one of the most popular pre-trained models for image classification. Introduced in the famous ILSVRC 2014 Conference, it was and remains the model to beat even today. Developed at the Visual Graphics Group at the University of Oxford, VGG-16 beats the then standard of AlexNet and was quickly adopted by researchers and the industry for their image classification tasks.

As you can see, the model is sequential in nature and uses lots of filters. At each stage, small $3 * 3$ filters are used to reduce the number of parameters, and all the hidden layers use the ReLU activation function. Even then, the number of parameters is 138 billion—which makes it a slower and much larger model to train than others.

Additionally, there are variations of the VGG-16 model, which are basically improvements to it, like VGG-19 (19 layers). You can find a detailed explanation (Figs. 3 and 4).

The following are the layers of the model:

- Convolutional layers = 13
- Pooling layers = 5
- Dense layers = 3.

Let us explore the layers in detail:

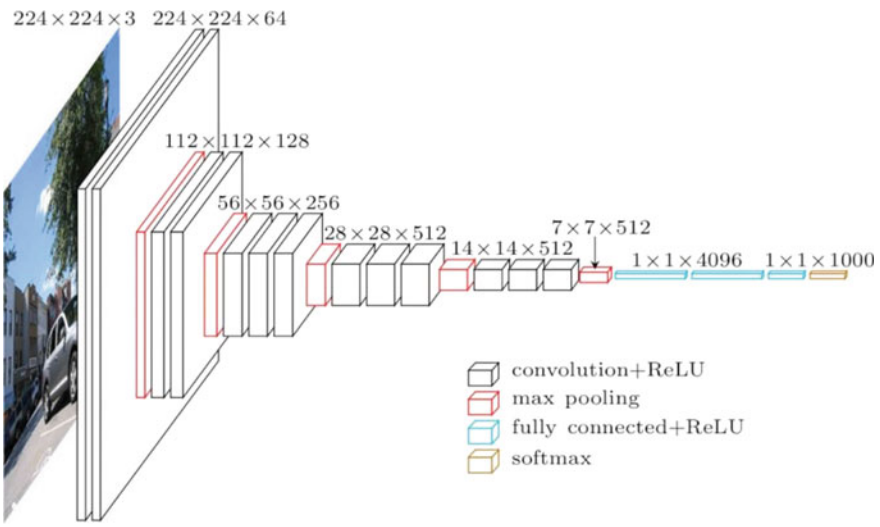


Fig. 3 Architecture of VGG-16 model

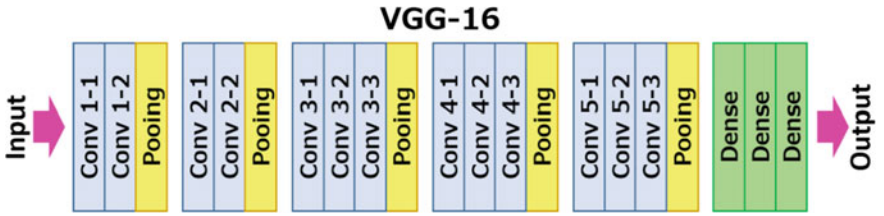


Fig. 4 Intuitive model of VGG-16 model

1. **Input:** Image of dimensions (224, 224, 3).
2. **Convolution Layer Conv1:**
   - Conv1-1: 64 filters
   - Conv1-2: 64 filters and max pooling
   - Image dimensions: (224, 224)
3. **Convolution Layer Conv2:** Now, we increase the filters to 128
   - Input image dimensions: (112, 112)
   - Conv2-1: 128 filters
   - Conv2-2: 128 filters and max pooling
4. **Convolution Layer Conv3:** Again, double the filters to 256, and now add another convolution layer
   - Input image dimensions: (56, 56)
   - Conv3-1: 256 filters
   - Conv3-2: 256 filters
   - Conv3-3: 256 filters and max pooling
5. **Convolution Layer Conv4:** Similar to Conv3, but now with 512 filters
   - Input image dimensions: (28, 28)
   - Conv4-1: 512 filters
   - Conv4-2: 512 filters
   - Conv4-3: 512 filters and max pooling
6. **Convolution Layer Conv5:** Same as Conv4
   - Input image dimensions: (14, 14)
   - Conv5-1: 512 filters
   - Conv5-2: 512 filters
   - Conv5-3: 512 filters and max pooling
   - The output dimensions here are (7, 7). At this point, we flatten the output of this layer to generate a feature vector.
7. **Fully connected/dense FC1:** 4096 nodes, generating a feature vector of size (1, 4096)

8. **Fully connected/dense FC2:** 4096 nodes generating a feature vector of size (1, 4096)
9. **Fully connected/dense FC3:** 4096 nodes, generating 1000 channels for 1000 classes. This is then passed on to a softmax activation function.
10. **Output layer**

b. **Inception**

At only 7 million parameters, it was much smaller than the then prevalent models like VGG and AlexNet. Adding to it a lower error rate, you can see why it was a breakthrough model. Not only was this, but the major innovation in this paper is also another breakthrough—the inception module.

As can be seen, in simple terms, the inception module just performs convolutions with different filter sizes on the input, performs max pooling, and concatenates the result for the next inception module. The introduction of the 1 * 1 convolution operation reduces the parameters drastically (Figs. 5 and 6).

The Inceptionv2 model was a major improvement on the Inceptionv1 model which increased the accuracy and further made the model less complex. In the same paper as Inceptionv2, the authors introduced the Inceptionv3 model with a few more improvements on v2 (Fig. 7).

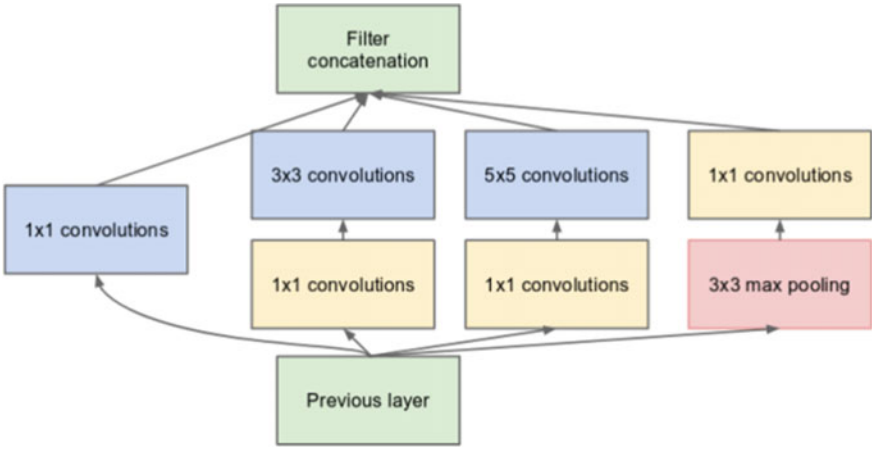


Fig. 5 Inception module with dimension reductions

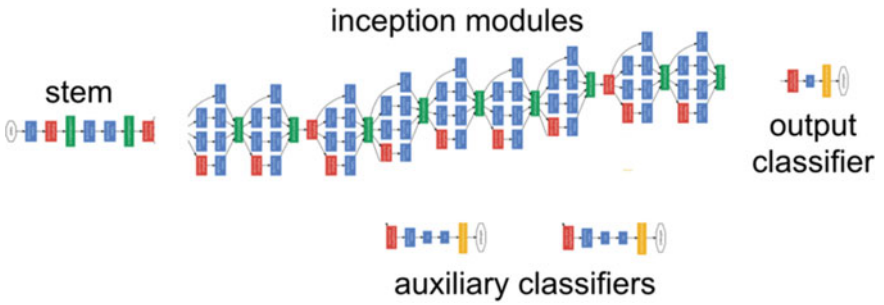


Fig. 6 Inceptionv2 model

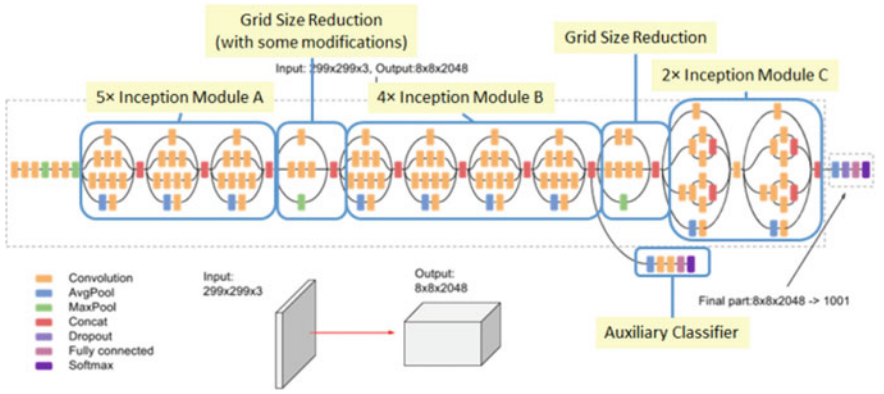


Fig. 7 Inceptionv3 model

The following are the major improvements included:

- Introduction of batch normalization
- More factorization
- RMSProp optimizer.

iii. **ResNet50**

Just like Inceptionv3, ResNet50 is not the first model coming from the ResNet family. The original model was called the residual net or ResNet and was another milestone in the CV domain back in 2015.

The main motivation behind this model was to avoid poor accuracy as the model went on to become deeper. Additionally, if you are familiar with gradient descent, you would have come across the vanishing gradient issue—the ResNet model aimed to tackle this issue as well. Here is the architecture of the earliest variant: ResNet34 (ResNet50 also follows a similar technique with just more layers) (Table 1).

Table 1 Architecture of ResNet family in terms of layers

Layer name	Output size	18-layer	34-layer	50-layer	101-layer	152-layer
Conv1	112 × 112	7 × 7, 64, stride 2				
Conv2..x	56 × 56	3 × 3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
Conv3..x	28 × 28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$
Conv4..x	14 × 14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 36$
Conv52..x	7 × 7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$
	1 × 1	Average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

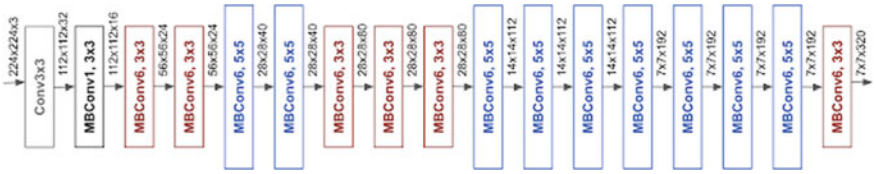


Fig. 8 EfficientNet model

iv. **EfficientNet**

We finally come to the latest model among these four that have caused waves in this domain, and of course, it is from Google. In EfficientNet, the authors propose a new scaling method called compound scaling. The long and short of it is this: The earlier models like ResNet follow the conventional approach of scaling the dimensions arbitrarily and by adding up more and more layers.

However, the paper proposes that if we scale the dimensions by a fixed amount at the same time and do so uniformly, we achieve much better performance. The scaling coefficients can be in fact decided by the user.

Though this scaling technique can be used for any CNN-based model, the authors started off with their own baseline model called EfficientNetB0 (Fig. 8).

MBConv stands for mobile inverted bottleneck convolution (similar to MobileNetv2). They also propose the compound scaling formula with the following scaling coefficients:

- Depth = 1.20
- Width = 1.10
- Resolution = 1.15.

This formula is used to again build a family of EfficientNets–EfficientNetB0 to EfficientNetB7.

The following is a simple graph showing the comparative performance of this family vis-a-vis other popular models (Fig. 9):

4 Theorems Used

Here we use two types of theorems:

A. Bayes theorem
B. Naive Bayesian classifier

A. *Bayes Theorem*

Bayes theorem is a simple mathematical formula used for calculating conditional probabilities.

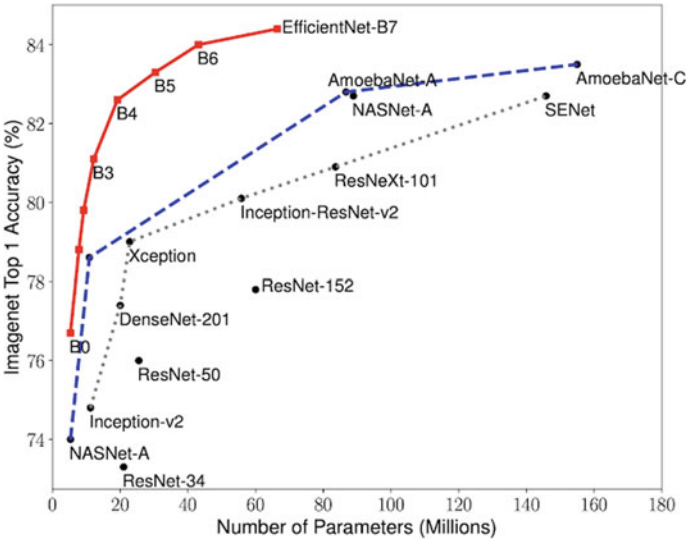


Fig. 9 Simple graph showing the comparative performance of this family vis-a-vis other popular models

Conditional probability is a measure of the probability of an event occurring given that another event has (by assumption, presumption, assertion, or evidence) occurred.

The formula is

$$P\left(A/B\right) = \frac{P\left(B/A\right).P(A)}{P(B)}$$

which tells us how often A happens given that B happens, written $P(A|B)$ also called posterior probability, when we know: how often B happens given that A happens, written $P(B|A)$ and how likely A is on its own, written as $P(A)$ and how likely B is on its own, written as $P(B)$.

In simpler terms, Bayes theorem is a way of finding a probability when we know certain other probabilities.

B. *Naive Bayesian Classifier*

Example: Given all the patients we have seen their symptoms and diagnosis (Table 2)

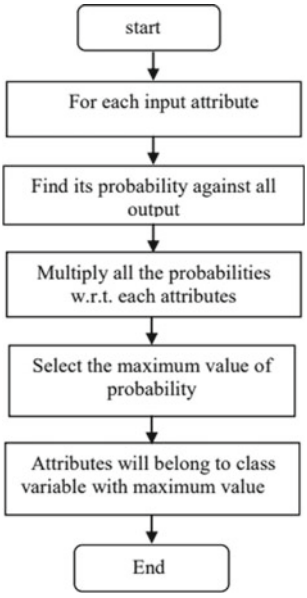
If there is a new entry of patient having following symptoms listed below in Table 3, so do we believe that a patient with following symptoms has the swine flu?

Table 2 Symptoms and diagnosis

Chill	Runny nose	Headache	Fever	Swine flu
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

Table 3 New entry in patients in the following symtoms

Chill	Runny nose	Headache	Fever	Swine flu
Y	N	Mild	N	?



So here we have seen above in figure we compute all possible individual probabilities conditioned on the target attribute of swine flu containing all probabilities attribute of swine flu

There is the flowchart of Naive Bayes classifier algorithm; by using this flowchart, we can easily conclude that the patient has been suffered from swine flu disease or not; this we will test from the incoming attribute which belongs to class variable with

maximum value. First, we compute all possible individual probabilities conditioned on the target attribute of swine flu containing all probabilities attribute of swine flu.

P (swine flu Y) = 5/8 = 0.625.

P (chills = Y | swine flu = Y) = 3/5 = 0.6.

P (swine flu = N) = 3/8 = 0.375.

P (chills = Y | swine flu = N) = 1/3 = 0.333.

Just like above, we can simply compute the possible probabilities for all condition, and these probabilities are enlisted in Table 3, and then we decide that p has split up into two cases: one for Y and second for $P1 \Rightarrow$ argmax P (swine flu = Y) * P (chills = Y | swine flu = Y) * P (runnynose = N | swine flu = Y) * P (headache = Mild | swine flu = Y) * P (fever = N | swine flu = Y)

= 0.625 * 0.6 * 0.2 * 0.4 * 0.2 = 0.006

$P1 = 0.006$

$P2 \Rightarrow$ argmax P (swine flu = Y) * P (chills = Y | swine flu = N) * P (runnynose = N | swine flu = N) * P (headache = Mild | swine flu = N) * P (fever = N | swine flu = N)

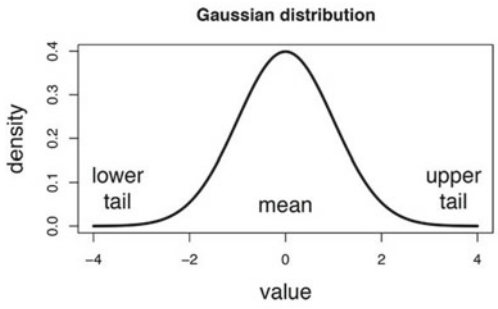
= 0.375 * 0.333 * 0.666 * 0.333 * 0.666

$P2 = 0.0185$

Therefore, the argument of probability of $P2$ seems maximum than $P1$ so that patient is not having the swine flu.

Types of Naive Bayes classifier:

- **Multinomial Naive Bayes:** Feature vectors represent the frequencies with which certain events have been generated by a multinomial distribution. This is the event model typically used for document classification.
- **Bernoulli Naive Bayes:** In the multivariate Bernoulli event model, features are independent Booleans (binary variables) describing inputs. Like the multinomial model, this model is popular for document classification tasks, where binary term occurrence (i.e., a word occurs in a document or not) features are used rather than term frequencies (i.e., frequency of a word in the document).
- **Gaussian Naive Bayes:** In Gaussian Naive Bayes, continuous values associated with each feature are assumed to be distributed according to a Gaussian distribution (normal distribution). When plotted, it gives a bell-shaped curve which is symmetric about the mean of the feature values as shown below:



The likelihood of the features is assumed to be Gaussian. Hence, conditional probability is given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Now, what if any feature contains numerical values instead of categories in the Gaussian distribution.

One option is to transform the numerical values to their categorical counterparts before creating their frequency tables. The other option, as shown above, could be using the distribution of the numerical variable to have a good guess of the frequency. For example, one common method is to assume normal or Gaussian distributions for numerical variables.

The probability density function for the normal distribution is defined by two parameters (mean and standard deviation).

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{Mean}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2} \quad \textit{S}\text{tandard Deviation}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{Normal distribution}$$

Text Preprocessing

The next step is to preprocess text before splitting the data set into a train and test set. The preprocessing steps involve removing numbers, removing punctuations in a string, removing stop words, stemming of words, and lemmatization of words.

Constructing a Naive Bayesian Classifier

Combine all the preprocessing techniques and create a dictionary of words and each word's count in training data.

1. Calculate probability for each word in a text and filter the words which have a probability less than threshold probability. Words with probability less than threshold probability are irrelevant.
2. Then for each word in the dictionary, create a probability of that word being in insincere questions and its probability insincere questions and then find the conditional probability to use in Naive Bayesian classifier.
3. Prediction using conditional probabilities.

5 Conclusion

The data mining technique can be used in collaboration with a Naive Bayesian classifiers algorithm which is used in diagnosing swine flu disease. The proposed approach showed promising results which may lead to further attempts to utilize information technology for diagnosing patients for swine flu diseases. Here we used Naïve Bayesian classification rules which are easy to interpret. In the future, we will try to get more the accuracy for the swine flu disease patient by increasing the various parameters suggested from the doctors by using different data mining techniques. The artificial intelligent system for efficient swine flu detection is tested and found to be genuine in detection compared to manual predictions. As time is progressing, the neural networks as well as the techniques for object detection are also progressing rapidly. By using Naive Bayesian classifier and CNN techniques, we can predict swine flu very quickly.

Bibliography

1. The Swine Influenza. http://en.wikipedia.org/wiki/Swine_influenza
2. The Swine Influenza. http://www.medicinenet.com/swine_flu/article.htm
3. Thakkar, Hasan, Desai, Health care decision support system for swine flu prediction using Naive Bayes classifier, in *International Conference on Advances in Recent Technologies in Communication and Computing* (IEEE, 2010)
4. WHO. www.tbevidence.org/documents/dxres/models/tb_diagnostics.pdf. World Health Organization
5. Swine flu paper. http://www.ijarcsse.com/docs/papers/Volume_4/6_June2014/V4I6-0235.pdf
6. G. Subbalakshmi, M.C. Rao, Decision support in heart disease prediction system using Naïve Bayes. IJCSE Indian J. Comput. Sci. Eng. ISSN: 0976-5166
7. I. Kononenko, Machine learning for medical diagnosis: history, state of the art and perspective
8. H. Wang et al., Medical knowledge acquisition through data mining, in *Proceedings of 2008 IEEE International Symposium on IT in Medicine and Education* 978-1-4244-2511-2/
9. H. Kaur, S.K. Wasan, Empirical study on applications of data mining techniques in healthcare. J. Comput. Sci. **2**(2), 194–200 (2006)
10. Health care decision support system for swine flu prediction using Naive Bayes classifier, in *Artcom International Conference*, 978-1-4244-8093-7
11. IBM, Data Mining Techniques. [Online] Available http://www.ibm.com/developerworks/opensource/library/ba-dataminingtechniques/index.html?ca=drs-. Downloaded on 04 Apr 2013

Correlation Analysis of Cognitive Regions in Automated Anatomical Labeling Atlas Using LSTM

Latha Gadepaka and Vinuthna Lingabathina

Abstract The main objective of our work in this paper is to infer musicianship of 36 participants (18 musicians and 18 non-musicians) who are continuously listening to music stimuli of three different genres using LSTM architecture, as LSTM works well for sequential data. Previous studies state that the region-wise classification is done by regression, we adapted LSTM in our study for region-wise classification and performed fivefold cross validation and came up with better accuracy scores than regression. We considered automated anatomical labeling (AAL) atlas regions which discriminated well and also found the correlation that exists between the regions using Pearson correlation coefficient. We also investigated in detail the cognitive functionality of the brain region and also the association that exists between them.

Keywords Classification · Correlation · LSTM · Pearson correlation coefficient

1 Introduction

Introduction to LSTM

LSTM networks are well-suited to classify, process and make predictions based on time series data [1]. LSTM unit consists of cell state, input gate, output gate and forget gate. The input, output and forget gate regulate the flow of information. At time step t, the input vector is x_t, input gate i_t controls how much each unit is updated, the output gate o_t controls the exposure of the internal memory state, and the forget gate f_t controls the amount of which each unit of the memory cell is erased. The memory cell c_t keeps the useful history information which will be used for the next process [2]. h_t is internal hidden state vector which is the function of previous hidden state h_{t-1} and current input x_t. Figure 1 describes LSTM unit.

L. Gadepaka (✉)
CMR Technical Campus, Hyderabad, India
e-mail: latha.cse@cmrtc.ac.in

V. Lingabathina
St. Peter's Engineering College, Hyderabad, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_52

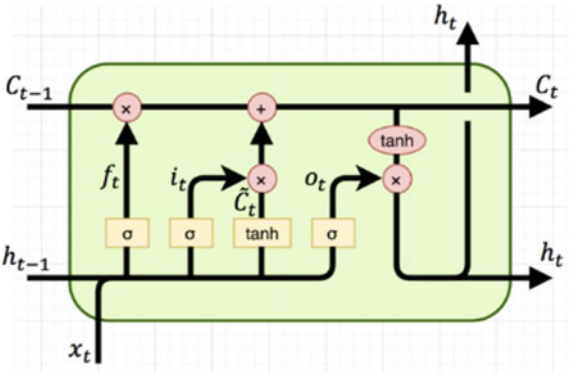


Fig. 1 LSTM unit

Pearson Correlation Coefficient

This is a measure of linear correlation between two sets of data. A coefficient of correlation is generally applied in statistics to calculate a relationship between two variables. The correlation shows a specific value of the degree of a linear relationship between the X and Y variables, say X and Y. There are various types of correlation coefficients. However, Pearson's correlation (also known as Pearson's R) is the correlation coefficient.

Karl Pearson's coefficient of correlation is the one in which the numerical representation is applied to measure the level of relation between linearly related variables. The coefficient of correlation is expressed by "r". The value of correlation coefficient lies between -1 and 1. When a correlation coefficient is 1, that means for every increase in one variable say X, there is a positive increase in the other fixed proportion say Y. When a correlation coefficient is -1, that means for every positive increase in one variable say X, there is a negative decrease in the other fixed proportion say Y. When a correlation coefficient is 0 for every increase that means there is no positive or negative increase, and the two variables are not related.

$$r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2} \sqrt{(Y - \overline{Y})^2}}$$

where,

$\overline{X}$ = mean of X variable

$\overline{Y}$ = mean of Y variable

Background Work

FMRI data was re-analyzed from a previously published dataset of 18 musicians and 18 no musicians [1, 3–5] who were scanned during continuous listening to pieces of real music related to three distinct musical genres. Musical features were

extracted computationally from the music pieces and used to encode regional neural activations in each participant. Log-likelihood ratios between musician and non-musician models describing neural activation patterns arising from musical feature processing were used as input to a classifier that was trained to infer the musicianship of the participants.

Model training and testing were done using cross validation to avoid overfitting and to obtain a realistic estimate of the model performance on novel participants [6]. The present study we replaced the linear regression with long short-term memory (LSTM), artificial neural network used in field of deep learning. LSTM is preferred over the linear regression as it can handle long sequences of data efficiently.

2 Proposed Method

Long Short-Term Memory (LSTM)

The main objective is to introduce long short-term memory (LSTM) over linear regression as LSTM works well for sequential data. LSTM will be able to classify efficiently by learning from the time series data.

Architecture

The model is built to classify the participant as musician or a non-musician based on the time series data which is collected while they are continuously listening to the music. The input is the 714 time points for a region for each participant. The model comprises of two hidden LSTM layers which are meant for learning to classify the participants. The first hidden layer LSTM2 contains 16 nodes and has a recurrent dropout of 0.5. The second hidden layer LSTM2 contains eight nodes and recurrent dropout of 0.5. The second hidden layer is connected to a dense layer with one output node which discriminates musician or non-musician. Binary cross entropy is used as loss function for this model. The hidden state vector of 250 dimensionality of LSTM2 layer is used as input to MLP layer (Fig. 2).

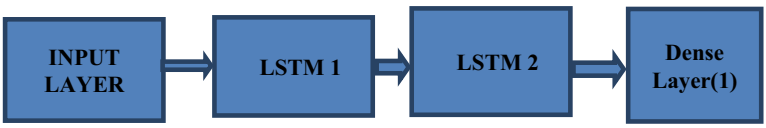


Fig. 2 LSTM architecture

3 Experimentation and Results

Dataset for LSTM Model: 18 musicians and 18 non-musicians were scanned during continuous listening to pieces of real music belonging to three different genres. Six musical features extracted computationally from music pieces are used to encode regional neural activations for each participant. To reduce computational load, dimensionality of fMRI data is reduced by parcellating into 116 whole brain anatomical regions based on automated anatomical labeling atlas (AAL) [6]. For each participant, the averaged BOLD responses were recorded at 714 time points.

- Number of samples = 36(18 musicians and 18 non-musicians)
- Number of features = 6 musical features
- Number of time points for each participant = 714
- Number of ROI's = 116.

Classification Accuracy: Fivefold cross validation is performed on the LSTM architecture and accuracy of each of the 116 brain regions is plotted as shown in Fig. 3. Among the 116 regions, found that 23 regions were able to classify the musician/non-musician with accuracy score greater than 75% (Fig. 3).

Among those 23 regions, nine regions which were matched with the previous work were mentioned in Table 1.

Inter Region Similarity

For all the above-mentioned nine brain regions which individually discriminated musicianship, found Pearson correlation coefficient for every region to the other eight regions for all the participants. We found about 45 correlations between the

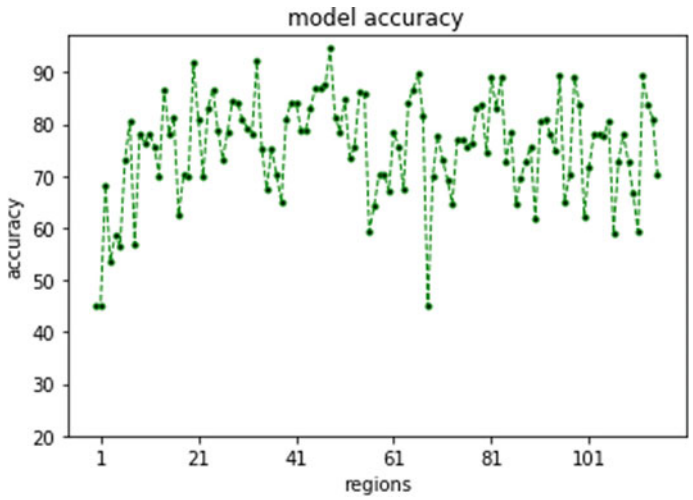


Fig. 3 Accuracy versus AAL atlas regions

Table 1 Accuracy (% mean) for nine regions of AAL atlas

ROI's	Accuracy (% mean)
Caudate nucleus (L)	73.21428657
Caudate nucleus (R)	80.71428657
Middle frontal gyrus, orbital part (L)	78.92857194
Middle frontal gyrus, orbital part (R)	83.21428657
Anterior cingulate and paracingulate gyrus (L)	73.21428657
Anterior cingulate and paracingulate gyrus (R)	78.57142925
Inferior frontal gyrus, opercular part (R)	67.50000119
Inferior frontal gyrus, triangular part (R)	81.07142925
Superior temporal gyrus (R)	80.71428657

regions among which for few correlations the correlation coefficient is > 0.6 is picked and a box plot is plotted for musicians and non-musicians to compare the association between the two regions.

1. **Anterior Cingulate and Paracingulate Gyrus (L) → Anterior Cingulate and Paracingulate Gyrus (R)**

Anterior cingulate gyrus is involved in a number of functions including emotional processing and vocalization of emotions. It has connections with speech and vocalization areas in frontal lobes which controls motor functions**.** It is observed that the average of Pearson correlation coefficient for all the musicians is 0.932 and for non-musicians is 0.93 from which we can conclude that there exists a positive linear association between anterior cingulate and paracingulate gyrus left and right hemispheres (Fig. 4).

2. **Middle Frontal Gyrus, Orbital Part (L) → Middle Frontal Gyrus, Orbital Part (R)**

Middle frontal gyrus is made up about one-third of frontal lobe of human brain. MFG is a convolution located in the frontal lobe between the superior and inferior frontal sulci. It is observed that association between MGF left and right hemisphere for musicians is 0.79 and for non-musicians is 0.72.The association is more in musicians (Fig. 5).

3. **Caudate Nucleus (L) → Caudate Nucleus (R)**

The caudate nucleus is a paired, "C"-shaped subcortical structure which lies deep inside the brain near the thalamus. The caudate nucleus receives topographic visual input from cortical association areas and has receptive fields in the contralateral visual field. The caudate nucleus integrates visual input and inhibits the substantia nigra, disinhibiting the superior colliculus to enable the coordination of eye movement,

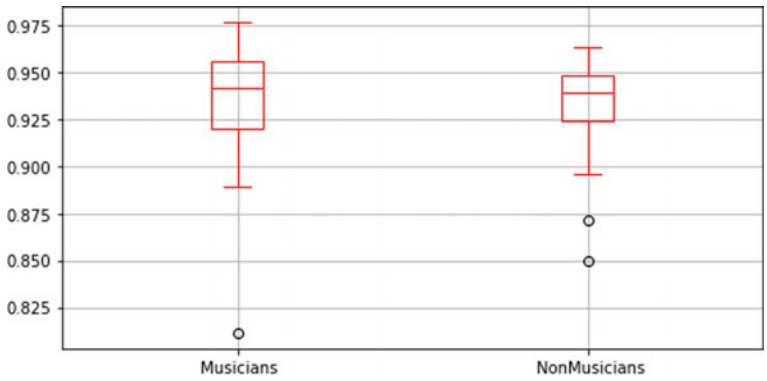


Fig. 4 Correlation between anterior cingulate and paracingulate gyrus (L) and anterior cingulate and paracingulate gyrus (R)

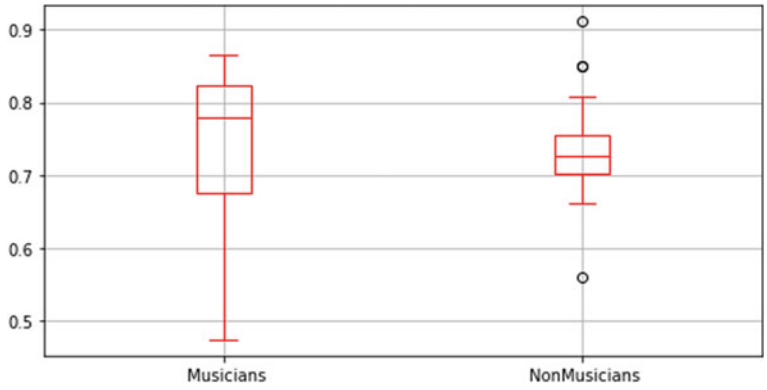


Fig. 5 Correlation between middle frontal gyrus, orbital part (L) → middle frontal gyrus, orbital part (R)

and is important in voluntary saccadic eye movement [7]. It is observed that the correlation between Caudate nucleus in left and right hemisphere is 0.88 for both the musicians and non-musicians. From this, we can say that association learning (connecting visual stimuli with motor responses) [8] which is the key function of caudate nucleus is the same for both trained and untrained individuals (Fig. 6).

4. **Inferior Frontal Gyrus, Opercular Part (R) → Superior Temporal Gyrus (R)**

Frontal operculum with a right hemisphere asymmetry constitutes a sensitive area involved at least partly in processing of musical syntax using chord sequence and melody paradigms. From group, comparisons have determined that bilateral IFGoper together with the anterior rSTG become more strongly activated in musically trained than untrained individuals [9]. Similar to previous work with an extension by finding

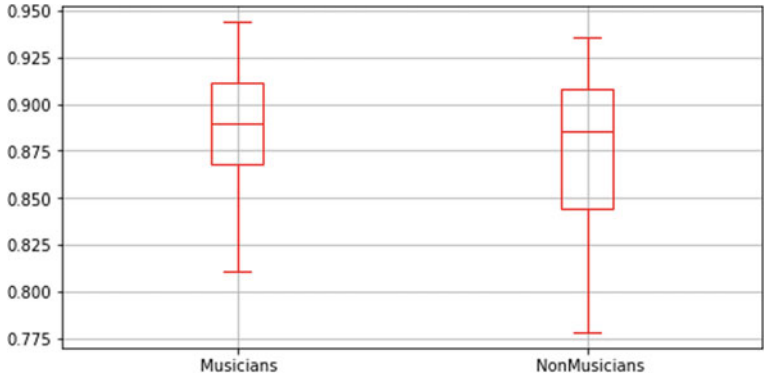


Fig. 6 Correlation between caudate nucleus (L) and caudate nucleus (R)

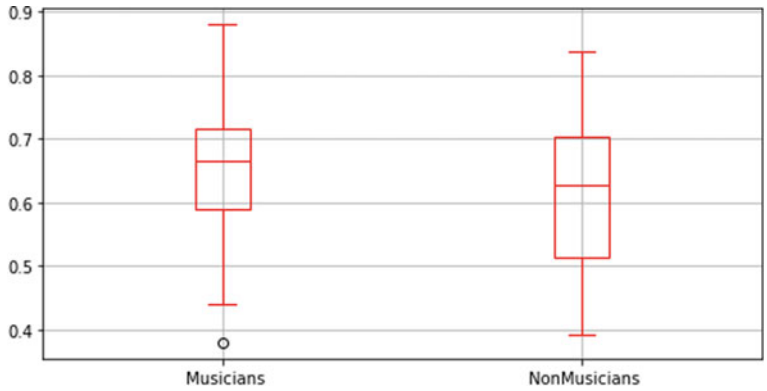


Fig. 7 Correlation between inferior frontal gyrus, opercular part (R) → superior temporal gyrus (R)

the correlation between IFGoper and rSTG, observed that the association between the two regions is slightly more in musicians when compared with non-musicians as depicted in Fig. 7.

4 Conclusions

From the study, we are able to conclude that by using LSTM classification of musician and non-musician is done region wise. We picked up 14 regions which discriminated well musicianship with accuracy. Caudate nucleus (L) discriminated with 73% accuracy. Superior temporal gyrus (R) discriminated with 80% accuracy. In addition to the region-wise classification, also found the Pearson correlation coefficient between

the selected regions and found the strength of correlation and how it varied among musicians and non-musicians.

5 Future Scope

We can keep track of the hidden state vectors when region-wise classification is performed using LSTM architecture. We can develop a multimodal deep learning MLP by feeding the hidden state vectors from the LSTM architecture as input to MLP with associated musicianship class.

References

1. I. Burunat, V. Tsatsishvili, P. Toiviainen, E. Brattico, Coupling of action-perception brain networks during musical pulse processing: evidence from region-of-interest-based independent component analysis. Front. Hum. Neurosci. **11**, 230 (2017)
2. P. Qian, X. Qiu, X. Huang, Bridging LSTM architecture and the neural dynamics during reading
3. I. Burunat et al., Action in perception: prominent visuo-motor functional symmetry in musicians during music listening. PLoS ONE **10**, e0138238 (2015)
4. V. Alluri et al., Musical expertise modulates functional connectivity of limbic regions during continuous music listening. Psychomusicology Music Mind Brain **25**, 443 (2015)
5. V. Alluri et al., Connectivity patterns during music listening: evidence for action-based processing in musicians. Hum. Brain Mapp. **38**, 2955–2970 (2017)
6. P. Saari, I. Burunat, E. Brattico, P. Toiviainen, Decoding musical training from dynamic processing of musical features in the brain
7. O. Hikosaka, Y. Takikawa, R. Kawagoe, Role of the basal ganglia in the control of purposive saccadic eye movements. Physiol. Rev. (2000)
8. M. Yanike, V.P. Ferrera, Representation of outcome risk and action in the anterior caudate nucleus. J. Neurosci. (2014)
9. B. Tillmann et al., Cognitive priming in sung and instrumental music: activation of inferior frontal cortex. Neuroimage **31**, 1771–1782 (2006)
10. https://en.wikipedia.org/wiki/Long_short-term_memory
11. https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Machine Learning-Based DoS Attack Detection Techniques in Wireless Sensor Network: *A Review*

Hanjabam Saratchandra Sharma, Moirangthem Marjit Singh, and Arindam Sarkar

Abstract In diverse areas, the wireless sensor network (WSN) has an important role to play and has its application in various domains. Even though WSN has become very popular in recent times but is also vulnerable to several security threats. The Denial of Service (DoS) attacks and their variant, Distributed Denial of Services (DDoS) attacks are the most commonly observed attacks in WSN. The DoS attacks can hamper the normal functioning of a WSN by repeatedly flooding the communication channel of a sensor node with request messages by attackers to deprive the network of essential services and applications. Therefore, robust techniques to detect DoS attacks in WSNs need to be developed and investigated. Attack detection schemes using machine learning techniques will be a highly promising area of research. A review of several DoS attack detection schemes in WSN using machine learning has been reviewed and presented in the paper.

Keywords Wireless sensor networks · Machine learning · DoS · DDoS

1 Introduction

Wireless sensor network is an assembly of sensor nodes that are scattered throughout a place or a region where involvement of humans in data collection is either considered hazardous or deemed to be unnecessary. The main role of these sensor nodes is to continuously collect data of the environment such as humidity, temperature, air pollution, etc., and forward the collected data to the sink nodes or base station via a series of sensor nodes. Each wireless sensor nodes consist of a processor or a

H. S. Sharma (✉) · M. M. Singh
Department of Computer Science and Engineering, North Eastern Regional Institute of Science and Technology, Papum Pare, Arunachal Pradesh, India
e-mail: sarat.hanjabam@gmail.com

A. Sarkar
Department of Computer Science and Electronics, Ramakrishna Mission Vidyamandira, Howrah, West Bengal, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_53

microcontroller, a transceiver to either receive or send data, a memory to store data, a sensor, and a battery to power up the node.

Even though the data are collected successfully, it is very important to keep the collected data, network components, and entire network free from various attacks. There are various types of attacks such DoS and DDoS attacks [1] that disrupt the normal flow of the network, rendering the network useless, and unfit for data collection. DoS and DDoS attacks are one of the most common type of attack found in WSN [2]. In this paper, a critical review on ML-based techniques to detect DoS and DDoS attacks have provided. Sections 2 and 3 present research works on ML-based DoS and DDoS attack detection methods, respectively, that were proposed by several researchers. Conclusions and references are given at the end.

2 Research Works on ML Based DoS Attack Detection in WSN

Denial of Service (DoS) is a condition that hinders the normal functioning of a networks such as traffic flow, resource exhaustion, etc., such conditions may lead to obstacle in network functionality. These conditions may affect the availability and functionality of any type of service in the network. But when this conditions are created intentionally by an attacker than this is called a DoS attack. DoS attack may reduce or destroy the network functionality. DoS attacks modify the infrastructure configuration, devour resources, and may physically destroy the network components [3].

An anomaly-based IDS using the Bayesian classifier in WSN was proposed by Klassen and Yang [4]. Both DoS and Blackhole attacks are detected by the proposed system. DoS and blackhole attacks are detected but unable to detect malicious attacks. A secure mechanism to detect energy exhaustion attacks based on ANN was proposed by Alrajeh et al. [5]. The scheme detects flooding attacks which prove to be more severe in nature and capable of draining the energy of sensor nodes. 98% detection rate was observed for flooding attacks, 95% detection rate in case of routing loop attacks, and detection rate of 55% in case of data link layer attack. KNN-based new detection algorithm in WSN was proposed by Li et al. [6]. The proposed system mainly detects flooding attacks. Results show that normal data flow in large networks are affected by flooding attack. Flooding attack can be detected and prevented by this proposed system.

An entropy-based decision tree for detecting flooding attacks in hierarchical WSNs was proposed by Kim et al. [7]. The proposed system used Distributed key management system, in the detection approach. Approximate entropy estimation is applied to the keys attached to the packets. To reduce computational head, the use of logarithm is replaced by the use of multiplication in the calculation of entropy. The proposed system provided increased precision in comparison with traditional calculation for entropy. An AI technique called SYN Flood Attack Detection Based on

Bayes Estimator (SFADBE) for Mobile Ad-hoc Network for detecting SYN flood attack was proposed by Hussain et al. [8]. Here, in this system, all the nodes in the network collects and congregates the channel's information, and SAFDBE then designate a reliable and channel free from congestion for the traffic. By employing this proposed system, the probability of detection of SYN flood attack is increased as compared with others. And it is less expensive and vigorous in comparison with existing approaches. An IDS that detect and prevent intrusion in WSN including DoS attacks was proposed by Chandre et al. [9]. The dataset is classified into three categories, i.e., structured, semi-structured, and unstructured data. The system engages the decision tree technique to the structured data. Existing techniques such as Automated Validation of Internet Security Protocols and Applications (AVISPA) is used to detect intrusion.

