Lecture Notes in Electrical Engineering 788

Jawar Singh
Sudhir Kumar
Umakanta Choudhury *Editors*

Innovations in Cyber Physical Systems

Select Proceedings of ICICPS 2020

Springer

Lecture Notes in Electrical Engineering

Volume 788

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

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

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

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

China

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

India, Japan, Rest of Asia

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

Southeast Asia, Australia, New Zealand

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

USA, Canada:

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

All other Countries:

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

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

More information about this series at http://www.springer.com/series/7818

Jawar Singh · Sudhir Kumar ·
Umakanta Choudhury
Editors

Innovations in Cyber Physical Systems

Select Proceedings of ICICPS 2020

Springer

Editors
Jawar Singh
Department of Electrical Engineering
Indian Institute of Technology Patna
Patna, India

Sudhir Kumar
Department of Electrical Engineering
Indian Institute of Technology Patna
Patna, Bihar, India

Umakanta Choudhury
Department of Electrical and Electronics
Rajkumar Goel Institute of Technology
Ghaziabad, India

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-16-4148-0 ISBN 978-981-16-4149-7 (eBook)
https://doi.org/10.1007/978-981-16-4149-7

Preface

This book is a compilation of the accepted papers of International Conference on Innovations in Cyber Physical Systems (ICICPS 2020), held at HMR Institute of Technology and Management, New Delhi, India, during October 22–23, 2020.

The growing integration of cyber (computing and communication) and physical (sensing and actuation) components has paved the way for the creation of sophisticated designed systems. These technologies (also known as Cyber Physical Systems or CPS) have enabled a wide range of applications such as smart power grids, self-driving cars, telehealth and smart cities. Advances in CPS will offer new capabilities, as well as greater adaptability, scalability and usability, considerably exceeding what is currently available in embedded systems. The major goal of the book is to provide latest research findings on fundamental tools, applications, systems, test-beds and field deployments of Cyber Physical Systems.

Industry 4.0 will continue to be an intrinsic element of many technological, managerial and business practices. ICICPS 2020 was organized to stimulate researchers, academicians and students towards it. The papers were selected after careful examination by the Technical Programme Committee after obtaining the views of the reviewers. The technical programme committee, reviewers, general chair, programme chair, data presenter, proof readers and designers, all of them contributed significantly in publishing the research findings in a single document for ICICPS 2020 post proceeding book.

The editors would like to offer their heartful gratitude to all members of the ICICPS 2020 committee for their tremendous effort. Finally, the editors would really like to thank all of the authors for his or her contribution to this book.

Patna, India — Prof. Jawar Singh
Patna, India — Prof. Sudhir Kumar
Ghaziabad, India — Prof. Umakanta Choudhury

Acknowledgements

This book is the result of papers presented at the International Conference on Innovations in Cyber Physical Systems (ICICPS 2020), which took place during October 22–23, 2020, at HMR Institute of Technology and Management, New Delhi, India.

On behalf of the editors of this Conference, we desire to thank the keynote speakers and the reviewers for their time, hard work and dedication to this Conference. Without their services, the editors could not maintain the standards in the field of cyber physical systems.

Our sincere thanks to Publication partner—Springer Nature, Collaborators—International Association of Professional & Fellow Engineering, Delaware, USA, and Indo-UK Confederation of Science, Technology & Research Ltd., London, for acknowledging the significance of this conference to explore the possible areas of interdisciplinary research in the field of Cyber Physical Systems.

We express our sincere gratitude to General Chair Prof. B. K. Panigrahi, IIT Delhi, and Programme Chair Prof. M. P. S. Bhatia, NSUT, Delhi, for their motivation and support in hosting ICICPS 2020. We express our sincere gratitude to the management of HMR Institute of Technology and Management, New Delhi, India, for their kind support and motivation.

We would like to thank Mr. Md. Ehsan Asgar for assisting to give shape to the presented papers in the form of a book. Last but not the least, we are thankful to the authors for their timely submission of the manuscripts.

Corresponding Editor
Dr. Umakanta Choudhury
RKGIT, Ghaziabad, India

Contents

About the Editors

Jawar Singh is currently an Associate Professor in the Department of Electrical Engineering, Indian Institute of Technology, Patna, Bihar, India. He received his PhD degree from the Department of Computer Science and Engineering, University of Bristol, UK. He obtained his MTech degree from the Department of Electrical Engineering, IIT Roorkee, India. His research interests include low-power integrated circuits, embedded systems and RF energy harvesting. Dr. Singh holds two US patents related to computer memories and is the recipient of the prestigious Indo-US fellowship BHAVAN 2016. Dr. Singh has 35 international journal publications, 40 conference publications, two book chapters and one full book from Springer. He was the visiting researcher at the Pennsylvania State University, USA, 2009, and the University of North Texas, USA, 2016, each for 6 months. Dr Singh is a Senior Member IEEE, Associate Editor of IET Electronics Letters, Associate Editor of IETE Technical Review, and IEEE VLSI Circuits and Systems Letters. Dr. Singh has supervised 4 PhDs and 21 MTech dissertations, and 7 PhDs are ongoing.

Sudhir Kumar received his PhD degree from the Department of Electrical Engineering, Indian Institute of Technology Kanpur, India, in 2015. He is currently an Assistant Professor with the Department of Electrical Engineering, Indian Institute of Technology, Patna, India. He has worked as a Scientist with TCS Research, Kolkata, India, and is an Erasmus Mundus Fellow with the Department of Computer Science, University of Oxford, U.K. He published more than 45 research articles in prestigious journals and conference proceedings. His broad research interests include wireless sensor networks and Internet of Things. Dr. Kumar is a recipient of several awards and fellowships. He has guided several BTech. and MTech. projects which won national awards as well.

Umakanta Choudhury has been working as Professor and Head of the Department Electrical & Electronics Engineering (EEE) at Rajkumar Goel Institute of Technology. Before this, Dr. Choudhury worked as a Professor and Head of EEE Department at HMR Institute of Technology, Delhi, India. He is an educationalist as well as an administrator with 22 years of teaching along with industrial experience. He worked at various level in reputed Institutes like Galgotias College of

Engineering, Ansal University, Pusa Institute (Under Govt. of Delhi), etc. He had worked as Director of Delhi Technical Campus, affiliated to GGSIP University, Delhi. He graduated in Electrical Engineering from University College of Engineering (presently (VSSUT), Odisha), and post-graduated from Delhi College of Engineering, Delhi, and got his PhD from Jamia Millia Islamia (Central University), Delhi. He published several papers in national and international journals. His research interest areas are application of artificial intelligence techniques to power systems, power system stability and active filters. He has guided many undergraduate projects and is currently guiding PhD scholars also.

Cyber Physical System of Smart Three-Phase Induction Motor and Its Security Measures

Kundankumar Rameshwar Saraf and P. Malathi

Abstract Three-phase induction motor is widely used in industry due to its ruggedness, reliability, efficiency, and less cost and small maintenance feature. This motor can be affected by various electrical, mechanical, and environmental faults. This fault causes performance degradation, unexpected heating, direction reversal, heavy shock problem in the motor. This research presents the Cyber Physical System of smart three-phase induction motor to detect certain faults and operate the safe switching of motor on occurrence of fault. This system automatically detects the temperature, vibration, humidity, and imaginary input voltage. Various sensors are connected in the vicinity of motor to detect these parameter changes. This system can be controlled by SMS and by accessing the webpage over a remote location through the Iinternet. If any of these parameter goes beyond the predefined limits, it will immediately switch OFF the three-phase induction motor. This system will then send the reason of turn OFF to the user by SMS. When the disturbed parameter achieves its specified acceptable range the three-phase induction motor will be restarted immediately. User can also perform the switching operation of three-phase induction motor using internet over a remote location. This research also discovers the possible security threats on each layer of this system. The security measures to overcome all these threats are presented in detail.

Keywords Application control layer · Aware execution layer · Data transport layer · SMS · Wi-Fi

K. R. Saraf (✉)
Ph.D. Research Scholar, Department of Electronics and Telecommunication Engineering, D.Y. Patil College of Engineering, AkurdiPune, Maharashtra 411 018, India

P. Malathi
Head of Electronics and Telecommunication Engineering Department, D.Y. Patil College of Engineering, AkurdiPune, Maharashtra 411 018, India
e-mail: viceprincipal@dypcoeakurdi.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_1

1 Introduction

Cyber Physical Systems (CPS) monitors the behavior of physical processes and generates the required actuating actions in order to execute the application correctly. CPS consists of physical components such as sensors, actuators, RFID Tags, etc. Aware execution layer contains all these components [1]. These components can communicate with each other through the Internet. Data transport layer allows the communication between these components. To control the behavior of these components, application control layer generates the commands.

1.1 Smart Three-Phase Induction Motor

Three-phase induction motor has high ruggedness, reliability, efficiency, and less cost and maintenance. Due to these features, this motor widely used in industry [2]. During the operation of this motor, various faults can be encountered. It may include electrical faults such as over and under voltage, over loading, single phasing, phase reversing, etc. This motor can be susceptible to mechanical faults such as rotor or stator winding failure; bearing fault, etc. The external fault can occur in motor such as over vibration, external moisture, over or under temperature, drastic humidity change. All these faults may affect the performance of three-phase induction motor [3]. It may cause a fire in the motor. It may also cause the temporary or permanent failure of operation. Hence, to overcome these faults, it is better to design a smart three-phase induction motor that has various sensors to detect these parameters and switch OFF the induction motor on the occurrence of a fault.

This paper presents the Cyber Physical System of the Smart Three Phase Induction motor. This system is controlled by PIC 18 F452 microcontroller. This system can detect multiple faults in the motor.