A System that employs both decision tree and SVM to discover DoS attack in WSN was proposed by Al-issa et al. [10]. They also make use of selective forwarding attacks. According to the result, decision tree yields 99.86% true positive rate and 0.05% false positive rate, and SVM techniques yields 99.62% true positive and 0.09% false positive. An IDS based on neural network and SVM for detecting Denial of Service attacks was proposed by Raj et al. [11]. The results shows that SVM gives 97% accuracy and NN gives 91% accuracy in DoS attack detection. The results also shows that SVM takes computing time of 0.25 s and NN takes computing time of 0.75 s for identifying the attacks probabilities in WSN. An IDS in wireless sensor network based on Improved AdaBoost-RBFSVM (IABRBFSVM) was proposed by Jianjian et al. [12]. The proposed system detects DoS attacks in WSN. The comprehensive performance of the network is improved by the proposed system. The results showed that high detection rate was obtained in less time while taking care and separating malicious nodes.

An Mean Shift Clustering Algorithm (MSCA) and SVM-based knowledge-based intrusion detection system (KBIDS) that expose blackhole and DoS attacks in WSN was proposed by Qu et al. [13]. The MSCA recognized the abnormal patterns of the attacked WSN from a normal WSN. And SVM is used to enlarge the margin between normal and abnormal characteristics to reduce the error classification. They showed lower false alarm rate, and a higher detection rate is obtained by KBIDS in comparison with other IDS model. An adaptive chicken swarm optimization algorithm which is an efficient clustering technique, the algorithm that reduces time consumption along with an IDS based on SVM was proposed by Borkar and Patil [14]. Various types of attacks such as Probe, DoS, U2R, and R2L were exposed. A security technique was engage to provide reliable flow of packets from one node to another. The proposed algorithm yields better results in comparison with other techniques. A detection system that detects DoS attacks based on random forest was proposed by Le et al. [15]. The proposed system is compared with ANN and found to be more efficient. Ramesh et al. [16] proposed an algorithm based on optimized deep neural network for detecting DoS attacks in wireless multimedia network. For detection accuracy, multiple layer of neurons was used. The proposed method proves to be more efficient in comparison with the RAS-HO, TMS, and SVM-DoS.

3 Research Works on ML-Based DDoS Attack Detection in WSN

Distributed Denial of Service (DDoS) are DoS attacks that are done from many different location using many different systems. DDoS attacks are those kinds of attacks which main objective is to disturb the normal functioning of a network by maximizing the resource consumption. Attackers send multiple request messages in the network to increase network congestion and reduce the lifespan of the network [17]. Furthermore, the DDoS attacks may be divided into two types, i.e., direct attack and reflector attack.

A new Game-based Fuzzy Q-Learning (FQL), cooperative game-theoretic defense after discussion the interactivity between sink nodes, base stations, and adversaries was proposed by Shamshirband et al. [18]. Cooperative game theory is amalgamated with FQL algorithm elements in the system. To secure the network from an incoming DDoS attack, the interaction between detection sink node player and repose base station players is administered. The proposed model is set up as a two-player game strategy which in fact is a triple-player game strategy, providing double layer security against one attacker. The proposed system teach itself of the past attacks in the FQL decision-making process to conquer incoming attacks. An Intrusion Prevention System (IPS) security and a FQL-based approach was presented by Sherazi et al. [19]. The analysis had been carried out on the basis of 4 tuple variables. Q-Learning and Fuzzy Logic algorithms are implemented in the proposed system. The result shows that DDoS caused traffic congestion as a result of packets flooding. The scheme proves to be effective for maximizing security in the heterogeneous networking against various attacks.

A SOM neural network, a centralized approach based on self-organizing maps was proposed by Baig and Khan [20]. The system is trained with network traffic conditions which includes both compromised and normal situation. The SOM application is applied on the sink nodes, thereby perceiving the network flow from all the clusters in WSN along with the data from the individual detector nodes of the network. The system proves to be efficient for detecting DDoS attacks. An ANN-based IDS to detect and secure networks from DDOS attacks was proposed by Aljumah and Ahamad [21]. A six-step algorithm is designed and chaos theory is deployed to perceive DDoS attacks in the network. To start the learning process, a replica image of the actual network environment is used. Several DDoS attacks are launched while there is a normal flow of authentic traffic in the network. Supervised and non-supervised techniques of neural networks are used to identify DDoS attacks and normal traffic. Lyapunav coefficient is employed to achieve the best result in identifying the authentic traffic and DDoS attacks. More than 95% accuracy is achieved in detecting DDoS attacks.

A detection system that detect DDoS attacks in SDN using various ML techniques, such as Naive Bayes, K-means, K-Nearest neighbors, and K-medoids was proposed by Barki et al. [22]. They are able to show that the use of ML techniques in detecting DDoS attacks yields better results. The proposed system is executed using

POX controller and Mininet. An IDS suitable with the wireless sensor network was proposed by Ifzarne et al. [23]. The model is built on information gain ratio (IGR) and online passive aggressive classifier (OPGC). The relevant features of the sensor data are selected by using the IGR and to detect and classify several types of DoS attacks the OPGC algorithm is trained. The detection accuracy for scheduling attacks is 86%, gray hole attacks is 68%, flooding attacks is 63% and for blackhole attacks is 46%. The method detects the appearance of an intrusion and determines the attack type in a cluster WSN network topology. Minimum resource wastage is reported in this model. A three IDS for DDoS attacks detection based on deep learning, namely deep neural networks (DNN), convolutional neural networks (CNN), and recurrent neural networks was proposed by Ferrrag et al. [24]. Each IDS's performance is discussed within two classification types (binary and multiclass) using two traffic datasets (CIC-DDoS2019 and TON_IoT dataset) that include several types of DDoS attacks. Deep learning techniques shows better results in comparison with other techniques.

Aysa et al. [25] proposed a system to detect abnormal security activities in IoT networks. Here, they used standard datasets to detect Mirai and BASHLITE attacks. The proposed framework discloses any unfamiliar movements by surveying the traffic and ordering this action by comparing with the recorded normal traffics of IoT gadget. ML algorithms such as LSVM, NN, and decision tree were used to detect DDoS attacks. And came to the conclusion that combination between the random forest and decision tree yields better accuracy in detecting attacks. Bhale et al. [26] proposed an edge-based machine learning enables IDS to detect DDoS attacks in IoT networks. The result of the proposed method shows computational and storage efficiency. The memory utilization, energy preservation, response time, and accuracy are much better as compare other recent works.

Mohd et al. [27] proposed an IDS based on SVM that is used to detect distributed denial of sleep attack in WSN. The model has been developed with various functions such as linear, radial, and sigmoidal functions. Opnet modeler 17.5 has been used to simulate the proposed work and association rule set was used for training the dataset and training it. The use of radial function yields greater accuracy. Gite et al. [28] proposed a detection scheme that identifies the malicious node by analyzing the traffic pattern. They employed C4.5 and CART classifier to detect various types of attacks such as blackhole, gray hole, DDoS, and wormhole attacks. C4.5 performs better when the number of nodes increases. But CART classifier accuracy is consistent and increases linearly while C4.5 accuracy fluctuates (Tables 1 and 2).

4 Conclusion

Various attack detection schemes using machine learning techniques for detecting DoS and DDoS attacks in WSN have been reviewed in this paper. In order to provide a very sound and robust security mechanism to WSN, development of attack detection techniques using machine learning is essential. It can be seen from the current

Table 1 Summary of research on DoS attacks detection in WSN using machine learning

Cited	ML technique uses	Limitation	Remarks
[4]	Bayesian	Need to minimize host and destination nodes numbers and to increase network size. Fails to classify between normal and malicious node	DoS attacks including blackhole attacks are detected but malicious nodes are not detected
[5]	Neural networks	Fails to detect flooding in datalink and network layer	Detection rate of 98% achieved
[6]	KNN	–	Best for dynamic network. Detect fresh attacks in the absence of previous knowledge
[7]	Entropy-based decision tree	–	Higher detection accuracy in comparison with entropy calculation
[8]	Bayes estimator (SFADBE)	Consumes more energy and not efficient in preventing SYN flood attack	Works well in detection of attacks before the destruction of the entire network
[9]	Decision tree	Not efficient in preventing DoS attacks	Detect and prevent DoS attacks
[10]	VM and decision tree	Data mining techniques and classifiers are missing. Can only detect selective forwarding attack	Decision tree gives true positive rate (99.86%) and false positive rate (0.05%), and SVM methods give true positive (99.62%) and false positive (0.09%)
[11]	SVM and NN	Fails to defend DoS attacks	SVM gives better performance than NN
[12]	SVM (IABRBFSVM)	Only DoS attacks can be detected	Malicious nodes are removed in less time
[13]	Unsupervised learning (MSCA) and SVM	–	Detection rate is highest with lower false alarm rate
[14]	SVM	–	Higher detection rate with lower time consumption
[15]	Random forest	Can only predict and detect DoS attacks	Detection rate of blackhole (99%) and gray hole (96%)
[16]	DNN	–	More efficient in comparison with the RAS-HO, TMS, and SVM-DoS

literature that very few number of DoS and DDoS attack detection schemes have been developed using machine learning as of now. The need to develop new ML-based techniques to increase the attack detection accuracy could be essential future research direction. As we look into the future, the role of ML-based security attack

Table 2 Summary of research on DDoS attacks detection in WSN using machine learning

Cited	ML technique uses	Limitation	Remarks
[18]	Fuzzy Q-learning	Lacks the power to detect new or prevailing attacks. Need to include data of other attacks types	Detecting attacks and accuracy in defense yield a larger improvement than available methods, continuously self-learned
[19]	Q-learning	Fails to detect other attacks	Security, detection, and defense is improved
[20]	SOM neural network	Not suited for WSN due to energy constraint nature of the victim nodes	Increase detection efficiency
[21]	ANN	–	More than 95% detection accuracy achieved
[22]	Naive Bayes, KNN, K-means, and K-medoids	–	Use of ML in detecting DDoS attacks yields better results
[23]	Online passive-aggressive	Need to combine with other algorithms to increase detection accuracy	96% detection accuracy as compared to offline models
[24]	Deep learning	–	Deep learning techniques provide better results as compared to other techniques
[25]	LSVM, NN, and decision tree	–	Combination of decision tree and random forest yields better detection accuracy
[26]	Edge-based ML	–	Scalable and efficient with respect to storage and computation
[27]	SVM	–	Radial function yields better accuracy in compassion with linear and sigmoidal functions
[28]	C4.5 and CART classifier	–	C4.5 outperformed CART classifier when the node density increases. But CART accuracy is consistent when compared with C4.5

detection schemes will be highly important and promising to secure the wireless sensor networks from several types of security threats and attacks in general.

References

1. M.M. Singh, N. Dutta, Security issues in wireless sensor networks. Int. J. Distrib. Cloud Comput. **5**(2) (2017). ISSN: 2321-6840
2. A.P. Abidoye, E.O. Ochola, Denial of service attacks in wireless sensor networks with proposed countermeasures, in *Information Technology—New Generations*, (2018), pp. 185–191. https://doi.org/10.1007/978-3-319-77028-4_27
3. S. Ghildiyal, A.K. Mishra, A. Gupta, N. Garg, Analysis of denial of service (DOS) attacks in wireless sensor networks. IJRET **3**, 2319–1163 (2014)
4. M. Klassen, N. Yang, Anomaly based intrusion detection in wireless networks using Bayesian classifier, in *2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI)*. https://doi.org/10.1109/icaci.2012.6463163
5. N.A. Alrajeh, S. Khan, J.L. Mauri, J. Loo, Artificial neural network based detection of energy exhaustion attacks in wireless sensor networks capable of energy harvesting. Ad-Hoc Sens. Wireless Netw. **22**, 109–133
6. W. Li, P. Yi, Y. Wu, L. Pan, J. Li, A new intrusion detection system based on KNN classification algorithm in wireless sensor network. J. Electr. Comput. Eng. **2014**, 1–8 (2014). https://doi.org/10.1155/2014/240217
7. M. Kim, I. Doh, K. Chae, Denial-of-service (DoS) detection through practical entropy estimation on hierarchical sensor networks, in *2006 8th International Conference Advanced Communication Technology*. https://doi.org/10.1109/icact.2006.206283
8. K. Hussain, S.J. Hussain, N. Jhanjhi, M. Humayun, SYN flood attack detection based on Bayes estimator (SFADBE) for MANET, in *2019 International Conference on Computer and Information Sciences (ICCIS)*. https://doi.org/10.1109/iccisci.2019.8716416
9. P.R. Chandre, P.N. Mahalle, G.R. Shinde, Machine learning based novel approach for intrusion detection and prevention system: a tool based verification, in *2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN)*. https://doi.org/10.1109/gcwcn.2018.8668618
10. A.I. Al-issa, M. Al-Akhras, M.S. ALsahli, M. Alawairdhi, Using machine learning to detect DoS attacks in wireless sensor networks, in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*. https://doi.org/10.1109/jeeit.2019.8717400
11. A.B. Raj, M.V. Ramesh, R.V. Kulkarni, T. Hemalatha, Security enhancement in wireless sensor networks using machine learning, in *2012 IEEE 14th International Conference on High Performance Computing and Communication and 2012 IEEE 9th International Conference on Embedded Software and Systems*. https://doi.org/10.1109/hpcc.2012.186
12. D. Jianjian, T. Yang, Y. Feiyue, A novel intrusion detection system based on IABRBFSVM for wireless sensor networks. Procedia Comput. Sci. **131**, 1113–1121 (2018). https://doi.org/10.1016/j.procs.2018.04.275
13. H. Qu, Z. Qiu, X. Tang, M. Xiang, P. Wang, Incorporating unsupervised learning into intrusion detection for wireless sensor networks with structural co-evolvability. Appl. Soft Comput. **71**, 939–951 (2018). https://doi.org/10.1016/j.asoc.2018.07.044
14. G.M. Borkar, L.H. Patil, A novel clustering approach and adaptive SVM classifier for intrusion detection in WSN: a data mining concept. Sustain. Comput. Inform. Syst. https://doi.org/10.1016/j.suscom.2019.06.002
15. T.-T.-H. Le, T. Park, D. Cho, H. Kim, An effective classification for DoS attacks in wireless sensor networks, in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)* (2018). https://doi.org/10.1109/icufn.2018.8436999
16. S. Ramesh, C. Yaashuwanth, K. Prathibanandhi, A.R. Basha, T. Jayasankar, An optimized deep neural network based DoS attack detection in wireless video sensor network. J. Ambient. Intell. Humaniz. Comput. (2021). https://doi.org/10.1007/s12652-020-02763-9
17. R. Upadhyay, S. Khan, H. Tripathi, U.R. Bhatt, Detection and prevention of DDOS attack in WSN for AODV and DSR using battery drain, in *2015 International Conference on Computing and Network Communications (CoCoNet)*. https://doi.org/10.1109/coconet.2015.7411224

18. S. Shamshirband, A. Patel, N.B. Anuar, M.L.M. Kiah, A. Abraham, Cooperative game theoretic approach using fuzzy Q-learning for detecting and preventing intrusions in wireless sensor networks. Eng. Appl. Artif. Intell. **32**, 228–241 (2014). https://doi.org/10.1016/j.engappai.2014.02.001
19. H.H.R. Sherazi, R. Iqbal, F. Ahmad, Z.A. Khan, M.H. Chaudary, DDoS attack detection: a key enabler for sustainable communication in internet of vehicles. Sustain. Comput. Inform. Syst. **23**, 13–20 (2019). https://doi.org/10.1016/j.suscom.2019.05.002
20. Z.A. Baig, A.I. Khan, DDoS attack modeling and detection in wireless sensor networks. Mob. Intell. 595–626 (n.d.). https://doi.org/10.1002/9780470579398.ch26
21. A. Aljumah, T. Ahamad, A novel approach for detecting DDoS using artificial neural networks. Int. J. Comput. Sci. Netw. Secur. **16**, 132–132 (2016)
22. L. Barki, A. Shidling, N. Meti, D.G. Narayan, M.M. Mulla, Detection of distributed denial of service attacks in software defined networks, in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. https://doi.org/10.1109/icacci.2016.7732445
23. S. Ifzarne, H. Tabbaa, I. Hafidi, N. Lamghari, Anomaly detection using machine learning techniques in wireless sensor networks, in *The International Conference on Mathematics and Data Science (ICMDS)* (2020). https://doi.org/10.1088/1742-6596/1743/1/012021
24. M.A. Ferrag, L. Shu, H. Djallel, K.-K.R. Choo, Deep learning-based intrusion detection for distributed denial of service attack in agriculture 4.0. Electronics **10**, 1257 (2021). https://doi.org/10.3390/electronics10111257
25. M.H. Aysa, A.A. Ibrahim, A.H. Mohammed, IoT DDoS attack detection using machine learning, in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (2020). https://doi.org/10.1109/ismsit50672.2020.9254703
26. P. Bhale, S. Biswas, S. Nandi, ML for IEEE 802.15. 4e/TSCH: energy efficient approach to detect DDoS attack using machine learning, in *2021 International Wireless Communications and Mobile Computing (IWCMC)* (2021), pp. 1477–1482. https://doi.org/10.1109/IWCMC51323.2021.9498637
27. N. Mohd, A. Singh, H.S. Bhadauria, A novel SVM based IDS for distributed denial of sleep strike in wireless sensor networks. Wireless Pers. Commun. (2019). https://doi.org/10.1007/s11277-019-06969-9
28. P. Gite, K. Chouhan, K. Murali Krishna, C. Kumar Nayak, M. Soni, A. Shrivastava, ML based intrusion detection scheme for various types of attacks in a WSN using C4.5 and CART classifiers. Mater. Today Proc. (2021). https://doi.org/10.1016/j.matpr.2021.07.378

Comparison of Cell Nuclei Classification in Cytological Breast Images Using Machine Learning Algorithms

Vrushali Ailawar and Vibha Bora

Abstract Breast cancer is the most common cancer among Indian women. In every 4 min, one lady is diagnosed with breast cancers, and in every 13 min, one woman dies of breast cancer in India. Hence detection of the breast cancer at an early stage is very important. Fine needle aspiration cytology and histopathological biopsies are golden truth for detection of breast cancer. To increase the accuracy of detection, computer-aided diagnosis is used as a second opinion. In computer-assisted pathology analysis, breast cytological image classification is a most important task as it helps in understanding the structure and distribution of nucleus for the prediction of breast cancer. In this paper, comparative analysis is done using five different machine learning algorithms on publicly available Wisconsin diagnostic dataset from UCI machine learning repository. The accuracy of support vector machine, random forest, K-nearest neighbour, decision tree and logistic regression was found to be 97.36%, 95.61%, 96.49%, 94.73% and 97.36%, respectively. The authors conclude that support vector machine and logistic regression perform equally better in the terms of accuracy.

Keywords Support vector machine · K-nearest neighbour · Decision tree · Random forest · Logistic regression

1 Introduction

One of the most predominant types of cancer among women is breast cancer. Breast cancer accounts for more than 40% of all cancer in women before the age of 40 year, and around 7% of women with breast cancer are detected in this age group [1]. In Fig. 1, *X*-axis indicates the different age groups from 20 to 30 years, 30 to 40 years and so on, and the *Y*-axis indicates the percentage of cases. The light blue colour represents the 25 years back incidence, and brown colour represents situation in 2020. Twenty-five years back, out of each a hundred breast cancers cases, 2% has

V. Ailawar (✉) · V. Bora
Department of Electronics Engineering, G H Raisoni College of Engineering, Nagpur, India
e-mail: vrushali.ailawar@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_54

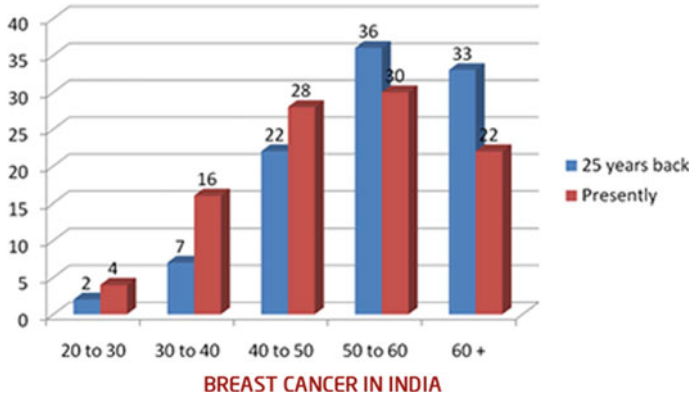


Fig. 1 Graphical representation of breast cancer [2]

been in 20–30 age group, 7% had been in 30–40 and so on. Sixty-nine percentage of the cases were above 50 years of age. Currently, 4% are in 20–30 years age, 16% are in 30–40 age, and 28% are in 40–50 age. About 48% patients belongs to 50 years of age. The rising number of cases is within 25–40 years of age, and this is an alarming situation [2]. The survival rate of breast cancer patient can be increased if it is detected at an early stage. In detection of breast cancer, important role is played by fine needle aspiration cytology (FNAC). FNAC test is recommended to check whether the lump is cancerous or not. In this test, a small hypodermic needle is inserted into the lump. The needle is inserted and drawn in and out for a period of about five seconds and then withdrawn. A sample of the lump is obtained. The sensitivity of the FNAC depends on the quality of image and pathologist experience. Pathologist examination is extremely specialized and time-consuming work disposed to inter and intra-observer incongruity.

To overcome this error, computer-aided diagnosis (CAD) can be used to assist the pathologists with the help of different machine learning algorithms [3, 4]. In this paper, five machine learning algorithms were compared on the basis of accuracy, precision, and recall to predict the breast cancer in FNAC images.

2 Related Work

Various researches have been conducted to detect the breast cancer in histopathological and FNAC images using machine learning (ML) [5] and deep learning-based algorithms [6]. Reis et al. [7] had done automatic classification of breast carcinoma from slides of H&E stained breast tissue. Random decision tree in combination with multiscale basic image features was used and archived the classification accuracy of 84%.

Saha et al. [8] suggested deep learning-based HER2 deep net (Her2Net) on cell nuclei from human epidermal growth factor receptor-2 (HER2)-stained images of breast cancer. They had done segmentation and classification of cell membranes. They used convolutional and deconvolutional network to obtain the cell membrane and nucleus segmentation task. Their proposed Her2Net attained precision (96.64%), recall (96.79%), F1-score (96.71%), negative predictive value (93.08%), accuracy (98.33%), and a false positive rate (6.84%).

3 Methodology

3.1 Dataset

The Wisconsin Breast Cancer Diagnostic (WBCD) dataset is from the UCI machine learning repository [9]. The WBCD dataset consists of 569 × 32 fine needle aspirate biopsy samples of humane breast tissues [10] as shown in Fig. 2. There are 569 samples and 32 features in datasets. Among 569 samples, benign samples are 357 and malignant samples are 212. Out of 32 features, first feature is identification number, and second feature is target or output variable which is the diagnosis as malignant or benign. Remaining 30 features are considered as input variables and are computed for each cell nucleus samples such as standard error, mean value, and extreme value of ten important attributes like radius, texture, perimeter, smoothness, area, compactness, concave points, concavity, symmetry, and fractal dimensions. Dataset does not contain any null or missing value. The dataset is split into training and testing instances of 80:20.

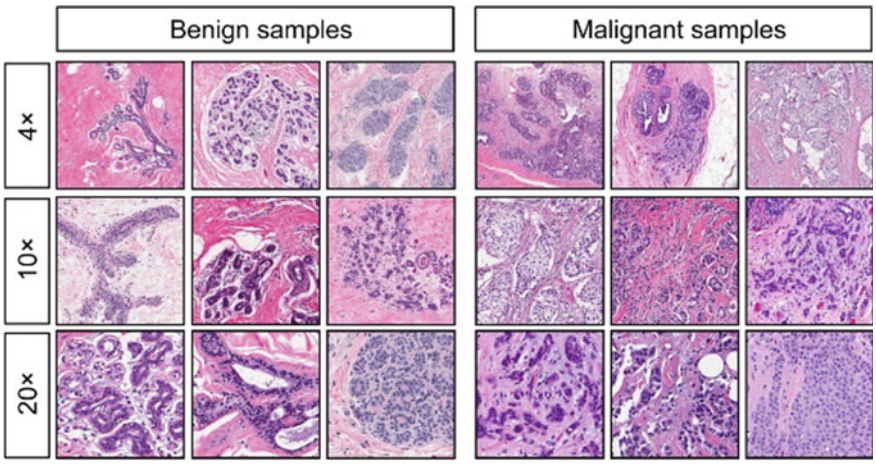


Fig. 2 Benign and malignant breast specimens [10]

3.2 *Algorithms Used*

Decision Tree: For classification and regression tasks, decision tree is a versatile ML algorithm. [11]. A structure like tree is used in decision tree as shown in Fig. 3, in which population start dividing into various feature from root node. Decision node is tree node or parent node that splits into two or more child node, and it represents the feature of dataset. Leaf node is a bottom node that cannot split any further, and it represents the outcome. The small portion of this tree is called as sub-tree.

Logistic Regression: It is type of supervised ML algorithm. It is a statistical analysis technique used to predict data value based on previous observation of dataset [11]. The output of an express structured variable may be either Yes or No, 0 or 1, true or false and many others. It gives the probabilistic values which lies between 0 and 1. In logistic regression, we used "s" shape logistic function, rather than fitting a regression line which predicts two maximum values (0 or 1). It is used to estimate the possibilities of event and can effortlessly determine the handiest variables used for the classification. Figure 4 indicates the logistic function.

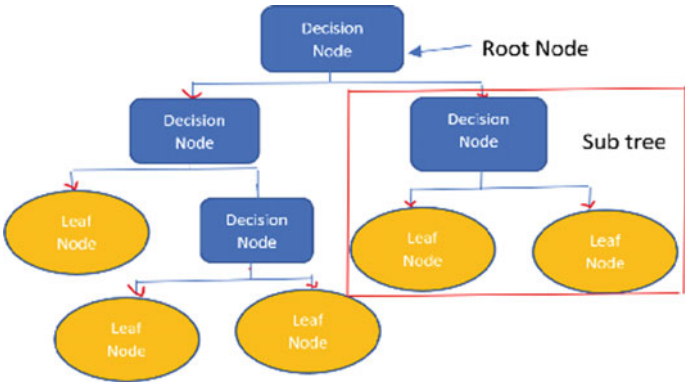


Fig. 3 Decision tree

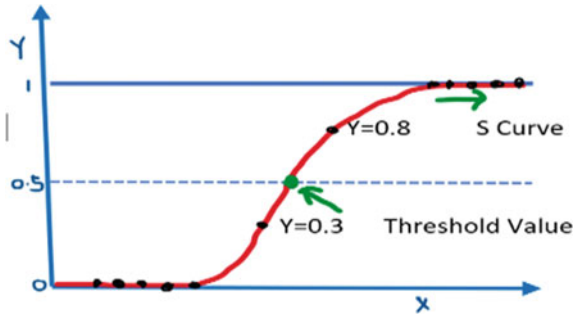


Fig. 4 Logistic regression

Random Forest: Another supervised ML algorithm is random forest used for solving both classification and regression problems [11]. It makes use of the idea of ensemble learning, which is a method of mixing more than one classifier to clear up a complicated problem and to improve the performance of the model. As the name indicates, random forest is a classifier that contains variety of decision tree. It draws some samples to build a model, then draws some samples again to build another model, and so on. All of these sampling and modelling are done independently and simultaneously. The final outcome is the average of the predicted values of all the models. The greatest range of tree inside the forest ends in higher accuracy and stops the hassle of overfitting. Figure 5 imitates random forest algorithm.

Support Vector Machine: Support vector machine or SVM falls under the class of ML algorithms which is supervised and used for solving classification and regression problems [11]. Support vector machine creates a centre line called as hyperplane to classify the data points. After creating a hyperplane, it creates two planes which are parallel to the hyperplane and passing through one of the nearest data point as shown by dotted lines in Fig. 6. The distance between two parallel planes is called as margin. The data point through which parallel plane passes is known as support vectors.

K-Nearest Neighbour: It is one of the classification algorithms which belongs to supervised ML category [11]. In this, K denotes the number of nearest neighbour. The number of neighbours is the core deciding factor. By using feature similarity, new data point is classified by majority votes from its K-nearest neighbour. As shown in Fig. 7, the new data point which is denoted by red dot is assigned to class pink as among these three points two belongs to pink class. Selection of right value of K is termed as parameter tuning, and it is significant for better precision. The value of K that gives the best accuracy for both training and testing is selected. Generally, odd value of K is considered to avoid the state of equality between the two classes.

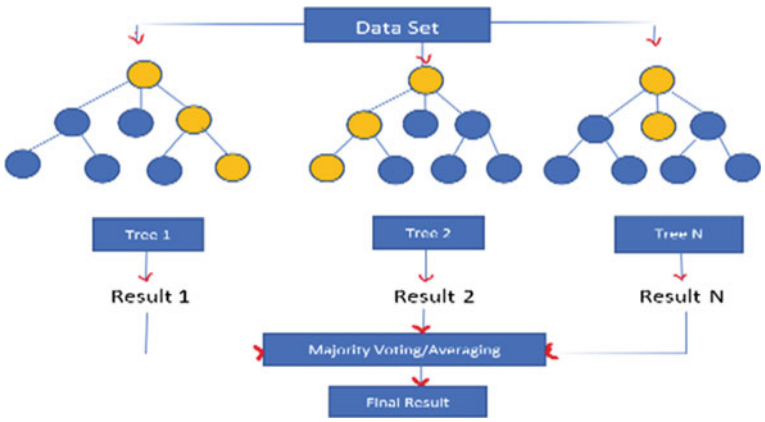


Fig. 5 Random forest

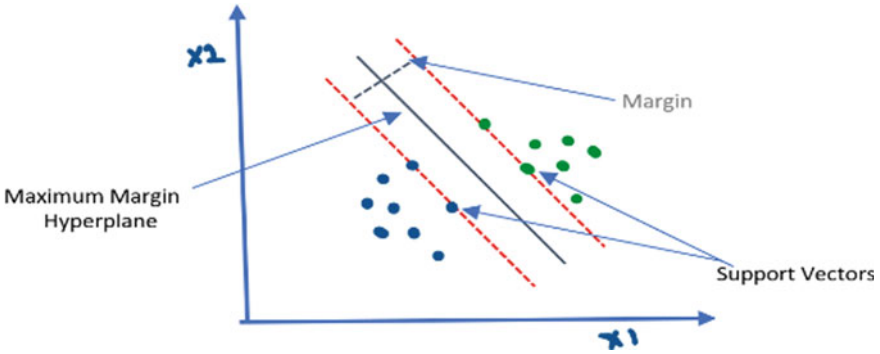
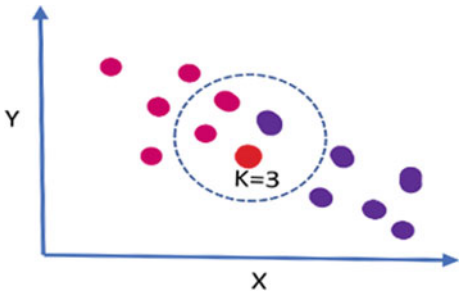


Fig. 6 Support vector machine

Fig. 7 K-nearest neighbour



4 Experiments Results

For training and testing different ML models, Anaconda3 Jupyter Notebook and Python programming language are used [12]. The performance is measured in terms of accuracy, precision, ROC curve and R2-score, so as to enhance the early detection of breast cancer. Overall performance measures are defined as follows:

Accuracy: Higher value of accuracy is the indication of more efficient model. It is calculated as a fraction of correctly anticipated samples to the total no. of samples.

$$\text{Accuracy} = (\text{TP} + \text{TN}) \big/ (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Precision: It is expressed as a ratio of successfully anticipated positive samples to the total positive samples.

$$\text{Precision} = \text{TP} \big/ (\text{TP} + \text{FP})$$

Recall: It gives the percentage of appropriate results of data that model identifies accurately. Recall is also called as sensitivity of the model.

Table 1 Comparative analysis of machine learning algorithms

Name of algorithm	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Accuracy	Precision	Recall	R2-score
Logistic regression	66	45	2	1	97.36	97.05	98.5	0.89
Decision tree	63	42	5	4	94.73	92.64	94.02	0.78
Random forest	65	44	3	2	95.61	95.58	97.01	0.81
SVM	64	46	4	0	97.36	94.11	100	0.89
K-NN	66	45	2	1	96.49	97.05	98.5	0.85

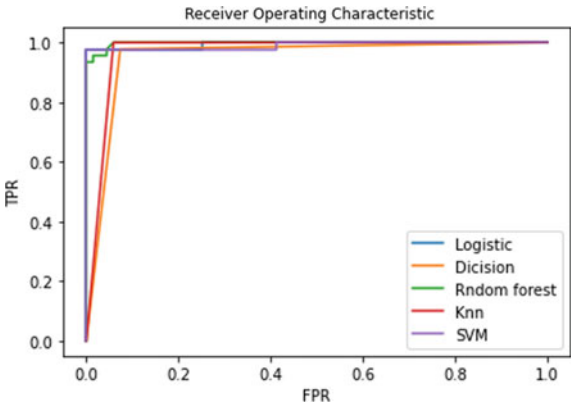


Fig. 8 ROC of five ML algorithms

$$\text{Recall} = \text{TP} \big/ (\text{TP} + \text{FN})$$

R2-Score: It gives the idea about how good your model fits in a dataset. It takes the value between 0 and 1.

Table 1 gives the comparative measurement analysis of five different machine learning algorithms: 1. logistic regression, 2. decision tree, 3. random forest, 4. support vector machine and 5. K- nearest neighbour. Figure 8 shows the receiver operating characteristic (ROC) curve of the five ML algorithms.

5 Conclusion

The comparative analysis of five machine learning algorithm on the basis of accuracy, precision, recall and R2-score was performed. After comparing all the applied algorithm such as support vector machine, random forest, K-nearest neighbour, decision

tree and logistic regression, the accuracy was found to be 97.36%, 95.61%, 96.49%, 94.73% and 97.36%, respectively. Logistic regression and support vector machine perform equally better in the terms of accuracy. The overall performance of support vector machine algorithm in the terms of precision, R2-score and F1-score is better as compared to other ML algorithms.

References

1. C.K. Anders, R. Johnson, J. Litton, M. Phillips, A. Bleyer, Breast cancer before age 40 years. Semin. Oncol. **36**(3), 237–249 (2009)
2. https://www.breastcancerindia.net/statistics/trends.html. Last Accessed 26 May 2020
3. Kaushal, S. Bhat, D. Koundal, A. Singla, Recent trends in computer assisted diagnosis (CAD) system for breast cancer diagnosis using histopathological image. IRBM **40**(4), 211–227 (2019)
4. M. Saha, R. Mukherjee, Chakraborty, Computer-aided diagnosis of breast cancer using cytological images: a systematic review. Tissue Cell **48**(5), 461–474 (2016)
5. M. Kumar, J. Divya Udayan, A survey of machine learning techniques for cancer disease prediction and diagnosis. Indian J. Public Health Res. Dev. **10**(4), 157 (2019)
6. A. Saber, M. Sakr, O.M. Abo-Seida, A. Keshk, H. Chen, A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. IEEE Access **9**, 71194–71209 (2021)
7. S. Reis, P. Gazinska et al., Automated classification of breast cancer stroma maturity from histological images. IEEE Trans. Biomed. Eng. (9), 2344–2352 (2017)
8. M. Saha, C. Chakraborty, Her2Net: a deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. IEEE Trans. Image Process. **27**(5), 2189–2200 (2018)
9. W.H. Wolberg, W.N. Street, O.L. Mangasarian, Breast cancer Wisconsin (Diagnostic). UCI Machine Learning Repository (1995)
10. R.M. Levenson, E.A. Krupinsk, Pigeons (Columba livia) as trainable observers of pathology and radiology breast cancer image. PLOS ONE **10**(11) (2015)
11. C. Bishop, *Pattern Recognition and Machine Learning* (Springer-Verlag, New York, 2006)
12. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd edn. (O'Reilly Media, 2019)

A Survey on Machine Learning Based Approaches for Stock Market Prediction

Dinesh Singh Dhakar and Savita Shiwani

Abstract As investing in stock markets is growing in popularity, predicting the nature of stock market and the associated trends has become a critical aspect of financial forecasting. Conventional statistical techniques have been used previously to analyze trends in the share prices as a function of time and related variables to extract temporal correlation. This however, is challenging due to the fact that stock market trends and share prices depict extreme volatility in values and depend on multiple parameters which can be non-numeric in nature such as social and political conditions in the country where the company is listed, government policies, etc. This makes the data fluctuate randomly as a time series function. Machine Learning based techniques have been used off late to analyze complex time series data sets such as stock market trends due to their ability to find patterns in extremely large and complex datasets. This paper presents a comprehensive survey on existing contemporary machine learning techniques used for stock market forecasting citing the important findings. The paper also presents the experimental results obtained from the applying the baseline techniques on benchmark datasets to draw a comparison among techniques commonly used.

Keywords Stock market prediction · Data pre-processing · Machine learning · Error · Prediction accuracy

1 Introduction

Stock market forecasting has long remained one of the most challenging yet sought after prediction problems for mathematical and computational analysis due to the extreme importance it carries for financial firms. The decision regarding investments, disinvestments and acquisitions in the financial sector critically depends upon the market trends of stocks [1]. Off late, machine learning and deep learning techniques

D. S. Dhakar (✉) · S. Shiwani
Department of Computer Science, Engineering, School of Engineering and Technology (SOET), Jaipur National University, Jaipur, India
e-mail: dineshhodcsemit@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_55

have taken over from conventional statistical techniques due to the increase in the accuracy of prediction of machine learning-based approaches as opposed to conventional statistical techniques [2]. The major challenge in forecasting trends in the stock markets is the lack of correlation among the governing variable and the stock prices and the difficulty in deciding the features or parameters, which actually govern the stock prices [3]. The feature selection in drastically different for various forecasting scenarios such as short-term forecasting, mid-term forecasting and long-term forecasting [4]. The data moreover, is generally noisy in nature which leads to inaccuracies in pattern recognition and hence forecasting the future prices with high accuracy [5]. Hence most techniques use some data cleaning and pre-processing technique prior to pattern recognition in the time series analysis of stock market data. This at times may increase the complexity of the algorithm though, which is applied to several time series data analysis models [6]. Subsequent to data cleaning and pre-processing, pattern recognition has seen two major approaches in recent literature which happen to be machine learning-based techniques and deep learning-based techniques, along with other standard regression models [7]. The remaining paper is organized as: Sect. 2 introduces the fundamentals of feature extraction and data filtration. Section 3 cites the noteworthy contribution in the domain of related work. Section 4 presents the experimental results obtained by applying the baseline techniques to benchmark datasets along with an analysis on the performance. Section 5 concludes the findings and presents possible approaches which can be adopted in future works to improve upon the performance of the existing techniques.

2 Data Pre-processing

Data pre-processing plays a crucial role in the forecasting performance of the system after data acquisition. The pre-processing generally consists of identifying features that are effective in pattern recognition termed as feature extraction or feature selection [8]. This is typically followed by some filtration technique that is applied to filter out the random fluctuation around the noise floor [9].

2.1 Feature Selection

The first step is to collect and analyze benchmark datasets available for stock market prediction. Typically, several datasets are available online on open-source platforms. In general, the stock prices of the company under consideration may be listed on more than one stock market. For instance, a company may be listed in New York Stock Exchange (NYSE) as well as London Stock Exchange (LSE) and Shanghai Stock Exchange (SSE). Collecting stock prices from multiple stock exchanges give a broader picture of the stock performance on the global scale. Moreover, some of the stock exchanges may have time differences due to different time zones [10].

Thus the changes or movements in a stock exchange can be used as a metric for prediction for a stock exchange that is time-lagged. Global influencing parameters of one of the stock exchanges may be used for other stock exchanges as well. This approach may be useful for designing a more holistic approach for prediction [11]. The subsequent step is generally the design of a data cleaning or filtration technique that would remove the noisy component of the data to facilitate pattern analysis [12].

2.2 Filtration

As stock time series data exhibits sudden fluctuations, hence it is necessary to filter out local disturbances to facilitate pattern recognition. One of the most effective and commonly used techniques is the discrete wavelet transform and the maximal overlap discrete wavelet transform [13, 14]. They are suitable for filtration and analysis of fluctuating data as the base functions are fluctuating in nature [15]. The use of the wavelet decomposition of raw time series data for filtration is depicted in Fig. 1.

The wavelet transform can be viewed as a recursive filtering tools which act as a sieve to separate low and high-resolution coefficients of the data often termed as the approximate and detailed coefficients (C_A and C_D), respectively. The DWT is the down-sampled version of the continuous wavelet transform (CWT) and it is suited for discrete time series analysis of large and noisy datasets. The DWT is mathematically defined as:

$$\mathrm{DWT}(X, l, k) = \delta_0^{l\frac{-1}{2}} \sum_i X(i) F\left[\frac{k - i s_0^L}{s_0^l}\right] \tag{1}$$

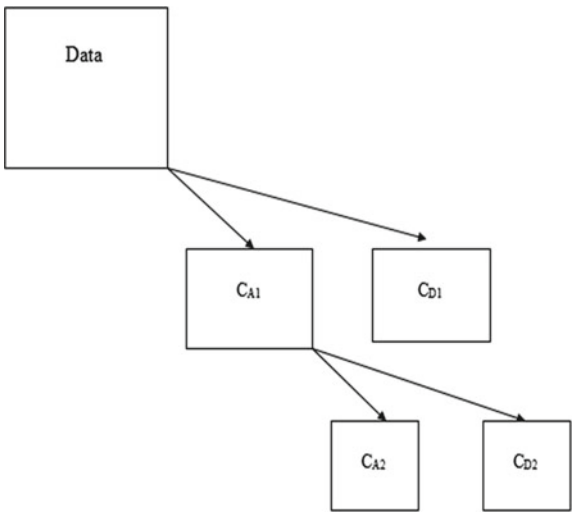


Fig. 1 Wavelet decomposition of data

Here,

$X(i)$ is raw data.

s_0^l denotes the scaling parameter.

is_0^l denotes the shifting parameter.

F denotes the wavelet family.

$s, \delta \in R$ denote the scaling and shifting operations constrained to $\delta \neq 0$.

There are several wavelet families depending upon the nature of the base function of the Kernel of the transform which governs the transform as:

$$Y \xrightarrow{\text{DWT}} \sum_{i,k,l} X_i^{k,l}, F \tag{2}$$

Here,

Y is the data in the transform domain.

DWT denotes the Discrete Wavelet Transform operation.

X denotes the raw data.

F is the wavelet family differentiated based on the Kernel of the transform.

Typical transforms, which are utilized for time series analysis, are the Haarlet, Symlet, Meyerlet, etc., which are basically different variants of the kernel function which are used for multi-resolution analysis of raw and noisy data. A major criterion for truncation of the wavelet decomposition is the energy or information content in the coefficients at different levels [16]. Due to the down-sampling nature of the DWT, another approach used is the maximal overlap discrete wavelet transform, which computes the maximum overlap of the wavelet function and the serial data stream and avoids down sampling or decimation of samples, even for low-resolution coefficient values [17]. Often data optimization and minimization techniques such as the principal component analysis (PCA) and independent component analysis (ICA) are utilized to facilitate training [18, 19].

3 Previous Work

This section discusses the common techniques used for stock market prediction problems, which are modelled as time series problems given by:

$$\text{stock prices} = \text{function (time, governing features)} \tag{3}$$

Different techniques have evolved over the course of time for the analysis of time series data, with its own pros and cons. The most commonly used models which have been used recently are explained subsequently. Next, a chronological analysis of noteworthy contribution in the field are also cited. To analyze the performance of the existing benchmark techniques commonly used, simulation results on the following benchmark models and their corresponding results have also been presented.

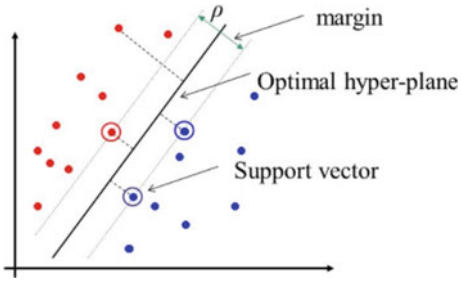


Fig. 2 Illustration of the SVM

Support Vector Regression (SVR)

The support vector regression is one of the most used and empirical regression models. It is fundamentally a variation of the support vector machine (SVM) with a regression-based cost function such as mean squared error or mean absolute error [20]. The SVM approach is depicted in Fig. 2.

The support vector regression can be designed as a least squares optimization (LS optimization) as:

for $(i = 1: n)$.

{

Update weights and bias

and

$$\text{Minimize}\left\{\frac{e_1^2 + e_2^2 + \cdots + e_n^2}{n}\right. \tag{4}$$

}

The SVR can be used for both linearly separable and non-linearly separable data. The Kernel is generally a non-linear function. In the context of the SVR model, the radial basis function Kernel (RBF) is the similarity between two points in the transformed feature space. Mathematically, it is defined as:

$$K\left(X, X'\right) = e^{-\gamma\left|X - X'\right|^2} \tag{5}$$

$$\gamma = \frac{1}{2\sigma} \tag{6}$$

Here,

γ is called the free parameter of RBF.

σ is called the feature factor.

K represents the RBF Kernel.

X and X' are the samples in an input feature space.

$\left|X - X'\right|$ is termed as the Euclidean Distance.

3.1 Neural Networks and Deep Neural Networks

Artificial Neural Networks (ANN) are one of the most effective techniques for time series or regression problems [21]. The output of the neural networks is given by:

$$y = f\left(\sum_{i=1}^{n} x_i w_i + \varphi\right) \tag{7}$$

Here,

y is the output.
x is the inputs.
w is the weights.
φ is the bias.
f stands for the activation function.

The commonly logic or activation functions used are the sigmoid, log sigmoid, tangent-sigmoid, rectified linear (ReLu), step or hard-limiting function, etc. The mathematical model for a neural network is depicted in Fig. 3.

A modified version of the neural network is the deep neural network which contains multiple hidden layers and is used for extremely complex pattern recognition problems. Neural networks in which data traverses a forward path from the output layer towards the output layer is termed as a feed-forward neural network. An alternative approach is to feedback the errors of each iteration as exogenous inputs for

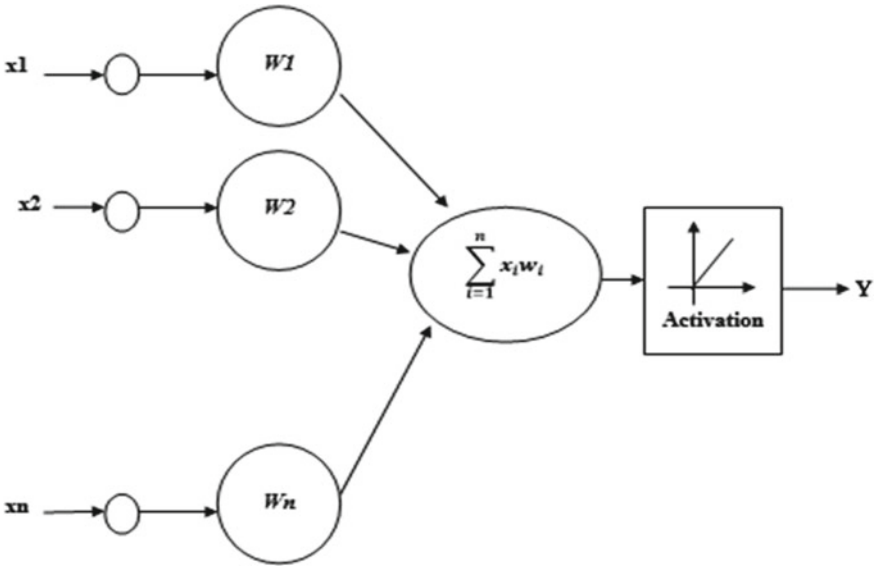


Fig. 3 Mathematical equivalent of neural network

training, which is called back propagation. A back propagation-based neural network (BP-ANN).

Neural networks with multiple hidden layers (used to create a dense hidden layer) is called a deep neural network (DNN) and training such neural networks is termed as deep learning. One of the most common forms of deep learning for time series prediction problems happen to be the recurrent neural networks (RNNs) and long short-term memory (LSTM) neural networks. RNNs are a category of neural networks which utilize the cascading feedback mechanism wherein the output of a previously hidden layer becomes the input to the subsequent hidden layer. Figures 4 and 5 depict the structure of the RNN and LSTM, respectively [22, 23].

The LSTM primarily has three gates:

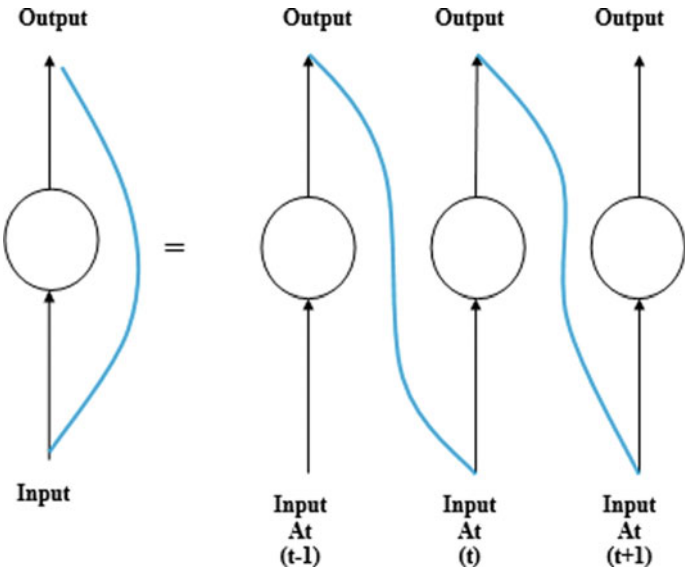
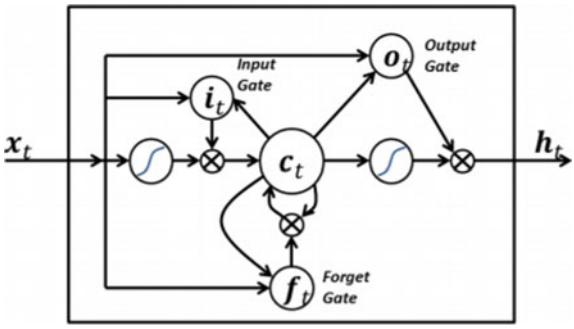


Fig. 4 Structure of RNN

Fig. 5 Structure of LSTM



(1) Input gate: This gate collects the present inputs and also considers the past outputs as the inputs.
(2) Output gate: This gate combines all cell states and produces the output.
(3) Forget gate: This is an extremely important feature of the LSTM which received a cell state value governing the amount of data to be remembered and forgotten.

Typically, the RNNs suffer from the vanishing gradient problem which is effectively sating that the gradient values near the input layer don't change as much as the gradient values close to the output layers. This issue is addressed in LSTMs. The LSTM primarily has three gates [24].

(1) Input gate: This gate collects the present inputs and also considers the past outputs as the inputs.
(2) Output gate: This gate combines all cell states and produces the output.
(3) Forget gate: This is an extremely important feature of the LSTM which received a cell state value governing the amount of data to be remembered and forgotten.

3.2 ARIMA

In an autoregressive integrated moving average model commonly known as the ARIMA model assumes that the future value of a variable can be linearly modelled as a function of previous samples of the variables and errors of prediction [25].

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_p y_{t-p} + \cdots + \theta_q \varepsilon_{t-q} \tag{8}$$

Here,

y_t is the value of the output variable at time 't'.
ε is the prediction error.
θ and φ are called the model parameters.
p and q are called the orders of the model.

A chronological analysis of the recent noteworthy contributions to the field of study is presented in Table 1. The analysis identifies the relevance and findings along with the research gap in the work. The research gap can be utilized to design algorithms by future researchers.

4 Experimental Results

The experimental results illustrate the results obtained by employing the benchmark techniques. The datasets used in this work are TCS, SBI, Reliance and Infosys share prices for a period of five years from 2016 to 2021. The features used are day, opening price, closing price, average of the day and volume. The data division has been done in the ratio of 70:30 for training and testing. The summary of the obtained results

Table 1 Summary of recent noteworthy contributions in the field, chronologically

Author(s)/year	Source/publisher	Findings and relevance	Research gap
S Kim, S Ku, W Chang, JW Song. 2020 [26]	IEEE Access 2020, Vol-8, pp. 111660–111682	Effective transfer entropy (ETE) used in conjugation with existing ML algorithms such as LR, MLP and LSTM	Lack of data optimization and dimensional reduction
S Bouktif, A Fiaz, M Awad, Amir Mosavi. 2020 [27]	IEEE Access 2020, Volume-8, pp. 40269–40282	Opinion mining and sentiment analysis was used along with historical stock prices for market prediction	No data filtering or optimization approach used
Jithin Eapen, Doina Bein, Abhishek Verma. 2019 [28]	2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)	A pipelined approach of CNNs along with a bi-directional LSTM model	Dimensional reduction and data optimization not used. Opinion mining not employed
Min Wen, Ping Li, Lingfei Zhang, Yan Chen. 2019 [29]	IEEE Access 2019, Volume 7, pp. 28299–28308	Stock market trend prediction using high-order information of time series. Cross entropy-based approach is used	Optimization in feature selection not applied which may increase computational cost for large datasets
Y Guo, S Han, C Shen, Y Li, X Yin, Y Bai. 2018 [30]	IEEE Access 2018, Volume-6, pp. 11,397–11,404	Adaptive support vector regression (SVR) model with weight updating mechanism based on particle swarm optimization	Opinion mining and filtering of data not employed. Performance SVR generally saturates after which adding more training data doesn't increase accuracy of the system
MS Raimundo, J Okamoto. 2018 [31]	2018 International Conference on Information and Computer Technologies (ICICT), IEEE 2018, pp. 111–114	Combination of discrete wavelet transform (DWT) and the support vector regression. The DWT was used as a data filtering tool	SVR's performance doesn't improve above a threshold. Opinion mining data not considered
Y Baek, HY Kim. 2018 [32]	Journal of Expert System and Applications, Elsevier 2018, Volule-113, pp. 457–480	An amalgamation of two LSTM layers performed. The first layer avoided the chances of overfitting while the second LSTM block was used for prediction	No data optimization and sentiment analysis data used

(continued)

Table 1 (continued)

Author(s)/year	Source/publisher	Findings and relevance	Research gap
S Selvin, R Vinayakumar, E. A Gopalakrishnan, Vijay Krishna Menon, K. P. Soman. 2017 [33]	International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE 2017, pp. 1643–1647	The approach used made predictions based on the daily closing price. The models used ARIMA, GARCH, LSTM, RNN and Sliding window CNN	Only daily closing price chosen as the time-dependent feature Very restricted feature set
M Billah, S Waheed, A Hanifa. 2016 [34]	International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), IEEE 2016, pp. 1–4	The Levenberg Marquardt (LM) back propagation algorithm was used to forecast closing stock prices of stocks for the Dhaka Stock Exchange	No data optimization and filtering mechanism used
GRM Lincy, CJ John. 2016 [35]	Journal of Expert Systems with Applications, Volume-44, Issue-C, ACM 2016	Multiple fuzzy sets were or multiple fuzzy inference systems. The system performed better than the adaptive neuro fuzzy inference system (ANFIS)	No data pre-processing and optimization. Global impacting parameters not considered
YE Cakra, BD Trisedya. 2015 [36]	2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), IEEE 2015, pp. 147–154	Opinion mining and sentiment analysis done based on Bayes classifier. The sentiment polarity was used along with previous stock prices as the input to linear regression model for prediction	Linear regression's accuracy doesn't increase for large data sets. No data filtration mechanism used
KN Devi, VM Bhaskaran. 2015 [37]	International Conference on Informatics, Electronics and Vision (ICIEV), IEEE 2013, pp. 1–6	Support vector regression (SVR) used for stock market prediction	SVR's saturation. No data filtration, optimization and sentiment analysis parameters included
Phayung Meesad, Risul Islam Rasel. 2013 [38]	International Conference on Informatics, Electronics and Vision (ICIEV), IEEE 2013, pp. 1–6	Support vector regression (SVR) used for stock market prediction	SVR's saturation. No data filtration, optimization and sentiment analysis parameters included

for the four different data sets is summarized subsequently. The performance of the system is evaluated in terms of the following metrics:

The parameters which can be used to evaluate the performance of the ANN design for time series models are given by:

1. Mean Absolute Error (MAE)
2. Mean Absolute Percentage Error (MAPE)
3. Mean square Error (MSE).

The above-mentioned errors are mathematically expressed as:

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|Y_t - \hat{Y}_t\right| \tag{9}$$

Or

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}|e_t| \tag{10}$$

$$\text{MAPE} = \frac{100}{N}\sum_{t=1}^{N}\frac{|Y_t - \hat{Y}_t|}{V_t} \tag{11}$$

$$\text{MSE} = \frac{1}{N}\sum_{t=1}^{N}e_t^2 \tag{12}$$

Here,

N denotes the number of samples in prediction.
Y is the predicted value of the variable.
$\widehat{Y}_t$ is the actual value of the variable.
e is the error value in each prediction.
The accuracy is generally computed as:

$$\text{Accuracy} = 100 - \text{error}(\%) \tag{13}$$

The actual and moulded prediction for the SBI dataset is shown for four prediction models which are:

(1) SVR
(2) ANN with back propagation
(3) LSTM
(4) ARIMA.

A comparative analysis of the approaches is presented subsequently (Figs. 6, 7, 8 and 9; Tables 2, 3, 4 and 5).

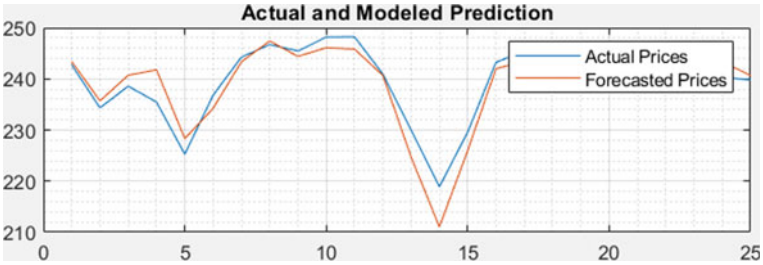


Fig. 6 Prediction using SVR

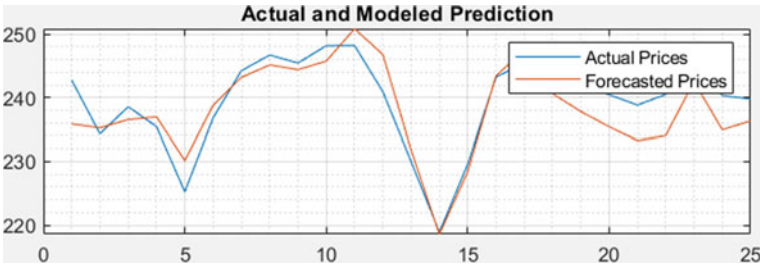


Fig. 7 Prediction using ANN with back propagation

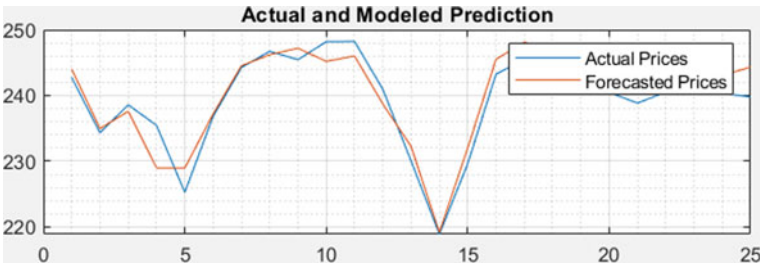


Fig. 8 Prediction using LSTM

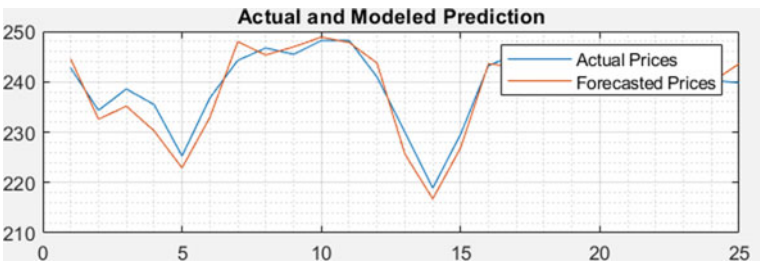


Fig. 9 Prediction using ARIMA

Table 2 Summary of results for SBI dataset

S. No.	Technique	MAPE (%)	Accuracy (%)	Regression
1	SVR	28.4	71.6	0.76
2	BP-ANN	23.72	76.28	0.82
3	LSTM	21.91	78.09	0.854
4	ARIMA	24.5	76.5	0.0837

Table 3 Summary of results for TCS dataset

S. No.	Technique	MAPE (%)	Accuracy (%)	Regression
1	SVR	16.2	83.8	0.91
2	BP-ANN	13.7	86.3	0.932
3	LSTM	12.81	87.19	0.944
4	ARIMA	15.2	84.8	0.928

Table 4 Summary of results for Infosys dataset

S. No.	Technique	MAPE (%)	Accuracy (%)	Regression
1	SVR	21.6	78.4	0.84
2	BP-ANN	19.7	80.3	0.862
3	LSTM	19.35	80.65	0.866
4	ARIMA	16.42	83.48	0.854

Table 5 Summary of results for Reliance dataset

S. No.	Technique	MAPE (%)	Accuracy (%)	Regression
1	SVR	24.2	75.8	0.815
2	BP-ANN	21.72	78.28	0.83
3	LSTM	19.6	80.4	0.861
4	ARIMA	18.66	81.34	0.879

A similar analysis has been done for the other stock prices. The summary of the results have been tabulated next. The parameters analysed are mean absolute percentage error (MAPE), accuracy of prediction and regression.

From the results obtained, it can be observed that the LSTM outperforms the other techniques for three datasets while the ARIMA attains the highest accuracy for the fourth dataset. Out of the SVR and back propagation BP-ANN, the latter exhibits higher prediction accuracy. Thus the preferred choice happens to be the LSTM deep neural network followed by the ARIMA, BP-ANN and SVR.

The results furnished are for the purpose of analyzing the working and performance of baseline techniques on benchmark datasets. The baseline techniques chosen

in this paper are SVR, BP-ANN, LSTM and ARIMA. The selection of the baseline techniques is based on the review of the literature which cites these four approaches as the commonly used approaches along with some of its derivatives. The natural progression to the existing baseline techniques can be:

(1) Adding feature optimization techniques to facilitate the training process [39].
(2) Applying data filtering so as to remove the random local fluctuations in the data thereby aiding pattern recognition [40].
(3) Applying a sliding window approach so as to add additional exogenous inputs to the training data to capture the latest trends in the data which may remain in-evident due to computation of large means [39, 41].

5 Conclusion

This paper presents a comprehensive review on stock market prediction as a time series problem which can be modelled using different regression techniques. An analysis of empirical models has been presented. Noteworthy contributions in the field are also presented along with the identified research gaps. The results presented are based on the application of benchmark methods such as ANN, SVR LSTM and ARIMA. The results exhibit the performance of the empirical methods. A comparison of common baseline techniques and their analysis has been furnished for the sake of comparison. The future directions of work may be design of approaches which can couple both filtration and deep neural networks. The research gap cited in the analysis of the contemporary work can be leveraged to design an approach which can address the shortcomings of the existing techniques. This two-fold approach comprising of data pre-processing/feature optimization and pattern recognition is can prove to be more effective compared to the existing baseline techniques.

References

1. S.P. Chatzis, V. Siakoulis, A. Petropoulos, E. Stavroulakisb, N. Vlachogiannakisb, Forecasting stock market crisis events using deep and statistical machine learning techniques. Expert Syst. Appl. **112**, 353–371 (2018)
2. R. Ray, P. Khandelwal, B. Baranidharan, A survey on stock market prediction using artificial intelligence techniques, in *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2018, pp. 594–598
3. D. Kumar, P.K. Sarangi, R. Verma, A systematic review of stock market prediction using machine learning and statistical techniques. Proc. Mater. Today (2021)
4. J. Shen, M.O. Shafiq, Short-term stock market price trend prediction using a comprehensive deep learning system. J. Big Data **7**, 66 (2020)
5. G.G. Rajput, B.H. Kaulwar, A comparative study of artificial neural networks and support vector machines for predicting stock prices in National Stock Exchange of India, in *2019 International Conference on Data Science and Communication (IconDSC)*, 2019, pp. 1–7
6. X. Wang, C. Wang, Time series data cleaning: a survey. IEEE Access **8**, 1866–1881 (2020)

7. M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, A. Mosavi, Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. IEEE Access **8**, 150199–150212 (2020)
8. J. Wen, J. Yang, B. Jiang, H. Song, H. Wang, Big data driven marine environment information forecasting: a time series prediction network. IEEE Trans. Fuzzy Syst. **29**(1), 4–18 (2020)
9. S. Majumdar, A.K. Laha, Clustering and classification of time series using topological data analysis with applications to finance. Expert Syst. Appl. **162**, 113868 (2020)
10. B. Weng, M.A. Ahmed, F.M. Megahed, Stock market one-day ahead movement prediction using disparate data sources. Expert Syst. Appl. **79**, 153–163 (2017)
11. M.R. Vargas, C.E.M. dos Anjos, G.L.G. Bichara, A.G. Evsukoff, Deep learning for stock market prediction using technical indicators and financial news articles, in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–8
12. S. Pryima, R. Vovk, V. Vovk, Using artificial neural networks to forecast stock market indices, in *2019 XIth International Scientific and Practical Conference on Electronics and Information Technologies (ELIT)*, 2019, pp. 108–112
13. Y. Xu, L. Chhim, B. Zheng, Y. Nojima, Stacked deep learning structure with bidirectional long-short term memory for stock market prediction, in *International Conference on Neural Computing for Advanced Applications NCAA 2020: Neural Computing for Advanced Applications* (Springer, 2020), pp. 447–460
14. H. Liu, Z. Long, An improved deep learning model for predicting stock market price time series. Digit. Signal Process. **102**, 102741 (2020)
15. B. Doucoure, K. Agbossou, A. Cardenas, Time series prediction using artificial wavelet neural network and multi-resolution analysis: application to wind speed data. Renew. Energy **92**, 202–211 (2016)
16. J. Quilty, J. Adamowski, Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. J. Hydrol. **563**, 336–353 (2017)
17. E.A. Maharaj, P. Teles, P. Brito, Clustering of interval time series. J. Stat Comput. **29**, 1011–1034 (2019)
18. Y. Wen, P. Lin, X. Nie, Research of stock price prediction based on PCA-LSTM model, in *IOP Conference Series: Materials Science and Engineering, 2nd International Conference on Communication, Network and Artificial Intelligence*, Guangzhou, China, 2019, vol. 790, pp. 1–7
19. J. Henriquez, W. Kristjanpoller, A combined independent component analysis–neural network model for forecasting exchange rate variation. Appl. Soft Comput. **83**, 105654 (2019)
20. R.K. Dash, T.N. Nguyen, K. Cengiz, A. Sharma, Fine-tuned support vector regression model for stock predictions. Neural Comput. Appl. **377**, 1–16 (2021)
21. M. Firdaus, S.E. Pratiwi, D. Kowanda, A. Kowanda, Literature review on artificial neural networks techniques application for stock market prediction and as decision support tools, in *2018 Third International Conference on Informatics and Computing (ICIC)*, 2018, pp. 1–4
22. R. Achkar, F. Elias-Sleiman, H. Ezzidine, N. Haidar, Comparison of BPA-MLP and LSTM-RNN for stocks prediction, in *2018 6th International Symposium on Computational and Business Intelligence (ISCBI)*, 2018, pp. 48–51
23. A. Moghar, M. Hamiche, Stock market prediction using LSTM recurrent neural network. Procedia Comput. Sci. **170**, 1168–1173 (2020)
24. F.A. Gers, D. Eck, J. Schmidhuber, Applying LSTM to time series predictable through time-window approaches, in *Neural Nets WIRN Vietri-01*. Perspectives in Neural Computing (Springer, 2020), pp. 193–200
25. F. Wang, M. Li, Y. Mei, W. Li, Time series data mining: a case study with big data analytics approach. IEEE Access **8**, 14322–14328 (2020)
26. S. Kim, S. Ku, W. Chang, J.W. Song, Predicting the direction of us stock prices using effective transfer entropy and machine learning techniques. IEEE Access **8**, 111660–111682 (2020)
27. S. Bouktif, A. Fiaz, M. Awad, A. Mosavi, Augmented textual features-based stock market prediction. IEEE Access **8**, 40269–40282 (2020)

28. J. Eapen, D. Bein, A. Verma, Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction, in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, 2019, pp. 0264–0270
29. M. Wen, P. Li, L. Zhang, Y. Chen, Stock market trend prediction using high-order information of time series. IEEE Access **7**, 28299–28308 (2019)
30. Y. Guo, S. Han, C. Shen, Y. Li, X. Yin, Y. Bai, An adaptive SVR for high-frequency stock price forecasting. IEEE Access **6**, 11397–11404 (2018)
31. M.S. Raimundo, J. Okamoto, SVR-wavelet adaptive model for forecasting financial time series, in *2018 International Conference on Information and Computer Technologies (ICICT)*, IEEE, 2018, pp. 111–114
32. Y. Baek, H.Y. Kim, ModAugNet: a new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. J. Expert Syst. Appl. **113**, 457–480 (2018)
33. S. Selvin, R. Vinayakumar, E.A. Gopalakrishnan, V.K. Menon, K.P. Soman, Stock price prediction using LSTM, RNN and CNN-sliding window model, in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, 2017, pp. 1643–1647
34. M. Billah, S. Waheed, A. Hanifa, Stock market prediction using an improved training algorithm of neural network, in *2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*, IEEE, 2016, pp. 1–4.
35. G.R.M. Lincy, C.J. John, A multiple fuzzy inference systems framework for daily stock trading with application to NASDAQ stock exchange. J. Expert Syst. Appl. **44**(C) (2016)
36. Y.E. Cakra, B.D. Trisedya, Stock price prediction using linear regression based on sentiment analysis, in *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, IEEE, 2015, pp. 147–154
37. K.N. Devi, V.M. Bhaskaran, Cuckoo optimized SVM for stock market prediction, in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, IEEE, 2015, pp. 1–5
38. P. Meesad, R.I. Rasel, Predicting stock market price using support vector regression, in *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, IEEE, 2013, pp. 1–6
39. J.J. Kurian, M. Dix, I. Amihai, G. Ceusters, A. Prabhune, BOAT: a Bayesian optimization AutoML time-series framework for industrial applications, in *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*, 2021, pp. 17–24
40. A.U. Haq, A. Zeb, Z. Lei, D. Zhang, Forecasting daily stock trend using multi-filter feature selection and deep learning. Expert Syst. Appl. **168**, 114444 (2021)
41. J. Li, B. Yang, H. Li, Y. Wang, C. Qi, Y. Liu, DTDR–ALSTM: extracting dynamic time-delays to reconstruct multivariate data for improving attention-based LSTM industrial time series prediction models. Knowl. Based Syst. **211**, 106508 (2021)

Stock Market Prediction Employing Discrete Wavelet Transform and Moving Average Gradient Descent

Dinesh Singh Dhakar and Savita Shiwani

Abstract Stock market behavior is extremely volatile and complex in nature due to the randomness of the governing parameters. This makes attaining a high degree of accuracy in stock market forecasting extremely challenging. Several techniques have been explored to forecast stock market behavior while aiming to mitigate the effects of noisy data sets, nonlinearity and randomness in the data. In general, the inherent nature of the stock data to be noisy containing sudden fluctuations and spikes makes pattern recognition extremely challenging and prone to errors. In this paper, the complex values wavelet transform along with the gradient descent approach has been proposed for stock market forecasting. The wavelet transform has been employed as an iterative sampling filter to remove the effects of noise in the data. The gradient descent algorithm has been used to train a deep neural network using the filtered data. The neural network is fed with approximate and detailed coefficient values of the decomposed data so as to find the patterns in each of the levels of decomposition while leaving out the noisy effects at each level of decomposition. This multi-level input vector augments the capability of the gradient decent algorithm to evaluate the actual effects of noise in the data set. The day-wise data for shares has been considered with opening prices, closing prices, average price, and volume as the temporal parameters. It has been shown that the proposed approach is capable of analyzing noisy data samples and attains an accuracy higher than contemporary approaches, which makes it a promising technique for stock market forecasting.

Keywords Stock market prediction · Transform · Multi-level decomposition · Moving average · Gradient descent · Accuracy

D. S. Dhakar (✉) · S. Shiwani
Department of Computer Science and Engineering, School of Engineering and Technology (SOET), Jaipur National University, Jaipur, India
e-mail: dineshhodcsemit@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_56

1 Introduction

Stock markets have been a vital area of research with respect to economy. The evaluation of stock prices and stock markets is important to understand the changing dynamics of market investments. It is crucial for the investors and policy makers to predict the stock prices and gauge the movement of stock values. It is essential for investment planning and investment-based profits. Stock investment is a substantial financial activity that can lead to huge profits or losses. Stock market predictions based on the stock price prediction are a key aspect for share market analysis and planning [1]. The volatile nature of the share market is a major challenge that needs to be focused. Stock market prediction is of paramount importance in this context, which aims at predicting the future stock values. Accuracy in prediction of stock market prices can aid in creating profitable investment strategies. A robust prediction model can help in forecasting accurate stock values.

There have been rapid developments in this area of research with the use of artificial intelligence and machine learning methods [2]. The domain of artificial intelligence has helped in understanding and analyzing the financial markets in an effective manner. There has been a widespread increase in the use of machine learning backed methods for efficient evaluation of financial markets. The technical analysis in prediction of stock market investments based on past data of share market can be done using efficient machine learning-based methods [3]. This domain has been an active area of research for accurate prediction of stock values based on machine learning-based methods. While different machine learning approaches have been explored to forecast stock market behavior accurately, the nonlinearity, extremely high randomness and non-stationarity of the noisy data sets render inaccuracies in forecasting for conventional machine learning algorithms used for time series forecasting problems [4]. The paper aims at devising a technique for stock market forecasting employing statistical data filtering coupled with machine learning. For that purpose, the wavelet transform is utilized, as a transform domain iterative filter along with principal component analysis for data optimization. The gradient descent approach is employed subsequently for training.

2 Methodology

The methodology presented aims at fulfilling the objectives of the research which are data filtration and optimization followed by training a machine learning algorithm. The data preprocessing and optimization are done using the principal component analysis (PCA) and discrete wavelet transform (DWT). The gradient boost algorithm is used for training. The methodology and contribution are discussed subsequently [5].

2.1 Data Preprocessing

The data utilized for the present experimental study is the SBI, Infosys and Reliance stock prices listed in the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE), India. The data preparation and processing techniques bolster the training algorithms for stock market forecasting. The techniques used in this paper are the principal component analysis (PCA) and the discrete wavelet transform (DWT). The PCA has been used as data optimization and dimensional reduction tool for time series forecasting problems. In the context of this paper, it is particularly useful to segregate the fluctuating component of the data from the predictable component of the data, which is effective to filter out noisy data samples from the data samples of predictable coherent nature [6]. The application of the PCA on the raw data allows to dimensionally reduce the data into a more coherent and compact data vector with lesser dimensions. This allows to lessening the perturbations in the data employing truncation of the dimensions which exhibit relatively lesser coherence. The data representation of the target vector can be given by

$$T = T(t), T(t-k), \ldots, T(t-nk) \tag{1}$$

Here,

T is the composite target vector,
t is the time variable,
k is the delay or lag variable, and
n is the number of lags.

The target vector is generally influenced by several governing parameters or features which are often nonlinear and non-stationary in nature. The PCA can be used to generate a dimensionally reduced set of the data vector and in the process remove the uncorrelated data variables in the noise spectrum of the data [7]. The filtered data can be expressed as follows:

$$T_r = P_r^{n,m} S_r^{n,m} \Delta_r \tag{2}$$

Here,

T_r is the truncated and dimensionally reduced data vector,
$P_r^{n,m}$ is the principal components,
$S_r^{n,m}$ is the reduced singular matrix, and
Δ_r is the PCA matrix.

2.2 The Discrete Wavelet Transform

The next step is the decomposition of the PCA reduced data using the discrete wavelet transform. The discrete wavelet transform is a sampled version of the continuous

wavelet transform which can be used for filtering time series data. The DWT acts as a multi-level filter, with even length scaling and shifting operations acting as high- and low-pass filtering operations [8]. DWT is a recursive pyramidal filter which breaks down the function to be transformed into the approximate and detailed coefficient values [9]. While the detailed coefficient values would contain the noisy disturbances, the approximate coefficient values would contain the intrinsic patterns in the data except the noisy component. Thus, iteratively applying the DWT, discarding the detailed coefficient values and retaining the approximate coefficient values would allow to filter out the noisy component of the data. While some data loss may be incurred in the process, it would outweigh the detrimental effects of the disturbances in the data [10, 11]. Mathematically, it can be summed up as follows:

$$X(t) \xrightarrow{S,W} \text{Coefficients(Approx, Detailed)} \tag{3}$$

Here,

Approx. represents the approximate coefficient values,
Detailed represents the detailed coefficient values,
S and W are the scaling and wavelet filters of the DWT, and
X is the time domain samples of the data.

2.3 Training Algorithm

The raw data is structured and processed using the PCA and the DWT which yields the actual training data. The data features used in this study are date, previous day closing price, present-day opening price, volume (swing), and highest and lowest price of the day [13, 14]. The training algorithm employed here is the back propagation-gradient descent. Following a standard convention, 70% of the data is utilized for training the neural network and 30% is used for testing. The back propagation mechanism is chosen since it is found to yield relatively good results for temporal prediction problems. In this approach, the data vector is applied to the network to be trained, and the error or objective function is computed [15]. Recursive trainings or iterations are performed based on the following constraints:

(a) The objective function is minimized, and the values become stationary for multiple iterations consecutively.
(b) The objective function does not attain stationarity, but the maximum number of iterations defined has been reached.

Satisfying either of the two conditions results in the termination of the training process. The back propagation fundamentally utilizes the chain rule given by

$$w_{k+1} = w_k - \alpha \frac{\partial e}{\partial w} \tag{4}$$

Here,
w_{k+1} is the weight of the next iteration,
w_k is the weight of the present iteration,
e is the error, and
α is the learning rate.

$$\frac{\partial e}{\partial w} = \frac{\partial e}{\partial y} \cdot \frac{\partial y}{\partial w} \tag{5}$$

The chain rule in relation (12) can be used for computing the error gradient. Summarizing the back propagation algorithm employed, the following steps are to be employed:

$$T^1 = P \tag{6}$$

Here,
P is the initial training vector,
T is the feed-forward vector, and
The superscript 'i' represents the initial state of the network.
Next, the propagation of the training vector through the network would yield:

$$T^{m+1} = f^{m+1}\left(W^{m+1}a^m + b^{m+1}\right) \forall m \in 0, \quad 1, \ldots, M-1 \tag{7}$$

At each step, the following vector is updated:

$$a = a^m \tag{8}$$

Here,
m is the iteration number,
f is the activation function,
W is the weight of the network, and
b is the bias of the network.
Subsequently, the sensitivities of the network are to be fed back to the next iteration as a composite training vector, given by [16]:

$$S^M = -2F^M n^M (t-a) \tag{9}$$

Here,
S denotes the sensitivity,
t denotes the time metric, and
a denotes the delay or lag after which the sensitivity is computed in the network.
Typically, the error of iteration 'm' along with the error gradients is utilized as the variables for the sensitivity vector in feedback, i.e.,

$$e_m, \frac{\partial y}{\partial x} \in S \tag{10}$$

The values of the weight and bias are generally updated as follows:

$$W_k = W_{k-1} - \alpha S^m a^{m-1^T} \tag{11}$$

$$b_k^m = b_{k-1}^m - \alpha S^m \tag{12}$$

Here,

α denotes the learning rate,

k denotes the subsequent iteration, and

$k - 1$ denotes the previous iteration.

A parallel search for the learning rate is carried out to minimize the objective function along the direction of search of the error gradient in each iteration. Mathematically, it is given by

$$R_k = R_{k-1} = \alpha P_k \tag{13}$$

Here,

R is the scaling step,

α is the learning rate, and

P_k is the search vector for iteration 'k'.

The scaled version of the steepest descent converges in 'n' iterations for a quadratic objective function typically the mean squared error. However, after the 'n' iterations are over, a repetitive cycle of scaled gradients is employed.

As data is fed to a neural network for pattern recognition, the weights keep updating. However, it has been found that in case of time series problems, the latest data samples have the maximum impact on the latest output. Hence, it is logical to calculate a moving average of latest (previous) data and apply it to the neural network [17]. This is also called a moving average. Mathematically,

$$I_k = X_{1,k}, \text{Mean}(X)_{k,k-n}, Y_k \tag{14}$$

Here,

I_k is the kth input sample to the neural network,

$X_{1,k}$ are the data samples from the first to the kth sample,

$\text{Mean}(X)_{k,k-n}$ is the mean of the data samples form $k - n$ to k; i.e., it is a moving average depending on the value of k, and

Y_k is the target.

Thus, a moving average of the C_A and C_A values can be computed after the application of the PCA. The next step would be creating a new training vector compising the follwing variables:

$$\text{Tr} = \left[X1_{\text{CA,CD}}, X2_{\text{CA,CS}}, \ldots, Xn - 1_{\text{CA,CD}} Xn_{\text{CA,CD}}, \text{Avg}_{n-k}, Y\right] \quad (15)$$

Here,

Tr is the training vetcor,

Y is the target vector,

$X1_{\text{CA,CD}}, X2_{\text{CA,CS}}, \ldots, Xn - 1_{\text{CA,CD}} Xn_{\text{CA,CD}}$ are the individual decomposed values of the features using the DWT iteratively, and

Avg_{n-k} is the moving average of the variables.

The moving average would allow the machine learning algorithm to find recent patterns in the data along with overall historical data [18, 19]. The final data vector would thus contain the DWT coefficients of the decomposition along with the moving average exogenous input. Gradient boosting is then proposed to compute the final prediction output as a summation of the outputs from the individual learning models [20]. The performance evaluation of the proposed model is done based on the evaluation of the following parameters [21, 22]:

(1) Mean absolute percentage error (MAPE)
(2) Mean square error (MSE)

$$\text{MAPE} = \frac{100}{N} \sum_{t=1}^{N} \frac{|V_t - \hat{V}_t|}{V_t} \quad (16)$$

$$\text{MSE} = \frac{1}{N} \sum_{t=1}^{N} e_t^2 \quad (17)$$

Here,

N is the number of predicted samples,

V is the predicted value,

$\hat{V}_t$ is the actual value, and

e is the error value.

The next section discusses the obtained results.

In general, the steepest descent is the simplest to implement, but is slow in convergence. Hence, we employ the scaled conjugate variant of the steepest descent to speed up the convergence of the objective function [23, 24]. This needs the computation of a search vector for each iteration 'k' which would point to a scaled gradient with the maximum error change rate with respect to weights [25]. It is given by

$$g_k = \max\{g\} \forall\ W, \quad e: \text{ at iteration } k \quad (18)$$

The truncation of the algorithm is based on the fulfillment of either of the two conditions (whichever occurs earlier) [26]:

1. The cost function is stable for at least 5–6 epochs.
2. The maximum epochs pre-defined are exhausted.

3 Experimental Results

The data has been extracted from https://in.finance.yahoo.com/quote. Three data sets have been used for the analysis of the proposed algorithm, which are the SBI, Infosys and Reliance. The features selected are the date, opening price, closing price, mean price of the day, maximum price and minimum price. The day-wise parameters are modeled to affect and govern the final share prices. The data processing and prediction methodologies are explained subsequently using the obtained results (Fig. 1).

After the dimensional reduction of the data using PCA, the DWT is applied iteratively to obtain further filtering of the data. The data is decomposed into the approximate coefficient and the three-level detailed coefficient values. Here '*s*' represents the approximate coefficient value, while d1, d2, d3 and d4 represent the detailed coefficient values. The noise spectrum can be analyzed using the histogram analysis of the data. An analysis of the recursive wavelet decomposition is given in Table 1. Table 1 clearly indicates that there is a distinct similarity between the histogram of the original data and the approximate coefficient values of the data, whereas the detailed coefficient values bear a clear difference in both magnitude, distribution and the polarity of the values. This clearly indicates that the noise and disturbance in the noise floor affect the detailed coefficient values much more compared to the approximate coefficient values. A similar analysis has been done for the Infosys and Reliance data sets.

The *R* (regression) values are depicted in Fig. 2 for the training, testing, validation and average overall cases. It can be observed that the proposed system attains an average regression of 93%. A high value of regression indicates the closeness in the forecasted and actual values (Fig. 3).