1.2 Block Diagram of Smart Three Phase Induction Motor

As shown in Fig. 1, the three-phase induction motor is connected with three-phase relays. These three-phase relays are controlled by a PIC microcontroller. This system has two supplies in which a single-phase supply controls the operation PIC microcontroller, GSM modem, and sensors. Three-phase supply is connected to three-phase induction motor. The 5 V constant D.C. power supply is provided to the PIC microcontroller and ESP 8266 Wi-Fi controller. This can be implemented using a single-phase A.C. supply, step down transformer, bridge rectifier, filter capacitor, and LM7805 regulator. Microcontroller is further interfaced with 16 * 2 LCD display, GSM modem SIM 800, temperature sensor LM 35, vibration sensor ADXL 335,

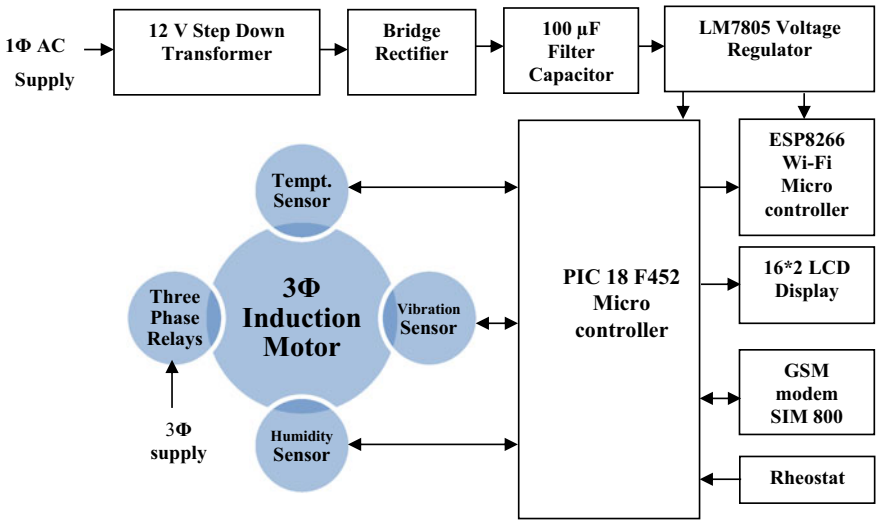


Fig. 1 Block diagram of cyber physical system of smart three-phase induction motor

humidity sensor SY HS 230. ESP 8266 Wi-Fi controller is connected with a PIC microcontroller to control the motor operation through the Internet.

1.3 Operation of System

This system can be operated in two ways: either by SMS or by the Internet. This system has a toggle switch by which it is operated in two modes of operation. Initially, user can select the mode GSM module. To turn OFF the motor user can send SMS 2# to the GSM module. After safe starting of motor, user can change the position of the toggle switch to move into mode 2. In this mode, motor can be started and stopped through the Internet over a remote location as shown in Fig. 2. In this mode, all microcontroller detects physical fitness of motor. For measuring this fitness, temperature sensor LM35 detects the temperature near the motor. Vibration sensor ADXL335 detects the vibrations of the motor. Humidity sensor SY HS 230 detects the humidity near the motor. Rheostat is connected to a PIC microcontroller to change the imaginary voltage of motor. In actual environment, motor voltage can only be changed by changing the voltage of each phase of three-phase supply. This will increase the cost and size of this system. Hence in this system, only virtual voltage change is considered to show over voltage and under voltage fault in the motor.

The top view of the proposed Cyber Physical System is shown in Fig. 3. This figure shows the fault-free condition of this Cyber Physical System. In this condition the temperature, humidity, and imaginary voltage are measured by sensors and shown on LCD display, webpage www.threephaseinductionmotorcps.in.

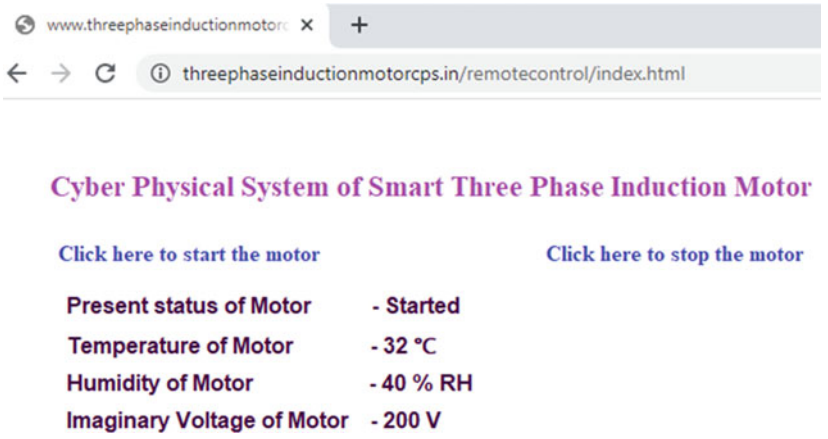


Fig. 2 CPS control through internet

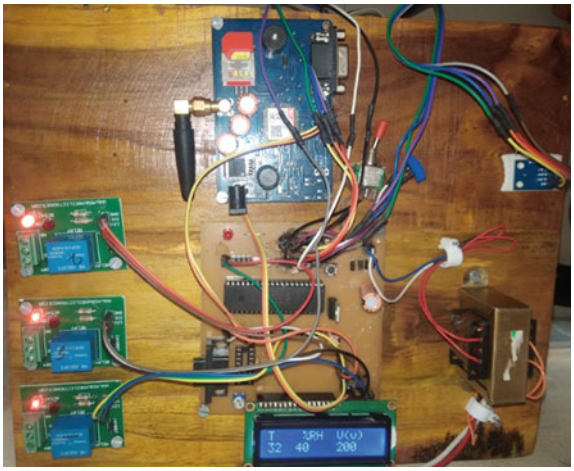


Fig. 3 Normal condition of the system in the second mode

The 3Φ induction motor will automatically turn OFF when temperature, vibration, humidity, and imaginary voltage go beyond the predefined limit. GSM modem SIM800 sends SMS to the user in following cases,

(1) If the temperature goes below 5 °C and above 55 °C the temperature sensor LM35 will detect the under temperature and over-temperature condition and it will send the response to the PIC microcontroller. Figure 4 shows the condition of over temperature.
The SMS received by user in this condition is shown in Fig. 5.

(2) In case of over vibration of motor vibration, sensor ADXL 335 will detect it and send the response to the PIC microcontroller. Figure 6, shows the condition

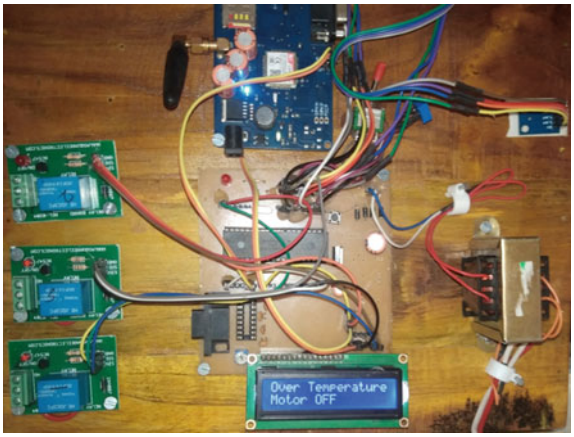


Fig. 4 Condition of motor off due to over temperature

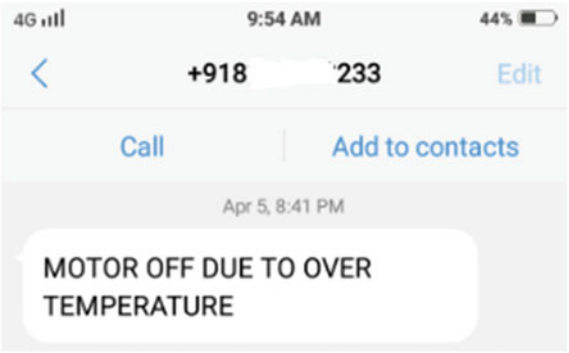


Fig. 5 SMS received by the user in case of over temperature

of over vibration.
The SMS received by the user in over vibration fault is shown in Fig. 7.

(3) In a similar manner the condition of under temperature, under humidity, over humidity, under voltage, and over voltage occurs, if the value of these parameters near the motor goes beyond the acceptable limits.
The user-defined acceptable ranges of input voltage, vibration, temperature, and humidity are shown in Table 1. In each case, motor will send the SMS of turn OFF condition to the user. When the parameter arrives, its optimum value motor will automatically turn ON and send SMS of motor ON condition to the user. The range of various parameters in this system are given in Table 1.

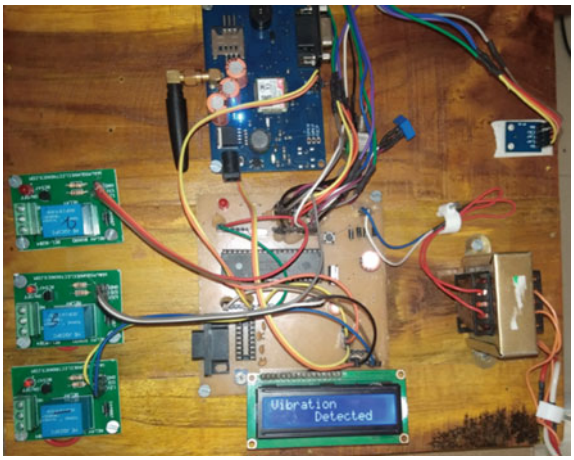


Fig. 6 Condition of Motor OFF due to over vibration

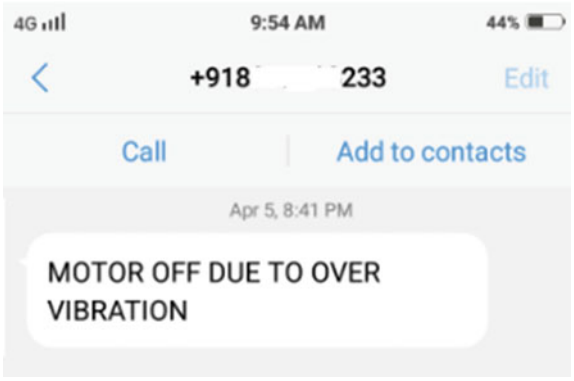


Fig. 7 SMS received by user in case of over vibration

Table 1 Range of various parameters

Sr. no	Name of parameter	Maximum acceptable value	Minimum acceptable value
1	Input voltage	239 V	161 V
2	Vibration	29º angle	–
3	Temperature	5º Celsius	50 °C
4	Humidity	2º Celsius	60 °C

2 Three Tier Architecture of CPS

Cyber Physical System has three tiers of architecture which includes the following layers [4]. Figure 8 shows three-tier architecture of CPS.