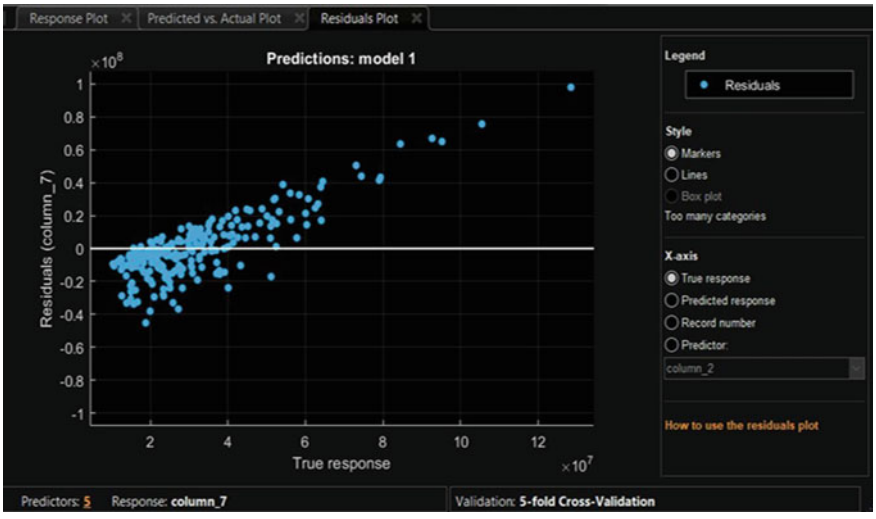


Fig. 1 Scatter plot of the data after application of PCA (SBI data set)

Table 1 Statistical analysis of data

S. No.	Parameter	Values	Class
1	Minimum	8.628×10^4	Original data
2	Maximum	8.807×10^7	
3	Mean	2.756×10^6	
4	Median	2.068×10^6	
5	Standard deviation	4.308×10^6	
6	Mean absolute deviation	6.861×10^6	
7	Minimum	3.383×10^6	Approximate coefficient values
8	Maximum	3.858×10^7	
9	Mean	7.796×10^6	
10	Median	6.501×10^6	
11	Standard deviation	5.152×10^6	
12	Mean absolute deviation	1.194×10^6	
13	Minimum	-3.304×10^7	Detailed coefficient values
14	Maximum	7.22×10^6	
15	Mean	-8.081×10^5	
16	Median	-2.9×10^5	
17	Standard deviation	4.732×10^6	
18	Mean absolute deviation	9.876×10^5	

Three different data sets have been used in the proposed work which are that of SBI, Infosys and Reliance share prices, obtained from [1]. The actual and modeled forecasting values are depicted in Fig. 3. The mean absolute percentage error of the system is found to be 8.72% for the SBI data set. This yields an accuracy of 91.28% which is relatively high compared to the existing literature [1–3] and more recent hybrid techniques in [12–14]. A similar analysis has been adopted for the Infosys and Reliance data sets. The MAPE and regression values for the Infosys and TCS data sets are 11.22, 0.915, and 13.15, 0.91, respectively. The values of the MAPE and accuracy are suggestive of the fact that the proposed system is capable of filtering out the noisy values from the original noise floor and forecasting the stock prices with relatively high accuracy. This statistical analysis of the approximate and detailed coefficient values also indicates the same as the histogram of the detailed coefficient values is significantly w.r.t. to the original data, while the detailed coefficient values are convergent with the actual data. This indicates that the noise effects have been removed by iterative filtering employing the DWT. The scaled version of the neural network used has ten hidden layers with a log-sigmoid activation function trained with

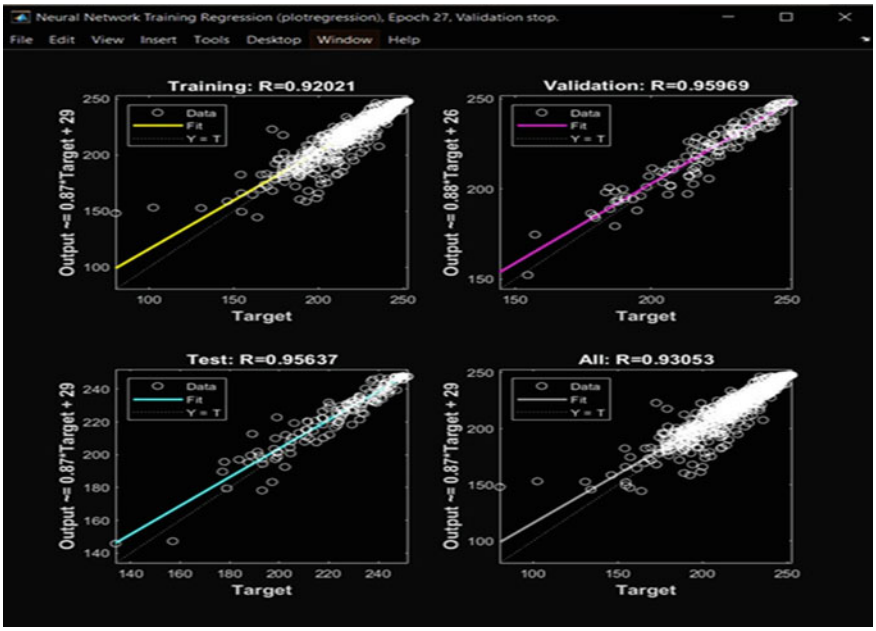


Fig. 2 Regression of the proposed model

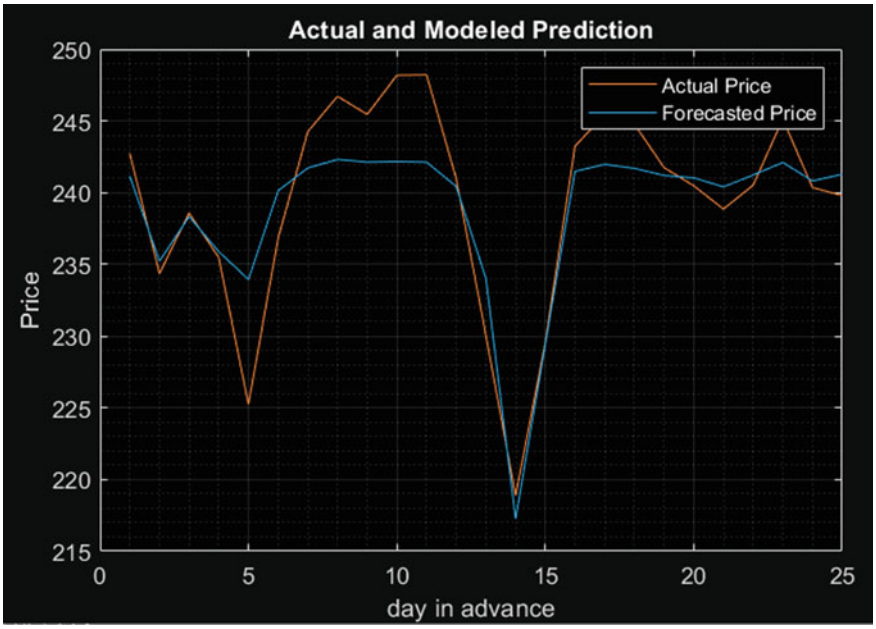


Fig. 3 Forecasting for SBI data set

Table 2 Summary of MAPE and regression values

S. No.	Data set	MAPE	Iterations to convergence	Regression (overall)
1	SBI	8.72	27	0.93
2	Infosys	11.22	38	0.915
3	Reliance	13.15	31	0.91

Table 3 Summary of comparative average accuracy

S. No.	Technique	Accuracy (%)
1	Transfer entropy and machine learning [1]	57
2	Augmented textual feature-based learning [2]	60
3	LSTM with sentiment analysis [3]	49.6
4	HFS-based XBoost [12]	79
5	Hybrid red deer-gray algorithm [13]	85.2
6	Variational autoencoders (VAE) [14]	67
7	**Proposed technique (mean accuracy)**	88.97

the scaled version of the gradient descent. This allows faster convergence compared to the conventional gradient descent (Table 2).

A comparative analysis with existing work in the domain is presented in Table 3.

4 Conclusion

This paper presents a stock market forecasting model based on recursive DWT decomposition and gradient boost algorithm. A third-level decomposition of the data is done, and subsequent statistical analysis is performed to correlate the noise floor with the actual data to be analyzed. It has been successfully shown that discarding the detailed coefficient values helps in data cleaning, and retaining the approximate coefficient values results in subsequent accurate pattern recognition in the data. The performance of the system has been evaluated in terms of the regression, mean absolute percentage error and accuracy of the system. The system attains an average accuracy of 88.97 which is significantly higher compared to existing literature. The novelty of the approach lies in the fact that the approach uses the gradient boosting methodology for all the coefficient values of the DWT decomposition as training values. Moreover, the moving average acts as an additional input the algorithm to find recent patterns in the data.

References

1. S. Kim, S. Ku, W. Chang, J.W. Song, Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques. IEEE Access **8**, 111660–111682 (2020)
2. S. Bouktif, A. Fiaz, M. Awad, A. Mosavi, Augmented textual features-based stock market prediction. IEEE Access **8**, 40269–40282 (2020)
3. X. Li, P. Wu, W. Wang, Incorporating stock prices and news sentiments for stock market prediction: a case of Hong Kong. Inf. Process. Manag. **57**(5), 1–19 (2020)
4. G. Bansal, V. Hasija, V. Chamola, N. Kumar, M. Guizani, Smart stock exchange market: a secure predictive decentralized model, in *2019 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2019, pp. 1–6
5. P.D. Yoo, M.H. Kim, T. Jan, Machine learning techniques and use of event information for stock market prediction: a survey and evaluation, in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, IEEE, 2006, 2005, pp. 835–841
6. B.M. Henrique, V.A. Sobreiro, H. Kimura, Literature review: machine learning techniques applied to financial market prediction. J. Expert Syst. Appl. **124**, 226–251 (2019)
7. L. Zhu, Y. Wang, Q. Fan, MODWT-ARMA model for time series prediction. J. Appl. Math. Model. **38**(5), 1859–1865 (2014)
8. M. Awad, R. Khanna, Support vector regression, *Efficient Learning Machines* (Springer, 2015), pp. 67–80
9. X. Zhang, Y. Hu, K. Xie, S. Wang, E.W.T. Ngai, M. Liu, A causal feature selection algorithm for stock prediction modeling. J. Neurocomput. **142**, 48–59 (2014)
10. L. Hong, Multiresolutional filtering using wavelet transform. IEEE Trans. Aerosp. Electron. Syst. **29**(4), 1244–1249 (1993)
11. K.A. Althelaya, S.A. Mohammed, E.-S.M. El-Alfy, Combining deep learning and multiresolution analysis for stock market forecasting. IEEE Access **9**, 13099–13111 (2021)
12. N. Naik, B.R. Mohan, Novel stock crisis prediction technique—a study on Indian stock market. IEEE Access **9**, 86230–86242 (2021)
13. S.S. Alotaibi, Ensemble technique with optimal feature selection for Saudi stock market prediction: a novel hybrid red deer-grey algorithm. IEEE Access **9**, 64929–64944 (2021)
14. H. Gunduz, An efficient stock market prediction model using hybrid feature reduction method based on variational autoencoders and recursive feature elimination. Financ. Innov. **7**(28) (2021)
15. Y. Du, Application and analysis of forecasting stock price index based on combination of ARIMA model and BP neural network, in *2018 Chinese Control and Decision Conference (CCDC)*, 2018, pp. 2854–2857
16. M.K. Kim, Y.S. Kim, J. Srebric, Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: artificial neural network vs. linear regression. Sustain. Cities Soc. **62**, 102385 (2020)
17. D. Shah, H. Isah, F. Zulkernine, Stock market analysis: a review and taxonomy of prediction techniques. Int. J. Financ. Stud. **7**(2), 1–26 (2019)
18. M. Naeini, H. Taremian, H. Hashemi, Stock market value prediction using neural networks, in *2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, IEEE, 2010, pp. 132–136
19. F.A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: continual prediction with LSTM, in *9th International Conference on Artificial Neural Networks: ICANN '99* (IET Publications, 1999), pp. 850–855
20. Q. Gu, Y. Chang, N. Xiong, L. Chen, Forecasting nickel futures price based on the empirical wavelet transform and gradient boosting decision trees. Appl. Soft Comput. **109**, 107472 (2021)
21. W. Chen, Y. Zhang, C.K. Yeo, C.T. Lau, B.S. Lee, Stock market prediction using neural network through news on online social networks, in *2017 International Smart Cities Conference (ISC2)*, 2017, pp. 1–6

22. H. Gunduz, Z. Cataltepe, Y. Yaslan, Stock market direction prediction using deep neural networks, in *2017 25th Signal Processing and Communications Applications Conference (SIU)*, 2017, pp. 1–4
23. X. Pang, Y. Zhou, P. Wang, W. Lin, V. Chang, An innovative neural network approach for stock market prediction. J. Supercomput. **76**, 2098–2118 (2020)
24. M. Borhani, Corpus analysis using relaxed conjugate gradient neural network training algorithm. Neural Process. Lett. **50**, 839–849 (2019)
25. P.K. Upadhyay, C. Nagpal, SCG backpropagation based prediction of stressed EEG spectrum, in *2020 Advances in Science and Engineering Technology International Conferences (ASET)*, 2020, pp. 1–5
26. I. Kuzborskij, C. Szepesvári, Nonparametric regression with shallow overparameterized neural networks trained by GD with early stopping, in *Conference on Learning Theory, 2021, Proceedings of Thirty Fourth Conference on Learning Theory*, PMLR, vol. 134, pp. 2853–2890 (2021)

Smart Monitoring and Quality Control in Leather Processing Industry

N. P. G. Bhavani, B. Deepalakshmi, K. Sujatha, and C. Tamilselvi

Abstract There is currently no trustworthy method in place to provide continuous monitoring of all process conditions and leather qualities inside closed reactors during leather production, implying that there is no genuine control. Only by pausing the reactor and sampling the solution and leather can conditions inside the reactors be checked. The leather industry's automation systems are based on a dated centralised architecture, which creates a critical point of failure and an operational bottleneck. To increase the efficiency of the leather manufacturing process, the paper offers a fault-tolerant multi-agent system (MAS) architecture that delivers the high flexibility and agility required by the leather industry's turbulent environment. The quality of leather can be determined by image structure, and the quality can be estimated using image processing.

Keywords Multi-agent systems · Leather processing · Tanning · Leather industry · Image processing · Smart monitoring

1 Introduction

The transformation of decayable raw skins and hides into a stable substance called leather, which is used to make a wide range of commodities, is the main activity of the leather industry (shoes, bags, garments, upholstery). The quality, cost, and

N. P. G. Bhavani (✉)
Department of ECE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India
e-mail: bhavaninpg.sse@saveetha.com

B. Deepalakshmi
Department of EEE, Ramco Institute of Technology, Chennai, India

K. Sujatha
Department of EEE, Dr. MGR Educational and Research Institute, Chennai, India

C. Tamilselvi
Department of IT, Dr. MGR Educational and Research Institute, Chennai, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_57

environmental effect of leather processing are all determined by a series of chemical and mechanical activities that are all interconnected. Leather production necessitates a considerable amount of water in a precise volume, at the correct temperature, and at the correct time. During leather processing, a variety of chemicals, both solid and liquid, are employed to produce a uniform grade leather. In Fig. 1, a systemic overview of leather production processes is depicted, including inputs, outputs and main disturbances.

The leather industry is listed amongst the "World's Worst Pollution Problems" [1]. During leather processing activities, poor absorption and excessive use of chemicals produce pollutants that are hazardous to both human health and the environment, negatively impacting all forms of life. Chromium (III), the "ideal" tanning agent in terms of leather performance [1], is the most widely used tanning agent, accounting for roughly 85% of all leather produced worldwide [2], although it can be transformed into chromium (VI), which is poisonous and carcinogenic [3] due to complex mechanisms.

Tanneries, or leather processing plants, are hostile and corrosive environments with a high level of complexity and diversity, containing a significant number of

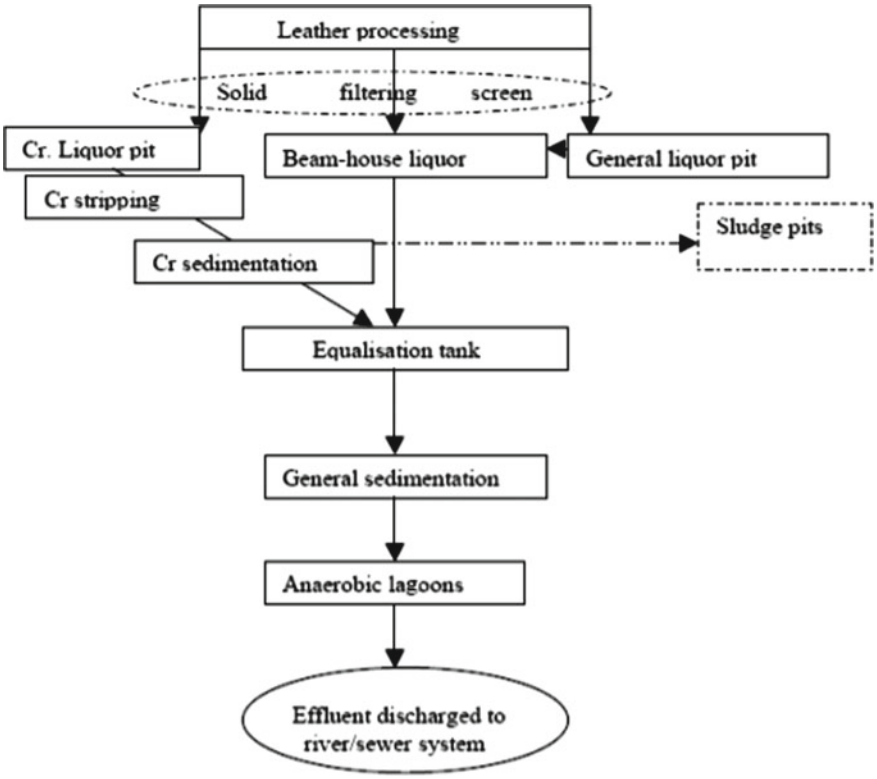


Fig. 1 Flow diagram illustrating effluent treatment in a typical leather processing

subsystems that are geographically dispersed. Traditional tannery control systems are built on a monolithic structure [4], which combines all control functions into a single central controller that is in charge of global production planning and scheduling. In the case of stable, deterministic industrial systems, this strategy enables long-term production optimisation, but it lacks agility and adaptability in the face of unanticipated disruptions [5]. However, disturbances are unavoidable in tanneries due to the structure and features of raw hides and skins, the complexity of processing processes, demand volatility, equipment malfunction, and raw material availability, to name a few factors. The author claims in [6] that it is possible to forecast the outcome of leather processing procedures to a large extent. As a result, a control system that closely monitors processes and takes corrective action is required. The application of AI approaches is necessitated by the lack of mathematical models that adequately reflect reality, the nature and diversity of raw materials, discontinuity in processing processes, and, last but not least, the nonlinear behaviour and time-variance of leather processing procedures. The study proposes a multi-agent method to process management with the goal of eliminating the drawbacks of rigid, centralised production control and providing the following benefits: (1) automated, adaptable, and optimal production planning, scheduling, and monitoring, (2) reduced chemical and water use, (3) less exposure to human operators to hazardous chemicals, and (4) consistent quality standards. Reference [7] provided an overview of leather manufacturing processes and the motivation for using an agent-based strategy in the leather sector; this paper is a continuation, describing an agent architecture and how agents can be used to regulate processes. References [8–11] have reported on the use of multi-agent technology in batch process automation. Multi-agent systems are based on the decentralisation of control functions across a group of dispersed autonomous, cooperative entities known as agents (Fig. 2).

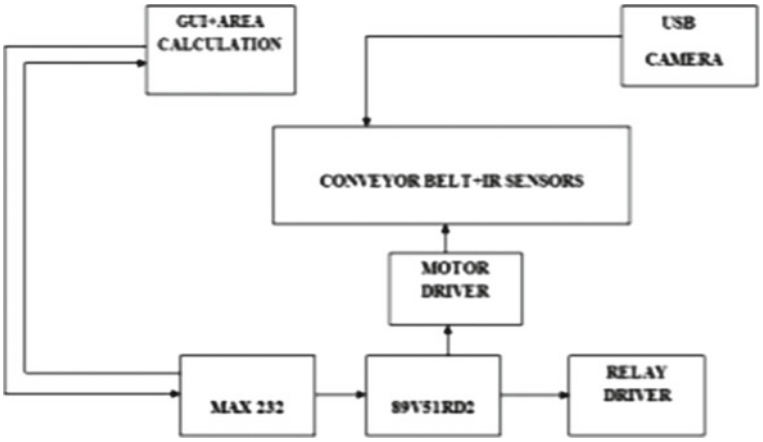


Fig. 2 Leather monitoring system

2 Materials and Methods

Processes and technologies for leather processing Leather processing technologies and standards differ dramatically from tannery to tannery. Leather processing procedures have been passed down through generations, and each tannery has its own set of recipes for producing the highest quality leather [12]. Using non-standard recipes and techniques based on the knowledge and talent of leather technicians is a fairly frequent practise, particularly in poor nations, but not exclusively so. The range of hides and skins ranging from light to heavy, unknown preservation methods, questionable provenance, and "blind" process monitoring of closed reactors have all contributed to this scenario. Chemical dosing, water batching, and pH value adjustment are still done by hand in many tanneries throughout the world [13], resulting in operational inefficiencies, variable leather quality, and human workers being exposed to harmful chemicals. The chemical reactor, also known as a drum, is the most typical piece of equipment used in the leather industry for wet leather processing (pre-tanning, tanning, and post tanning). Programmable Logic Controllers (PLCs) are used in modern chemical reactors to perform the following fundamental functions: (1) automatic float temperature measurement, (2) automatic pH value measurement, (3) management of all processing durations, (4) reactor rotations at various speeds these automatic functionalities can be supplemented using third-party solutions. The SCADA technique, for example, is a centralised control system that monitors and operates chemical reactors via an industrial server providing processing instructions to the reactor's PLCs via an industrial network. According to [4, 14], the SCADA system provides information on plant productivity, time of run, time interval of stop, scheduled actions, unscheduled occurrences (errors), breakdowns, machine maintenance reminders, and the trend of production parameters such as pH, temperature, pressure, volumes, number of leathers processed, and area thickness. SCADA systems, on the other hand, lack flexibility, scalability, and resilience to failure or assault [14]. This is mostly due to SCADA's centralised approach, which acts as a single point of failure and creates operational bottlenecks in the system. As a result, if the industrial server fails, the entire system will suffer. Such type of faults can avoid by using image processing and smart monitoring.

3 Results and Discussion

Automated testing will become even more successful in the future as quality criteria continue to rise and image processing technologies improve in performance. The design of a test system like this differs depending on the type of product and the needs. Two- to four-line cameras with specific lenses and as many illuminations are typically utilised. A camera usually works in incident light, whereas the other cameras work in the dark area, with the direction, angle, and colour of the illumination varying. A powerful evaluation unit is required because to the massive volume of

data (depending on configuration, several terabytes per hide). PCs with multi-core processors presently meet these standards the best. Robust assessment algorithms are critical for textured surfaces. The evaluation can be broken down into the four steps below: Preprocessing, Defect Detection, Clustering, and Defect Classification are all steps in the defect classification process.

3.1 Preprocessing

Defect identification on leather surfaces is complicated by the uneven natural structure of the leather. Extensive pre-smoothing operations are required to discern between defects (such as scratches) and natural imperfections (such as veins). This makes it easier to spot these flaws in later steps and boosts the data's robustness against data disturbances (e.g. noise). Shock filters, on the other hand, are employed to preserve and strengthen surface flaws in the image.

3.2 Defect Detection

Finding edges is a crucial step in spotting problems. An edge detector can be used to find them in the image data. More complicated edge detection algorithms have been created because basic gradient-based edge detectors do not deliver the required result. The foundation is a wavelet-scaled image. Several scales are used to detect the edges. Structures in the image are not only localised, but their size and shape are also determined. Topological gradients, particularly tailored erosion and dilatation, and geodetic reconstruction are also employed.

3.3 Clustering

The approaches mentioned are so-called detection methods, which are algorithms that detect possible hidden problems. Due to the surface structure, these methods also produce pseudo-defects. Potential flaws must be classified in a subsequent phase in order to be eliminated. Because direct classification is too inaccurate, another method, known as "clustering," was developed. Clustering methods create a similarity hierarchy of all probable flaws discovered using descriptors. This is a totally automated process that does not require any human input. The agglomerative hierarchical clustering compares and organises all discovered flaws in pairs.

3.4 Defect Classification

The clustering result is now used for classification, with the created hierarchy being included in the list of descriptors. The quality of the categorisation is governed by the descriptors, which characterise the defect characteristics determined from the image (Fig. 3).

The classification is usually done using a training record in which the flaws are manually recorded. The procedure of creating the training data record is complicated because the data record must be sufficiently large on the one hand and as consistent as possible on the other. Consistent indicates that only genuine flaws are noted, but that no flaws are overlooked.

There are many different types of classifiers, but the most common are support vector machines. The classifiers are run numerous times and not only identify between defects and faux defects, but also provide information about the defect kind, such as scratches, pigmentation, warts, and filth (Fig. 4).

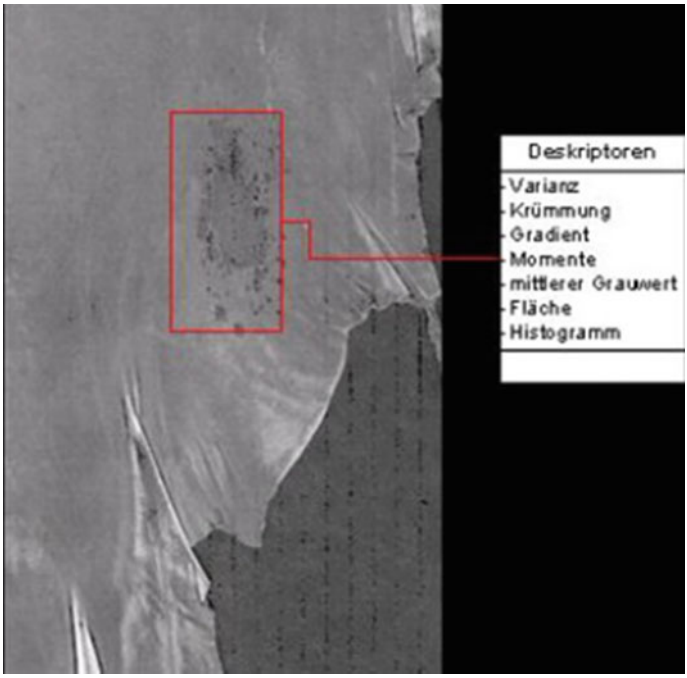


Fig. 3 Image of defect classification

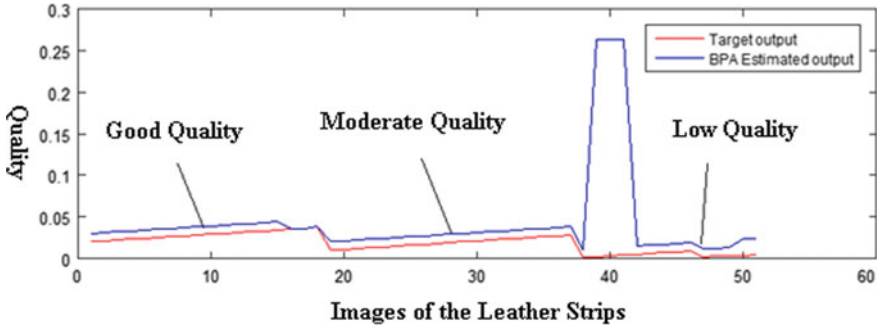


Fig. 4 Results of leather quality with different images

4 Conclusion and Future Work

For leather processing monitoring and control, a distributed multi-agent architecture was presented. The role of autonomous agents in the handling and processing of skins and hides is discussed. The interaction mechanisms are considered, which include agent language communication, ontology, and the coordination protocol (CNP). The research tries to demonstrate the architecture's great capability by taking into account the wide range of materials, the complexity of processing procedures, and the difficulties of fitting precise mathematical models of various processes. In comparison to traditional automation solutions, the architecture based on the agent idea highlights its advantages in terms of control reconfigurability, adaptation to changing industrial conditions, resilience, modularity, and flexibility. Using an agent development environment and production data from an existent tannery, preliminary steps were performed to replicate the correctness of the described technique.

References

1. W. Lepuschitz, G. Koppensteiner, M. Barta, T.T. Nguyen, C. Reinprecht, Implementation of automation agents for batch process automation, in *2010 IEEE International Conference on Industrial Technology*, IEEE, 2010, pp. 524–529
2. T. Hamaguchi, T. Hattori, M. Sakamoto, H. Eguchi, Y. Hashimoto, T. Itoh, Multiagent structure for batch process control, in *Control Applications, 2004. Proceedings of the 2004 IEEE International Conference on*, 2004, vol. 2, pp. 1090–1095
3. P. Thanikaivelan, J.R. Rao, B.U. Nair, T. Ramasami, Recent trends in leather making: processes, problems, and pathways. Crit. Rev. Environ. Sci. Technol. **35**, 37–79 (2005)
4. V. Deselnicu, L. Albu, TehnologiiModerne de Prelucrare a PieilorşiBlănurilor (2007)
5. M. Renner, E. Weidner, G. Brandin, Sustainable and environmentally friendly production of leather (2003). Available at: http://www.aidic.it/CISAP3/webpapers/43Renner.pdf. 04 Mar 2013
6. R.C. Panda, P.G. Rao, Chemical engineering applications on unit operations of tanning industry–case studies. Int. J. Chem. Eng. Appl. **2** (2011)

7. Reliance—Industrial SCADA/HMI system. Available at: http://www.reliance-scada.com. 02 Mar 2013
8. S. Calderwood, W. Liu, J. Hong, M. Loughlin, An architecture of a multi-agent system for SCADA-dealing with uncertainty, plans and actions, in *Proceedings of the 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO'13)*, Iceland, 2013
9. M. Wooldridge, *An Introduction to Multiagent Systems*, 2nd ed. (Wiley, 2009)
10. C. Plăişanu, D. Clapa, Developments in the holonic control of production flow. Sci. Bull. Ser. C **73** (2011)
11. D.C. McFarlane, S. Bussmann, Holonic manufacturing control: rationales, developments and open issues, in *Agent Based Manufacturing—Advances in the Holonic Approach* (Springer, 2002)
12. P. Vrba, P. Tichy, V. Mrik, K.H. Hall, R.J. Staron, F.P. Maturana, Rockwell automation's holonic and multiagent control systems compendium. IEEE Trans. Syst. Man Cybern.—Part C: Appl. Rev. **41** (2011)
13. FIPA—Agent Communication Language Specification. Available at: http://www.fipa.org/repository/fipa2000.html. 10 Feb 2014
14. S.A. Guta,Multi-agent systems for modelling and controlling leather manufacturing processes. Automatic Control and Systems Engineering Department, University POLITEHNICA of Bucharest, 2014, in press

Analysis of SWCET Student's Results Using Educational Data Mining Techniques

K. Shilpa and T. Adilakshmi

Abstract Educational data mining (EDM) is a process of determining meaningful and useful knowledge from large amounts of data that are dug up from educational locations. The purpose of EDM is to discuss tools, techniques, and design research for gleaning meaning from large data sets on the fly. EDM is a research area that applies to machine learning, data mining, and pattern recognition techniques. Predicting student performance is an application of EDM to analyse student results. The paper aims to predict students' results and compare different classification algorithms, considering educational data from Shadan Women's College of Engineering and Technology from the past four years in a directive to predict and analyse students' performance by identifying different data mining techniques such as J48 and logistic regression algorithms using the WEKA tool. This paper will contribute to constructing classification models for the 'SWCET' data set consisting of 627 different instances with eight different attributes for the first year, second year 376 instances, third year 360 instances, and final year 344 instances. It evaluates and compares implementation results to improve prediction accuracy. This study's findings, in particular, provide more insight for evaluating student performance with 100% accuracy using J48 and the logistic regression algorithm, which yields 99.84%. The student's performance will be advantageous to developing the quality of education and knowledge to drop the failure rate. All these things will help to improve the quality of the college.

Keywords Data mining · EDM · Classification · J48 · Logistic regression · Weka tool

K. Shilpa (✉)
Department of CSE, University College of Engineering, Osmania University, Hyderabad, India
e-mail: shilpamtech555@gmail.com

T. Adilakshmi
Department of CSE, Vasavi College of Engineering, Osmania University, Hyderabad, India
e-mail: t_adilakshmi@staff.vce.ac.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_58

1 Introduction

In the educational process, instructors' and students' interactions are done directly through lectures and discussions. The assessment process is based on tracking students' interactions and exam performance. These are very difficult for a large number of students. Evolving educational tools are built on computers and the Internet, enabling such things as observing and the process of analysis.

Data preprocessing is one of the data mining techniques that includes converting raw data into an understandable format.

Classification is the process of building models (functions) that describe and distinguish classes or concepts to predict the class of objects whose class label is unknown [1]. Examples of classification are decision trees, classification rules, neural networks, Naïve Bayesian classification, support vector machines, and K-nearest neighbour classification [1].

Educational data mining has various applications to report on different educational problems. Most scientific research has instances of educational applications. Usually, EDM methods are familiar to improve the understanding of students and the settings. The focus of this paper is to predict student performance and model the student characteristics from diverse viewpoints that are founded on the type of available data and predict student performance. These models use the system resources to the student needs and requirements [2–4].

1.1 Motivation

- To improve the results of the average students,
- To analyse students' performance and decrease their percentage of failure, and
- To develop the quality of the education.

1.2 Objectives

- To analyse students' performance using educational data mining,
- To examine the performance of students from Shadan Women's College of Engineering and Technology of IT Department (SWCET),
- To predict student performance using classification, and
- To analyse the growth of students modelling.

2 Review of Literature

Gurmeet Kaur, Williamjit Singh (2016)—Prediction of Student Performance Using Weka Tool: In this paper, prediction of student performance is done by applying Naïve Bayes and J48 decision tree classification techniques using Weka. By applying data mining techniques to student data, they have obtained knowledge that describes the student's performance. This knowledge will help to increase the quality of education, student performance, and decrease the failure rate. In this paper, they have shown this accuracy, Naïve Bayes provides 63.59% accuracy, and J48 provides 61.53% accuracy [5].

Mustafa Agaoglu (2016)—Predicting Instructor Performance Using Data Mining Techniques in Higher Education: In this paper, four different classification techniques to build classification models, decision tree algorithms, support vector machines, artificial neural networks, and discriminant analysis are used. Their performance is compared over a data set composed of the responses of students to real course evaluation questionnaires. Using accuracy prediction, recall, and specificity performance metrics although all of the classifier models perform comparably well in terms of accuracy, precision, and specificity [6], the C5.0 classifier performs the best in terms of accuracy, precision, and specificity [6].

Abelardo Pardo, Feifei Han, and Robert A. Ellis (2017): Combining University Student Self-Regulated Learning Indicators And Engagement With Online Learning Events To Predict Academic Performance: In this paper, they explore how to combine data about self-regulated learning skills with observable measures of online activity in a blended learning course to increase the predictive capabilities of students' academic performance for the purpose of informing reaching and task decision. A case study on a course with 145 student's shows that the variation of the student's final score for their course is better explained when factors from both approaches are considered [7].

Ayesha Anzer, Hadeel A. Tabaza, Jauhar Ali (2018): Predicting Academic Performance of Students in the UAE Using Data Mining Techniques—In this, the author implemented an approach to predict final exam scores from students' course assessments during the semester. In this paper, a linear regression model is used to check what factors affect the final exam scores. This paper explains the origins of data mining in education after preprocessing and preparing for the task. This paper implemented the linear regression model. The results of this work, like quizzes, show that they are the most accurate predictors of final exam scores compared to other kinds of assessment. The main aim of academic institutes is to provide quality education to students. The most important metric of quality education is students' academic performance. Identifying factors that affect the academic performance of students is of ultimate priority for any academic organization. In this paper, the author used a linear regression model to select and avoid some attributes of the same data set. It can be concluded that total assignments and quizzes have more impact on the final

exam. When students perform well in all these factors, those students can achieve good marks on the final exam [8].

3 Research Methodology

The aim of the study is to analyse and assess the effectiveness of a student's performance when using preprocessing and classification algorithms.

The present research is based on primary as well as secondary data. Secondary data was collected from different sources like books, IEEE journals, research papers, articles, and websites.

Primary data was collected directly from the IT Department in the Shadan Women's College of Engineering and Technology. In this paper, the Weka tool is used for preprocessing, classification, regression, and visualization.

Proposed Algorithms: **J48** is a decision tree like J48 on that data set would allow you to predict the target variable of a new data set. Decision tree J48 is the implementation of algorithm ID3 (Iterative Dichotomiser) developed by the Weka project team. **Logistic regression algorithm** is a classification algorithm used when the value of the target variable is categorical in nature. Logistic regression is most commonly used when the data in question has a binary output, so when it belongs to one class or another, or is either a 0 or 1.

4 Input Sample Data (SWCET)

See Fig. 1.

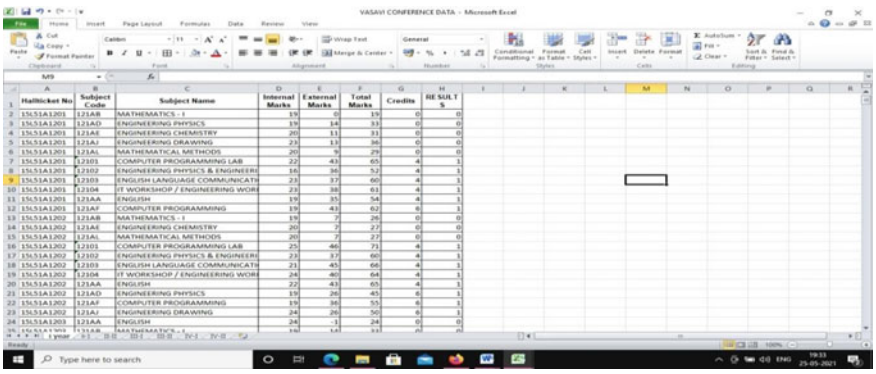


Fig. 1 Input sample data (SWCET)

5 Results and Discussion

Experimental tool: This paper used Waikato Environment for Knowledge Analysis (WEKA) version 3.9.5, the tool which is preferred in research to investigate SWCET college student data and estimate the accuracy and performance of data mining techniques practical to these sets. The particular data mining methods are taken as attributes for analysis. Additionally, the measures of model performance are presented, which are the basis for the comparison of methods' effectiveness and accuracy. To conclude, the visualization of each algorithm's accuracy is shown. This is built based on personal experience with the Weka environment. This paper extracted SWCET-IT Department student's data.

I year—preprocessed data—it applied data mining algorithms like J48, logistic regression, and visualization as shown in Figs. 2, 3, 4, 5, 6 and 7.

5.1 Figures

II year—preprocessed data—it applied data mining algorithms such as J48, logistic regression, and Visualization as shown in Figs. 8 and 9.

III year—preprocessed data—it applied data mining algorithms such as J48, logistic regression, and visualization as shown in Figs. 10 and 11.

IV year—preprocessed data—it applied data mining algorithms such as J48, logistic regression, and visualization as shown in Figs. 12 and 13.