Aware execution layer: This layer contains all types of physical equipment including RFID tags and readers, actuators, and sensors. This layer is responsible for access perception to the physical data environment. It also implements control commands of the system. In this system, aware execution layer will contain three sensors such as temperature sensor LM35, vibration sensor ADXL 335, humidity sensor SY HS 230, and one rheostat.

Data transport layer: This layer performs information exchange and real-time communication through the Internet. It uses the Internet of private network, local area network, or communication network. It can also use other data transmission and interaction networks for communication. It has massive information processing and management ability. In this system, instead of the Internet wired connection of sensors with PIC controller is used to transmit the sensing signal to PIC microcontroller.

Application control layer: This layer receives information from the data transport layer. It generates control commands based on preset rules and judgment of control semantics. This layer also provides feedbacks in the form of control commands to the

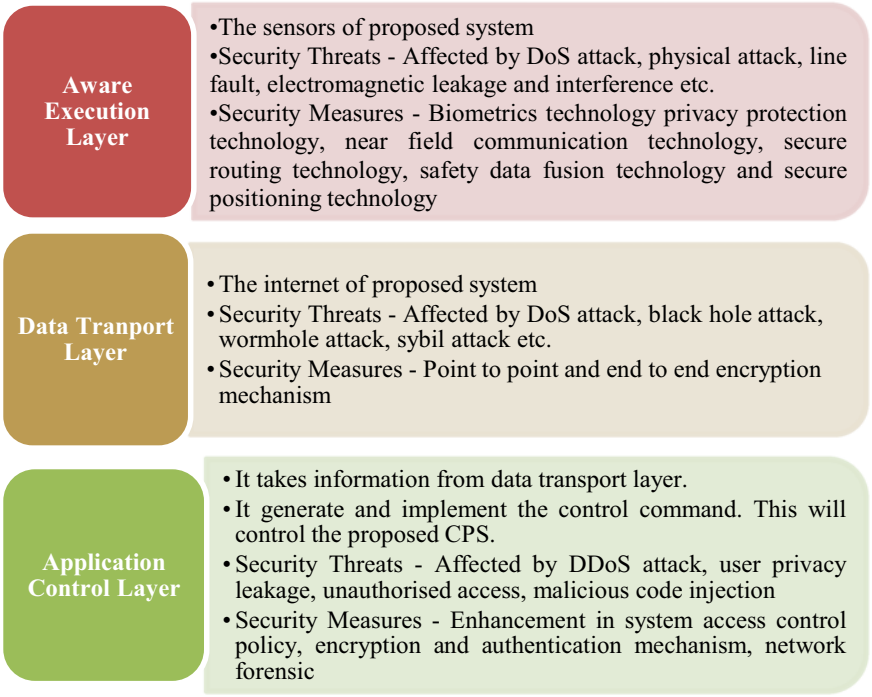


Fig. 8 Security threats and security measures of three-tier architecture of CPS

physical unit through the data transport layer. This layer combines the Cyber Physical System with industrial applications. It is used to generate intelligent solutions such as smart transport, smart grid smart agriculture, etc. In this system, control commands to switch the motor ON and OFF are provided by the PIC microcontroller through three-phase relays connected with three-phase induction motor.

The details of all the attacks mentioned in Fig. 8 are given below,

Physical attacks: In this attack, the attacker causes any physical damage to CPS. For example, if in proposed physical damage to any sensor will stop the sensing of the corresponding physical parameter.

Equipment failure: It results in abnormal behavior of equipment due to external unaware force or unexpected environmental conditions or aging effect. It also partially or fully reduces the performance of equipment.

Line fault: It occurs due to the short circuit connection of conductors.

Electromagnetic Leakage: It creates electromagnetic signal radiation through physical equipment, ground, signal lines, or power lines.

Electromagnetic interference: This attack causes disturbance in communication due to electromagnetic signals.

Denial of service attacks: By this attack network bandwidth is highly consumed by an attacker. Target system flooded by fake requests and it declines to provide its services.

Routing attacks: In this attack, attacker sends fake routing information to forging the routing process. This attack produces wrong routing interference.

Aggregation node attacks: Attack on this core node of data transmission can create a threat to the confidentiality of data.

Direction misleading attacks: Packet received by malicious node changes source and destination address of the packet. Due to these packets delivers to the wrong path which results in network routing confusion.

Blackhole attack: In this attack, the attacker creates a fake channel to receive the routed packet.

Trap doors: In any network, trap door is a secrete way of gaining access by exploiting its vulnerabilities or any loopholes in a network. Attackers can perform the attack using such trap doors.

Sybil attack: In this attack malicious node appears as a multiple node. This will damage the system network.

Wormhole attack: In this attack, multiple malicious nodes are used for data tapping. Data transmission can also be blocked by this attack. Leakage of user privacy: The privacy collector in this layer collects various private data of the user to provide optimum performance. This data leakage will lead to attack on confidentiality.

Unauthorized access: In this attack, attacker gains unauthorized access to the network which leads to private data loss.

Malicious code: Unnecessary code in the system which creates a reverse connection to the malicious server causes loss of user's private data.

Flood attacks or Distributed Denial of Service: In this attack, multiple compromised systems are used to attack the target by flooding it with false requests.

These will create a denial of service for genuine requests. It blocks normal network communication.

The control command forged attack: Attacker generates the false control command in the application control layer. This command will provide complete system access to the attacker and performs the breaching of data.

3 Conclusion

This research creates a CPS of smart three-phase induction motor. This system detects various physical faults and performs the switching of the motor accordingly. This system also protects the CPS against various possible threats. This system is surrounded by various sensors such as temperature sensor to measure the temperature of three-phase induction motor, humidity sensor to measure the humidity near the motor, vibration sensor to detect the over vibration of the motor. Imaginary voltage switch (rheostat) is placed in this system. The maximum and minimum acceptable limits of all the parameters are predefined for this system. This motor can be operated using the Internet or through SMS. Once the limit of any parameter crosses above or below the acceptable value, this system will automatically turn OFF and the corresponding SMS is sent to its user. This SMS will explain the reason of turn OFF the system. Once the disturbed parameter reaches its acceptable limit, the system will automatically turn ON. This motor will automatically perform the switching based on varying environment and physical conditions. Also, this system sends the corresponding SMS to the user using a GSM modem. Users can also control this system through the Internet over a remote location. In the later part of the paper three-tier architecture of Cyber Physical System along with all possible threats to each layer are described. This paper also describes security measures to prevent or overcome all these threats.

4 Future Scope

This research considers only environmental faults and vibration faults in three-phase induction motor. The improved system can be designed to detect and overcome the other electrical and mechanical faults in this motor. It will improve the performance of the motor and reduces the heating and direction reversal problems.

References

1. CPS Steering Group (2008) Cyber-physical systems executive summary. http://precise.seas.upenn.edu/events/iccps11/doc/CPS-Executive-Summary.pdf.
2. Kazemzadeh E, Banadkooki MZ (2015) Smart starting of induction motor by changing core magnetic flux value. In: 2015 30th international power system conference (PSC). IEEE, pp 203–209
3. Shnibha, R. A., and A. S. Albarabar, "Smart technique for induction motors diagnosis by monitoring the power factor using only the measured current", In Journal of Physics: Conference Series, vol. 364, no. 1, p. 012062. IOP Publishing, 2012.
4. Zhang L, Qing W, Bin T (2013) Security threats and measures for the cyber-physical systems. J. China Univ Posts Telecommun 20(2013):25–29
5. Giraldo J, Sarkar E, Cardenas AA, Maniatakos M, Kantarcioglu M (2017) Security and privacy in cyber-physical systems: a survey of surveys. IEEE Design Test 34(4):7–17
6. Fink GA, Edgar TW, Rice TR, MacDonald DG, Crawford CE (2017) Security and privacy in cyber-physical systems. In: Cyber-physical systems. Academic Press, pp 129–141
7. Sakr YS (2015) Security and privacy in cyber-physical systems: physical attacks and countermeasures, PhD diss., UCLA
8. Ahmad I, Zarrar MK, Saeed T, Rehman S (2018) Security aspects of cyber physical systems. In: 2018 1st international conference on computer applications and information security (ICCAIS). IEEE, pp 1–6
9. Chelliah BJ, Ajith AP, Samtani CG, Paul D, Bachhav C (2019) Security implications in cyber physical systems. Int J Innovat Technol Explor Eng (IJITEE). 8(6S). ISSN: 2278–3075
10. Lu T, Zhao J, Zhao L, Li Y, Zhang X (2015) Towards a framework for assuring cyber physical system security. Int J Secur Appl 9(3):25–40

Conventional and Fuzzy Proportional Integral (PI) Speed Controller of Induction Motor Drive

S. Madunisha, P. R. Dharshani, M. J. Archana, R. Saranraj, and P. Parthasarathy

Abstract Electrical drives assume a significant job in quickly developing in the modern time, and speed control of drive is additionally significant. Speed varying drives were a significant piece of intensity framework, microelectronic, power plant, etc. Ordinary controlling strategies for electrical drives required scientific formulae of the process which was exceptionally monotonous, that fathomed by man-made consciousness which consists of hard and delicate calculation. Man-made consciousness has discovered huge application in most practical frameworks (i.e., motor drive). As no scientific framework is required for fake astute framework, productivity and unwavering quality of drives increment, cost and weight of drives decline. Soft starting devices are generally utilized with the electronic engine, gives better working attributes, and improved control. Delicate starters additionally diminishes mileage impacts on the engine and their related drive frameworks which thusly lessen upkeep, preserve vitality, and have a huge influence in improving framework execution so the lifts, siphons, lifts, and transport lines can be worked all the more adequately with a delicate starter.