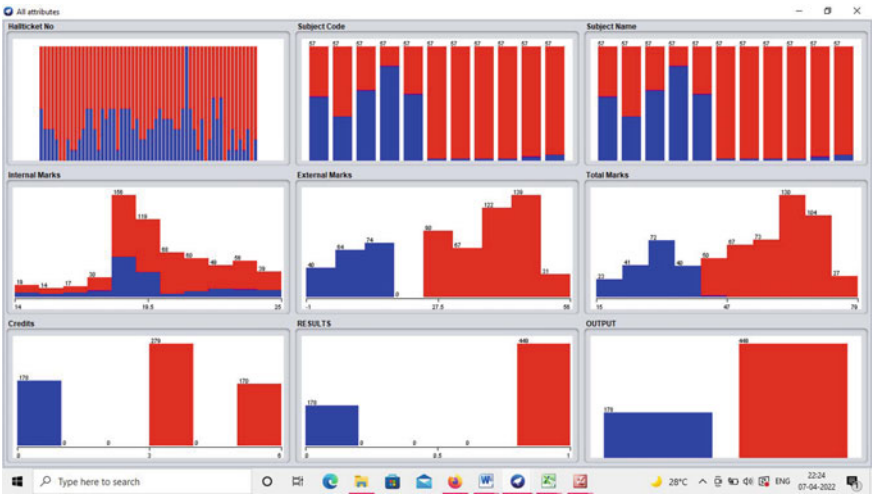


Fig. 2 Visualization of all attributes

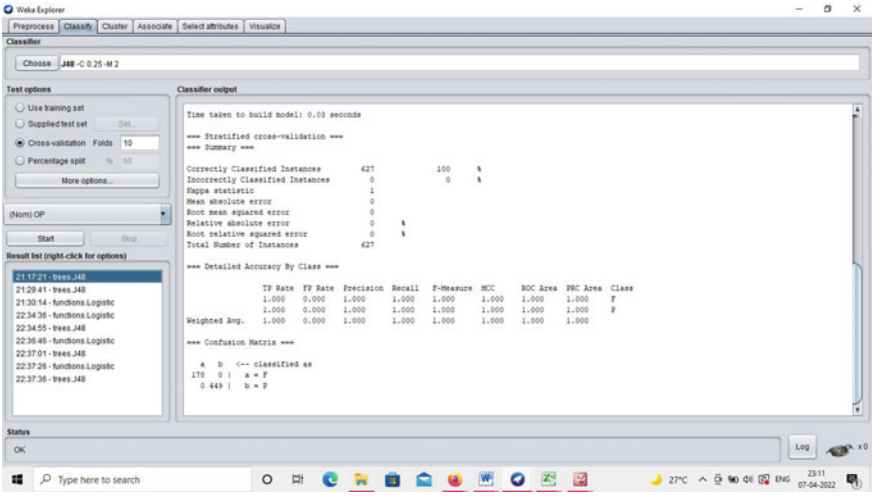


Fig. 3 Output of J48

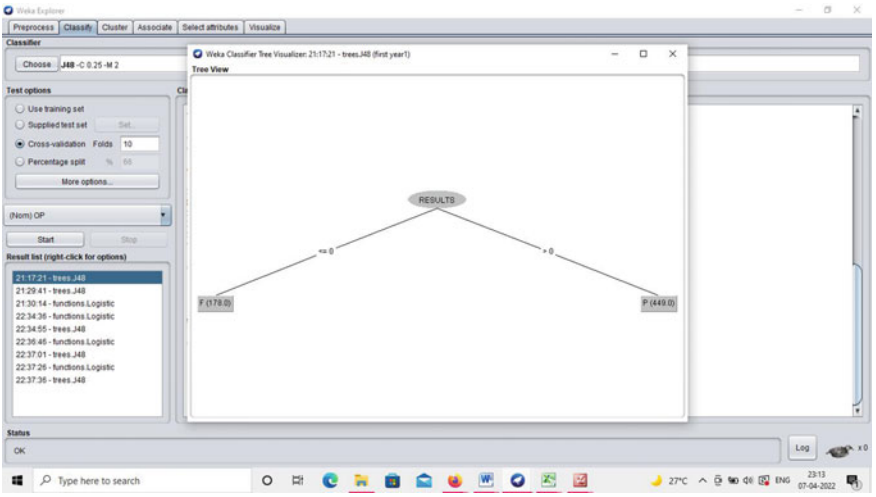


Fig. 4 Output of J48

5.2 Execution Steps

The process of data mining in academic performance and the data mining task's steps:

Collect data: Data has been collected from Shadan Women's College of Engineering and Technology for the past four years, from 2015 to 2019. This data is gathered

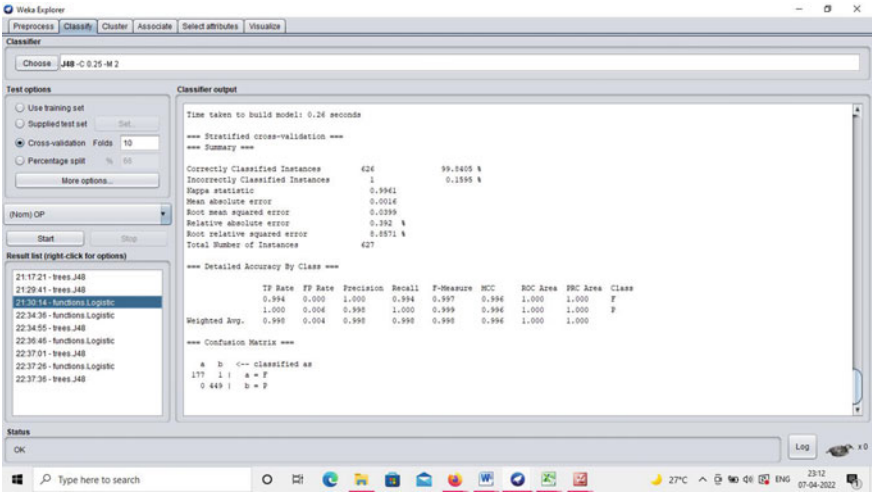


Fig. 5 Output of logistic regression

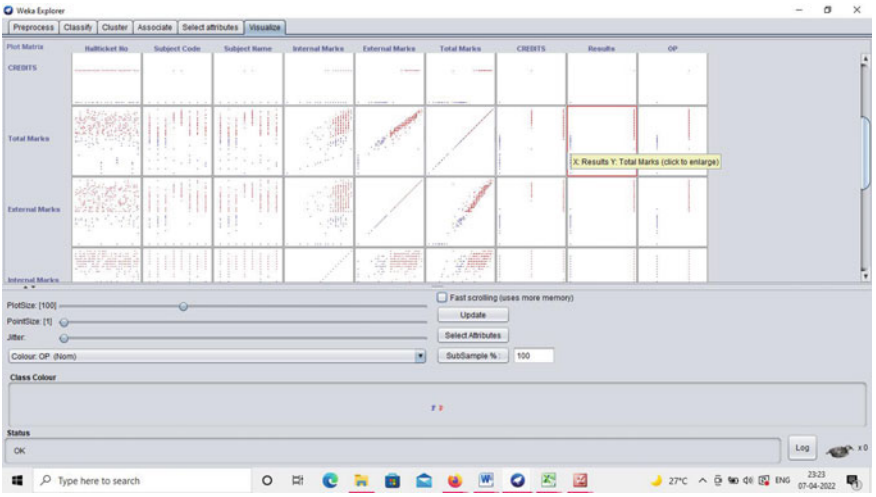


Fig. 6 Visualization of output

from the IT Department. There are eight attributes such as student, hall ticket no, subject code, subject name, internal marks, external marks, total marks, credits, and results. The total number of instances in the first year's results is 627, and the number of attributes is 8. The total number of instances in the second year-I semester is 376, and the number of attributes is 8. The total number of instances in the third year-I semester results is 360, and the number of attributes is 8. Total instances in the fourth year-I semester results are 344, and attributes are 8. Next, preprocess the data: In the

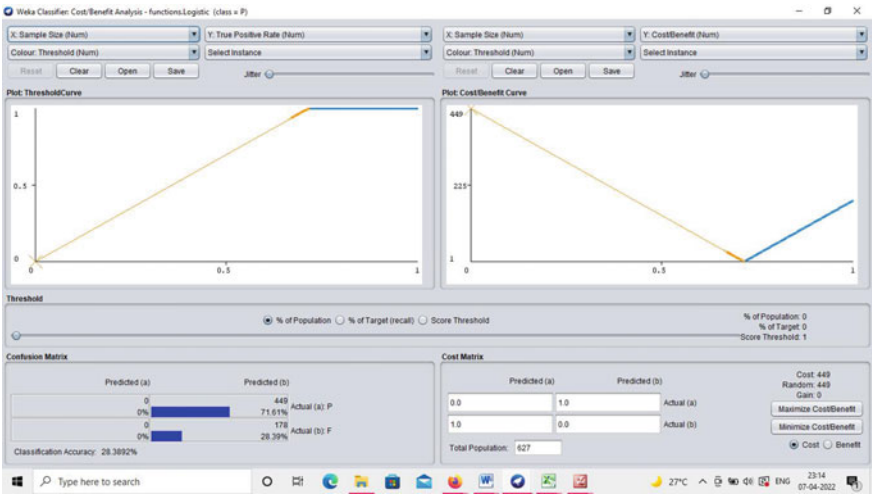


Fig. 7 Visualization of cost/benefit curve

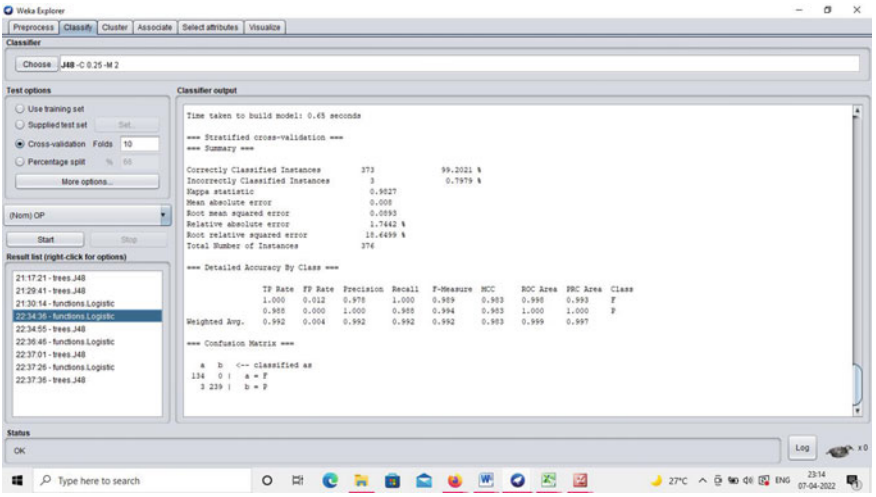


Fig. 8 Output of logistic regression

weka tool, go to the preprocess tab, choose Open File, and choose SWCET Data. Next, apply the data mining algorithm.

Apply data mining: In this paper, classification algorithms are used to predict students' performance using logistic regression algorithm, J48, and finally, results are visualized using visualization tab (Figs. 14, 15 and 16; Tables 1, 2 and 3).

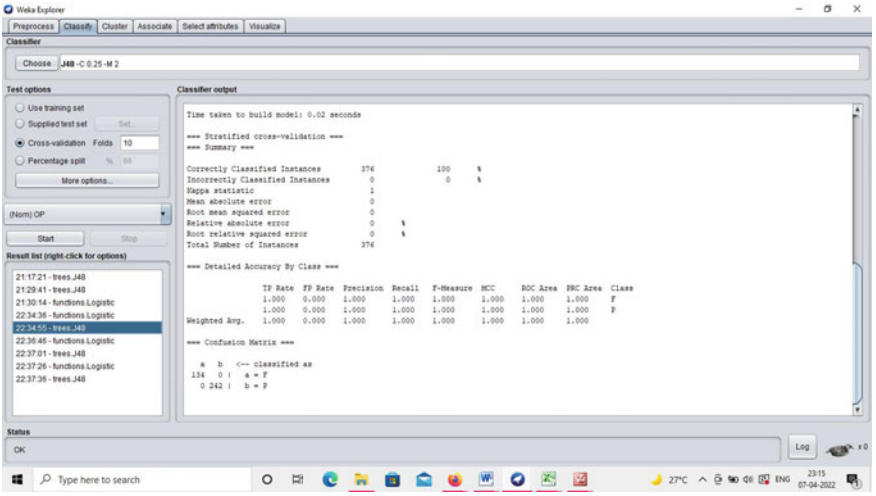


Fig. 9 Output of J48

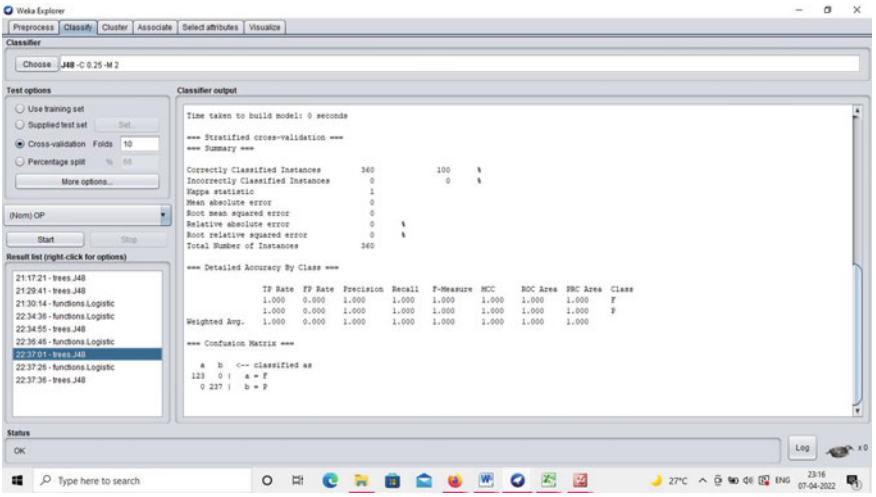


Fig. 10 Output of J48

$$\text{Accuracy} = \frac{\text{Total correct predictions}}{\text{Total number of instances}}$$

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

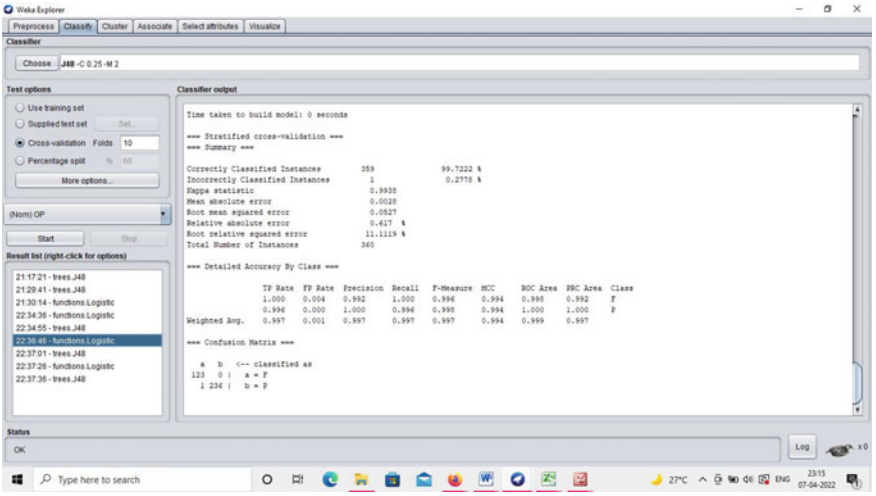


Fig. 11 Output of logistic regression

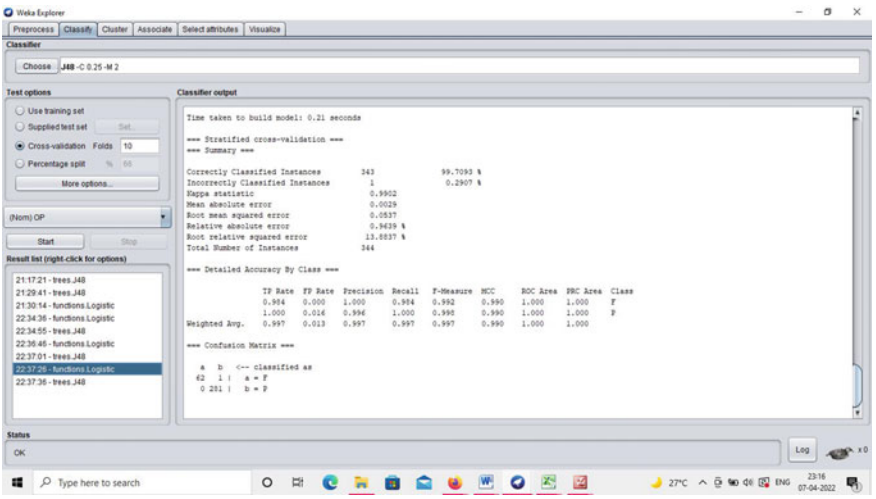


Fig. 12 Output of logistic regression

6 Conclusion

This paper considered past four years data from Shadan Women's College of Engineering and Technology—2015 to 2019—from the IT Department and applied classification algorithms like J48, logistic **regression** algorithms, and analysis of student performance useful in recognizing average students and above-average students; performance shows that the J48 classification algorithm shows best algorithm with

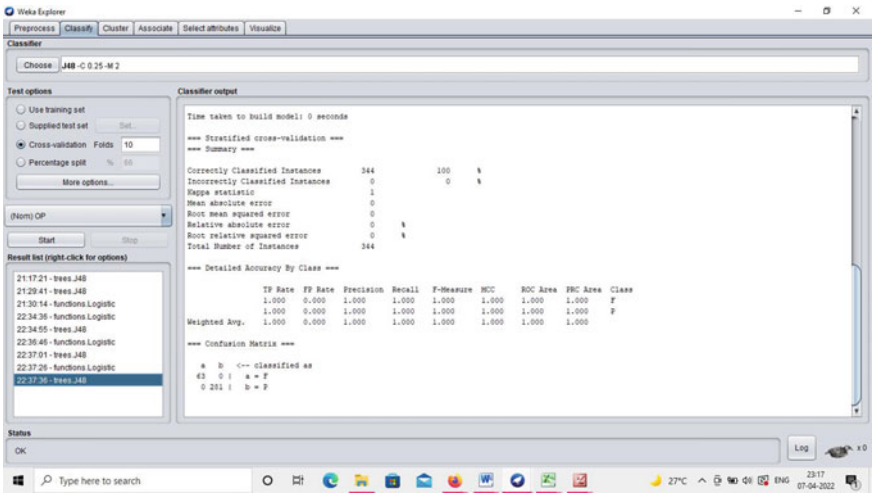


Fig. 13 Output of J48

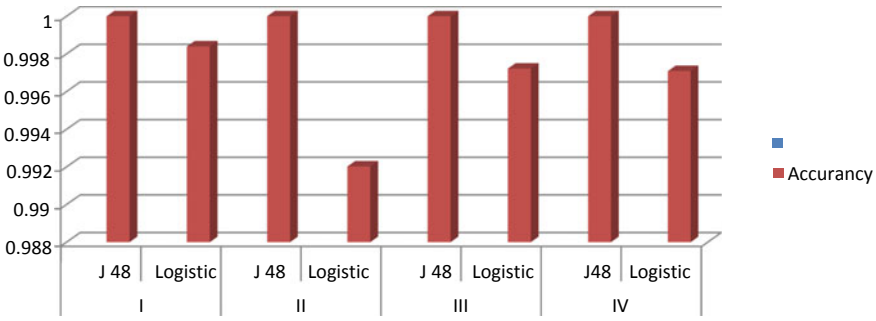


Fig. 14 Performance and accuracy comparison of J48 and logistic algorithm

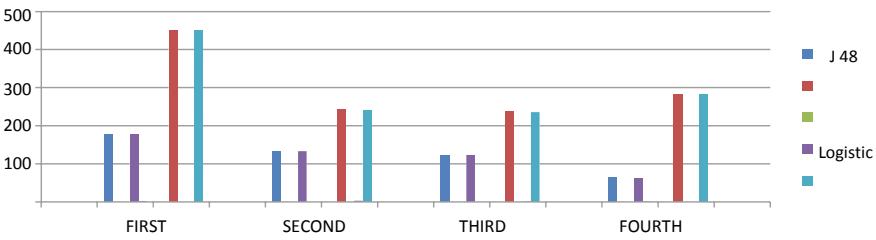


Fig. 15 Confusion matrix of four year students-1 sem

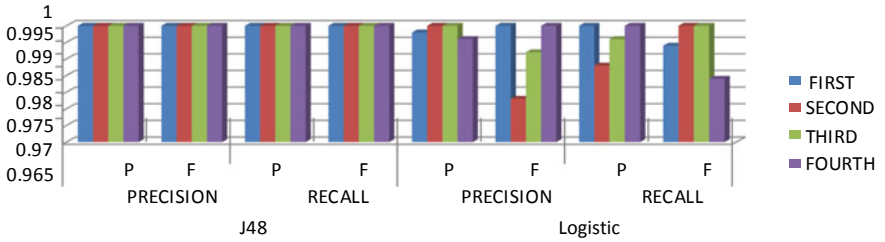


Fig. 16 Precision and recall for J48 algorithm and logistic algorithm

Table 1 Performance and accuracy comparison of J48 and logistic algorithm

Years	I year		II year		III year		IV year	
	J 48	Logistic	J 48	Logistic	J 48	Logistic	J48	Logistic
Time taken to build model	0.03	0.18	0	0.02	0	0	0	0.03
Kappa statistic	1	0.9961	1	0.9827	1	0.9938	1	0.9902
Mean absolute error	0	0.0016	0	0.008	0	0.0028	0	0.0029
Root mean squared error	0	0.0399	0	0.0893	0	0.0527	0	0.0537
Relative absolute error	0	0.39%	0	1.74%	0	0.62%	0	0.96%
Root relative squared error	0	8.86%	0	18.65%	0	11.11%	0	13.88%
Total number of instances	627	627	376	376	360	360	344	344
Accuracy (%)	100	99.84	100	99.20	100	99.72	100	99.71

Table 2 Confusion matrix of four year students-1 sem

Confusion matrix				
Years	J48		Logistic	
	a = F	B = P	a = F	B = P
First	178	0	177	1
	0	449	0	449
Second	134	0	134	0
	0	242	3	239
Third	123	0	123	0
	0	237	1	236
Fourth	63	0	62	1
	0	281	0	281

100% when compared to the logistic regression algorithm. This paper aims to predict students' results and compare different classification algorithms. In this paper, for constructing classification models for the 'SWCET' data set, consisting of 627

Table 3 Precision and recall for J48 algorithm and logistic algorithm

Methods	J48				Logistic			
	Precision (%)		Recall (%)		Precision (%)		Recall (%)	
Years	P	F	P	F	P	F	P	F
First	100	100	100	100	99.8	100	100	99.4
Second	100	100	100	100	100	97.8	98.8	100
Third	100	100	100	100	100	99.2	99.6	100
Fourth	100	100	100	100	99.6	100	100	98.4

different instances with eight different attributes, after implementing these algorithms on the data set, we evaluate and compare the implementation results for better accuracy of prediction. The result of this study particularly provides a greater insight into evaluating the student performance with 100% accuracy using the J48 algorithm and the logistic regression algorithm, which gives 99.84% accuracy.

References

1. J. Han, M. Kamber, *Data Mining: Concepts and Techniques* (Morgan Kaufmann, San Francisco, 2006)
2. N. Khodeir, Student modeling using educational data mining techniques, in *IEEE EGYPT Section* (2019), pp. 7–14
3. W. Chango, R. Cerezo, C. Romero, Predicting academic performance of university students from multi-sources data in blended learning, in *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems* (ACM, 2019)
4. R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K.U. Sarker, M.U. Sattar, Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. Appl. Sci. 1–20 (2020)
5. G. Kaur, W. Singh, Prediction of student performance using Weka tool. Res. Cell: Int. J. Eng. Sci. **17** (2016)
6. M. Agaoglu, Predicting instructor performance using data mining techniques in higher education. IEEE Access 2379–2387 (2016)
7. A. Pardo, F. Han, R.A. Ellis, Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. IEEE Trans. **10**(1) (2017)
8. A. Anzer, H.A. Tabaza, J. Ali, Predicting academic performance of students in UAE using data mining techniques, in *ICACCE-2018* (IEEE, 2018), pp. 179–183
9. L.M. Nkomo, M. Nat, Discovering students use of learning resources with educational data mining, in *2016 Honet-Ict 2016* (IEEE, 2016), pp. 98–102

A Novel Approach to Project Employee's Performance and Confinement

Gangarapu Shirisha and M. Raja Sekar

Abstract Employees are an organization's most precious asset. The graphical representation of anticipating pay is a process that attempts to build a computerized system that can follow all daily salary growth graphs in any location and predict revenue after a certain period of time. This application may access the company's staff database and generate a graph based on the information contained therein. Before importing a graph to aid in the visual representation, it will validate the fields. Every company devotes time and resources to grooming new workers and preparing them for the workplace. If such individuals leave after completing their training, the company will suffer a complete loss. Most businesses assess retention rates on an annual basis, but you may measure it in smaller increments for speedier results. If it is 80%, it signifies that just 20% of employees leave the company during a certain time period. Here we will compare and choose the best algorithm with accurate results and analysis.

Keywords Employee retention · Logistic regression · And machine learning

1 Introduction

Employees have always been valuable assets to any company. They might be referred to as an organization's life-blood because of their crucial nature. Knowing where you are now and where you want to go next is one of the most significant aspects of the job search. A job seeker's perspective is similar to that of a product manager, who is looking for the best product and market fit. However, because technology requires human resources to operate, this circumstance does not diminish the worth of employees in a business.

G. Shirisha (✉) · M. R. Sekar
Department of Computer Science, VNR Vignana Jyothi Institute of Engineering Technology, Hyderabad 500090, India
e-mail: g.shirisha88@gmail.com

M. R. Sekar
e-mail: rajasekar_m@vnrvjiet.in

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_59

Overall, there is a great degree of uncertainty about the best strategy to restore and maintain employee engagement. Employers who are astute understand the value of retaining top talent. In the Indian context, retaining talent has never been more vital; however, things have altered in recent years. Companies have never placed a greater emphasis on retaining key personnel and addressing attrition issues.

1.1 Employee Confinement: The Three R's

You must execute each of the three pillars of Employee Retention: respect, recognition, and rewards, in order to keep staff and maintain high levels of satisfaction. RESPECT is a term that refers to a person's esteem, special regard, or unique treatment. Respect is the cornerstone of keeping your employees, as the pyramid illustrates. If you do not appreciate your employees, RECOGNITION and REWARDS will have minimal impact. "Special notice or attention" and "the act of clearly perceiving" are two definitions of recognition.

1.2 Objectives of the Study

Our predictive approach differs from many others since it is not constrained by sample numbers. Rather, it forecasts salaries based on over 600,000 variables connected to job titles, locations, and talents, as well as patterns and values for the elements that go into your individual wage forecast. Our salary ranges are shown in our public salary estimates.

1.3 Scope of the Study

This will assist management in understanding their employees' attitudes toward their jobs. The study's recommendations and suggestions can be applied to similar projects. It will be beneficial for management to recognize employee needs in order to keep them in the company. This project can be used by students who are working on a project in a related field, as well as organizations looking to improve their retention techniques. You can keep your top performers and build a loyal, engaged, and goal-oriented team by employing the appropriate strategies. Note that the first paragraph of a section or subsection is not indented.

2 Literature Survey

Mita et al. [1] tells that it examine the findings of various authors' research articles in order to determine the elements that influence employee commitment and retention in the workplace. Existing studies, while some did, did not place enough attention on the type of employee, the sector of the economy, or the sort of firm that is most affected by one element or another. More research is needed to better provide organizations with the knowledge they need to improve their retention capability.

Kumar et al. [2] explains the purpose of this paper was to propose a conceptual framework that explains the mediating role of employee brand equity and the moderating role of demographic characteristics in the link between employee value proposition and intention. The findings point to the need for more empirical study to explain why higher-level employees leave and to discover effective remedies.

Reddy Asi [3] this focuses mostly on employee compensation, which is a crucial factor for any firm to address in order to keep its most valuable asset, its personnel. Employee loyalty and retention will improve if they are fairly compensated for their contributions to the organization. Also, it sheds light on how to adequately recompense an employee's performance, ensuring that the employee stays with the company for the long haul.

Gorde et al. [4] demonstrate how important staff retention is in this day and age, and what the consequences will be if employers are not aware of the situation and do not take immediate action to that end, and how they will influence the organization and the industry.

Silva et al. [5] to determine which criteria are significant in keeping employees in firms and, as a result, not only avoiding turnover but also retaining the top individuals.

3 Methodology

Employers are fully aware that certain qualities, such as a strong work ethic, critical thinking, excellent communication, and teamwork, are desired by undergraduate students. This project will assist you in becoming that smart product manager by advising you on what attributes and top abilities the market seeks, as well as assisting you in navigating pay negotiations and making the best decision possible.

We examine multiple models to discover the classifier that best describes the data in a job recruitment situation characterized by high levels of noise, high dimensionality, and a restricted quantity of samples after phrasing the prediction issue as a classification challenge. Our main focus is on pay range prediction, as this might lead to improved categorization of job postings and hence easier navigation for end-users.

Retention Rate Formula

$$(\text{Remaining Headcount during Set Period} \div \text{Starting Headcount during Set Period}) \times 100 \quad (1)$$

Turnover Formula

$$(\text{Number of separations}/\text{average number of employees}) \times 100 = \text{turnover rate} \tag{2}$$

Employee Turnover Cost Formula

$$\text{Number of employees lost} * (\text{average cost of replacement} * \text{average employee salary}) = \text{turnover cost} \tag{3}$$

3.1 Dataset

This dataset contains various features to define how the employee performance is measured for future benefits. This helps organizations to use data to assess the effectiveness of a strategy. For this project:

1. The dataset has 14,249 observations for past/present employees.
2. The observations span 12 different departments.
3. Each observation includes the employee's current employment status.

This information is classified into 3 categories which are given as follows (Fig. 1). Below are some important features with % contribution toward attrition

- Satisfaction level: 34.77%
- Time spend in the company: 28.24%
- Average monthly hours: 12.05%
- Latest Evaluation: 11.46%
- Number of projects: 10.77.

These are the driving factors which leads an employee to leave the company. The describe() method is used to calculate statistical data for numerical values in a Series, such as percentile, mean, and standard deviation. It examines numeric and object series, as well as mixed data type DataFrame column sets.

	avg_monthly_hrs	department	filed_complaint	last_evaluation	n_projects	recently_promoted	salary	satisfaction	status	tenure
0	221	engineering	NaN	0.932868	4	NaN	low	0.829896	Left	5.0
1	232	support	NaN	NaN	3	NaN	low	0.834544	Employed	2.0
2	184	sales	NaN	0.788830	3	NaN	medium	0.834988	Employed	3.0
3	206	sales	NaN	0.575688	4	NaN	low	0.424764	Employed	2.0
4	249	sales	NaN	0.845217	3	NaN	low	0.779043	Employed	3.0

Fig. 1 Dataset for the model

```
avg_monthly_hrs          0
department             709
filed_complaint          0
last_evaluation       1351
n_projects               0
recently_promoted        0
salary                   0
satisfaction             0
status                   0
tenure                   0
dtype: int64
```

```
avg_monthly_hrs           0
department                0
filed_complaint           0
last_evaluation           0
n_projects                0
recently_promoted         0
salary                    0
satisfaction              0
status                    0
tenure                    0
last_evaluation_missing   0
dtype: int64
```

Fig. 2 Looking for missing values and filling them by feature selection

3.2 Data Preprocessing

Missing values are supplemented, outliers are identified and removed, and self-contradiction is resolved via data preprocessing. Because the sample code number has no bearing on illnesses, it has been removed from the dataset. In the dataset, there are 16 trait values that are missing. The class's missing traits are replaced by the mean. In addition, the dataset uses random selection to ensure that the data is circulated properly. An outlier is a statistician's term for an observation point that stands out from the others (Fig. 2).

3.3 Data Visualization

The graphical depiction of information and data is known as data visualization. Data visualization tools make it easy to examine and comprehend trends, outliers, and patterns in data by employing visual elements like charts, graphs, and maps. It is also crucial in large-scale data efforts. It is also part of the larger data presentation architecture discipline, which strives to find, package, and convey data in the most efficient way possible.

3.4 Training and Testing Phase

Adequate training necessitates the algorithm seeing the training data several times, which means that if the model is run over the same dataset, it will be exposed to the same patterns. To avoid this, you will need a new set of data to aid your algorithm's recognition of diverse patterns. However, because you require your testing data set for other purposes, you do not want to involve it before the training is completed. The training data sets can be obtained from a variety of sources, and the one you choose

Table 1 Accuracy for different machine learning algorithms

Algorithm	Cross-validation	Accuracy (%)
Logistic regression	0.90	90
SVM	0.96	97
Random forest	0.98	98
*K*NN	0.97	95
Decision tree	0.96	97

is determined by your goals, the requirements of your machine learning project, as well as your money, time, and manpower constraints. The training step extracts the dataset's features, whereas the testing phase determines how the appropriate model performs in terms of prediction.

3.5 Model Selection

Applying Machine Learning to any dataset is at its most interesting stage. For forecasting the best results, it is also known as Algorithm selection. When dealing with huge data sets, Data Scientists typically employ a variety of Machine Learning algorithms. To offer a learning framework for future data processing, input and output data are labeled for classification. This table describes the values for 3 fields as department, salary and status of the employees. Categorical data must be encoded as part of the feature engineering process. It is more vital to know the coding scheme to utilize, taking into account the dataset and model we will be using.

Random forest is a decision tree-based methodology for modeling predictions and behavior analysis. It contains a number of decision trees, each of which represents a unique classification of data fed into the random forest. They aim to maximize whatever assessment criteria you are using and they might take shortcuts to do so. For example, if you had two classes, one with 99 examples and the other with only one, a model could always correctly predict the first class 99% of the time! The model would have a high accuracy rating, but it would not be able to help you find examples of the smaller class (Table 1).

4 Recommendations and Thoughts

The following ideas are based on the findings: employees are dissatisfied with the organization's recognition and performance appraisal. Employees are feeling overworked as a result of the workload. As a result, the company should concentrate on ensuring a smooth workload in order to reduce employee stress. Employees and management must enhance their relationship. People want to enjoy their work, therefore make it enjoyable for them. Recognize that employees need to balance their

personal and professional lives, thus provide flexible start times and core hours. To encourage open communication, provide 360° feedback surveys and other questionnaires. Allow anonymous questionnaires on occasion so that staff can be more open and honest about their thoughts. Provide opportunities for career advancement and cross-training inside the organization.

5 Conclusion

It is self-evident that industries will not be able to survive in the long run unless they have predictive analytics skills in human resource management. To reduce attrition, the study finds that businesses should use new creative technologies and effective training programmers to provide possibilities for employees to progress within the corporation. In comparison to Logistic Regression and Support Vector Machine, the results showed that Random Forest Regression produces the best results. Instead, by identifying at-risk employees, analytics may be utilized to prevent this from happening in the first place. Predictive analytics can identify who is most likely to leave the company by analyzing criteria such as compensation, promotions, corporate culture, and relationships with managers.

References

1. M. Mita, K. Aarti, D. Ravneeta, Study on employee retention and commitment. Int. J. Adv. Res. Comput. Sci. Manag. Stud. **2**, 154–164 (2014)
2. S. Kumar, U. Devadas, K. Dhammika, in *Impact of Employee Value Proposition on Employee Retention: A Conceptual Paper* (2021). https://doi.org/10.5281/zenodo.5513384
3. V. Reddy Asi, *Compensation System: Influence on Employee Retention* (2021)
4. S. Gorde, *A Study of Employee Retention* (2019)
5. M. Silva, J. Carvalho, Á. Dias, in *Determinants of Employee Retention* (2019). https://doi.org/10.4018/978-1-5225-7888-8.ch004

Air Contamination Prediction and Comparison Using Machine Learning Algorithms

P. ArunaKumari, Y. Vijayalata, G. Susmitha Valli, and Y. Lakshmi Prasanna

Abstract Dealing with air contamination presents a significant environmental hazard in the urban environment. Constant monitoring of contamination data empowers local authorities to analyse the current traffic circumstance of the city and settle on choices likewise. Arrangement of the Internet of Things-based sensors has extensively changed the elements of foreseeing air quality. Existing exploration has utilized distinctive artificial intelligence learning tools for contamination prediction; however, near examination of these procedures is needed to have a superior comprehension of their handling time for numerous datasets. This paper tends to the test of foreseeing the air quality index (AQI), with the aim to predict the contamination on different studies, utilizing two machine learning algorithms: neural networks and support vector machines. The air contamination datasets downloaded from the Central Pollution Control Board (CPCB). The proposed machine learning (ML) model is used to predict the Delhi Air Quality Index (AQI) data and compare it with the actual and predicted data.

Keywords Machine learning · Neural networks · Support vector machines

1 Introduction

Air is the most fundamental characteristic asset for the presence and endurance of the whole species of earth. All types of species which include plants, insects, and creatures rely upon the air as it is their fundamental endurance. Hence, all organic

P. ArunaKumari (✉) · Y. Vijayalata · G. Susmitha Valli · Y. Lakshmi Prasanna
Department of Computer Science and Engineering, GRIET, Hyderabad, Telangana, India
e-mail: arunakumari.3610@gmail.com

Y. Vijayalata
e-mail: vijaya@ieee.org

Y. Vijayalata
Department of Computer Science and Engineering, KG Reddy College of Engineering & Technology, Hyderabad, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_60

entities need the fresh air which is liberated from unsafe gases to live their life. In 2008, the Blacksmith institute did research on air pollution and noticed the most contaminated spots, the two of the most noticeably terrible contamination issues on the planet are metropolitan air contamination and indoor air contamination [1]. The major causes of air contamination are industrial wastes, automobiles, agricultural activities, exhausts from factories, mining, etc., which causes long haul and transient wellbeing impacts. It is tracked down that the old and little youngsters are more affected by air contamination. Short period side effects are burning eyes, loose nose, itching in the throat, headaches, skin allergies, respiratory problems, and pneumonia. Few long-term side effects are cancer in the lungs, brain misfunctioning, damage to the liver, kidney, coronary illness, and respiratory infections such as emphysema [2]. Also, it affects the whole ecosystem like the formation of smog with solid and liquid particles which is a visible type of air pollution that also sometimes having different shapes and variant colours. Likewise adds to the ozone layer damage, which shields the planet from direct UV beams from the sun. Few more adverse consequences of air contamination particles fall on earth eventually which directly pollutes the water surface bodies and soil. Those particles can affect agriculture causes less yield. The particles of nitrogen oxide and sulphur dioxide in the air can cause acid rain formations while mixing with water in the upper atmosphere, which damages the planet's surface soil, wildlife, rivers, lakes, and trees. Thus, air contamination is quite possibly the most disturbing worry for us today. Tending to this worry, in the past years, numerous analysts have invested loads of time in considering and creating various models and strategies in air quality examination and assessment.

Air quality prediction took place by following traditional methodologies from all these years. These methodologies include a collection of raw manual data and evaluation. There are several conventional methodologies or strategies available for air quality forecast by using numerical and factual procedures [2]. In these strategies, at first an actual model is planned and information is coded with numerical conditions. Besides such strategies may also experience the ill effects due to the drawbacks like:

1. Less accuracy while prediction—they are unable to predict maximum and minimum levels of pollution
2. Inappropriate methodologies for output prediction
3. Implementation of complex algorithms
4. Amalgamation of old and new data.

In spite with the headway in innovation and exploration, options and in contrast to conventional strategies, several new strategies or methodologies are proposed which makes use of AI and big data approaches. Even though numerous scientists have created and utilized analytical models of big data and AI-based models to direct air quality assessment to accomplish better precision in assessment and prediction. This paper is mainly focused on few scientific researches and study on the available methodologies published on air pollution prediction. The significant target is to give a depiction of the tremendous exploration work and valuable audit on the present status of the workflow on AI, ML, deep learning, and big data approaches for air quality assessment and prediction (Table 1).

Table 1 Health statements for AQI categories (this table was adapted from the Central Pollution Control Board's National Air Quality Index Report)

AQI	Associated health impacts
Good (0–50)	Minimal impact
Satisfactory (51–100)	May cause minor breathing discomfort to sensitive people
Moderately polluted (101–200)	May cause breathing discomfort to the people with lung disease such as asthma and discomfort to people with heart disease, children and older adults
Poor (201–300)	May cause breathing discomfort to people on prolonged exposure and discomfort to people with heart disease
Very poor (301–400)	May cause respiratory illness to the people on prolonged exposure. Effect may be more pronounced in people with lung and heart disease
Severe (401–500)	May cause respiratory effects even on healthy people and serious health impacts on people with lung/heart diseases. The health impacts may be experienced even during light physical activity

The paper is structured as per the following. Section 2 shows the striking work done by the researchers to estimate air quality. Section 3 presents data and perceptions identified with the information and strategies utilized in this investigation. Section 4 talks about the investigations and results of this examination. At the end, future extension of the work is being discussed in the fifth segment.

2 Related Work

A deep learning algorithm uses various layers to automatically extract high-level features from raw data in a dynamic manner. Deep learning designs such as RNNs are used to model sequential data. This function contains cyclical connections wherein the preceding time steps taken into account as inputs to the present time. Several RNNs use a backpropagation and propagation through time (BPTT) procedure to propagate errors and update weights. In any event, gradients can disappear and explode while preparing an RNN. The slope yield in the neural network (NN) model causes massive estimations of the gradients, if more than one with numerous layers. It is due to this phenomenon that the trainable weights have significant differences in their quality with each cycle, which, in turn, increases the effect of the layer beneath the trainable weight. However, when they are less than 1, the gradients' effects on the underlying layers are negligible. To deal with these issues, LSTM has been presented as a way to reduce their impact.

The LSTM is a particular RNN type that is intended to achieve a higher level of accuracy than regular RNNs when modelling temporal sequences. During recurrent shrouded layers, the memory blocks contain extraordinary units. In the memory

blocks, self-replicating memory cells store the temporal state of the organization as well as unique multiplying elements called entryways that control the progress of data.

In 2004, the Department of Electronics addresses the prediction problems of Ozone and PM10 toxins [3]. Researchers proposed a statistical approach using feed-forward neural networks (FFNN) which is compared with pruned neural networks and lazy learning. The PNN is a constitute of parameter parsimonious approach which mainly depends on dumping the redundant parameters of FCNN. The lazy learning is an approach of local linear algorithm which performs the prediction using local learning procedure. Using these three algorithms, the authors trained and tested to predict the O3 and PM10 levels, but the prediction accuracies are does not make any difference [3].

In the year 2008, researchers from the Shandong University proposed an approach to find the quality of air using back propagation neural networks (BPNN) on rough set theory [4]. In this experiment, first they minimized the data monitoring of the contamination sources utilizing the hypothesis of the harsh set, from the clean guidelines. At that point, the topological design of the multi-facet BPNN and the nerve cells of the suggestive layer is characterized by these principles. From that point forward, the associated weight benefits of comparing hubs of the BPNN are found out. Utilizing BP math, the forecast model is prepared with the checking data of the contamination sources and air screen stations for acquiring its different boundaries.

In the year 2009, Prediction of Indoor Air Quality Using Artificial Neural Networks was published by the University of Science and Technology of Beijing. In this research work, they proposed an ANN-based prediction model to predict the air quality. They fed three toxic indoor air contamination gases out of six as input to the network, and an occupant symptom metric is used to measure the air quality and given as an output variable [5].

In the year 2010, from the State Environmental Protection Key Laboratory with the collaboration with Nankai University proposed a GA-ANN model to predict air quality. This model is an advanced ANN model called genetic algorithm and artificial neural networks. They have used genetic algorithm to select the factors from the dataset which is directly fed to ANN model for training and testing. In order to experiment, they have used Tianjin city air pollution data. Finally, they compared the GA-ANN with ANN and PCAANN models [6].

In the year 2015, C. Xiaojun et al. proposed an IOT-Based air pollution monitoring and detection system [7]. They used a huge volume of sensors to make sure the accuracy of monitoring also which reduces monitoring cost also makes the data systematic and accessed using IOT and telematics. The data received from the sensors need to be processed to find out the predicted value in order to do that they have used ANN model and Bayesian regularization model for analysing the data [7].

In the year 2017, Gazi University published a paper on air quality prediction using deep learning model [8]. The authors used RNN—recurrent neural network model-based long short memory (LSTM) networks model to predict future air quality index values in metropolitan cities.

In the year 2018, proposed a deep learning-based spatial temporal approaches to detect the air quality [9]. The authors tried to predict the quality of air for 48 h using different NN—neural networks as follows: (1) ANN, (2) CNN, and (3) LSTM model to extract spatial temporal relations [9].

In 2020, Daniel Schürholz et al. proposed an artificial intelligence-based approach to find out the air quality prediction [10]. The authors in their research experimented a novel setting expectation model that incorporates setting to computer concepts to consolidate an exact air contamination forecast calculation utilizing long short-term memory and deep neural networks with data.

3 Methodology and Dataset

3.1 Data Acquisition

Out of Delhi's 1484 square kilometres, 783 are devoted to the rural areas and 700 to the urban areas. There are currently 19 million people living in Delhi. By 2028, Delhi will be the most crowded city in the world, according to the World Urbanization Prospects of the United Nations 4. Although there have been numerous steps none of the measures taken to control air pollution have been particularly beneficial. The authors choose to explore Dwarka as a possible location for our investigation. In addition to being a sub-city in Delhi, it is not far from Gurugram, the world's most polluted city. A contamination hotspot in Delhi is Dwarka. A total of 170 stations (areas) (expresses) gathered from 18 states and 102 urban areas with a total of 102 urban communities were used to obtain pollution concentration and meteorological data. The data collected for the NSUT (formerly NSIT) station in Dwarka's Sector-3 has been analysed. When comparing NSUT data with data from different stations, there were a few little holes (non-designated values) that would equal a couple days up to half a month in a year. The authors chose 3.5 years as the period of data. The data is chronologically dated from 1 April 2015 to 31 March 2017, and from 1 October 2017 to 1 April 2019 and are unable to access data from April 1, 2017 to September 30, 2017 [11]. A breakdown of the air pollutants and meteorological boundaries that were used in this analysis can be found in Tables 2 and 3.

In order to conduct this study, air pollution data was collected from numerous monitoring stations. In Fig. 1, Anand Vihar can be seen. Secondly, polluted regions are where these observation stations are located. Furthermore, the choice of these stations demonstrated the complexity and heterogeneity of predicting city-level contamination patterns. For example, the pollution concentrations for NO_2, CO, O_3, SO_2, PM_{10}, and $PM_{2.5}$ were collected from the Central Pollution Control Board (CPCB) site and a "Noise and Air Contamination Observing Framework" to gather pollution levels. A Wi-Fi module transmits data to the cloud, and a SD card stores data on the device itself, and multiple noise and gas sensors work together as part of this framework. At the initial IoT stage, the data is saved on the cloud where it can

Table 2 List of pollutants considered

S. No.	Parameters	Unit
1	CO (carbon monoxide)	mg/m^3
2	NO (nitrogen oxide)	$\mu g/m^3$
3	NO_2 (nitrogen dioxide)	$\mu g/m^3$
4	Ozone	$\mu g/m^3$
5	$PM_{2.5}$ (particular matter 2.5 mm)	$\mu g/m^3$
6	SO_2 (sulphur dioxide)	$\mu g/m^3$

Table 3 Meteorological parameters

S. No.	Parameters	Unit
1	Temperature	°C
2	RH (relative humidity)	%
3	SR (solar radiation)	W/mt^2
4	WS (wind speed)	m/s

be accessed by anyone. Likewise, authors extracted data from the above-referenced sources for affecting elements such as wind speed, wind heading, temperature, and relative humidity.

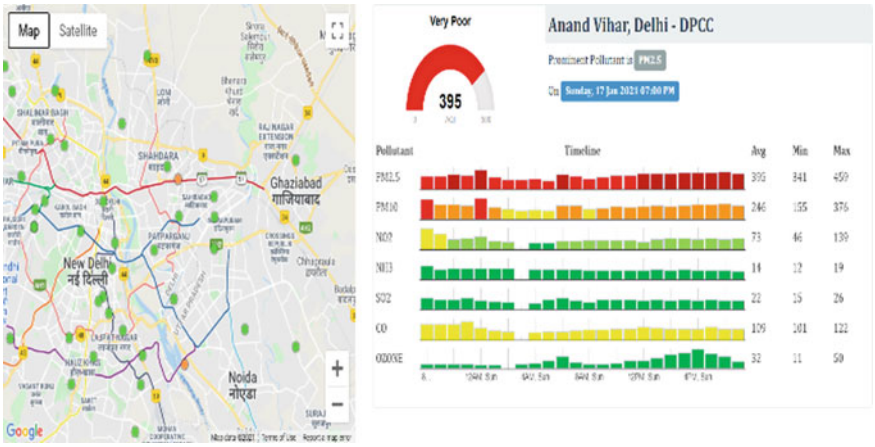


Fig. 1 National air quality index pollution monitoring centres

	City	Datetime	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	Xylene	AQI	AQI_Bucket
15	Delhi	01-01-2015 16:00	211.51	340.66	13.09	37.08	47.07	33.54	15.24	12.33	34.35	3.51	5.85	4.73	456.0	Severe
16	Delhi	01-01-2015 17:00	191.12	257.40	15.44	38.40	51.24	34.98	13.11	18.00	34.38	3.58	5.23	3.78	450.0	Severe
17	Delhi	01-01-2015 18:00	218.44	351.01	22.85	44.73	59.26	34.80	18.35	19.17	38.42	3.67	5.93	4.98	446.0	Severe
18	Delhi	01-01-2015 19:00	296.80	600.95	69.20	47.76	101.68	34.25	16.67	21.50	49.12	14.75	27.13	8.78	475.0	Severe
19	Delhi	01-01-2015 20:00	336.43	714.63	148.42	46.31	171.10	35.75	12.17	17.67	56.44	26.56	45.62	9.99	480.0	Severe

Fig. 2 Refined data

3.2 Data Pre-processing

3.2.1 Data Refinement

After cleaning the input data by removing any missing values, the analysed data was examined. An imputer function is used to perform the interpolation to determine whether a missing value (NaN) occurs if there arises an occurrence of the target object, i.e., the polluted gas. Here, mean values are utilized as a method of assessment (Fig. 2).

3.2.2 Data Visualization

Different pollutants can be pictured utilizing a chart that gives experiences about the ascent and fall of their concentration in air. The histogram graph for every one of the pollutants is plotted with *x*-axis speaking to the number of tests and *y*-axis speaking to concentration in $\mu g/m^3$ (Figs. 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12).

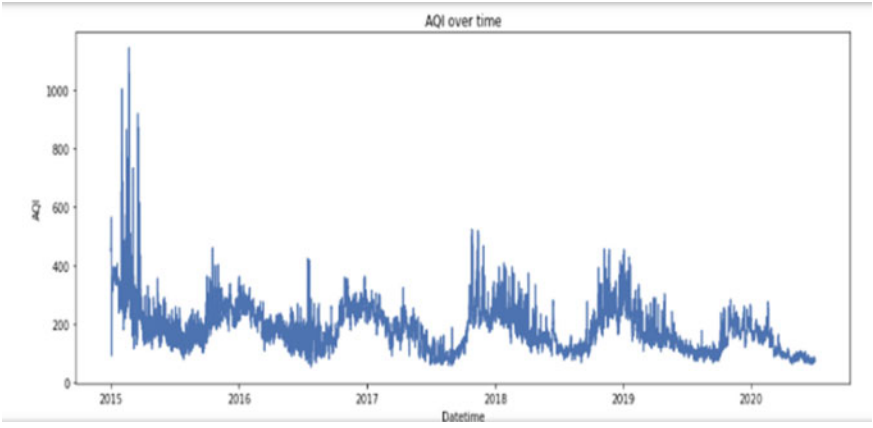


Fig. 3 AQI over time period

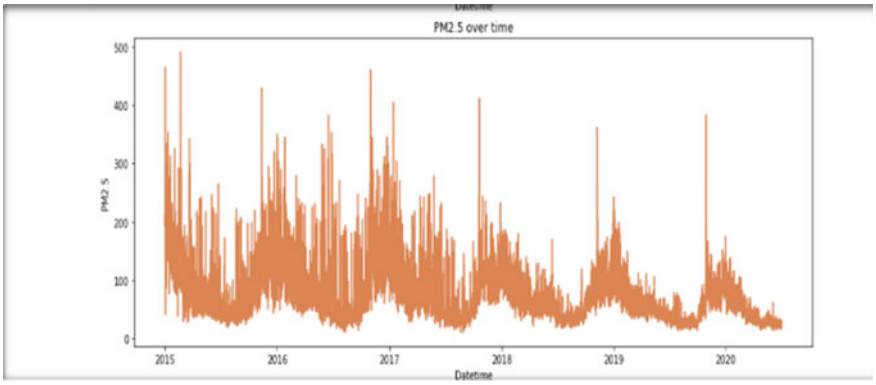


Fig. 4 $PM_{2.5}$ over time period

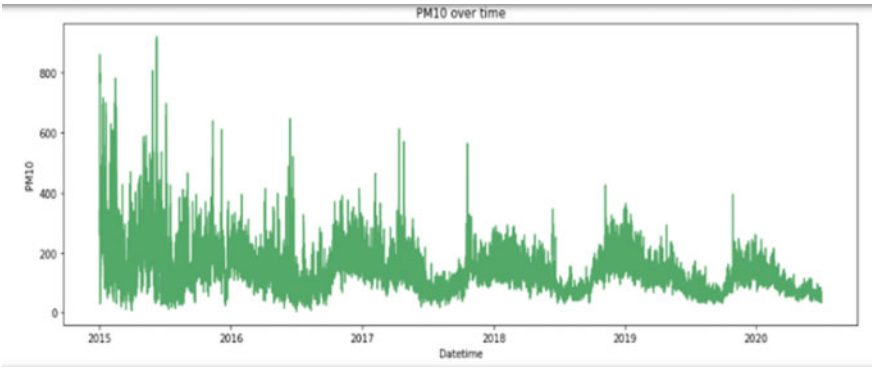


Fig. 5 PM_{10} over time period

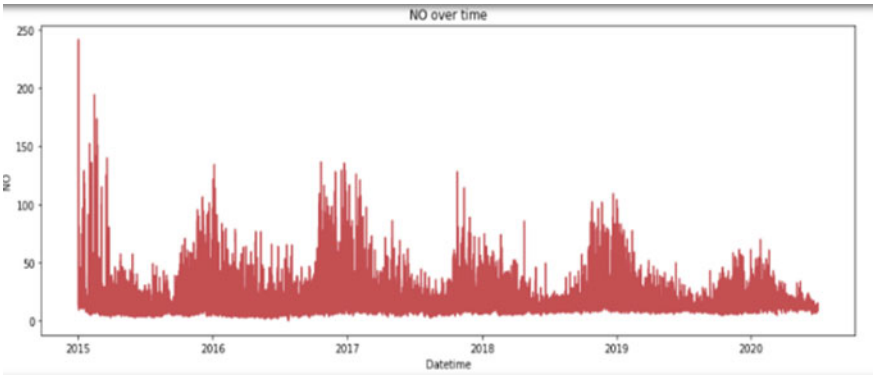


Fig. 6 Nitrous oxide (NO) over time period

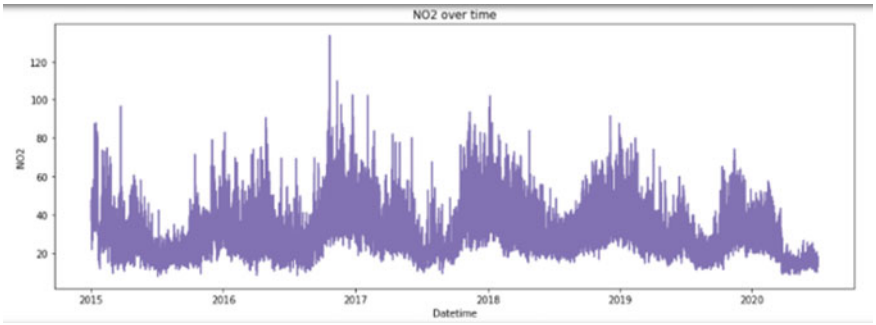


Fig. 7 NO_2 over time period

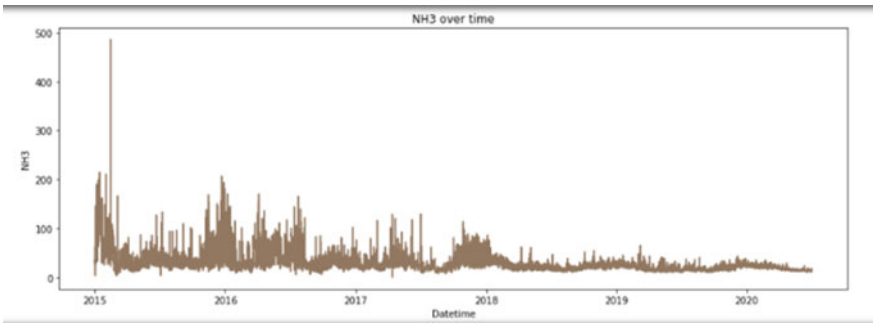


Fig. 8 NH_3 over time period

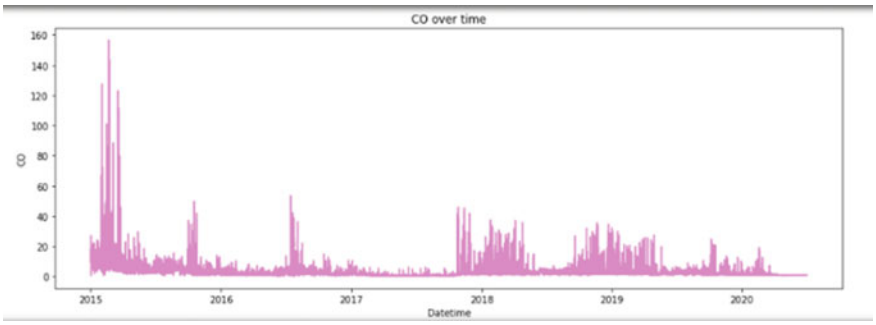


Fig. 9 CO over time period

3.2.3 Splitting Data

In order to evaluate the exhibition of a model, it is important to appropriately fit training, validation, and test data. Data does not have any standard way of being separated into preparations, validations, and tests. A half year's data has been selected

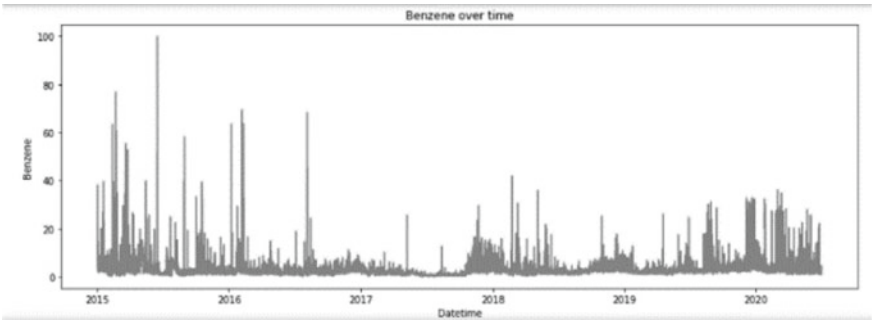


Fig. 10 Benzene over time period

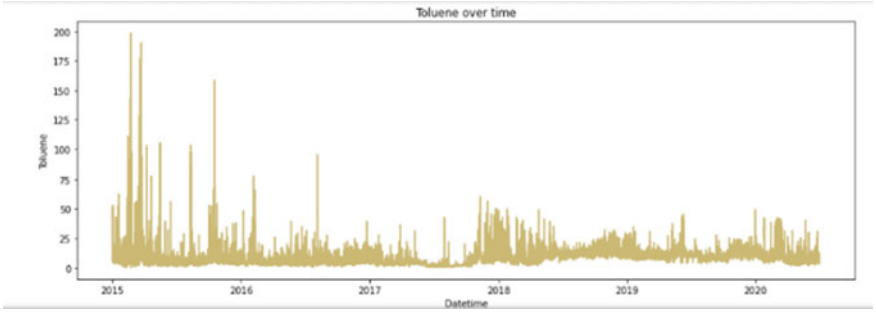


Fig. 11 Toluene over time period

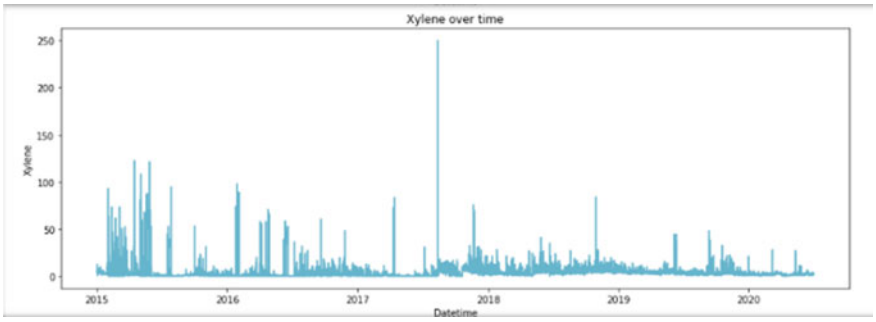


Fig. 12 Xylene over time period

for approval, a year's data has been chosen for testing, and 2 years' worth of data has been chosen for preparing. The training component contains a time-series dataset validation that includes a dataset from 1 April 2015 to 31 Walk 2017, validation includes a dataset from 1 October 2017 to 31 Walk 2018, and testing includes a

dataset from 1 April 2018 to 1 July 2020. Each data entry represents a time series data point and refers to pollutants and metrological boundaries.

Figure 2 reveals that the features in the study of the information belong to various reaches, means, and standard deviations. As a result of non-comparability of quality scopes among factors, gradients may slack and fade before meeting nearby/global minima. The Min–Max normalization allows different features to take on qualities in similar reach, eliminating the model learning issue. This allows gradients to merge more rapidly as gradients are standardized within a range of 0–1.

3.3 LSTM Model

3.3.1 Architecture

An LSTM graph is depicted in Fig. 3 as a part of the neural network architecture. There are four layers in the model: input layer, two LSTM layers, and dense layer. The LSTM layer relies on the input layer to create sequential data. The number of time steps determines how many sequences Seqi have. There are k feature vectors for each sequence. On the diagram, X represents a vector, where each of the features is represented by X_t. X_t represents the present time's concentration of pollutant p, thus Ctp. It also contains weather parameters such as temp, rh, sin, and s.r. In turn, the LSTM layer is fed with these sequences. There are several memory blocks in each LSTM layer. Gates are special multiplicative units that serve as multiplicative links between memory cells. An information gate controls the flow of information in a network which is stored in a memory cell. An LSTM cell is made up of three types of gates—output gate, input gate, and forget gate. An input gate is responsible for the flow of inputs into the memory cell, while an output gate is responsible for distributing outputs from the memory cell to the rest of the network. Forget gates address a weakness of LSTM models by allowing them to enter continuous input streams without segmenting them into sub sequences. Adaptively forgetting or resetting the memory of the cell the scaled internal state is added as input to the cell via the cell's recurrent connection. As shown in Fig. 13, an LSTM cell can be simplified to represent a box.

3.3.2 Sequential to Supervised Conversion

It is necessary to feed LSTMs marked input and yield data since it is a supervised learning algorithm. This allows sequential data to be transformed into supervised data prior to fitting the LSTM model.

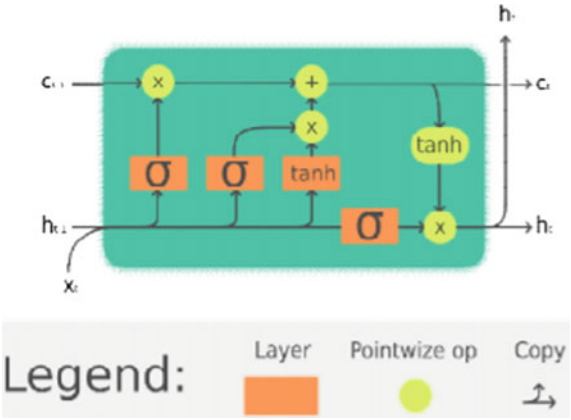


Fig. 13 LSTM cell

3.4 Model Fitting

Datasets must be prepared and tested before being used. Earlier in the model, which will determine network boundaries the preprocessed data satisfied to the model. To gauge the performance of a neural network, an optimizer is required. In neural networks, optimizers are methods or calculations used for setting various parameters such as weights, bias, and learning rates. There are different optimizers, and the decision is based on which one of them addresses the issue at hand. It is important to use different learning rates for different pollutants when predicting $PM_{2.5}$ concentration in the air (Fig. 14).

3.5 Model Evaluation Parameters

It is calculated for the model, the mean square error (RMSE), to measure how accurate the predictions are. An error of approximately 1% was produced by the proposed model.

3.5.1 Root Mean Square Error

To assess the model's performance, we calculated three error functions: the root mean square error (RMSE), the mean absolute error (MAE), and the efficiency coefficient (R^2). They can be expressed mathematically as root mean square errors (RMSEs).

$$\mathrm{RMSE} = \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(y_{i_{\mathrm{pred}}} - y_{i_{\mathrm{true}}} \right)^2 \right\}^{\frac{1}{2}} \tag{1}$$

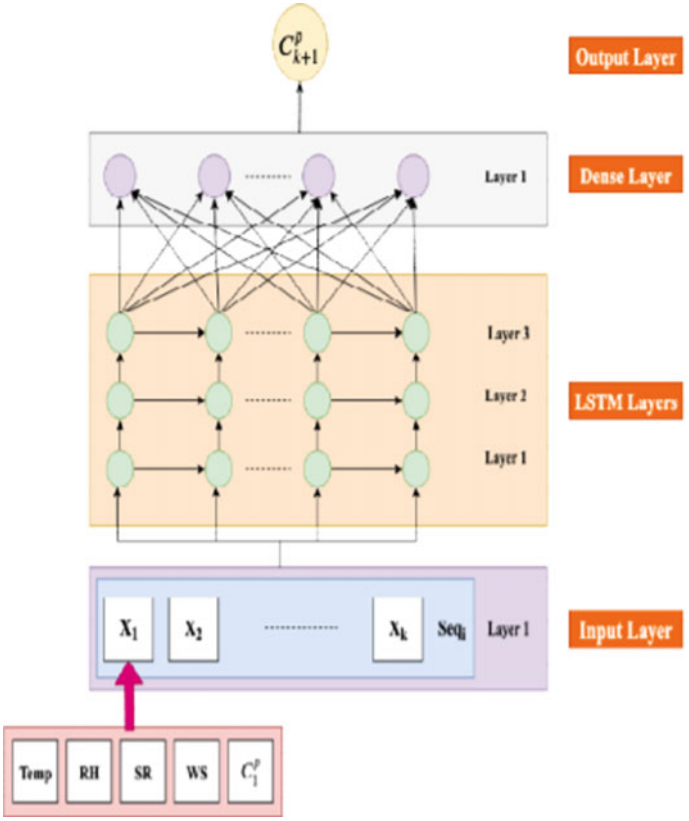


Fig. 14 Model architecture

3.5.2 Mean Absolute Error (MAE)

$$\text{MAE} = \frac{\sum_{i=1}^{n} \left| y_{i_{\text{true}}} - y_{i_{\text{pred}}} \right|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{2}$$

3.6 Training Model

Python and Scikit-learn, an open-source machine learning library were used to implement the regression strategies in segment 3. Getting to Jupyter Notebook (an open-source Python supervisor) for programming in Python was done with the Boa Constrictor Guide 5.1, which is a Python data science platform that uses open-source Python data. Each of the stations had three instances, the first being AQI of $PM_{2.5}$, the second being AQI of PM_{10}, and the last being AQI of gases. The nine training data

arrangements contained eight regression models prepared for eight sets of training data. According to the US Centers for Disease Control and Prevention, the following correlations exist between the average AQI for $PM_{2.5}$ and the actual AQI. For the remaining eight cases, comparative results were obtained.

4 Experimental Results

Multiple Python packages are used in the development of the proposed LSTM model, including Keras, Scikit-learn, and TensorFlow. In order to standardize data in the 0–1 range, Scikit-learn library uses its Min–Max scaler. For plotting all the graphs, Matplotlib is used in Python. LSTM architecture is governed by multiple variables, such as the number of epochs, batches, LSTM layers, and units per LSTM layer. By adjusting these parameters, the authors aim to balance overfitting and underfitting. Overfitting can also be prevented by using dropout layers. During training, dropout layers randomly remove units from the network, preventing too much coadaptation with the unit. The model is trained using an ADAM optimizer for 20 epochs and 70 batches. The ADAM algorithm is used to train deep-learning models based on stochastic gradient descent. With ADAM's optimization algorithm, dealing with meager slopes and uproarious issues using the most beneficial properties of both Adam Grad and RMSProp algorithms. Most issues with ADAM can be handled with the default setup parameters.

In order to assess the forecast model's health, it is essential to conduct a productive assessment. During the construction of a model, measurements are used for input and to make necessary changes until an appealing exactness is attained or no further advancement of that measurement is possible. In the light of this, the model must be evaluated earlier before it is executed on the test dataset. Different factual measurements are used to assess models as they depend on the objectives of the model, its tasks, and so forth. For example, the authors used mean square error (MSE), mean supreme error (MAE), and R^2 to assess relapse procedures. In Tables 1, 2 and 3, it is shown how the models are presented R. K. Puram, Punjabi Bagh, and Anand Vihar for each instance.

The number of epochs it must run in Fig. 15. As a result, it cycles through all of the data 20 times. In Fig. 16, authors implemented plotting the data to see how the model's loss is decreasing, which is a good indicator as the epochs increase. Figure 17 depicts their efforts in forecasting the data by considering the test data and used it to run the model effectively. As a result, the final result projections are proved to be extremely accurate.

```
Epoch 1/20
126/126 - 1s - loss: 0.0141 - val_loss: 0.0100
Epoch 2/20
126/126 - 1s - loss: 0.0054 - val_loss: 0.0112
Epoch 3/20
126/126 - 1s - loss: 0.0047 - val_loss: 0.0111
Epoch 4/20
126/126 - 1s - loss: 0.0042 - val_loss: 0.0106
Epoch 5/20
126/126 - 1s - loss: 0.0038 - val_loss: 0.0101
Epoch 6/20
126/126 - 1s - loss: 0.0035 - val_loss: 0.0093
Epoch 7/20
126/126 - 1s - loss: 0.0032 - val_loss: 0.0083
Epoch 8/20
126/126 - 1s - loss: 0.0030 - val_loss: 0.0076
Epoch 9/20
126/126 - 1s - loss: 0.0027 - val_loss: 0.0066
Epoch 10/20
126/126 - 1s - loss: 0.0026 - val_loss: 0.0057
Epoch 11/20
126/126 - 1s - loss: 0.0025 - val_loss: 0.0051
Epoch 12/20
126/126 - 1s - loss: 0.0024 - val_loss: 0.0044
Epoch 13/20
126/126 - 1s - loss: 0.0023 - val_loss: 0.0042
Epoch 14/20
126/126 - 1s - loss: 0.0023 - val_loss: 0.0037
Epoch 15/20
126/126 - 1s - loss: 0.0022 - val_loss: 0.0034
Epoch 16/20
126/126 - 1s - loss: 0.0022 - val_loss: 0.0033
Epoch 17/20
126/126 - 1s - loss: 0.0022 - val_loss: 0.0030
Epoch 18/20
126/126 - 1s - loss: 0.0021 - val_loss: 0.0029
Epoch 19/20
126/126 - 1s - loss: 0.0021 - val_loss: 0.0028
Epoch 20/20
126/126 - 1s - loss: 0.0021 - val_loss: 0.0027
```

Fig. 15 History of loss

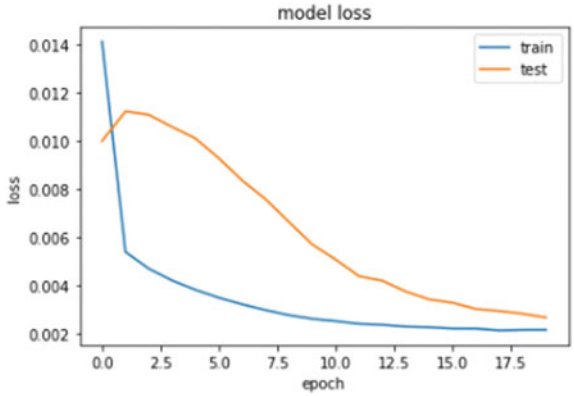


Fig. 16 Model loss graph

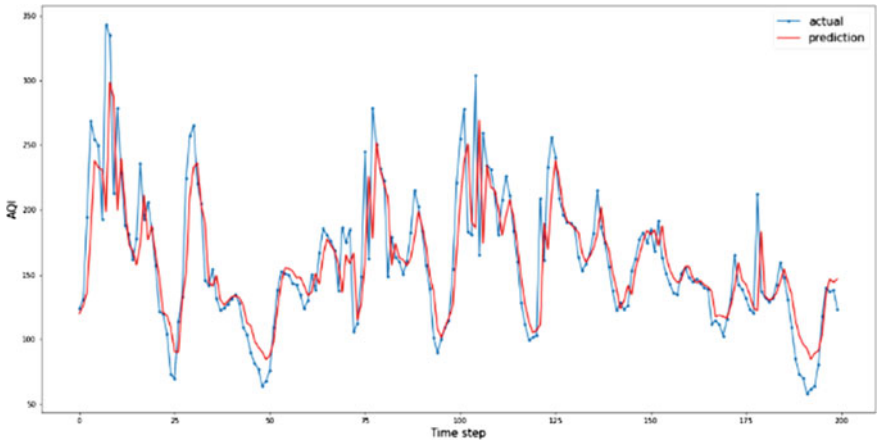


Fig. 17 AQI predicted output

5 Conclusion

Recent advances in the improvement of deep learning models have prompted a fast expansion in their application in scholarly and modern settings. In Delhi, the best natural concern is air pollution as fine PM, which comprises of fluid and strong molecule exacerbates that are hazardous to human wellbeing. Regardless of expanding levels of air toxins in Delhi, the quantity of estimation stations stays inadequate to acquire precise PM levels all through the nation. In this study, authors proposed prescient models of fine PM fixation utilizing LSTM and DAE approaches and thought about their RMSE esteems for 10-day PM10 and PM2.5fixations forecast results for Seoul. The foremost commitments of this investigation are as per the following: (1) As indicated by the trial results, authors have streamlined the LSTM and DAE model with a learning pace of 0.01, age of 20, and cluster sizes

of 70. Complete normal RMSE of expectation of PM_{10} and $PM_{2.5}$, the LSTM forecast model was more exact than the DAE model. The examination indicated that the proposed calculation can foresee and get suitable exactness among LSTM and DAE models. Later on, a future study on elective deep learning models can be done to get more precise outcomes with bigger informational collections. The proposed model can be further improvised by considering GIS-based spatial information.

References

1. G.K. Kang, J.Z. Gao, S. Chiao, S. Lu, G. Xie, Air quality prediction: big data and machine learning approaches. Int. J. Environ. Sci. Dev. **9**(1), 8–16 (2018)
2. A. Kashyap, V.K. Gunjan, A. Kumar, F. Shaik, A.A. Rao, Computational and clinical approach in lung cancer detection and analysis. Proc. Comput. Sci. **89**, 528–533 (2016)
3. G. Corani, Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning. Ecol. Model. **185**(2–4):513–529. ISSN 0304-3800
4. V.K. Gunjan, P.S. Prasad, S. Mukherjee, Biometric template protection scheme-cancelable biometrics, in *ICCCE 2019* (Springer, Singapore, 2020), pp. 405–411
5. Z. Jiang, X. Meng, C. Yang, G. Li, A BP neural network prediction model of the urban air quality based on rough set, in *2008 Fourth International Conference on Natural Computation* (2008), pp. 362–370. https://doi.org/10.1109/ICNC.2008.807
6. H. Xie, F. Ma, Q. Bai, Prediction of indoor air quality using artificial neural networks, in *2009 Fifth International Conference on Natural Computation* (2009), pp. 414–418. https://doi.org/10.1109/ICNC.2009.502
7. H. Zhao, J. Zhang, K. Wang, Z. Bai, A. Liu, A GA-ANN model for air quality predicting, in *2010 International Computer Symposium (ICS2010)* (2010), pp. 693–699. https://doi.org/10.1109/COMPSYM.2010.5685425
8. C. Xiaojun, L. Xianpeng, X. Peng, IOT-based air pollution monitoring and forecasting system, in *2015 International Conference on Computer and Computational Sciences (ICCCS)* (2015), pp. 257–260. https://doi.org/10.1109/ICCACS.2015.7361361
9. F. Shaik, A.K. Sharma, S.M. Ahmed, V.K. Gunjan, C. Naik, An improved model for analysis of diabetic retinopathy related imagery. Indian J. Sci. Technol. **9**, 44 (2016)
10. P.-W. Soh, J.-W. Chang, J.-W. Huang, Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. IEEE Access **6**, 38186–38199 (2018). https://doi.org/10.1109/ACCESS.2018.2849820
11. D. Schürholz, S. Kubler, A. Zaslavsky, Artificial intelligence-enabled context-aware air quality prediction for smart cities. J. Clean. Product. **271**, 121941 (2020). ISSN 0959-6526. https://doi.org/10.1016/j.jclepro.2020.121941

Drowsiness Detection Using IoT and Facial Expression

R. N. Ashlin Deepa, DontiReddy Sai Rakesh Reddy, K. Milind, Y. Vijayalata, and Kamishetty Rahul

Abstract In India, approximately 500,000 people are losing lives due to road accidents every year. With the rapid urbanization and development of big cities and towns, the graph of accidents is steadily increasing. Accidents due to driving after consuming alcohol and falling asleep behind the wheel have a share of about 8%. We have come up with a prototype to provide a solution to these crucial problems. This phenomenal rise in road accidents in cities is a matter of great concern and alarm to all of us. Drivers being drunk, sleeping, skipping signals, wrong route driving, and whatnot are the reasons for these frequent accidents. These recurring accidents have caused the citizens to fear driving around. They feel very insecure and vulnerable when stepping into their vehicles to travel in and around the city. So, we have designed a solution to detect drowsiness while driving. We are using image processing with the help of a camera to extract facial landmarks of the eye and image processing techniques that draw out the facial landmarks for detecting whether the driver is drowsy or awake. Additionally, we have planted an Arduino board with an MQ3 sensor to detect alcohol from the drivers' breath to check drivers' condition.

Keywords ATmega328 processor · MQ3 alcohol sensor · Arduino board · Facial recognition · Alcohol detection · Computer vision (CV) · Image processing · Facial landmark detection (FLD) · Drivers breath

R. N. Ashlin Deepa (✉) · D. Sai Rakesh Reddy · K. Milind · Y. Vijayalata · K. Rahul
Department of Computer Science and Engineering, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India
e-mail: deepa.ashlin@gmail.com

Y. Vijayalata
e-mail: vijaya@ieee.org

Y. Vijayalata
Department of Computer Science and Engineering, KG Reddy College of Engineering & Technology, Hyderabad, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_61

1 Introduction

Road accidents are one of every mother's worst fears, as well as the nightmares of every human on the road. Accidents that occur when a motorist is under the influence of alcohol or falls asleep behind the wheel account for around 10% of all accidents, affecting numerous families. According to the Ministry of Road Transport and Highway, there were 4552 accidents recorded in India each year, resulting in the deaths of thousands of people due to inattentive driving (road accidents in India 2016). Approximately, 20% of people have acknowledged falling asleep while driving, with 40% saying they had done it at least once throughout their driving careers. Research shows, in India, 40% of highway crashes or near-crashes occur due to drowsy driving. However, more than 50% of all deadly highway crashes which involve more than two cars are alcohol-related. Based on National Highway Safety Administration report (NHTSA's) [1] 2.2–2.6% of accidents happened every year because of drowsy drivers during 2005–2009. According to [2–5], alcohol consumption increases the severity of traffic accidents and direct medical health costs. This can be the reason that causes the road accident.

The research and studies done in the past were eclectic but could not solve the issue with the influence of alcohol. Implementing these innovative ways to keep a check on people using various means of transport can prevent road accidents by a great deal and ensure public safety. It symbolizes our initiative, which aims to make human driving safer and accident-free. There are a variety of sleepiness detection systems that measure driver tiredness levels while driving and alert them if they are not paying attention to the road. Facial expressions like yawning, eye closing, and head motions may all be used to infer how weary someone is in terms of behaviour-based methods, some related investigations [6] have been presented to determine the fatigue as a percentage of eyelid closure (PERCLOS) by detecting the eyelid closure frequency. As an unconscious behaviour caused by fatigue, yawning is also used for visual fatigue detection. References [7–9] achieve good results in fatigue testing using yawning facial video, which verifies the feasibility of fatigue detection through facial expressions. Computer vision techniques based on artificial neural networks (ANNs) have been successfully (and still being) applied to many road safety problems (e.g. traffic safety analysis of toll plazas and identifying behavioural changes among drivers [10]). The biological condition of the driver and the automotive behaviour are researched utilizing behavioural, vehicular, and physiological parameter-based techniques to determine driver tiredness. Some investigations have demonstrated that heart rate variability (HRV) is related to fatigue. HRV is used to measure the change features of heart rate [11] (Fig. 1).

While face recognition technology is rapidly evolving, we cannot still combine alcohol detection and facial identification. Our technology seeks to reduce the number of road accidents caused by intoxicated drivers soon. It can be used in any vehicle and is undetectable to criminals. This initiative intends to ensure the safety of those who are seated within a vehicle.

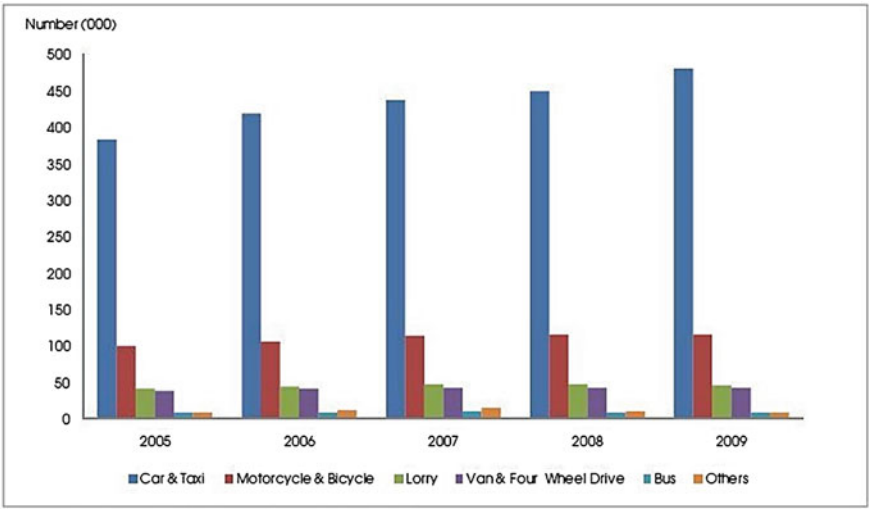


Fig. 1 Road accidents of various types of vehicles in Malaysia from the year 2005–2009

2 Related Work

In the United States, 2.5% of incidents were recorded in 2009. Throughout these five years, the number of accidents decreased slightly. But, the ratio of accidents caused by drowsiness tends to be stable from 2.3 to 2.7% [1]. Different AI procedures like PERCLOS calculation, HAAR-based course classifier, and OpenCV are used to decide the driver's condition including driver languor recognition [12]. In 2017, [13] used descriptive and binary logit model to identify the factors affecting the occurrence and severity of rear-end crash. They found that the likelihood of being involved in fatal rear-end crashes in rural roads is 2.889 times higher than in urban roads, and 1.923 times with speeding behaviour of drivers. However, the probability to be involved in rear-ends crash injuries increases in factors such as: tailgating; Abu Dhabi licensed drivers; driving experience between 0 and 4 years; and passenger car type. Also, it was found that the severity of rear-end crashes is less on relatively wider roads (road with four lanes or more). In instances where alcohol may be present, it is critical to evaluate the user's mental condition. To solve this, a hardware system that is based on infrared light, following the face detection step, the facial components that are more important and considered as the most effective for drowsiness, are extracted and tracked in video sequence frames [14]. Immediate detection of sleepiness during driving is important [15]. The system is capable of distinguishing between a standard eye blink and a drowsiness-related eye blink. It can work in low-light situations and even while the driver is wearing spectacles. This method was successful in distinguishing the normal eye blink from drowsiness-related eye blink, but it couldn't provide a solution when the user was observed wearing a pair of glasses. Iris recognition is one of the methods used to assess the user's state of mind. Iris recognition technology is continuously

growing over the years and could resolve drunken driving accidents worldwide. Lately, with a single camera view, the system now uses OpenCV and Raspberry Pi modules. Image processing methods determine the eye state. This study solely considers the condition of the eyes and ignores the frequency with which people yawn [16]. It is crucial to discover a strategy to assist the user's safety as the priority. The eye closure is detected using HAAR-based cascade classifier and an alcohol gas sensor that functions as a Breathalyser [17]. The downside to Haar cascades is that they tend to be prone to false-positive detections, require parameter tuning when being applied for inference/detection, and, in general, are not as accurate as the more modern algorithms we have today. A sleepy driver's driving pattern may differ from that of a regular driver. Driver sleepiness may be predicted by monitoring a few metrics such as lane departure, steering movement, abrupt changes in acceleration, gas pedal, and brake pedal [18].

But, adding too many parameters to monitor often introduces new problems. The prime causes of driver sleepiness include sleep deprivation, constant work for long hours, the use of sedative medications, working at odd hours, and untreated sleep problems [19]. A normal individual driving a car will have a predictable driving pattern; nevertheless, if he strays from his lane or any other vehicle-based metric, an unwanted alarm may be sent. The vehicle-based metric has only been employed by a few researchers since it may lead to more false positives [20]. The early detection of sleepiness is critical to avoid accidents and the loss of precious lives. Although considerable research has been done on detecting driver sleepiness, the majority of it falls into three categories: vehicle-based measurements, physiological-based measures, and behavioural-based measures [21]. Du et al. [22] proposed a multimodal fusion recurrent neural network (MFRNN) framework. They used RGB-D cameras and infrared video to extract the driver's eye-opening degree, mouth opening degree, and heart rate information, and at the same time, extracted time information related to each fatigue feature to improve the performance of driver fatigue detection. The MQ3 alcohol sensor was utilized to detect the presence of alcohol in human breath [23]. This paper proposes a method for detecting the drowsy state based on the time-series analysis of the angular velocity of the steering wheel. Thus, according to the literature, using an alcohol sensor attached to an Arduino board proposed system detects driver's drowsiness. It uses a camera and image processing algorithms to extract the facial landmarks of the eye and eye aspect ratio to determine if the driver is drowsy or awake. A framework was proposed using a convolutional neural network (CNN) to effectively extract the deep representation of eye and mouth-related fatigue features from the face area detected in each video frame, based on the factorized bilinear feature fusion model, we performed a nonlinear fusion of the deep feature representations of the eyes and mouth [24]. In Fig. 2 we can see various driver drowsiness detection methods use currently.

The system which is proposed in this paper covers cases where driver can be drunk but cannot be detected in the alcohol sensor, this is where the drowsiness detection algorithm comes to work which keeps a check on the driver whether he is falling asleep behind the wheel. Further, in the paper, we have given detailed flow of the system developed along with analysis and experimental results.

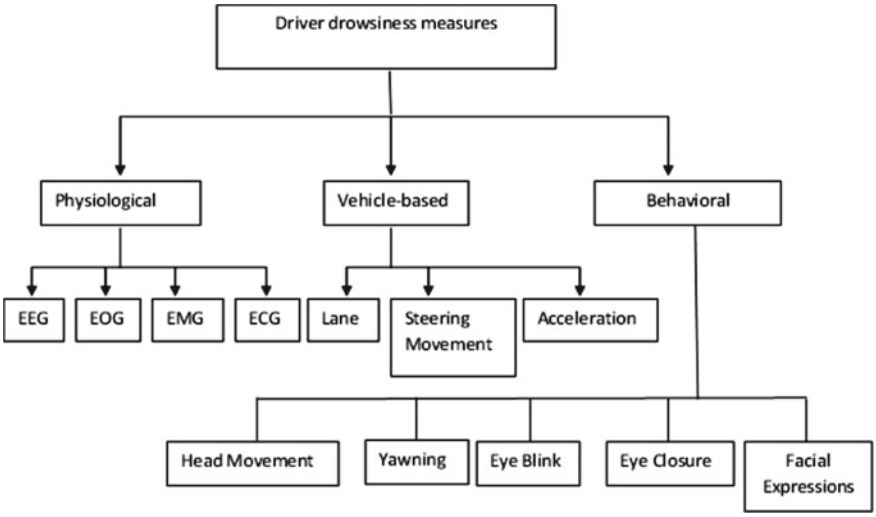


Fig. 2 Existing systems on driver drowsiness measures

3 Methodology

Detecting drowsiness of drivers is still ongoing research to reduce the number of such miss-happenings and accidents. This research work discusses advancements of using an alcohol detector, a device that detects a change in the alcoholic gas content of the surrounding air. These devices are also known as breath analysers because they analyse the alcohol level in the breath. When the analyser detects the presence of alcohol in the car, the engine is instantly locked. We introduced face detection algorithms and the Internet of Things (IoT), two highly advanced and now extensively used technologies, into our project to address this latter issue. Identifying the user's state of mind and taking immediate action is a complex task. There are various methods and techniques to assess the state of mind of the user.

We have taken up this project to minimize accidents percentage. Any vehicle out there can use this system and is hidden from suspects. It also ensures the safety of those who are seated within a vehicle. This paper aims to summarize the development of the system and gather the latest methods so it can become a base for researchers in future. When compared with the traditional procedure, this method provides a series of advantages [25]. This paper proposes a comprehensive and complete study of ongoing works identified with it additionally presents. Here, we use OpenCV and Dlib library to enhance detailed feature extraction of the eyes, the mouth, and positions of the head to estimate the driver's level of drowsiness.