Keywords PI controller · Induction Motor · Fuzzy sets

1 Introduction

The profitability of assembling industry, advancement in processing plant robotization are required to improve due to scientific technology development. In the assembling process, variable speed drives assume a significant job and possess a significant part as far as transport lines, and fiber handling units. In earlier days, DC motors are used but now AC motors are used as they are less expensive profoundly decreased size, and improved unwavering quality than DC drives. DC drives can be structured according to the application required with the wide assortment of AC-DC

S. Madunisha · P. R. Dharshani · M. J. Archana · R. Saranraj (✉)
K. Ramakrishnan College of Engineering, Samayapuram, Tamilnadu, India

P. Parthasarathy
V.S.B. Engineering College, Karur, Tamilnadu, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_2

converter accessible, anyway, enlistment engine drives have been demonstrated to be in excess of a counterpart for DC drives in superior applications. Typically, these motors are fixed speed drives, yet there are two techniques to change the turn speed of AC motors. The first is to utilize a frequency change converter to change the speed and the second is by utilizing separate windings for various paces. Utilizing separate windings for various speeds isn't a fitting technique as it won't give precise speed control and is likewise exorbitant. For exact control, a frequency change converter is prudent as it can run a three- and single-phase AC drive at a wide range; however, the exhibition of drive ordinarily corrupts as it is utilized external its ideal operational speed.

2 Soft Startıng

Beginning torque assumes a significant job if there should be an occurrence of driving high-inertial burdens. Generally switch on of the thyristor depends on the gate terminal signal but switch off is independent of the gate signal and dependent on circuit conditions. For thyristor drives, a commutation device is required in order to perform turn off action along with turn on gate signal [1]. A few investigations were done to upgrade the beginning charcterisitcs of motor [2–5]. In beginning condition, slip is equivalent to solidarity by considering the equal circuit of motor, supply-side impedance has the most reduced worth, as a result of that a high ebb and flow is drawn by the engine during turning over period, in this manner, it is constantly prescribed to utilize the drivers which limit the beginning ebb and flow [6–8], and when the engine turns over legitimately, the high torque issue were experienced by adding additional weight on the accessories of motor.

To overcome the beginning current and high torque issues, there are different kinds of frameworks accessible. Some of them are:

- Direct on-line (DOL) starter,
- Star-delta starter,
- Frequency change starter, and
- Soft starter.

Up to 8 kW, DOL starter is favored for higher flows and type of start-up decrease is required. In the star-delta course of action, the motor windings are connected in star type so that beginning rise current and torque are controlled. Star-delta recurrence change technique converts AC voltage to DC. Voltage of wanted recurrence is utilized, yet it is confounded and expensive too (Fig. 1).

A delicate beginning framework gives an amazing option requiring little to no effort with a basic structure. It gives step-less engine control for example it permits both the current and torque to be balanced with little augmentations. Shifting of time taken to run the engine up to its ordinary working pace can be done using soft starters. Figure 2 indicates the fundamental square outline of the delicate starter.

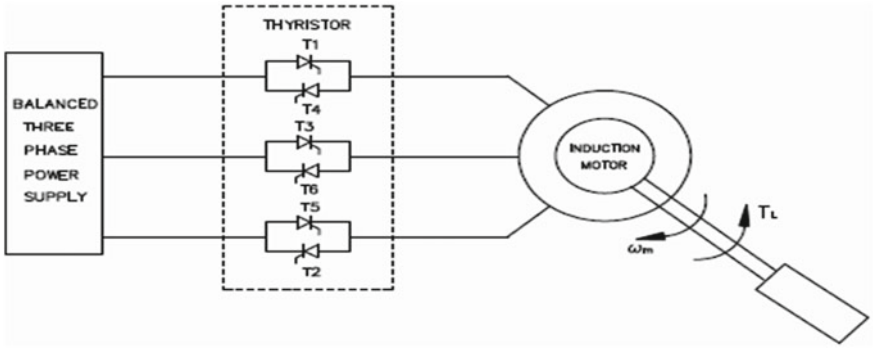


Fig. 1 Basic AC voltage controller (Soft Starter)

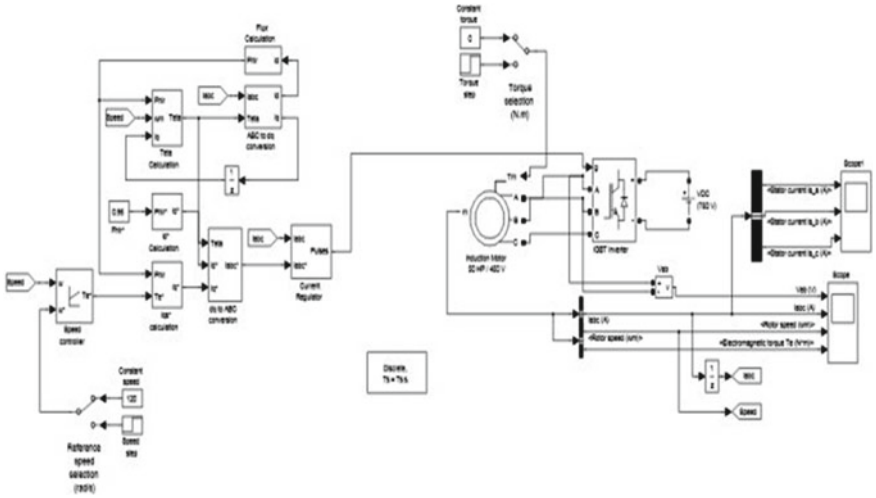


Fig. 2 Simulink model—conventional PI controller of induction motor drive

Delicate starters permit motor to begin, speed can be change and stop with least weights on the hardware, which should be possible by proper modification of the supply voltage. A few strategies for the shut circle control method to accomplish this have been created and introduced in [9]. Right now, a sensor is required to procure the sign criticism. A few specialists have proposed and built up a technique for ideal delicate beginning without sensing the speed for detecting of thyristors voltages is particularly required [10, 11]. Drives with solid-state starter incorporates the advantages of longer engine life, improved control torque; therefore, its area of developments regions are expanding quickly Arduino is an open-supply device with programming layout, client arranges the plans and makes microcontrollers and microcontroller packs for building propelled contraptions. Arduino board plans use a selection of microchips and controllers.

3 Control Principle

The squirrel-cage induction motor has a few focal points like ease, nearly support free, and solid activity; however, for high powerful execution mechanical application, their control is as yet not all that simple due to their non-direct trademark furthermore, the assortment of rotor resistance with working condition. Vector control prompts autonomous control of torque and transition, however, with significant detriment that is touchy to parameter varieties, thus the exhibition of customary controllers like PID or PI falls apart. In this way, these days a definitive objective is to expand the strength of the control framework, with the goal that it will turn out to be progressively resistant to vulnerability. In spite of the fact that customary controls have entrenched hypothetical foundations on solidness and permit diverse structure goals, for example, consistent state and transient qualities of shut circle framework, however, for that privilege numerical model of the system should be known. On the other hand, while using counterfeit control instruments it isn't essential to know the logical model of the system and besides the helplessness or cloud assortments in plant parameters and structure can be dealt with even more suitably. A couple of works added to the arrangement of crossbreed control plans [12].

Scaling factor in an intelligent controller is significant. Determination of appropriate qualities for scaling factors is dependent on the information about the cycle to be controlled and now and then through experimentation to accomplish the most ideal control execution. This is so on the grounds that not at all like regular non-fluffy regulators to date rationale regulator on the grounds that changing the scaling factors changes the standardized universe of talk, the areas, and the participation elements of info/yield factors of fluffy rationale regulator., there is no very much characterized technique for the good setting of scaling factors for fluffy rationale regulators. Yet, the scaling factors are the principle boundaries utilized for tuning the fluffy.

Right now, control techniques (traditional and fake insightful) are acquainted and applied with a backhanded field-arranged acceptance engine. In the first kind of approach, an ordinary PI controller is acquainted all together with accomplished speed control and beginning circumstance is watched. The structure of the PI controller utilized is shown in Fig. 3 which is a normal equal PI controller.

This control methodology with better outcome in light of the fact that fuzzy logic defeats the numerical challenges of demonstrating profoundly non-straight frameworks. It reacts in a progressively steady manner to lose input parameters, its change is exceptionally simple and entirely adaptable.

4 Simulation Results

Right now, contextual analyses have been performed. In both Simulink and Powersim tool compartments of MATLAB programming are utilized. In the main contextual investigation, a 38 KW motor is used.