The ATmega328 processor in the Arduino can perform more tasks than traditional microcontrollers. We are utilizing an MQ3 alcohol sensor to detect the presence of alcohol in human breath. Alerting the driver if he's too drunk can help prevent damage of all sorts. The face recognition algorithm works quite effectively

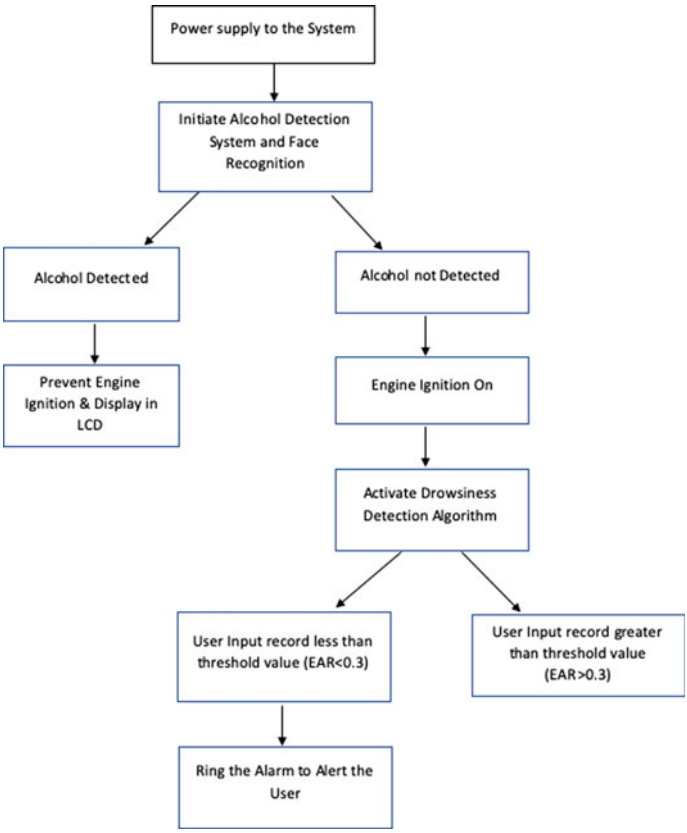


Fig. 3 Architecture of proposed methodology

in checking whether the driver is falling asleep, cautioning him to ensure safety to the fullest. Currently, there are no effective systems used in vehicles to tackle such kind of problem. However, this developed system can make a massive difference in saving lives and looking after their safety. The flow of the proposed methodology is given in Fig. 3 and the methodology is given in Algorithm 1.

3.1 Algorithm 1

Step 1: Load the alcohol detection algorithm integrated with the Arduino board along with the MQ3 sensor.

Step 2: Take readings from the user which are taken in by the MQ3 sensor and converted into analogue values.

Step 3: Compare the analogue values generated and processed by the controller with the threshold values set.

Step 4: If the input record is seen to be greater than the threshold, go to step 6.

Step 5: If the input record is seen to be less than the threshold value set, halt the alcohol detection algorithm and go to step 7.

Step 6: The engine ignition is system is locked since the input record of the user is greater than the threshold value set.

The driver drowsiness detection algorithm follows as:

Step 7: Activate the driver drowsiness detection algorithm.

Step 8: OpenCV library is activated for image processing, facial landmarks extraction is done continuously for a window of 48 frames.

Step 9: The eye aspect ratio threshold of 0.3 is set and compared with the user's records.

Step 10: NumPy extracts the eye-coordinates of user and therefore eye aspect ratio of the user's input record is calculated.

Step 11: The eye aspect ratio is compared with threshold and the counter for sounding the alarm is incremented.

Step 12: Alarm is activated when the user's eye aspect ratio is found to be less than 0.3 (EAR < 0.3) and repeat from step 7.

Step 13: Go to step 7 if the eye aspect ratio of the user is found to be greater than the threshold.

Step 14: Repeat this until the user is found to be drowsy and the engine ignition is locked.

The proposed method of the project includes both image processing and alcohol detection systems. The image processing project uses Python programming language. Since many image processing tools, functions, and libraries may be modified to create a robust environment, the programming language Python is being employed. There is no data collected for these projects. Both of our projects are developed in a real-time environment. For the drowsiness detection project, we are taking in the live video stream as our data. Since it is paramount to alert the driver at that very moment, we are using image processing techniques for live feed and not some pre-stored datasets. The same scenario applies to alcohol detection. The MQ3 sensor planted with Arduino board detects alcohol from the drivers' breath. Algorithms and approaches are used for proper implementation and functioning, and they are detailed below.

3.2 Drowsiness Detection

The drowsiness detection algorithm extracts the user's facial landmarks with precision to alert the user when he closes his eyes for longer than usual. We used various top-notch modules and functionalities in the implementation of our drowsiness detection algorithm. They are

- OpenCV is an open-source vision and machine learning applications library. This algorithm is applied to recognize objects identified, human activities classified in photographs, and camera motion tracked.
- To make working with OpenCV simple, we would need the imutils kit, my computer view sequence, and the image processing functions.
- SciPy: The SciPy kit is needed in an eye aspect ratio measurement to calculate the Euclidean gap between facial landmarks.
- The thread class ensures that our script will continue to execute even if the warning noises are played on a different thread from the main thread.
- We need the play-sound library to play our WAV/MP3 alarm.

We would need the Dlib library to detect and locate visual landmarks.

3.3 Alcohol Detection

In order to detect the alcohol levels of the operator, the MQ3 (alcohol) sensor is installed on the control unit. Analogue values are generated by the sensors and processed online by the controller. IoT updates sensor data to the server constantly. If the value of the alcohol sensor is at the limit, the gadget prevents accidents by stopping the automobile ignition system and keeps the intake data in the user's record. The control air is continuously checked in the cloud and updated with the assistance of the MQ3 (alcohol sensor). The mechanism interrupts the ignition system if the alcohol measurements surpass the threshold value. This is done by preventing the supply of fuel to the ignition system. The Arduino Uno is used for the entire system. Arduino Uno is responsible for the MQ3 module, LCD display. The interface of all modules is programmed in order to work the whole of the module. The panel may be connected to the personal computer, and the microcontroller programming can be done for the sensor to operate and to breathe. Reading is shown on the LCD board interfacing with the Arduino Uno board. Once the sensor detects, the information is transferred to the automotive ignition system that does not start the engine.

The MQ3 (alcohol) sensor is mounted on the driver's steering to detect the driver's alcohol level. The sensors produce analogue values and are interpreted by the controller via the internet. The sensor data is continuously updated with IoT to the server. When the alcohol sensor values reach the threshold, the device avoids accidents by halting the car ignition system and stores the alcohol intake values into the vehicle user's log. With the aid of the MQ3 (alcohol sensor) the air exhaled by the

controller is constantly monitored in the cloud and updated by the IoT. If detection of alcohol exceeds the threshold value, the system interrupts the ignition system. This is done by preventing the supply of fuel to the ignition system.

4 Analysis

The test is critical because it finds defects/bugs that assure software quality before sending the client. It makes the programme easier to use and more dependable. Software testing guarantees that software is reliable and efficient. When testing any product or project we need to check it on various bases and use all the possible test cases. To determine if a product is robust enough to function in several conditions, it must be tested that the system stands out in all situations and environments. According to previously reported approaches, there are very few public datasets currently available for comprehensive performance evaluations of different approaches for driver drowsiness detection, particularly those with driver attention information from real-world driving scenarios [26] (Fig. 4).

The above graph depicts eye aspect ratios of the user input records over a time period window frame. The frequency of the eye aspect ratio values is plotted against the actual values taken from the user. The pulse in the graph is observed to be constant over a period of time when the values of the user readings are found to be less than the threshold. Peak is observed at the median of the graph when the values of the eye aspect ratio cross the threshold of 0.3. The yellow region depicts the time frame in which the user is assumed to be drowsy due to the input recordings. The red region depicts the peak confirming the user's drowsiness state (Fig. 5).

TEST CASE 5.A: After running the image processing drowsiness detection algorithm, the algorithm extracts the facial landmarks and monitors the user's live stream

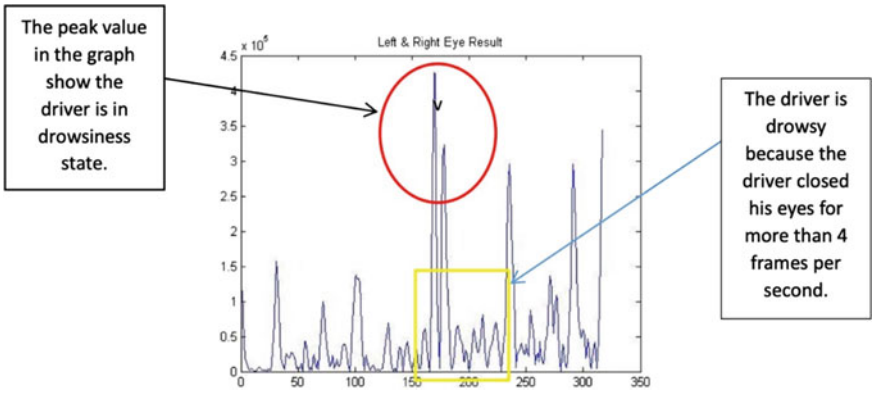


Fig. 4 Drowsiness and alcohol detection peak values

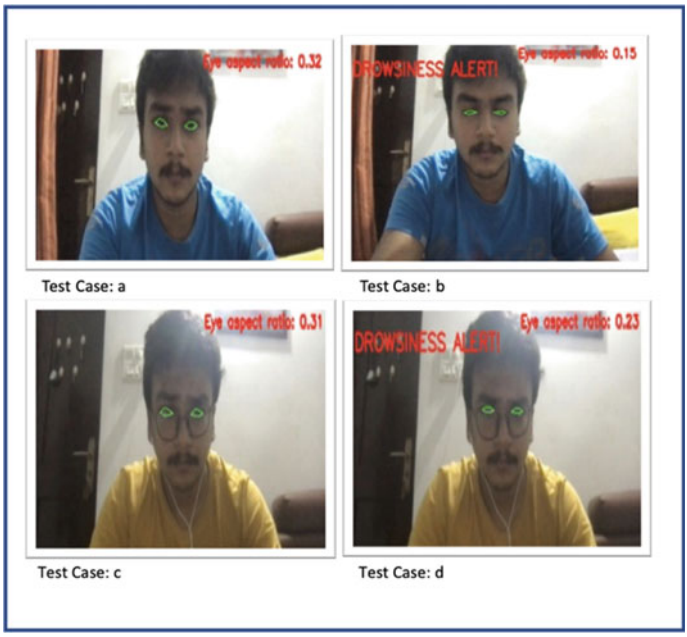


Fig. 5 Model test cases

for contiguous frames of 42. The average aspect ratio in this case, where the user keeps his eyes open for a certain number of frames is 0.3. Since the result of the drowsiness detection algorithm is seen to be greater than the threshold, No Drowsiness Alert was detected eye aspect ratio greater than 0.3(i.e. 0.32 > 0.3). This indication ensures a safe journey without alerting the alarm to halt the driver's journey.

TEST CASE 5.B: After running the image processing drowsiness detection algorithm, the algorithm extracts the facial landmarks and monitors the user's live stream for contiguous frames of 42. The average aspect ratio in this case, where the user keeps his eyes open for a certain number of frames is 0.3. Since the result of the drowsiness detection algorithm is seen to be less than the threshold the value is taken to be critical. Drowsiness Alert, the alarm is triggered which would help the driver get hold of the situation. This is when the eye aspect ratio was found less than 0.3 (i.e. 0.23 < 0.3).

To test our model further to its extremes, we used a pair of glasses to assess the performance of the model. Which resulted in:

TEST CASE 5.C: After running the image processing drowsiness detection algorithm, the algorithm extracts the facial landmarks and monitors the user's live stream for contiguous frames of 42. The average aspect ratio in this case, where the user keeps his eyes open for a certain number of frames is 0.3. Since the result of the drowsiness detection algorithm is seen to be less than the threshold the value is taken

to be critical. Drowsiness Alert, the alarm is triggered which would help the driver get hold of the situation. This is when the eye aspect ratio was found less than 0.3 (i.e. 0.31 > 0.3).

TEST CASE 5.D: After running the image processing drowsiness detection algorithm, the algorithm extracts the facial landmarks and monitors the user's live stream for contiguous frames of 42. The average aspect ratio in this case, where the user keeps his eyes open for a certain number of frames is 0.3. Since the result of the drowsiness detection algorithm is seen to be less than the threshold the value is taken to be critical. Drowsiness Alert, the alarm is triggered which would help the driver get hold of the situation. This is when the eye aspect ratio was found less than 0.3 (i.e. 0.19 < 0.3).

TEST CASE 3.A: We have tested the threshold values of the alcohol detection model by giving manual user inputs. When no alert alcohol consumption, when the input is less than threshold value, i.e. (215 < 500).

TEST CASE 3.B: We have tested the threshold values of the alcohol detection model by giving manual user inputs. When alert alcohol consumption, when the input is more than the threshold value, i.e. (560 > 500).

In our research, we used two-way analysis as shown in Table 1. Our first phase includes the information gathered by the model when the driver is facing the camera. The data from the first stage is used in the second phase, which comprises a thorough examination of the data using machine learning classifiers to determine whether the proposed approach is helpful. Naive Bayes, support vector machine, and random forest were the classifiers used in the empirical analysis. We evaluated the results obtained using performance measures to evaluate the classifiers' performance.

Table 2 depicts the values of the accuracy, precision, recall, TPR, FPR, classifier, and F-Measure against the various classifiers that we have used. We have collected a dataset of 750 images comprising of the images of various users who have been tested against the drowsiness detection algorithm in distinct circumstances. The dataset consists of the predictions made by the algorithm. The validity of the algorithm is compared with the actual conditions. These images are run through some of the best classifiers such as SVM, random forest, and Naive Bayes algorithm. SVM proves to have an accuracy of 80% when tested with a fresh set of images. The Naive Bayes algorithm considers a wide variety of possibilities choosing the best set of combinations. It proved to be 80% accurate similar to the SVM classifier. We have

Table 1 Results of user input records

Test case	Threshold value	User input record	Result (displayed message)
3.a	500	250	"Have a safe journey"—Engine ignition system on
3.b	500	560	"Alcohol level high, shouldn't drive!"—Engine ignition system locked

Table 2 Evaluation of classifier's performance

S. No.	Classifier	TPR	FPR	Accuracy	Precision	Recall	*F*-measure
1	SVM [13]	80	20.4	80	80.1	80	79.9
2	Random forest (proposed method)	84	16.1	84	84	84	84
3	Naive Bayes [27]	80	20.7	80	80.7	80	79.8

used the random forest algorithm as it proves to be the best among the classifiers used, having an accuracy of 84%.

From the above statistics, we can enumerate that the random forest classifier gives the best classification results with an accuracy of 84%.

5 Conclusion

An alternative perspective is utilizing a camera to determine the driver's drowsiness using the driver's facial areas. Based on the findings of image analysis, the driver is categorized as sleepy or not. Image data analysis methods are advantageous and properly represent raw data. The technology was also tested while in motion under various lighting conditions. A quantitative and a qualitative result backed up the proposed plan. Since the focus of our study is on facial recognition and alcohol detection, the entire system is in charge of preventing accidents. Our system acts more effectively in these cases. The model could be developed even more by supporting authorities in the event of an emergency. We have devised a very successful way of constructing an intelligent system for alcohol detection in automobiles with Arduino at its heart. One of the perks of this sensor is that its range is about 2 m, so it is possible to install it in secret. The system as a whole also has the advantage of being lightweight and dependable. The long-term purpose of the model is to reduce the number of alcohol-related and drowsiness-related accidents. Integrated measures such as combining physiological and behavioural-based or vehicle-based and behavioural-based can provide more accuracy than individual measures. Since this research is also dependent on road conditions, light effect, and traffic situation, further trial findings for other nations are necessary. This study is quite helpful in projecting accidents to the greatest extent possible. Prediction of accidents can assist people who are looking to stay safe and secure. So, they may arrange the necessary protocols accordingly once they know the possibility in future. Automobile organizations can benefit since they are aware of the trend and need in security and guarantee the safety of their customers.

References

1. NTHSA, Traffic safety facts, in *NHTSA's National Center for Statistics and Analysis* (Washington, 2011)
2. Q. Wu, G. Zhang, Analysis of driver injury severity in single vehicle crashes on rural and urban roadways. Accid. Anal. Prev. **94**, 35–45 (2016)
3. S.N.-L. Carlos Gomez-Restrepo, The influence of alcohol in traffic accidents and health care costs of it in Bogotá-Colombia. Clin. Epidemiol. **86**, 106–110 (2017)
4. C. Chen, G. Zhang, Examining driver injury severity outcomes in rural non-interstate roadway crashes using a hierarchical ordered logit model. Accid. Anal. Prev. 79–87 (2016)
5. N. Amarasingha, S. Dissanayake, Gender differences of young drivers on injury severity outcome of highway crashes. J. Saf. Res. **49**, 113.e1–120
6. V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
7. B. Mandal, L. Li, G.S. Wang, J. Lin, Towards detection of bus driver fatigue based on robust visual analysis of eye state. IEEE Trans. Intell. Transp. Syst. **18**(3), 545–557 (2017)
8. L. Zhao, Z. Wang, X. Wang, Q. Liu, Driver drowsiness detection using facial dynamic fusion information and a DBN. IET Intell. Transp. Syst. **12**(2), 127–133 (2018)
9. K. Fujiwara et al. Heart rate variability-based driver drowsiness detection and its validation with EEG. IEEE Trans. Biomed. Eng. **66**(6), 1769–1778 (2019)
10. J. Wijnands, J. Thompson, G. Aschwanden, M. Stevenson, Identifying behavioural change among drivers using long short-term memory recurrent neural networks. Transp. Res. Part F Traffic Psychol. Behav. **53**, 34–49 (2018)
11. J. Vicente, P. Laguna, A. Bartra, R. Bailón, Drowsiness detection using heart rate variability. Med. Biol. Eng. Comput. **54**(6), 927–937 (2016)
12. A. Sajikumar, A. Aji, J. Baby, A. Prasad, A.K. Stanly, N.M. John, Save the drowsy driver drowsy driver detection. IJERT (2021)
13. L. Pauly, D. Sankar, *Detection of Drowsiness Based on HOG Features and SVM Classifiers* (IEEE, 2015)
14. V.K. Gunjan, P.S. Prasad, S. Mukherjee, Biometric template protection scheme-cancelable biometrics, in *ICCCE 2019*, (Springer, Singapore, 2020), pp. 405–411
15. K.H. Lee, W. Kim, H.K. Choi, B.T. Jan, *A Study on Feature Extraction Methods Used to Estimate a Driver's Level of Drowsiness* (IEEE, 2019)
16. S.M. Ahmed, B. Kovela, V.K. Gunjan, IoT based automatic plant watering system through soil moisture sensing—a technique to support farmers' cultivation in rural India, in *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies* (Springer, Singapore, 2020), pp. 259–268
17. H. Wakana, M. Yamada, M. Sakairi, in *Portable Alcohol Detection System with Breath Recognition Function* (IEEE, 2018)
18. C.J. de Naurois, C. Bourdin, A. Stratulat, E. Diaz, J.L. Vercher, Detection and prediction of driver drowsiness using artificial neural network models. Accid. Anal. Prev. **126**, 95–104 (2019)
19. B. Reddy, Y.H. Kim, S. Yun, C. Seo, J. Jang (2017) Real-time driver drowsiness detection for embedded systems using model compression of deep neural networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition on Workshops*, pp. 121–128
20. A. Kashyap, V.K. Gunjan, A. Kumar, F. Shaik, A.A. Rao, Computational and clinical approach in lung cancer detection and analysis. Proc. Comput. Sci. **89**, 528–533 (2016)
21. P.S. Prasad, R. Pathak, V.K. Gunjan, H.V. Ramana Rao, Deep learning based representation for face recognition, in *ICCCE 2019* (Springer, Singapore, 2020), pp. 419–424
22. G. Du, T. Li, C. Li, P.X. Liu, D. Li, Vision-based fatigue driving recognition method integrating heart rate and facial features. IEEE Trans. Intell. Transp. Syst., 1–12 (2020)
23. P. Ingalepatil, P. Barhate, B. Nemade, V.D. Chaudhari, Alcohol detection system in vehicle using Arduino. IRJET (2017)

24. S. Chen, Z. Wang, W. Chen, Driver drowsiness estimation based on factorized bilinear feature fusion and a long-short-term recurrent convolutional network. Information **12**, 3 (2021)
25. B.G. Pratama, I. Ardiyanto, T.B. Adji, *A Review on Driver Drowsiness Based on Image, Bio-Signal, and Driver Behavior* (IEEE, 2017)
26. M. Ramzan, H.U. Khan, S.M. Awan, A. Ismail, M. Ilyas, A. Mahmood, A survey on state-of-the-art drowsiness detection techniques. IEEE Access **7**, 61904–61919 (2019)
27. S. Bakheet, A. Al-Hamadi, in *A Framework for Instantaneous Driver Drowsiness Detection Based on Improved HOG Features and Naïve Bayesian Classification* (MDPI, 2021)

Deep Learning-Based Suppression of Speckle-Noise in Synthetic Aperture Radar (SAR) Images: A Comprehensive Review

Ashwani Kant Shukla, Sanjay K. Dwivedi, Ganesh Chandra, and Raj Shree

Abstract In synthetic aperture radar (SAR) images, the speckle noise is inherent which is in granular form and oftenly defined as a multiplicative nature. Due to multiplicative nature not only the qualities of the SAR images get degraded but also all coherent images. In order to deals with granular form of noise, there are various despeckling techniques has been developed from the few decades. The primary objective of this paper is to majorly focus on the existing despeckling model with its emerging trends and greatly transform in mechanism. The nature cum characteristics of the speckle noise has been discussed in detail. Furthermore, the cons of the homomorphic filtering mechanisms are highlighted. The classification of various despeckling methods based on spatial and transform domain which can play the most significant role in the designing of the robust techniques has addressed. The efficiency of the despeckling methods gets analyzed using various performance metrics such as SNR, PSNR, SSIM, and UIQI. Also, the visual appearance of the any despeckling method can be analyzed using confusion matrix and Cohen's kappa method.

Keywords SAR images · Multiplicative nature · Speckle noise · Despeckling methods · Performance metrics

1 Introduction

Synthetic-aperture-radar (SAR) can acquire extreme resolution by coherent processing of scattered rays. However, this is hindered by the speckles appears in high resolution images that impair the ability of human observers to deal with

A. K. Shukla · S. K. Dwivedi · G. Chandra
Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Uttar Pradesh, Lucknow, India
e-mail: ashwanikant.shukla@gmail.com

R. Shree (✉)
Department of Information Technology, Babasaheb Bhimrao Ambedkar University, Uttar Pradesh, Lucknow, India
e-mail: rajshree.bbau2009@gmail.com

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_62

fine details. Unlike optical sensors, SAR active sensor is bound by speckle noise that outcomes from articulate imaging systems. In SAR images, the speckle noise obtains from the arbitrary deterrent of multiple primary scatters systems within a concentric block. Due to change in nature the characteristics of these noise is different from the noise acquired by passive sensors like optical sensors images [1]. The speckle noise in SAR images detected in granular form and has pros of a Rayleigh distribution. SAR images are employed by the observer to retrieve the information which is assist to detect defined region. But, the preliminary challenges for the SAR images, i.e., speckle-noise results as the quality images get degraded and this interference occurs when the signals are backscattered to the airplane [2]. So, the robust techniques reinforcement in suppressing the speckle noise is penetrating for the evaluation of the findings constituted in different SAR images. Precisely, tackle the granular form of noise, there are many despeckling techniques have been proposed over a decades. For example, in the Lee filter the multiplicative model is first estimated by a linear combination of the local mean and observed pixels. After that, the minimum mean-square error (MMSE) is applied to find the weighting constant to be employed to design the techniques. In spite of smartly dealing the speckle-noise in SAR images, these conventional techniques fundamentally show the cons in preserving the fine details such as edge, and texture [3]. To overcome these shortcomings, a multi-level geometric analysis technique known as wavelet transform has been proposed and is widely adopted in image processing. The wavelet-based despeckling techniques typically recover a despeckled image by treating the wavelet coefficient as some model and building an estimator based on a prototype or employ hard or soft threshold directly to single-scale wavelet coefficients [4]. Also, the outcomes have been shown that wavelet thresholding technique can give the better suppression in speckle noise than spatial-domain techniques. SAR technology has had a major breakthrough with the idea initiation for compound SAR imagery. The definition is relies on the concept of heterogeneity to gain multiple SAR images between the variance that one or more imaging elements. The compound SAR image model has opened several feasible vast amounts of emerging use cases. In fact that after making lots of enhancement in this process but still draws the attention of the researchers that multiform SAR data have the greater potential for quantitative radar remote sensing uses. The increasing number of SAR challenges has illuminated that speckle-noise is one of the preliminary cons of SAR technology [5].

The paper has been organized in the following manner: In Sect. 2, it has shown mathematically the multiplicative nature of speckle noise. The various conventional and non-convention existing despeckling methods with its characteristics, pros, and cons has been addressed in Sect. 3. In Sect. 4, the complementary theories which can assist in the development of the robust despeckling methods and how to use them intelligently have also been discussed. There are various performance metrics such as with-reference and without-reference indexes employ to analyze the efficiency of any speckle-noise suppression techniques has been discussed in Sect. 4. In Sect. 5, conclusion is drawn containing key findings that can be used in the development of speckle-noise suppression with the preservation of fine details.

2 Related Work

Over the years, there are various despeckling techniques have been developed to confront the granular form from SAR images. Lee [6], Kuan [7], Frost [8], enhanced Lee filter [9], enhanced Frost filter [9], and anisotropic diffusion filtering [10, 11] are single filters used to suppress speckle-noise from ultrasonography images. Because these despeckling techniques outcome in a enhancing aspects at the edges where as such mechanisms are not better for despeckling. Furthermore, result shows that not able to preserve the fine details of the images. In order to resolve the issues of single despeckling techniques there are several models have been developed with the integration of the different despeckling techniques.

In [12] has proposed the method to deal with speckle noise such as an enhanced Bayesian-based non-local mean. Such approach has employed through a piece-wise NLM technique. Later, the Pearson distance mechanism in the OBNLM technique has employed to determine the originality between two pixels which helps in despeckling.

In [13] developed the despeckling technique with the considering the multiple factors of speckle-noise organized details rely on the max probability analysis techniques. In results, close to the edge surface the smoothened image obtained with the left-over granular form using NLM-based despeckling techniques.

In [14] the block-matching ensemble despeckling (BM3D) and stochastic distance mechanism based method has been developed. In spite of the fact that the technology is more specialized than information security capability, speckle-noise persists in homogeneous areas of US images. Despeckling methods based on prior knowledge implement every diffusion of the sub-band images for better utilization the computational attributes of the subband images in the transform domain.

In [15] proposed a despeckling method employing the composition of Bayesian minimal squares and Gaussian mixture model. In [16], by taking the advantages of the Gaussian mixture model (GMM) a despeckling method has been developed. Furthermore, the related factors for decomposed images in the transform domain with diffusion have implemented. Later testing the method, the normalized Gaussian distribution [17] was employed to reduce the data's diffusion. In order to deal with speckle-noise, it has proposed the combination of several classified-based computational framework and probability density functions (PDF) like normal inverse Gaussian [18], Laplace distribution [19], and the Cauchy Rayleigh distribution [20].

Similarly some trials have performed by employing in transform domain for despeckling in SAR images. In [21] for digital images a despeckling method has been developed related to wavelet transform. The potential of the methods has praised due to its despeckling ability but still presents the artifacts in non-edge surface. In [22] based on the intra-scale correlation a despeckling mechanism has developed in wavelet domain. In [23] in the transform domain, a speckle-noise suppression model has proposed using logic of thresholding. In [24] given the recommendation for implementing Bayesian least mean square error (MMSE)-based despeckling method in wavelet transform. In [25] SRAD based despeckling method that can be better impact whiling dealing with granular form of the noise. Moreover, without

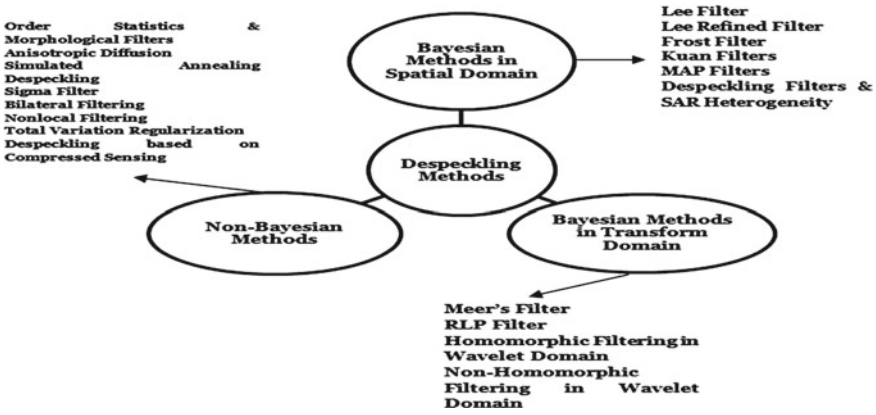


Fig.1 Classifications of the despeckling methods

using log-compression it has greater ability that can be employed directly to US image for despeckling the speckle-noise. Whereas, to remove the additive noise the wavelet transform play a quite significant role [26]. Hence, to transform from multiplicative to additive noise the SRAD homomorphic transformation employs before suppressing the speckle-noise in images. After applying the two-level decomposition the SRAD synthesize the seven high-frequency decomposed images and one low-frequency decomposed image from original image to make possible to preserve the each wavelet components and precisely suppress the granular form of noise using with the help of entropy. Using IDWT and exponential transformation to reconstruct the images and obtained the final despeckled images (Fig. 1).

3 Speckle-Noise Model

Multiple scatterers in an image formed in an image block produce speckle noise and have the characteristics of multiplicative noise. The result of diffuse scattering is that it has a granular appearance, which is difficult to remove. Although the multiplicative parameter is predominant in speckle noise, there is also a small amount of additive noise in SAR systems. The general form of speckle noise [27] in SAR images is given in Eq. (1).

$$I_n = I_D S_{\text{multiplicative}} + S_{\text{additive}} \tag{1}$$

where, I_n = Speckled image. $S_{\text{multiplicative}}$ = Multiplicative noise due to coherent interference. S_{additive} = Additive noise due to sensor noise.

4 Related Theories to Suppress the Speckle Noise

4.1 Speckle Reduction Anisotropic Diffusion (SRAD)

The diffusion framework for ultrasonic and radar imaging uses rely on partial differential equations (PDE) is known as speckle reduction anisotropic diffusion (SRAD).It is employed to suppress granular form, called speckles, without disturbing the significant characteristics of the image. Moreover, it facilitates several pieces of work such as image extraction, image compression, and constant repetitions over the image. The successive interrelationship between all these stages needs consistency after each stage where as every stages applies on the whole image. In addition, it is applies as one of the beginning phase in the heart wall use case. SRAD not only preserve edges but also enhance edges, although this is highly based on character or function edge detection precision. If the edge is not detected, the edge will not be stretched and smoothed either [28]. Similarly, if the noise is detected as edges, the noise will not be smoothed out and even amplified. Therefore the performance of the speckle reduction anisotropic diffusion is sensitive to the selection of the threshold value. It's affecting the spontaneous factor of divergence to suppress speckle-noise which is rely on the relationship among the Lee, Frost filters, and anisotropic diffusion. One of the greater characteristics of the SRAD that's it can suppress the speckle noise without distorting significant fine details of the despeckled images. In addition, it offers superior performance in terms of conventional anisotropic diffusion, advanced Lee filters and smoothing of homogeneous regions, and preserving fine details. In spite of the fact that it achieved great progress but it causes blocks due to the incomplete design of the diffusion coefficient [29].

4.2 Discrete Wavelet Transform

In despeckling process, to transform the images into subbands images of multiple size and direction are called the discrete wavelet transform (DWT). It is employed as a window function for the operations protocols produced by scaling and translation procedures. In DWT, the signal is decomposed into two levels are known as roughened approximation and detail parts. Scaling and wavelet function are two popular operations procedure in discrete wavelet transform. The scaling operation is practiced by the low pass filter (LPF), whereas wavelet function is practiced by the high pass filter (HPF). The backscattered signal in the time domain is decomposed into several frequency bands by constantly passing them through HPF and LPF. After that, the real signal is firstly gone through a half band high pass filter and then through a low pass filter. Further, the signal is sampled down to 2, whereas the every other sample gets ignored. For achieving the depth changes the HPF and LPF get repeated with constitutes one level of decomposition [30] (Fig. 2).

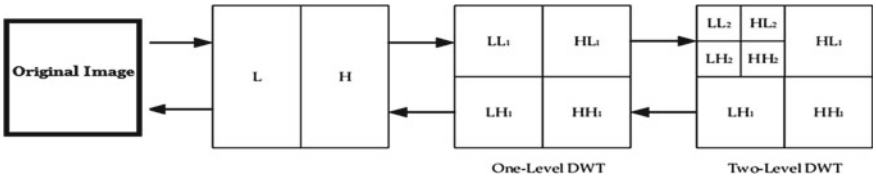


Fig. 2 Image decomposition by using DWT

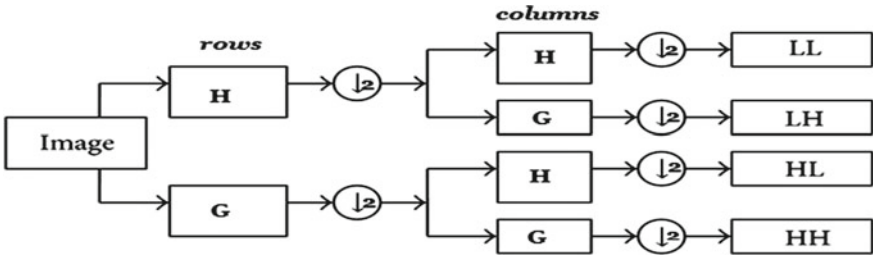


Fig. 3 2D-DWT image decomposition

The LL decomposition is obtained by restricted the HPF in horizontal and vertical angles both and is a closer to original image which is known as approximate part. The remaining is called the expanded or details components. The HL part has obtained by applying HPF in horizontal direction and LPF in vertical direction. The clear detail in a sub-image, e.g., edges has an absolute vertical angle because their adjustment is right angle of HPF known as vertical detail [31]. The details of the rest of the components are same (Fig. 3).

- **LL**: Passing through two simultaneous Low Pass Filters-Gives an Approximation
- **LH**: Passing through LPH and then High Pass Filter-Horizontal features (HPF along rows)
- **HL**: Passing through HPF and then Low Pass Filter-Vertical features (HPF along column)
- **HH**: Passing through two simultaneous High Pass Filters-Diagonal features (HPF distributed equally along both rows and column).

4.3 Method-Noise Thresholding

The method noise plays a very significant role in the process of removing the speckle noise in SAR images. It has capability to enhance the efficiency of granular pattern suppressing techniques and leads to obtain the precise outcomes. This is quite effective and general despeckling model for to enhance the outcomes. The speckle and despeckled images consists the few arbitrary details of the image that represents the differences because of inadequacy of the speckle-noise reduction techniques.

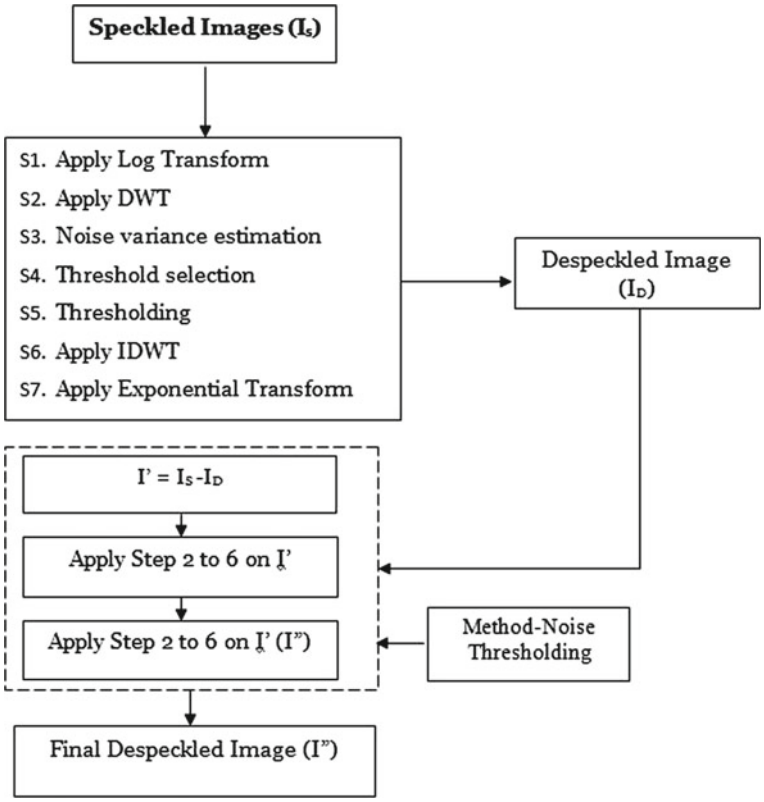


Fig. 4 Wavelet-based despeckling using method-noise thresholding

Figure 4 illustrates the typical model of method noise application based on wavelet transform. It is demonstrated how any filtering technique can be easily incorporated with the concept of method noise. Here, the residual component or the unfiltered component can be filtered using any despeckling method. The method is applied to the unfiltered component of the noise threshold image [32].

4.4 Deep Learning

Synthetic aperture radar (SAR) images can be widely informative due to their resolution and reachability. However, removing the speckle noise from these requires several pre-processing steps. In past few years, deep learning-based techniques have brought significant enhancements in the image processing especially in image restoration. In addition, the researches still has been hindered due to deficiency of organize data availability for training cum testing using deep neural network-based

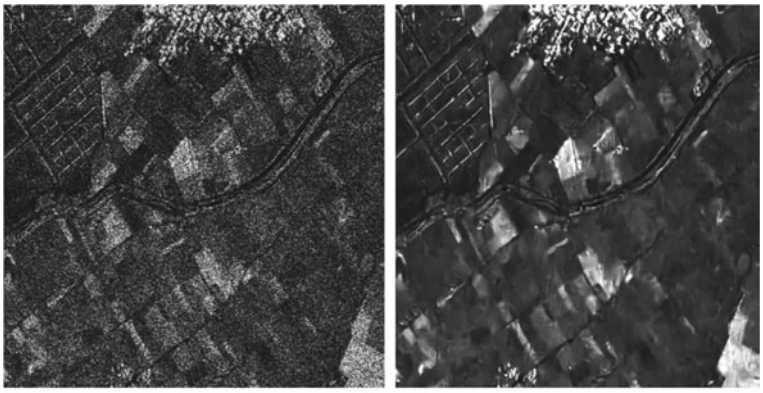


Fig. 5 **a** Noisy image, **b** denoized with supervised CNN

framework [33]. Increased data accessibility with the development of more robust computing devices which are drastic enhancements have been made in machine learning techniques enabling to produce better performances while performing quite complex processes such as image classification, object detection, and semantic segmentation. Specifically, the deep learning techniques have been utilized in satellite images processes due to the potential of deep neural networks to instinctively perceive requisite characteristics from images in a data-driven mechanism without manually setting the components of discrete techniques [34] (Fig. 5).

5 Analyzing Performance of the Despeckling Methods

In SAR image despeckling, the best-thought-out function is to suppress speckle noise while preserving fine details such as edges, and textures. The most common concern in this area is not detectable image factor because the ground reality is unperceived. In addition, the most significant topic is the relationship between the quality and reliability of SAR images. Furthermore, speckle-noise suppression characteristics such as the quality and reliability of the scattered SAR images are calculated by reducing them into identical surfaces that preserve fine detail in contrasting surfaces. An immediate and subjective approach to quality assessment is shown by visual appearance of the despeckled images. Visual appearance allows detection of key human visual features that characterize the behavior of speckle-noise suppression techniques. Similar characteristics include edge protection capability, degree of blurring, point target protection, as well as structural artifacts that are rarely detected by objective and direct measurement. In contrast, visual evaluation does not allow quantitative analysis between the performance of different disclosure filters or to estimate

effectively the bias introduced by the filters. To overcome the pros of visual comparison, several objective performance indices for the quality evaluation of deconvolutional filters have been proposed in the literature. The performance metrics has classified into two parts such as with-reference and without-reference indexes [35].

5.1 With–Reference Indexes

Many literatures are available to calculate the performance of the desperate SAR image in terms of reference indexes. The quantitative metric [36] used under this category uses reference SAR image information.

Mean square error is a performance metrics that estimates the mean error between a despeckled and reference SAR image. It analyzed the complete despeckled SAR image. It restricts to compute minute dedicated information of the SAR image. It is mathematically derived as,

$$\text{MSE} = \frac{1}{N}\sum_{j-0}^{N-1}(X - Y)^2 \tag{2}$$

where x and y = Despeckled and reference SAR images.

The SSIM is helps to compute the equivalence of the despeckled and the reference SAR image. It relies on three elements such as luminance, contrast, and shape. It is derived as,

$$\text{SSIM}\ (a,\ b)\ = \frac{(2\mu_a\mu_b + P_1)(2\sigma_{\text{ab}} + P_2)}{\left(\mu_a^2 + \mu_b^2 + P_1\right)\left(\sigma_a^2 + \sigma_b^2 + P_2\right)} \tag{3}$$

where $\mu_a, \mu_b, \sigma_a, \sigma_b, \sigma_{\text{ab}}$ = local means, standard deviation, and cross variance for images x, y

$$P1 = (0; 01 * L)^2 \quad \text{and} \quad P2 = (0; 03 * L)^2$$

L = specified dynamic scale value.

The SNR is a quantitative parameter employed to compute the discriminations of an imaging system.

$$\text{SNR} = 10 \cdot \log_{10}\left[\frac{V_{\text{ar}[g]}}{\text{MSE}}\right] \tag{4}$$

where $V_{\text{ar}}[g]$ = reference image variance.

PSNR is quite common performance metrics employed to verify the effectiveness of the despeckling methods. The greater value of PSNR represents the better results. It is calculated by:

$$\text{PSNR} = 10\log_{10}\left[\frac{255 * 255}{\text{MSE}}\right] \tag{5}$$

UIQI is derived by using three coefficients such as degree of linear correlation, mean proximity to light, and similarity of contrast in images. The scale of these three factors lies in from zero to one. Therefore, the last limit of UIQI lies 0–1. If the obtained UIQI value of the despeckled images is 1 means obtained the better image quality vice-versa 0 represents the least image quality.

$$Q = \frac{\sigma_{\text{xy}}}{\sigma_x \sigma_y} \cdot \frac{2\text{xy}}{x^2 + y^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \tag{6}$$

5.2 *Without–Reference Indexes*

The performance metric under a no reference index does not depend on the actual SAR data. These metrics depend on the arithmetic SAR data model as well as the core resolutions as well as the level of image asymmetry and symmetry. These matrices focus on the statistics of the arrangement of the pixel values of the actual speckled SAR image and target ratio images, coefficient of variation (CV), equivalent number of look (ENL), clutter ratio (TCR), and noise fraction (NV). Some discussions of the metrics used in the thesis are given below [37].