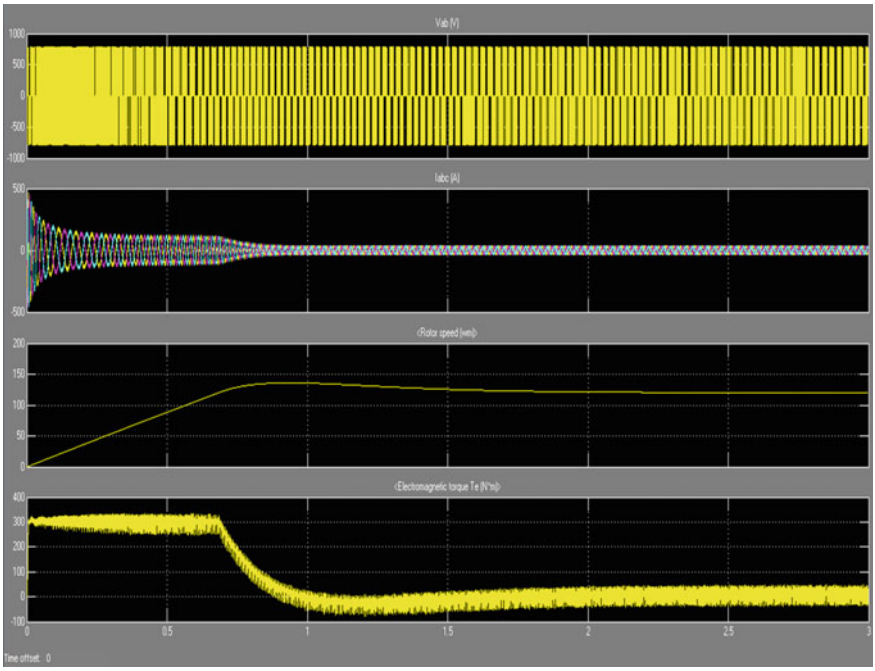


Fig. 3 Output waveforms—conventional PI controller of induction motor drive

Motor is turned over and constrained by a PI controller whose Simulink model is shown in Fig. 2. Three stage voltages and flows are estimated and plotted in the initial 3 s of its activity. Likewise speeding up bend and yield torque is examined. The charts are shown in Fig. 3. Similarly for Fuzzy Logic Based controller whose Simulink model and output is shown in Figs. 4 and 5.

Furthermore time reaction of quickening. For instance, the abundancy of current with an exemplary PI controller is around 480 A during start up while with fuzzy logic controller this worth decreased to 290A.

Fuzzy logic controller makes the framework innate to vulnerability in the event that we watch the speed waveform in Fig. 6. At time t = 1.1 s in Fig. 6 regarding load torque and rotor speed is analyzed in which fuzzy logic based PI controller performs well when contrasted with a conventional controller, along these lines fuzzy-based PI controller gives a powerful control.

Now a day's soft computing techniques are worldwide usage in controlling drive applications in that the more basic thing for the usage of controlling of the drive will be fuzzy logic controller. In that fuzzy logic, the choice based decision will be taken while setting the input based on normal output conditions.

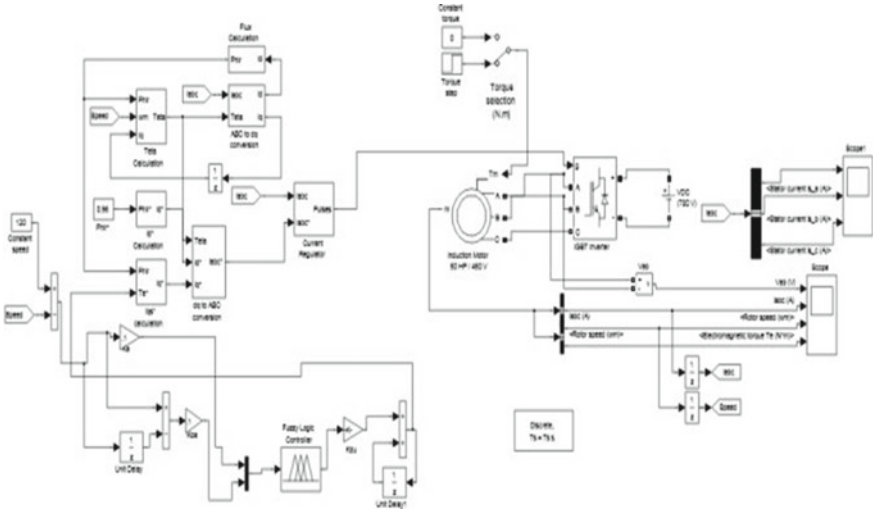


Fig. 4 Simulink model—fuzzy logic PI controller of induction motor drive

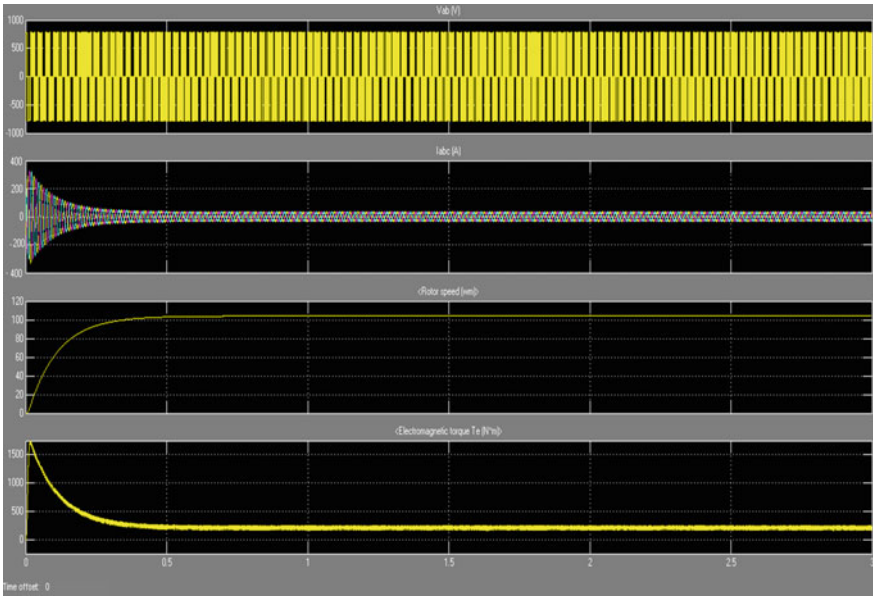


Fig. 5 Output waveforms—fuzzy logic PI controller of induction motor drive

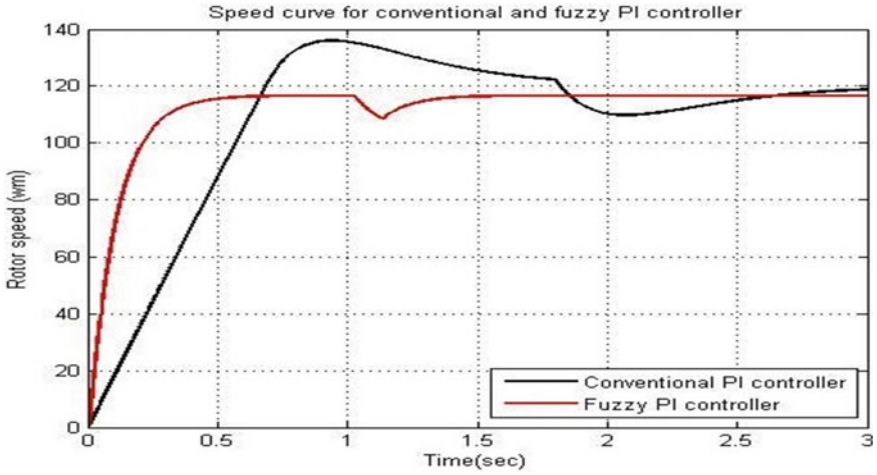


Fig. 6 Rotor speed comparison

5 Conclusion

An enlistment motor devours more vitality than it quite to play out its work not exactly at full burden situation, that overabundance vitality transmitted in this type of warmth. Presenting the man-made reasoning in framework can control plentifulness of beginning current, spare the vitality and new savvy control structure is harsh toward aggravation created outside or inside the framework. Furthermore, fuzzy controller based control methodology shows better outcomes in light of the fact that fluffy rationale beats the scientific challenges of displaying exceptionally non-straight frameworks, the expense and multifaceted nature of the controller is diminished and it reacts in an increasingly steady manner to uncertain readings from input parameters. In the future some improvement methods can be used to streamline the quantity of rule-based fuzzy logic controller, as time devoured in fuzzification and defuzzification are especially relies upon the quantity of rules which thusly impact the reaction of the controller.

References

1. Bose B (2006) Power electronics and motor drives advances and trends, 1st edn. Elsevier, Academic Press
2. Say M (1984) Alternating current machines, 2nd edn. Pitman, England
3. Badr MA, Abdel Halim MA, Alolah AI (1996) A nonconventional method for fast starting of three phase wound-rotor induction motors. IEEE Trans Energy Convers 11:701–707
4. Abdel Halim MA, Badr MA, Alolah AI (1997) Smooth starting of slip ring induction motors. IEEE Trans Energy Convers 12:317–322

5. Hamouda RM, Alolah AI, Badr MA, Abdel Halim MA (1999) A comparative study on the starting methodes of three phase wound-rotor induction
6. Motors part I. IEEE Trans. On Energy Conversion Vol. 14, pp. 918–922.
7. Zenginobuz G, Cadirci I, Ermis M, Barlak C (2001) Soft starting of large induction motors at constant current with minimized starting torque pulsations. IEEE Trans Ind Appl 37:1334–13479
8. Rashad EM, Radwan TS, Rahman MA (2004) Starting and vector control of series-connected wound- rotor induction motor in super synchronous mode. In: 39th IAS annual meeting conference on industry applications, vol 1, pp 32–39
9. Li WX, Lu JG, Liu MS, Zhao J (2004) Design of intelligent soft-start controller for induction motor. In: Proceedings of 3rd international conference on machine learning and cybernetics, vol 2, pp 908–912
10. Denai MA, Attia SA (2002) Fuzzy and neural control of an induction motor. Int J Appl Math Comput Sci 12
11. Cleland JG, Tumer MW, Fuzzy logic control of electric motors and motors drives feasibility study. In: United States air and energy engineering environmental protection research laboratory agency research triangle park
12. Shen JX, Zhu ZQ, Howe D, Buckley JM (2002) Adaptive fuzzy logic speed control of PM brushless AC drives. In: International conference on power electronics machines and drives, pp 68–73
13. Dannenberg A (2005) Fuzzy Logic Motor Control with MSP430x14x. Texas Instruments, Application Report, SLAA235
14. Dhayalini K (2016) Optimization of wind-thermal coordination dispatch using flower pollination algorithm. Int J Adv Chem 12(16):4963–4970
15. Manivasagam R, Saranraj R (2020) Wind turbine manufacturing supply chain. Int J Sci Technol Res 9
16. Manivasagam R, Dharmalingam V (2014) Power quality problem mitigation by unified power quality conditioner: an adaptive hysteresis control technique. Int J Power Electron 6:403–425
17. Manivasagam R, Parthasarathy P, Anbumozhi R (2019) Robust analysis of T-S fuzzy controller for non linear system using H-infinity. Adv Intell Syst Comput 949:643–651
18. Miloud Y, Draou A (2005) Performance analysis of a fuzzy logic based rotor resistance estimator of an indirect vector controlled induction motor drive. Turk J Elce Engine 13

Privacy Protection of Biometric Templates Using Deep Learning

Shefali Arora, M. P. S. Bhatia, Harshita Kukreja, and Soumya Jain

Abstract With the growing use of biometric authentication in real-world applications, there are growing concerns related to the privacy of biometric templates enrolled in the database. Many algorithms have been proposed for the protection of biometric templates, however, this leads to a trade-off between the security of templates and matching performance. In this work, we propose a robust framework for improving upon the existing biometric template protection techniques. The proposed method uses one-shot enrollment for the mapping of images against unique binary codes assigned to them. The binary codes are obtained by binarization of image pixels. This is followed by the use of the SHA-256 hash function, and the class labels are generated based on the hashes templates. This is followed by the extraction of features using a deep convolutional architecture. During verification, the predicted labels are matched against labels generated from the hashed templates of the users stored during enrollment. The proposed framework is capable of achieving state-of-the-art performance.

Keywords Biometric authentication · Template security · Hashing · Deep Learning · CNN · Binary codes · Privacy · Templates

1 Introduction

Biometric recognition involves the use of human traits that are unique to an individual. These traits can be the face, fingerprint, iris, voice, gait, and signature among others. Out of these, the face is the one being used in a large number of applications for recognition. Face recognition can take into account several features like facial measures, facial feature shapes, eye color, etc., to provide better performance. It also provides an advantage to those users who have lost fingerprints or have wrinkled ones. With

S. Arora (✉) · M. P. S. Bhatia
Division of Computer Engineering, Netaji Subhas Institiute of Technology, Delhi, India

H. Kukreja · S. Jain
Deparment of Computer Engineering, Indira Gandhi Delhi Technical University for Women, Delhi, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_3

the advancement in technology, the use of face recognition in providing authentication has increased rapidly due to more reliability than conventional methods like passwords, keys, and patterns. However, face templates demand a more secure database than the conventional methods because a security breach in these templates can reveal the identity of the concerned user. Also, it will turn out to be very costly and time-consuming to reissue another template for the same.

This paper focuses on providing security to face templates. In our proposed approach, we binarize the images and hash the converted strings. The hashed strings are then passed as labels for each user and a Deep Convolutional Neural network is applied using the proposed model/architecture. This is then evaluated on the Labeled Faces in the Wild (LFW) Dataset and the Casia Web Face Dataset. The resulting biometric template security system outperforms the state-of-the-art systems.

The rest of the paper is organized in the following sections. The related works based on the biometric template security are discussed in Sect. 2. The methodology used is presented in Sect. 3 and the results achieved using our framework being comprised in Sect. 4. Further, Sect. 5 concludes the paper.

2 Related Work

Biometric Template Protection approaches can be broadly categorized as biometric cryptosystems, transform-based biometrics, and a hybrid of the two methods. Biometric cryptosystems [1, 2] were initially developed to bind biometric traits with a cryptographic key or generate a key directly. However, with some amount of helper data about the templates in the biometric cryptosystem [3], it can also be used as a template protection system by matching the information to query biometric features and obtain a cryptographic key. Error-correcting codes are further used in biometric cryptosystems to handle the intra-class variations. The cancelable approach, on the other hand, applies a transform function to the biometric template, and only the transformed template is saved in the database [4]. Two possible approaches can be used, salting, and non-invertible transform. Salting is invertible; the original biometric template can be recovered with access to the key and the transformed template. A non-invertible transform is, however, one-way and is computationally hard to invert even if the key is known. A technique using more than one of the basic approaches out of these is called a hybrid scheme. The authors [5] proposed a cryptographic solution of a one-way hash to provide high template security and also addressed the re-enrollment problem in case the templates were compromised. In terms of matching performance, they outperformed the existing state-of-the-art systems by 7%. To overcome the trade-off between matching performance and template security, the authors [6] proposed deep CNN along with random projection to boost the inter-class variations, lower the extracted feature vector dimensionality and reduce the intra-class variations. This is an example of the hybrid approach for template security. In [7], a deep hashing component was used along with a neural network decoder to generate a binary code. A further step in this approach was taken by [8]

which proposed a multimodal authentication system using feature-level fusion to demonstrate finer results. To tackle the similarity attacks, Deep Secure Quantization was developed by [9].

3 Methodology Used

3.1 Convolutional Neural Networks

Convolutional Neural Networks (CNN) are one of the most suitable neural networks for images. Our CNN model employs several components like Kernel, activation function, Pool, Fully connected layers, and softmax function.

A kernel is a filter that is passed over the image to transform it. Subsequent feature map values are calculated consistent with the subsequent formula. Given input image f and kernel h, convolution can be represented as follows(m and n denote the indexes of matrices).

$$G[m,n] = (f*h)[m,n] \sum_j \sum_k h[j,k] f[m-j, n-k] \tag{1}$$

ReLu (The rectified linear unit) has been used as an activation function. It will return 0 in case of any negative input, and will return the positive value back.

$$f(x) = \max(0, x) \tag{2}$$

Pooling layers are used for downsampling feature maps by summarizing the presence of features in patches of that feature map. Two types of pooling layers are max-pooling and average pooling. Maximum pooling calculates the maximum values in each patch of each feature map.

FC Layer is a final connected layer that takes inputs from previous layers to calculate a final probability. The goal of convolution and pooling layers is to extract features from the images and the final layer(s) are used to classify those features. It has a typical equation:

$$\mathrm{f}\left(\mathrm{W}^{\mathrm{T}} \cdot \mathrm{X} + \mathrm{b}\right), \text{ where f is an activation function.} \tag{3}$$

Softmax functions are multiclass sigmoids, so they can be used to determine probabilities of multiple classes. It is defined as

$$\sigma\left(z_j\right) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{4}$$

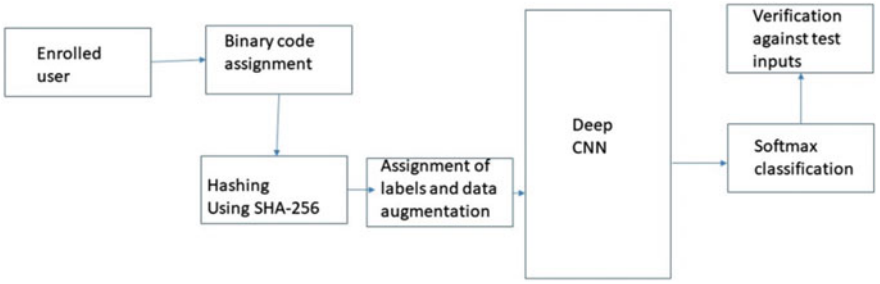


Fig. 1 Privacy-preserving framework

Here z is the input vector, and σ(z) is the output class probability. It is particularly useful in multiclass classification when the input has to be one, and only one class.

The methodology of privacy preservation of templates, followed by biometric authentication is depicted in Fig. 1.

3.2 Assignment of Binary Codes and Hashing Using SHA-256

At the time of user enrollment, unique codes are generated for each user by binarizing the image pixels. These binary codes are not exposed to the user and are further hashed using the SHA-256 algorithm. The use of hashing would ensure the security of the templates, thus preserving the privacy of enrolled users. The main steps involved in the SHA-256 algorithm are:

1. Given a message M to be hashed, the first step is to form a padded message M'. This consists of the message along with padding. The length is a multiple of 512.
2. The padded message M' is parsed into N blocks of size 512 bits. Each block is an input block of 32 bits, and there are 16 blocks in number.
3. Finally, a hash of length 256 bits is initialized (there are 8 blocks in number, each of size 32 bits). The initial hash value is set by taking the first 32 bits of the fractional parts of the square roots of the first eight prime numbers.

3.3 Data Augmentation

Data augmentation is useful in the case of biometric data because of the limited number of training samples and these would be fed to deep architectures to get a good performance. As one-shot enrolment is done, one binary code is assigned for each user. This is followed by data augmentation to increase the number of training

Table 1 CNN architecture used

Name	Type	Number of filters	Filter size	Activation
Conv1,2 Pool1	Conv Max-pooling	16	3×3 2×2	ReLu
Conv3,4 Pool2	Conv Max-pooling	32	3×3 2×2	ReLu
FC1	Fully connected	256	1	NA
FC2	Fully connected	1024	1	NA
FC3	Fully connected	Number of enrolled users	1	Softmax

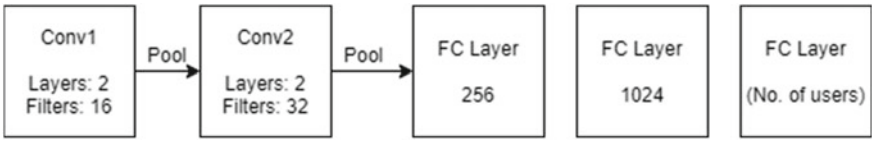


Fig. 2 CNN architecture in the proposed framework

samples. Data augmentation is done for each face image using various operations like flipping, re-scaling, shearing, etc. These would be fed to the deep Convolutional Neural Network for extraction of features.