The ENL is a performance evaluation metric of a depressed SAR image, which analyzes the smoothing factor during subsequent operations of image creation and exercising. This metric relies on the least number of effective looks after despeckling SAR image. It is computed on the non-heterogeneous region and is derived as the ratio between the mean (μ) and squared variance (σ).

$$\text{ENL} = \left[\frac{\mu}{\sigma}\right]^2 \tag{7}$$

The ratio is the image pixels that contain significant information in both homogenous and non-punitive regions, defined as pixels by pixel ratio between the homogeneous SAR ($a(m)$) and the lowly SAR image ($b(m)$).

$$R(m) = \frac{a(m)}{b(m)} \tag{8}$$

The coefficient of the variance is employed to estimates the texture of non-homogeneous areas of the despeckled SAR images. The statistical coefficient of the images described as the ratio of standard and mean values which is representing as a percentage.

$$\text{CV} = \left(\frac{\text{Standard deviation}}{\text{Mean}} \right) \tag{9}$$

The noise variance is employed to represents the amount of speckle-noise i.e. still in the filtered image. The minimum value of NV shows better results of the despeckling method in SAR images. It's not relies on the image intensity. NV can be mathematically expressed as:

$$\sigma^2 = \frac{1}{N} \sum_{j-0}^{N-1} u_j^2 \tag{10}$$

where $N =$ Image size.

6 Conclusions

There are various speckle-noise suppression techniques has been developed by the researcher which is dealing with the speckle-noise but to preserve the fine details such as edge, texture, etc. and maintain the originality of the image is still a challenging task. The majority of the despeckling methods only smoothened the edges, but not improved the edges. The primary objective of this paper that has studied about the various existing methods with the certain parameters such as working model, characteristics, pros, and cons. and the major findings will draw the attention of the researchers in the development of the robust despeckling techniques. Moreover, the other findings are that there are some techniques has greater potential to deal with speckle noise but still needs of improvements, so the integration of more than two speckle-noise suppression techniques can be produced the better results with the greater preservation of fine details of the images. The various performance metrics which are employ to analyze the robustness of the despeckling techniques. In the case of speckle-noise suppression, the with-reference indexes can be employed to determine the efficiency of the any despeckling methods because it contains the information of the previous images whereas the without-reference indexes does not contains. There are some with-reference indexes parameters are quite popular to analyze the efficiency of speckle-noise suppression techniques such as SNR, PSNR, UIQI, and SSIM. Finally, the Bayesian methods gives the better results in transform domain as compare to Bayesian methods spatial domain.

References

1. S. Wang, T.Z. Huang, X.L. Zhao, J.J. Mei, J. Huang, Speckle noise removal in ultrasound images by firstand second-order total variation. Numer. Algorithm **78**, 513–533 (2018)
2. D.M. Alex, A.H. Christinal, D.A. Chandy, A. Singh, M. Pushkaran, Speckle noise suppression in 2D ultrasound kidney images using local pattern based topological derivative. Pattern Recognit. Lett. **131**, 49–55 (2020)
3. P. Singh, R. Shree, A new homomorphic and method noise thresholding based despeckling of SAR image using anisotropic diffusion. J. King Saud Univ. Comput. Information Sci. (2017)
4. P.A. Hatwar, R. Kher, Analysis of speckle noise reduction in synthetic aperture radar images. Int. J. Eng. Res. Technol. **14**(1), 508–512 (2015)
5. R. Shree, A.K. Shukla, R.P. Pandey, V. Shukla, P. Singh, A critical review on despeckling methods in agricultural SAR image. Int. J. Appl. Exerc. Physiol. 1–9 (2019)
6. J.S. Lee, Refined filtering of image noise using local statistics. Comput. Graph. Image Process. **15**, 380–389 (1981)
7. D. Kuan, A. Sawchuk, T. Strand, P. Chavel, Adaptive noise smoothing filter for images with signal-dependent noise. IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-7**, 165–177 (1985)
8. V.S. Frost, J.A. Stiles, K.S. Shanmugan, J.C. Holtzman, A model for radar images and its application to adaptive digital filtering of multiplicative noise. IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-4**, 157–166 (1982)
9. A. Lopes, R. Touzi, E. Nezry, Adaptive speckle filters and scene heterogeneity. IEEE Trans. Geosci. Remote Sens. **28**, 992–1000 (1990)
10. E.K. Abd, A. Youssef, Y. Kadah, Real-Time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion. IEEE Trans. Biomed. Eng. **49**, 997–1014 (2002)
11. C. Tang, L. Wang, H. Yan, Overview of anisotropic filtering methods based on partial differential equations for electronic speckle pattern interferometry. Appl. Opt. **51**, 4916–4926 (2012)
12. P. Coupe, P. Hellier, C. Kervrann, C. Barillot, Nonlocal means-based speckle filtering for ultrasound images. IEEE Trans. Image Process. **18**, 2221–2229 (2009)
13. P.V. Sudeep, P. Palanisamy, J. Rajan, H. Baradaran, L. Saba, A. Gupta, J.S. Suri, Speckle reduction in medical ultrasound images using an unbiased non-local means method. Biomed. Signal Process. Control **28**, 1–8 (2016)
14. C.A.N. Santos, D.L.N. Martins, N.D.A. Mascarenhas, Ultrasound image despeckling using stochastic distance-based BM3D. IEEE Trans. Image Process. **26**, 2632–2643 (2017)
15. J. Portilla, V. Strela, M.J. Wainwright, E.P. Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Trans. Image Process. **12**, 1338–1351 (2003)
16. H. Rabbani, N. Vafadust, P. Abolmaesumi, S. Gazor, Speckle noise reduction of medical ultrasound images in complex wavelet domain using mixture priors. IEEE Trans. Biomed. Eng. **55**, 2152–2160 (2008)
17. H. Rami, L. Belmerhnia, A.E. Maliani, M.E. Hassouni, Texture retrieval using mixtures of generalized Gaussian distribution and Cauchy-Schwarz divergence in wavelet domain. Sign. Process. Image Commun. **42**, 45–58 (2016)
18. A.R. Hassan, M.I. Bhuiyan, An automated method for sleep staging from EEG signals using normal Gaussian parameters and adaptive boosting. Neurocomputing **219**, 76–87 (2017)
19. H. Rabbani, Image denoising in steerable pyramid domain based on a local Laplace prior. Pattern Recognit. **42**, 2181–2193 (2009)
20. P.R. Hill, A.M. Achim, D.R. Bull, M.E. Mualla, Dual-tree complex wavelet coe cient magnitude modeling using the bivariate Cauchy-Rayleigh distribution. Sign. Process. **105**, 464–472 (2014)
21. S. Zada, T. Yassine, M. Kumar, F. Mendoza-Santoyo, A. Nassim, Contribution study of monogenic wavelets transform to reduce speckle noise in digital speckle pattern interferometry. Opt. Eng. **58**(3), 034109 (2019)
22. F. Luisier, T. Blu, M. Unser, A new SURE approach to image denoising intrascale orthonormal wavelet thresholding. IEEE Trans. Image Process. **16**, 593–606 (2007)

23. A. Fathi, A.R. Naghsh, Efficient image denoising method based on a new adaptive wavelet packet thresholding function. IEEE Trans. Image Process. **21**, 3981–3990 (2012)
24. D. Sun, Q. Gao, Y. Lu, Z. Huang, T. Li, A novel image denoising algorithm using linear Bayesian MMSE estimation based on sparse representation. Sign. Process. **100**, 132–145 (2014)
25. H. Xie, L.E. Pierce, F.T. Ulaby, Statistical properties of logarithmically transformed speckle. IEEE Trans. Geosci. Remote Sens. **40**, 721–727 (2002)
26. J. Zhang, G. Lin, L. Wu, Y. Cheng, Speckle filtering of medical ultrasonic images using wavelet and guided filter. Ultrasonics **65**, 177–193 (2016)
27. Y. Hao, X. Feng, J. Xu, Multiplicative noise removal via sparse and redundant representations over learned dictionaries and total variation. Sign. Process. **92**(6), 1536–1549 (2012)
28. P. Singh, R. Shree, A new SAR image despeckling using directional smoothing filter and method noise thresholding. Eng. Sci. Technol. Int. J. **21**, 589–610 (2019)
29. C. Hyunho, J. Jechang, Speckle noise reduction in ultrasound images using SRAD and guided filter, in *Proceedings of the International Workshop on Advanced Image Technology, Chiang Mai, Thailand* (2018), pp. 1–4
30. P. Singh, R. Shree, Importance of DWT in despeckling SAR images and experimentally analyzing the wavelet based thresholding techniques. Int. J. Eng. Sci. Res. Technol. **5**(10) (2016)
31. Q. Gao, Y. Zhao, Y. Lu, Despeckling SAR images using stationary wavelet transform combining with directional filter banks. Appl. Math. Comput. **205**, 517–524 (2008)
32. R. Shree, A.K. Shukla, R. Pandey, V. Shukla, A new fusion-based agricultural synthetic aperture radar image despeckling by using anisotropic diffusion and discrete wavelet transform methods. J. Test. Eval. **49**(4), 2244–2261 (2021)
33. S. Dabhi, K. Soni, U. Patel, P. Sharma, M.K. Parmar, Virtual SAR: a synthetic dataset for deep learning based speckle noise reduction algorithms (2020)
34. L. Zhang, L. Zhang, B. Du, Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosci. Remote Sens. Mag. **4**, 22–40 (2016)
35. P. Singh, M. Diwakar, A. Shankar, R. Shree, M. Kumar, A review on SAR image and its despeckling. Arch. Comput. Methods Eng. 1–21 (2021)
36. G. Andria, F. Attivissimo, A.M.L. Lanzolla, M. Savino, A suitable threshold for speckle reduction in ultrasound images. IEEE Trans. Instrum. Meas. **62**(8), 2270–2279 (2013)
37. P. Singh, R. Shree, Analysis and effects of speckle noise in SAR images, in *IEEE 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA)* (2016), pp. 1–5

Counter Measures to Control and Reduce Fatal Accidents by Improving Driving Capabilities in Aged Adults in India

Syed Musthak Ahmed, P. Ramchandar Rao, Neelima Chakrabarty, and Vinay Kumar Pothula

Abstract The numbers of fatal accidents are increasing at an alarming rate in recent years in India. This could be attributed to many factors including the cognitive, sensory and motor skills of drivers. The causes are not restricted to external factors like road conditions, weather, night driving, inattentive pedestrians, cattle menace, etc., on highways, but also extended to the driver's cognitive skills. Aging causes progressive declination in their physical and cognitive health (Ahmed et al. in Lecture Notes in Electrical Engineering, LNEE. Springer book Series, pp. 1110–1117, 2020). This consequently leads to declination in their driving skills. Immediate attention to address this serious issue and countermeasures to reduce fatal accidents is required for the geriatrics population to maintain their mobility and independent living. The present paper discusses the multiple countermeasures to reduce the frequency of fatal accidents with a special highlight on the problems of elderly drivers as they are the most vulnerable for being involved in traffic accidents (Ahmed et al. in Lecture Notes in Electrical Engineering, LNEE. Springer book Series, pp 291–302, 2020; Ahmed et al. in Lecture Notes in Electrical Engineering, LNEE. Springer book Series, pp. 279–302, 2020; Ahmed et al. in International Conference on Data Science, Machine Learning and Applications (ICDSMLA-2020) at Pune, India. 2020; Ahmed et al. in International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA 2021). Goa, 2021).

Keywords Cognitive · Physical · Sensory · Aged Adults · Countermeasures · Independent living

S. M. Ahmed (✉) · P. R. Rao
S R Engineering College, Warangal, India
e-mail: syedmusthak_gce@rediffmail.com

N. Chakrabarty
Central Road Research Institute, Delhi, India

V. K. Pothula
SR University, Warangal, India

A. Kumar et al. (eds.), *Proceedings of the International Conference on Cognitive and Intelligent Computing*, Cognitive Science and Technology,
https://doi.org/10.1007/978-981-19-2358-6_63

1 Introduction

According to the 'Global Status Report on Road Safety by the WHO (2019), across the globe, road accidents are the major cause of death among people in the age range of 5–29 years. Sadly and notably, India has left behind all other countries since 2008 in the list of road crashes. In 2015, India signed Brasilia Declaration on road safety and commits to reducing the frequency of road accidents by 50% by 2020. But, this preventive commitment wasn't enough as approximately 5 lakh Indian people got involved in road crashes in 2018 that included the killing of 1.5 lakh people and serious injuries to others. These data reports are very alarming and a few serious steps should be taken to prevent further road crashes in India.

Driving is a very complex task that involves various cognitive and sensory-motor abilities, driving skills along with the interaction with various other external factors such as cattle menace, traffic conditions, dips and pits on roads, driving scenario (night time driving), and weather (driving in rainy condition). For this reason, these external factors can't be neglected or ignored.

The present paper mainly focuses on drivers' behavior responsible for accidents and the countermeasures to improve the driving capabilities of ageing adults in India. In order to inquire in detail about drivers' aberrant behavior, the first self-reported instrument known as the driving behavior questionnaire (DBQ) was formulated [1]. It identifies the three-factor structure of aberrant driving-lapses, error and violations. It is one of the widely accepted driving questionnaires to date. It measures both driving style [2] and the relationship of accidents and driving behavior [3]. Original DBQ consists of 50 items and the behavioral classifications of those items were based on the 'GEMS' theory of Reason [4]. DBQ has been globally accepted including major countries like UK [5], USA [6], Qatar and UAE [7], Sweden [8–10], Australia [11–14], China [15], Greece [16], etc. Items and factor structure of DBQ varied in between different countries. Both the number and factor structures of the original DBQ varied in different studies. Various studies uphold the original three-factor structure [1, 9, 13, 16] DBQ, many studies have identified other factors in addition to three and some eliminated one or two factors from the original DBQ. Some studies showed fewer factors [17, 18] and some showed more factors [16, 19]. The driving style varied among sub-groups as per age, gender, and mileage/year [1, 8, 9, 14]. Another study that included four-factor structures became widely used besides the original three- factor structure DBQ [15, 20–22]. However, as per the current literature search, there is a scarcity of driving behavioral related researches done in India that have applied original or modified [23] behavior questionnaire (except a few studies) to quantitatively test the factorial validity and reliability of the three-factor structure. We exactly can't predict the factors responsible for road crashes in the Indian scenario. But, a few factors that can give anticipation and causes can be predicted and the study support fatality is presented in [24–28]. A study on safe driving performance applying biofeedback therapy in drivers of Delhi is discussed in [29]. However, good strategies and training become essential in the Indian scenario [30] where proper

training, skills are to be provided using effective tools to reduce fatality and improve safety among drivers.

2 Methodology

Sample Characteristics

The present study was conducted on 101 participants of both genders (Table 1). Firstly, the subjects were administered essential sensory and cognitive tests (color blindness test, circle drawing test, and trail making test) followed by physical and motor action development (yoga, physiotherapy, and meditation), before placing them on to the car driving simulator (DBQ's). Certain systematic procedures were adopted for the selection criteria like driving experience, age of the participants. Based on data analysis, conclusions were drawn, which have been discussed in successive sections.

Study Design and Procedure

In this driving-simulator based observational study, all the participants were given explicit instructions not to use any kind of drugs and alcohol on and before the experimental day. The driving behaviors were observed by analyzing various factors while driving through pre and post driving assessments of driving performance of the same group of subjects, errors were noted and suggested for correction. During the driving skill test on the simulator, their saccadic eye movements were recorded. The intervention process between the pre and post experimentation involved different yoga techniques and physical exercises to improve the sensory, cognitive and motor deficits of the drivers.

Table 1 Subjects participated in the study

Age group	Gender	
	Male	Female
21–25	4	7
26–30	10	7
31–35	9	4
36–40	12	6
41–45	10	3
46–50	7	1
51–55	8	0
56–60	5	1
61–65	4	1
66–70	2	0
Total	71	30

Cognitive Abilities

The state of mind reflects many facts and that decides the driving behavior of the person the present study, was conducted with a few tests suggested by medical experts that reflect on behavior, correcting measures which intern help in suggesting corrective measures. The few tests are suggested to conduct on participants are listed in the following sections.

(i) **Circle Drawing Test**: CDT test (Fig. 1) was conducted on the subjects to observe the state of mind deficiency while they are on the drive. Figure 2 shows the participation of the subject in the CDT test. Table 3 is the justification table used to examine the subjects of various age groups. The results are tabulated in Table 2.

(ii) **Trail Making Test**: The TMT Test A and B are two cognitive ability tests that give the concentration and motor action abilities of the driver attention to self and surroundings. There are two conclusions drawn. Figure 3 signifies the behavior of a driver by following a particular path and making decisions to move forward, while Fig. 4 signifies the driving performance in the presence of other situations. Table 4 is the performance result of TMT Test A and B evaluated using the justification given in Table 5. While Figs. 5 and 6 shows the participation of the subject in the activity.

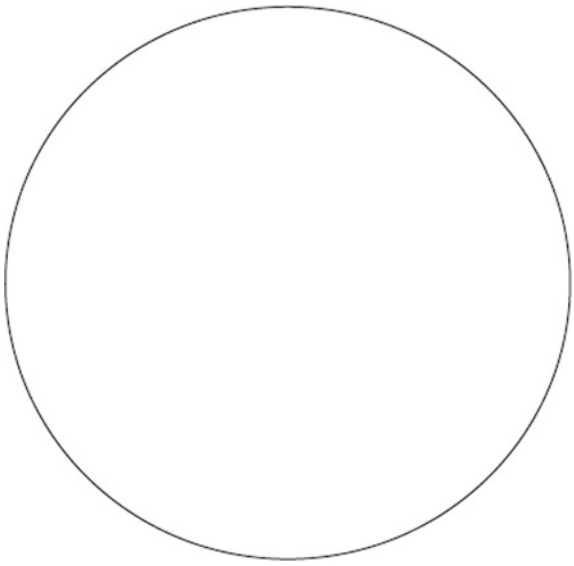


Fig. 1 Circle drawing test

Fig. 2 Participant on CDT test

Table 2 Participants performance on CDT test

CDT—circle drawing test			
Age group	Circle accuracy		
	1	2	3
21–25	6	7	3
26–30	3	5	8
31–35	5	5	3
36–40	3	7	7
41–45	3	7	4
46–50	1	3	4
51–55	1	3	5
56–60	1	3	1
61–65	0	2	1
66–70	0	2	1

Table 3 Description of scoring system of CDT

Description of scoring system of CDT		
Score	Error (s)	Impairment/deficiency
1	No errors in the task	Perfect
2	Mildly impaired spacing while drawing	Minor visuospacial error
3	More impaired spacing deviating in and around the circle	Cognitive deficit

(iii) **Color Blindness Test**: CBT test is very important for any person who performs a driving job while making the drive in traffic and signal light situations. Figure 7 gives the participation of the subject in the CBT Test. While Figs. 8 and 9 are the two visibility tests conducted to find color blindness, i.e., an individual

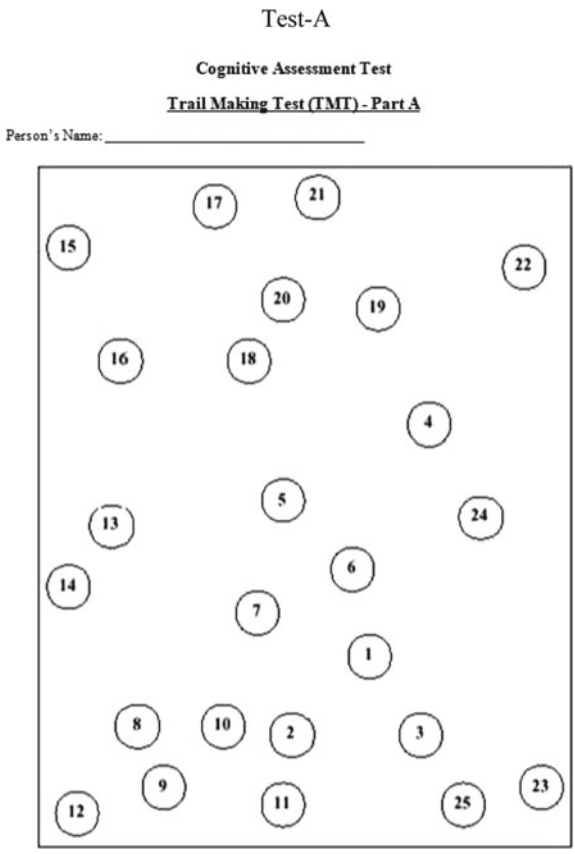


Fig. 3 TMT test-A to find cognitive level

having red/green blindness. Table 6 gives the performance measure of various age groups evaluated by considering Table 7.

Vision Test 1

See Table 6.

Vision Test 2

See Table 7.

Statistical Analysis: The physical and motor actions of a person is judged by a questionnaire listed in Table 9. The type of behavior performed can be categorized under several independent factors IF1–IF6 as listed in the Table. This evaluation helps in finding Errors and violations usually committed by different age group drivers in Indian scenarios.

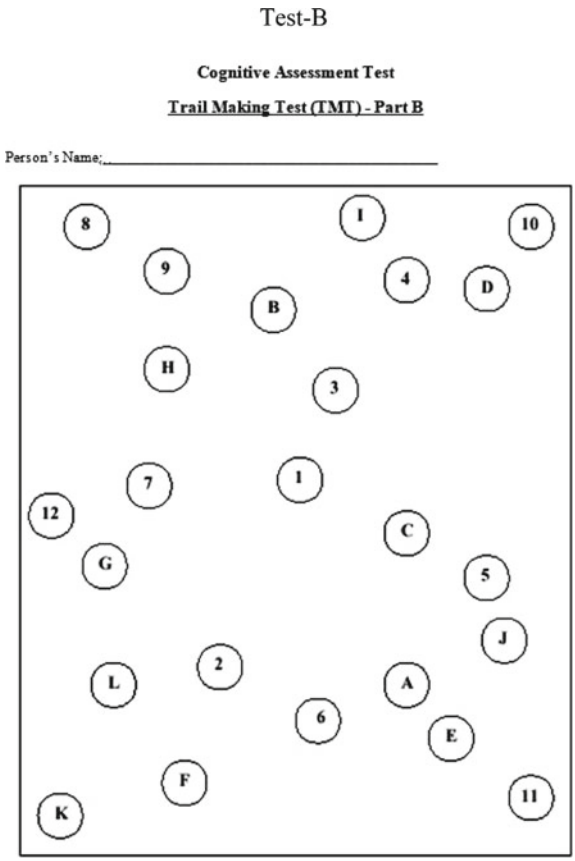


Fig. 4 TMT test-B to find cognitive level

Fig. 5 Participant on TMT test-A

Fig. 6 Participant on TMT test-B

Table 4 Participants performance on TMT A and B

TMT-A&B (Trail making test)				
Age group	Comment			
	Normal		Deficit	
	Trail A	Trail B	Trail A	Trail B
21–25	8	5	0	2
26–30	12	6	0	0
31–35	7	2	0	2
36–40	8	2	0	3
41–45	3	0	0	1
46–50	1	1	1	2
51–55	3	0	0	2
56–60	1	1	0	1
61–65	0	0	1	2
66–70	0	0	1	3

Table 5 Description of scoring on TMT test

Description of TMT test		
Trial	Normal time (s)	Deficiency (s)
Trial A	29	> 78
Trial B	41	> 82

Fig. 7 Participant on CBT test

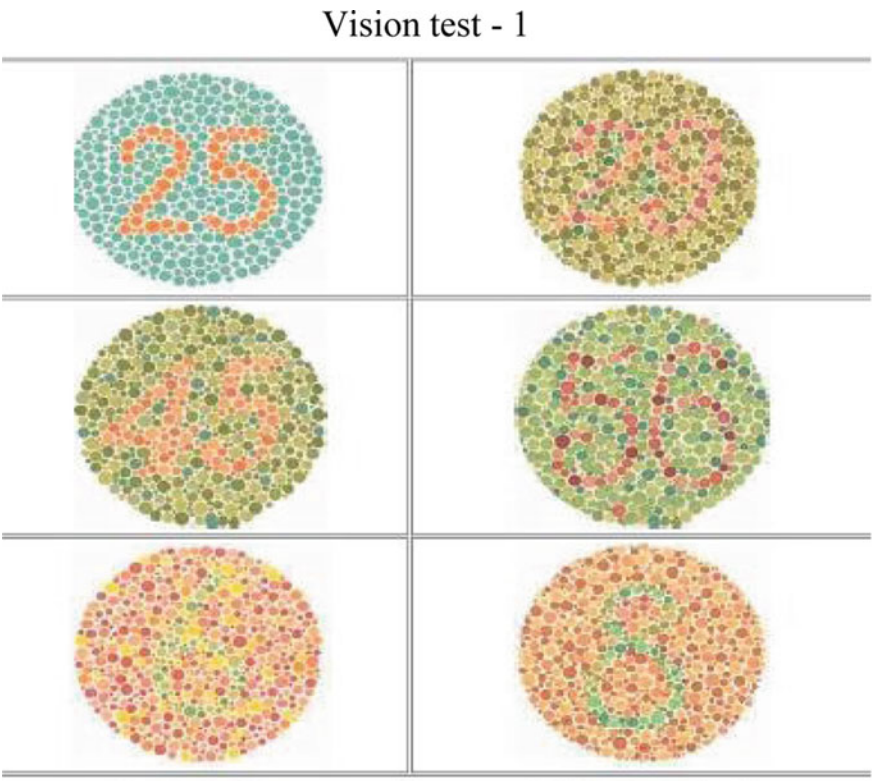
Data Analysis

Every participant was put under study on the simulator. The driving performance was tested under all geographical conditions such as daylight, dark, city driving, busy area driving, rainy, foggy, drizzling, thunder, traffic, lonely, and crossing. The performance of every subject was recorded and the errors and violations committed by them were noticed which was useful for providing them better advice during the intervention. The set-up used and the sample of participation is shown in Fig. 10. And, the performance of all participants for pre and post-simulation is recorded in Table 10.

The mean and standard deviations were calculated for both pre and post simulator study with an intervention to improve the driving conditions, using Statistical Package for Social Sciences software (SPSS). Table 10 is the consolidated analysis carried out on the subjects who participated in the experimentation process.

In the above table, the values of H indicate that after intervention the subject performance has shown maximum improvement, M values were found that there is an improvement but the intervention suggested has to be followed for some more time period which helps to improve driving skills and behavior, while L value marked in table shows no improvement or less improvement. But the factors DBQ3 and DBQ15 fall under dangerous and inattention errors. This is a normal phenomenon where the cognitive performance has slowed down by age. However, the intervention provided has to be made a regular practice in lifetime to overcome the errors committed while on drive. The continuous practice of interventions (yoga, sensory and cognitive tests) creates greater changes in drivers' behavior while driving.

Based on the experiment conducted and the analysis carried out it is inferred that the attitude of a driver and his performance can be modified by following a regular follow-up as suggested and make our country free from accidental disasters.



Normal Color Vision			Red-Green Color Blind		
	Left	Right		Left	Right
Top	25	29	Top	25	Spots
Middle	45	56	Middle	Spots	56
Bottom	6	8	Bottom	Spots	Spots

Fig. 8 CBT vision test-1 to find color blindness

3 Inferences

From the analysis carried out in the Pre Simulation-Intervention-Post simulation, it is concluded that the DBQ's 10–14, 17, 18, 23–27, and 39 has shown maximum improvement by intervention. The DBQ's 2, 4, 5, 9, 19–28, 29–37 shown moderate improvement. DBQ38 has shown no change. Few DBQ's need more periods of intervention to improve their performance. If the suggested exercises, Yoga and other selective measures suggested pertaining to age is followed, the driving can be made safe and secure with a healthy mind and physic thereby can make their driving life happy and safe. The follow-up suggested helps every individual who performs

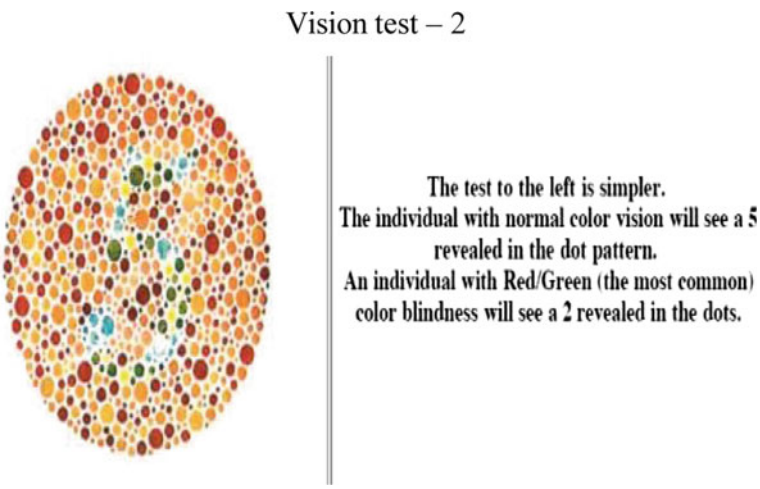


Fig. 9 CBT vision test-2 to find visibility of red/green light

Table 6 Participants performance on CBT test

CBT—color blindness test				
Age group	Vision test 1		Vision test 2	
	Normal color vision	R-G color blind	Normal color vision	R-G color blind
21–25	13		12	1
26–30	16		15	1
31–35	12	1	8	5
36–40	17		13	4
41–45	14		13	1
46–50	8		5	3
51–55	9		5	4
56–60	5		3	2
61–65	3		2	1
66–70	3		1	2
Total	100	1	77	24

Table 7 Description of CBT vision test 1

Normal color vision			Red-green color blind		
	Left	Right		Left	Right
Top	25	29	Top	25	Spots
Middle	45	56	Middle	Spots	56
Bottom	6	8	Bottom	Spots	Spots

Table 8 Description of CBT vision test 2

Normal vision	Color blindness
5	2

Table 9 Factors considered to check the performance

Independent factors	Symbols	DBQ No.	DBQs
Dangerous errors	IF1	3	Releasing the hand brake before putting the vehicle into gear
		4	Forgetting to take the correct exit on the road
		5	Failing to understand a few road signs
		6	Diverging on the wrong road
		7	Forgetting to notice the moment light turns green
		8	Missing the green arrowhead on a stop-light
		36	Driving while being tired
		37	Driving under the influence of sleeplessness from last night
		38	Driving under the influence of medication
		39	Avoiding prescription spectacles while driving
Inattention errors	IF2	1	Adjust the sitting posture before starting the vehicle
		2	Forgetting to wear the seat belt
		9	Not caring about the speed limit (even on an empty road)
		10	Not caring the speed limit in a residential area
		11	Forgetting to change the lane suitable for your turning
		12	Not maintaining a safe distance from the car ahead
		13	Off-roading for tight turns
		14	Ignoring light which is about to turn red
		15	Not trying to find the meaning of a new sign which you just noticed
Inexperience errors	IF3	16	Trying to stay clear and allowing other drivers to pass
		17	Reducing speed to help the overtaking vehicle
		18	Taking the car aside from the road so as not to block other drivers

(continued)

Table 9 (continued)

Independent factors	Symbols	DBQ No.	DBQs
		19	Prevent splashes from the vehicle's tire on others
		20	Use the rightmost lane for overtaking only
		21	Avoiding high beam whenever possible
		22	Not tailing other cars thereby avoiding misleading suspicion
Ordinary violations	IF4	23	Honking to convey your anger
		24	Avoiding honking to prevent inconvenience
Aggressive violations	IF5	25	Wrongly predicting the distance and speed required wrongly while overtaking
		26	Failing to notice a vehicle that is trying to overtake
		27	Miscalculating the speed of an oncoming vehicle while overtaking
		28	Making a right-hand turn and misjudging the approach speed
Positive behavior	IF6	29	Mistaking the presently engaged gear
		30	Trying to shift into the next gear when you are already in that gear
		31	Committing mistakes while shifting gears
		32	Forgetting to shift the gear up and revving the engine unnecessarily
		33	Overtaking a slower vehicle on the left to move fast
		34	Using emergency breaks to avoid collision with a pedestrian
		35	Using emergency brakes to avoid colliding with an animal

driving to make his profession comfortable, positive thinking removes arrogance and overcome mistakes thereby making their driving performance suitable for safe and healthy living.

4 Conclusion

This paper is novel in terms of listing out the errors and violations committed by the drivers on Indian roads. These errors and violations were then mitigated through the intervention techniques like yoga, physical exercises and other selective measures suggested pertaining to age. After the stipulated time of intervention, participants

Fig. 10 Driving simulator setup and testing **a** Driving simulator setup (Front view with screen). **b** Driving simulator setup (Back view with projectors). **c** Female subject participation. **d** Male subject Participation

Table 10 Improvement factors with mean and SD

Factor No.	Q. No.	Pre-simulator performance		Post-simulator performance		Improvement factor
		Mean	S.D	Mean	S.D	
IF1	DBQ3	1.67	0.471	1.33	0.471	L
	DBQ4	1.70	0.459	1.79	0.408	M
	DBQ5	1.77	0.421	1.79	0.408	M
	DBQ6	1.55	0.500	1.71	0.455	L
	DBQ7	1.74	0.439	1.80	0.400	M
	DBQ8	1.74	0.439	1.77	0.421	M
	DBQ36	1.72	0.450	1.76	0.428	M
	DBQ37	1.79	0.408	1.86	0.347	M
	DBQ38	1.73	0.445	1.73	0.445	N
	DBQ39	1.79	0.408	1.91	0.286	H
IF2	DBQ1	1.34	0.475	1.20	0.400	M
	DBQ2	1.17	0.376	1.88	0.325	M

(continued)

Table 10 (continued)

Factor No.	Q. No.	Pre-simulator performance		Post-simulator performance		Improvement factor
		Mean	S.D	Mean	S.D	
	DBQ9	1.73	0.445	1.90	0.300	M
	DBQ10	1.78	0.415	1.39	0.489	H
	DBQ11	1.78	0.415	1.89	0.313	H
	DBQ12	1.74	0.439	1.95	0.218	H
	DBQ13	1.80	0.400	1.93	0.255	H
	DBQ14	1.78	0.415	1.94	0.238	H
	DBQ15	1.57	0.497	1.65	0.478	L
IF3	DBQ 16	1.28	0.450	1.04	0.196	H
	DBQ17	1.14	0.347	1.02	0.140	H
	DBQ18	1.19	0.393	1.09	0.286	H
	DBQ19	1.30	0.459	1.21	0.408	M
	DBQ20	1.38	0.487	1.13	0.337	M
	DBQ21	1.34	0.475	1.11	0.313	M
	DBQ22	1.37	0.484	1.11	0.313	M
IF4	DBQ23	1.61	0.489	1.99	0.100	H
	DBQ24	1.46	0.500	1.04	0.196	H
IF5	DBQ25	1.69	0.464	1.94	0.238	H
	DBQ26	1.83	0.376	1.96	0.196	H
	DBQ27	1.77	0.421	1.97	0.171	H
	DBQ28	1.80	0.400	1.89	0.313	M
IF6	DBQ29	1.64	0.481	1.80	0.400	M
	DBQ30	1.61	0.489	1.80	0.400	M
	DBQ31	1.60	0.492	1.80	0.400	M
	DBQ32	1.78	0.415	1.88	0.325	M
	DBQ33	1.47	0.501	1.11	0.313	M
	DBQ34	1.48	0.502	1.20	0.400	M
	DBQ35	1.33	0.471	1.18	0.385	M

Note H—High, M—Moderate, L—Low, N—No change

showed improvement in the errors and violations committed by them in the post-assessment session. The errors and violations were reflected largely in the older drivers in comparison to the younger drivers. As age progresses, there is a declination in the cognitive, sensory and motor skills which consequently affects the driving abilities of the older populations. We can't control the biological ageing of an individual, but with the help of the proposed training strategies, one can extend the span of driving and exhibit improvement in their abilities to make driving comfort and

confidence with an improved state of mind sod that they drive safely without any fear and live an independent life.

Acknowledgements The authors acknowledge the Department of Science and Technology, Government of India for financial support vide Reference No: DST/CSRI/2017/457 under Cognitive Science Research Initiative (CSRI) to carry out this work. The authors acknowledge the Centre for Creative Cognition, S R Engineering College, Warangal. The authors also acknowledge the Management and Principal of SR Engineering College, Warangal Urban for their continuous support by providing all the necessary facilities.

The authors extend their gratitude and support of Dr. B. Jagadeesh Babu, Psychiatrist, ESI Hospital, Warangal, Dr. Mohammed Abdulla, Ortho Physiotherapist, Certified Mulligan Practitioner, and Dr. Shruthi, Neuro Physiotherapist, V Care Physio Center, Warangal, and Dr. P Harish Chandra Reddy, Associate Professor of ENT, Government Medical College, Nizamabad for their continued support and assistant in the intervention process.

References

1. J. Reason, A. Manstead, S. Stradling, J. Baxter, K. Campbell, Errors and violations on the roads: a real distinction? Ergonomics **33**(10–11), 1315–1332 (1990)
2. A. Bener, T. Lajunen, T. Özkan, D. Haigney, The effect of mobile phone use on driving style and driving skills. Int. J. Crashworthiness **11**(5), 459–466 (2006)
3. J.C.F. De Winter, D. Dodou, The driver behaviour questionnaire as a predictor of accidents: a meta-analysis. J. Saf. Res. **41**, 463–470 (2010)
4. J.T. Reason, *Human errors* (Cambridge University Press, Cambridge, 1990)
5. D. Parker, J.T. Reason, A. Manstead, S.G. Stradling, Driver errors, driving violations, and accident involvement. Ergonomics **38**(5), 1036–1048 (1995)
6. C. Owsley, G. McGwin Jr., S.F. McNeal, Impact of impulsiveness, venturesomeness, and empathy on driving by older adults. J. Saf. Res. **34**, 353–359 (2003)
7. A. Bener, T. Özkan, T. Lajunen, The driver behaviour questionnaire in Arab gulf countries: Qatar and United Emirates. Accid. Anal. Prev. **40**, 1411–1417 (2008)
8. P.-A. Rimmö, L. Hakamies-Blomqvist, Older drivers' aberrant driving behaviour impaired activity, and health as reasons for self-imposed driving limitations. Transp. Res. Part F Traffic Psychol. Behav. **5**(1), 47–62 (2002)
9. L. Åberg, P.-A. Rimmö, Dimensions of aberrant driver behaviour. Ergonomics **41**(1), 39–56 (1998)
10. L. Åberg, H. Warner, Speeding- deliberate violations or involuntary mistake? Eur. Rev. Appl. Psychol. **58**(1), 23–30 (2008)
11. P.N. Blockey, L.R. Hartley, Aberrant driving behavior: errors and violations. Ergonomics **38**(9), 1759–1771 (1995)
12. J. Davey, D. Wishart, J. Freeman, B. Watson, An application of the driver behaviours questionnaire in an Australian organizational fleet setting. Transport. Res. F Traffic Psychol. Behav. **10**(1), 11–21 (2007)
13. A. Dobson, W. Brown, J. Ball, J. Powers, M. McFadden, Women drivers' behavior, sociodemographic characteristics and accidents. Accid. Anal. Prev. **31**(5), 525–535 (1999)
14. R. Lawton, D. Parker, A. Manstead, S.G. Stradling, The role of affect in predicting social behaviours: the case of road traffic violations. J. Appl. Soc. Psychol. **27**, 1258–1276 (1997)
15. C.-Q. Xie, D. Parker, A social psychological approach to driving violations in two Chinese cities. Transp. Part F Traffic Psychol. Behav. **5**(4), 293–308 (2002)
16. T. Kontogiannis, Z. Kossiavelou, N. Marmaras, Self-reports of aberrant behavior on the roads: error and violations in sample of Greek drivers. Accid. Anal. Prev. **34**(3), 381–399 (2002)

17. A. Hennessy, D.L. Wiesenthal, Driving vengeance and willful violations: clustering of problem driving attitudes. J. Appl. Soc. Psychol. **35**(1), 61–79 (2005)
18. N. Sümer, Personality and behavioral predictors of traffic accidents: testing a contextual mediated model. Accid. Anal. Prev. **35**(6), 949–964 (2003)
19. D. Parker, L. McDonald, P. Rabbit, P. Sutcliffe, Elderly drivers and their accidents: the aging driver questionnaire. Accid. Anal. Prev. **32**(6), 751–759 (2000)
20. J. Mesken, T. Lajunen, H. Summala, Interpersonal violations, speeding, violations and their relation to accident involvement in Finland. Ergonomics **45**(7), 469–489 (2002)
21. T. Lajunen, D. Parker, H. Summala, The Manchester driver behaviour questionnaire: a cross-cultural study. Accid. Anal. Prev. **36**(2), 231–238 (2004)
22. P.-A. Rimmö, Aberrant driving behaviour: homogeneity of a four-factor structure in samples differing in age and gender. Ergonomics **4**(8), 569–582 (2002)
23. S. Mehrotra, P. Sudhir, M. Sharma, N. Chakrabarthy, Aggressive driving and anger on roads among two wheeler riding college youth. CRRI, New Delhi (Funding by CSIR) 2013–2014
24. S.M. Ahmed, A. Shireen, B. Jagadeesh Babu, Shruthi, Powered wheelchair for mobility with features to address physical strength, cognitive response, and motor action development issues, Published in *Lecture Notes in Electrical Engineering, LNEE*, vol. 601 (Springer book Series, 2020), pp. 1110–1117
25. S.M. Ahmed, M. Sheshikala, A. Maurya, V.K. Gunjan, Sensory–motor deterioration in older drivers and their amelioration through various training strategies: a study, Published in *Lecture Notes in Electrical Engineering, LNEE*, vol. 643 (Springer book Series, 2020), pp. 291–302
26. S.M. Ahmed, S. Jha, A. Maurya, Analyzing driving skills in adults: a cognitive measure using eye tracker, Published in *Lecture Notes in Electrical Engineering, LNEE*, vol. 643 (Springer book Series, 2020), pp. 279–302
27. S.M. Ahmed, S. Jha, R. Karthik, Human behaviour while driving: a supplementary review, in *International Conference on Data Science, Machine Learning and Applications (ICDSMLA-2020)* at Pune, India, 21th–22nd Nov 2020. (Accepted)
28. S.M. Ahmed, P. Ramchandar Rao, N. Chakrabarty, V.K. Gunjan, Self-assessment report of licensed Indian drivers on their driving performance—a survey, in *International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA 2021)*, Goa, 21st–22nd Aug 2021
29. N. Chakrabarty, K. Sumit, K. Gupta, A study on safe driving performance using biofeedback therapy among drivers in Delhi. INDIAN J. Health Sci. Biomed. Res. KLEU **11**(3) (2018)