3.4 Feature Extraction Using CNN

The different layers of the CNN architecture are depicted in Table 1. Figure 2 visualizes the architecture.

3.5 Performance Metrics and Verification

The model is trained and tested using the validation set approach. We reserve 60% of the dataset for training and the remaining 40% for validation. We also repeat the experiments on other splits (80–20 and 20–80, however, results obtained using 60–40 split are the best).

To evaluate the performance of the system, the trained model is tested on input images and the results obtained in this paper are evaluated in terms of False Acceptance Rate (FAR) and Genuine Acceptance Rate (GAR).

4 Results

4.1 Datasets Used

The following datasets have been used to evaluate the performance of the privacy-preserving framework:

(1) Labeled Faces in The Wild: The database contains around 13,000 images of faces collected from the web. The faces have been given labels i.e. the names of those persons. 1680 of the people pictured have two or more distinct photos in the data set.
(2) CASIA Web face dataset: It is annotated with 10,575 unique people with 494,414 images in total. This is the second largest public dataset available for face verification and recognition problems.

4.2 Performance Achieved

The experimental results are shown in Tables 2 and 3. The Genuine Acceptance Rate of the proposed method with one-shot enrolment is evaluated in Figs. 3 and 4. These graphs help to visualize the values achieved using different optimization algorithms like Adam and Stochastic Gradient Descent.

The results achieved can be compared in terms of the performance metric used, with frameworks proposed in the existing studies. Table 4 shows this comparison.

Table 2 Results achieved using the proposed framework and Adam optimization

Database	Enrolment type	GAR (in %) (Adam)	FAR (in %) (Adam)
CASIA–Web face	One-shot	99.56	1.2
LFW	One-shot	99.31	1.5

Table 3 Results achieved using the proposed framework and SGD optimization

Database	Enrolment type	GAR (in %) (SGD)	FAR (in %) (SGD)
CASIA–Web face	One-shot	99.20	1.6
LFW	One-shot	98.97	1.82

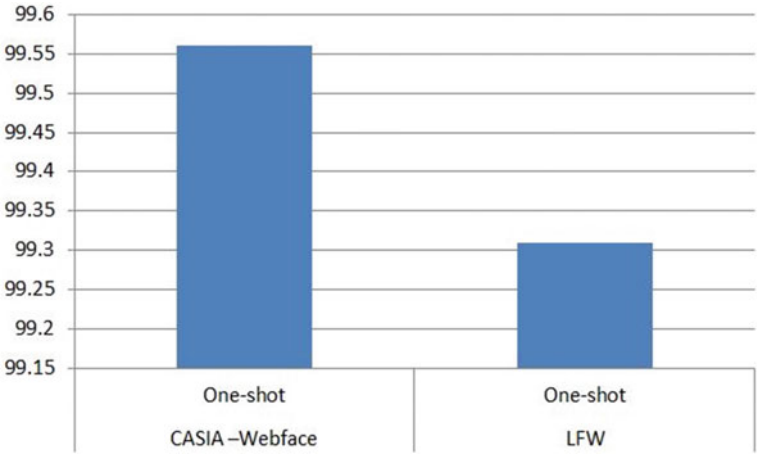


Fig. 3 Results achieved using Adam optimization

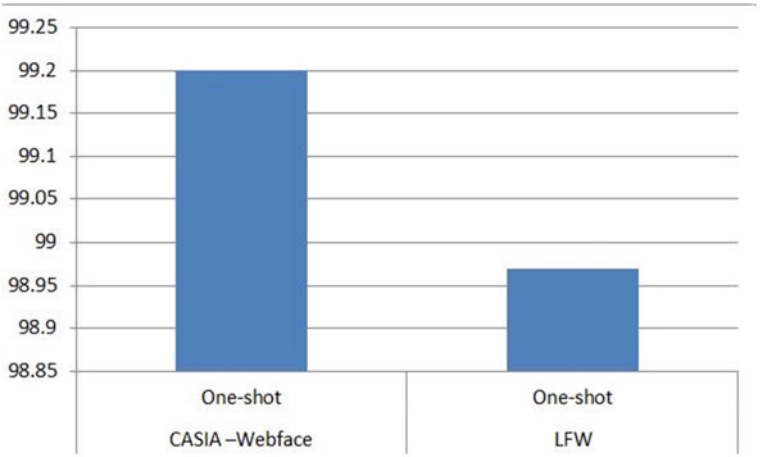


Fig. 4 Results achieved using SGD optimization

Table 4 Comparison with existing studies

Method	GAR
Hybrid method [2]	90.61%@1FAR
MEB [3]	90.13%@1FAR
DCNN [4]	96.53%@0FAR
Deep belief network [10]	94.3%@1.2FAR
PCA [10]	95%@4.3FAR
Hausdorff distance [10]	92.4%@5.2FAR
Adversarial learning without perturbations	99.90%@2.75FAR

5 Conclusion and Future Work

Cancellable biometric template generation using deep learning and cryptography fulfills the requirement of securing user privacy. Securing biometric templates over the cloud is important as a number of applications these days make use of these services, with a lot of user data involved. In this paper, the improvised DeepCrypt framework combines the use of deep convolutional neural networks along with a secure cryptographic hash function to store biometric templates on the cloud server. This technique is useful in obtaining cancellable and secure transforms, without degrading the performance of the authentication system. The results achieved are superior as the genuine accept rates of up to 99 percent are achieved. This ensures the privacy of enrolled individuals.

As a part of future work, we would extend our work on the protection of other biometric modalities. We will extend the provision of template security in other environments capturing biometric traits in uncontrolled environments. We plan to work on more such algorithms using deep networks on traits like gait, dynamics of keystrokes, etc. We will also work on improving the security of existing cryptographic functions for hashing of biometric templates.

References

1. Uludag U, Pankanti S, Prabhakar S, Jain A (2004) Biometric cryptosystems: issues and challenges. Proc IEEE 92(6):948–960. https://doi.org/10.1109/jproc.2004.827372
2. Vetro A, Memon N (2007) Biometric system security. In: Second international conference on biometrics.
3. Cavoukian A, Stoianov A (2007) Biometric encryption. Biometric Technol Today 15(3):11. https://doi.org/10.1016/s0969-4765(07)70084-x
4. Jain A, Nandakumar K, Nagar A (2008) Biometric Template security. EURASIP J Adv Signal Process 2008(1):579416. https://doi.org/10.1155/2008/579416
5. Jami S, Chalamala S, Jindal A (2019) Biometric template protection through adversarial learning. In: 2019 IEEE international conference on consumer electronics (ICCE). https://doi.org/10.1109/icce.2019.8661905
6. Jindal A, Rao Chalamala S, Jami S (2019) Securing face templates using deep convolutional neural network and random projection. In: 2019 IEEE international conference on consumer electronics (ICCE). https://doi.org/10.1109/icce.2019.8662094
7. Talreja V, Soleymani S, Valenti M, Nasrabadi N (2019) Zero-shot deep hashing and neural network based error correction for face template protection. arXiv:1908.02706
8. Talreja V, Soleymani S, Valenti M, Nasrabadi N (2019) Learning to authenticate with deep multibiometric hashing and neural network decoding. In: ICC 2019 - 2019 IEEE international conference on communications (ICC). https://doi.org/10.1109/icc.2019.8761480
9. Chen Y, Wo Y, Xie R, Wu C, Han G (2019) Deep Secure Quantization: On secure biometric hashing against similarity-based attacks. Signal Process 154:314–323. https://doi.org/10.1016/j.sigpro.2018.09.013
10. Liu Y, Ling J, Liu Z, Shen J, Gao C (2017) Finger vein secure biometric template generation based on deep learning. Soft Comput 22(7):2257–2265. Available: https://doi.org/10.1007/s00500-017-2487-9.

11. Jindal A, Chalamala S, Jami S (2018) Face template protection using deep convolutional neural network. In: 2018 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). https://doi.org/10.1109/cvprw.2018.00087
12. Talreja V, Valenti M, Nasrabadi N (2017) Multibiometric secure system based on deep learning. In: 2017 IEEE global conference on signal and information processing (Globalsip). https://doi.org/10.1109/globalsip.2017.8308652
13. Sadhya D, Singh S (2017) Design of a cancelable biometric template protection scheme for fingerprints based on cryptographic hash functions. Multimed Tools Appl 77(12):15113–15137. https://doi.org/10.1007/s11042-017-5095-x
14. Jain A, Nandakumar A, Nagar A (2008) Biometric template security. EURASIP J Adv Signal Process 1(579416). https://doi.org/10.1117/2.1200911.001590

A Comparative Analysis of Clustering and Biclustering Algorithms in Gene Analysis

Diksha Mehta, Shriya Sehgal, Tanupriya Choudhury, and Tanmay Sarkar

Abstract While dealing with high-dimensional biological data, subsets of genes that have similar behaviour under some conditions but behave independently under others, are frequently found. Discovering such co-expressions can aid in uncovering genomic knowledge such as gene networks or gene interactions. A great deal of research is being carried out on the algorithms of these approaches for decades, especially for array mining in gene expression analysis in the field of computational biomedical. In this paper, we compare the results of K-means, traditional clustering algorithm and OPSM (Order-preserving submatrix problem), a biclustering algorithm on a gene expression dataset on the basis of different parameters and examine the benefits of co-clustering over traditional clustering methods in different applications.

Keywords Genomic knowledge · Array mining · Gene expression analysis · OPSM · K-means

1 Introduction

In today's era, associated with increased computational power and the abundance of readily available data, data analysis and its interpretation is a major challenge faced. Data mining is an effective tool to gather relevant insights from raw data. It means finding unknown relationships, like similarities and anomalies, and getting patterns from the humongous batches of data to glean meaningful knowledge. This extracted information can further be utilized in getting a wholesome meaning from large databases particular to an area, where it can guide in decision making.

D. Mehta · S. Sehgal · T. Choudhury (✉)
Department of CSE, University of Petroleum and Energy Studies(UPES), Dehradun, Uttarakhand, India

T. Sarkar
Department of Food Technology and Biochemical Engineering, Faculty of Engineering and Technology, Jadavpur University, Kolkata 700032, India

Government of West Bengal, Malda Polytechnic, Bengal State Council of Technical Education, Daulatpur 732102, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_4

There have been numerous data mining techniques developed and being researched for years now. Of these techniques, Clustering, Biclustering and Triclustering are some of the most prominent ones. The advent of these tools and their comprehensive utilization in a variety of applications like the stock market, medicines, computational biology, image processing, food process technology and many more fields has led to their massive popularity. In this article, we compare the performance of a clustering algorithm, K-means [2] with that of a biclustering algorithm OPSM (Order-preserving submatrix problem) [6], by identifying potentially meaningful relationships among genes and/or experiments in the genome DNA microarray dataset [10].

Section 2 gives a brief theoretical background on the fundamental concept of both the data mining techniques. A detailed description of both the algorithm of interest and the dataset used can be found in Sect. 3. Whereas Sect. 4 expresses the results examined on gene expression datasets for all the aforesaid algorithms, the conclusion of our analysis is hence presented in Sect. 5.

2 Background

2.1 Clustering

Clustering is a method of grouping the data points based on similarity into functional and relevant subclasses called clusters [1]. Objects in one cluster have more analogous characteristics with each other than with those in the other clusters. It is an unsupervised learning technique with no predefined classes and training examples. The quality of the cluster is determined by the selection of the clustering algorithm. Figure 1 summarizes the classification of clustering algorithms.

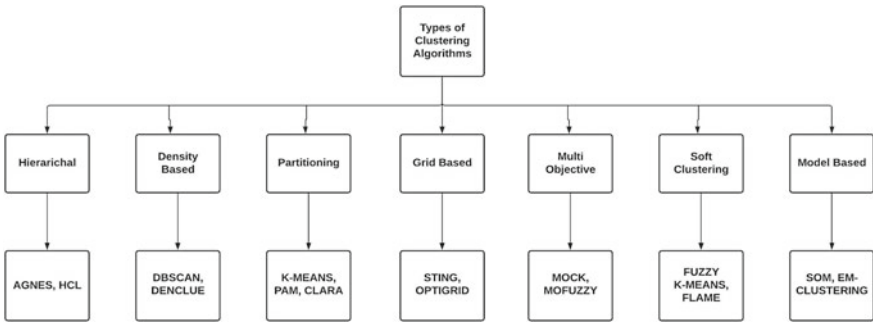


Fig. 1 Classification of clustering algorithm

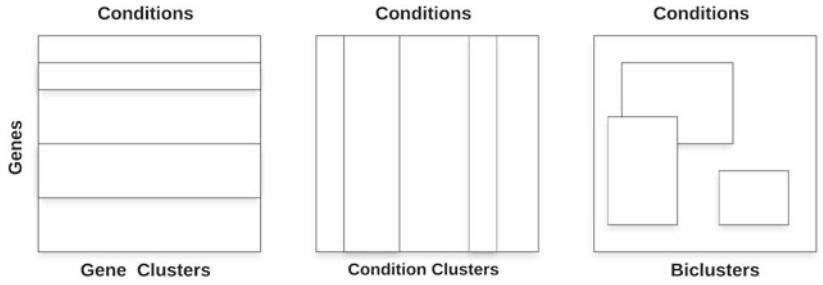


Fig. 2 Basic concepts of the clustering and biclustering technique

2.2 Biclustering

Biclustering is the concurrent clustering of rows and columns in a data matrix to identify data points that are co-regulated across a subset of conditions but not necessarily across other conditions. Figure 2 explains the fundamental distinction between clustering and biclustering techniques.

3 Methodology

3.1 Clustering: K-Means

K-means[13, 14] is an unsupervised, deterministic and iterative technique [2] that divides the dataset into K distinct clusters. It is one of the most traditional and widely used clustering algorithms for its simplicity and speed [3, 4, 15, 16]. The fundamental idea behind this technique is to define K centroids placed at different locations randomly and associate every data point closest to its closest centroid. Once that is done, the new centroids [17–20] are calculated and the data points are re-binded to their closest centre. This step is iterated until the centroids do not change their positions anymore. The main aim of the algorithm is to minimize the squared error function shown in (1) [5]. The flowchart of the algorithm is shown in Fig. 3.

$$J(Y) = \sum_{i=1}^{c} \sum_{j=1}^{ci} \left(\left\| x_i - y_j \right\| \right)^2 \tag{1}$$

where $\|x_i - y_j\|$: Euclidean distance between x_i and y_j.
c_i: The number of data points in the ith cluster,
c: The number of centroids.

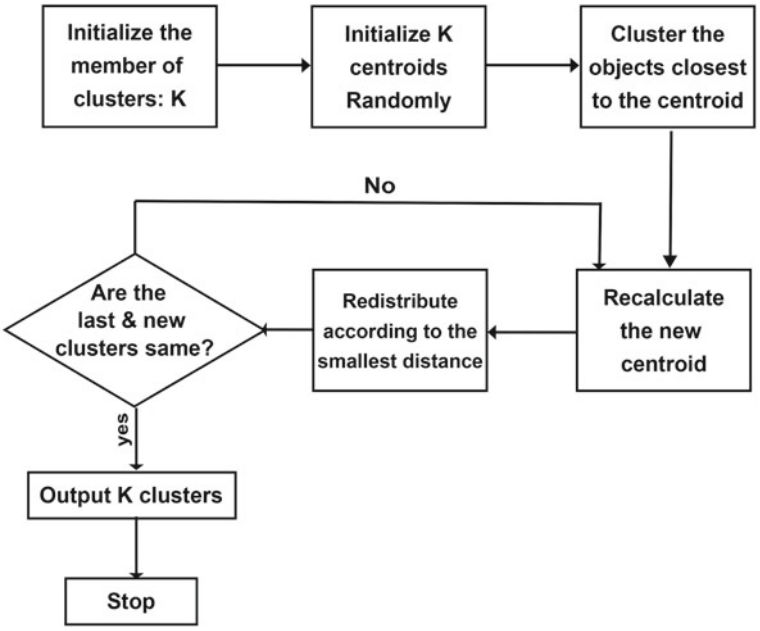


Fig. 3 Flowchart of the K-means algorithm

3.2 Biclustering: Order-Preserving Submatrix Problem

Order-preserving submatrix problem [6], or OPSM, is a greedy algorithm that is used to search for biclusters with ordered rows by defining them as order-preserving submatrices [7]. It constructs them by iteratively making partial biclusters, scoring by probabilities, and growing until a fixed size. It is an NP-complete problem [8].

Definition 1 Taking a matrix A with n x m data, R as rows, and C as a column, $D = (R_D, C_D)$ is a cluster which is a submatrix of A is considered to be an OPSM if there is any permutation of C_D. The fundamental aim is to find all the significant OPSMs in the data matrix A.

All Common Sequences (ACS), which is a variation of Longest Common Sequence (LCS) [9], from every two sequences in the dataset, are determined to find the similarity between them and to form the candidate patterns with random length. The algorithm to find ACS is shown in Algorithm 1. The database is scanned for the calculation of the row support for the patterns whose length is 2 for finding every frequent sequential pattern of length 2. Then a prefix tree is constructed and frequent sequential patterns, with length 2, are stored [6]. The tree is traversed and the node is inserted in the branch based on the A priori principle. The support is calculated again for obtaining frequent sequential patterns with length 3. This algorithm runs iteratively till all OPSMs satisfying the minimum support threshold could

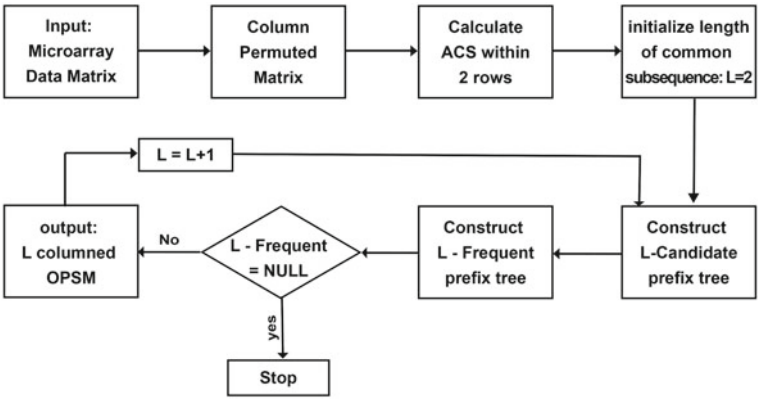


Fig. 4 Flowchart of order-preserving submatrix problem algorithm

be found. In this process, if a larger support threshold is not used to prune, then the results will contain all the deep OPSMs. Figure 4 shows the overview diagram of the algorithm.

Algorithm 1: To find deep OPSMs in the gene expression dataset

Input: Sequences P and Q
Output: A - Set of ACS of P and Q

```
1.  for i = 0 to |P| do
2.      N[i] = 0
3.      IN[i]=-1
4.  end
5.  for i = 1 to |P| do
6.      for j=1 to |Q| do
7.          if Pi=Qj then
8.              IN[i]=j
9.          end
10.     end
11. end
12. for i in |P| do
13.     for j in i do
14.         if IN[i] != -1 & IN[i] > IN[j] then
15.             N[i] = N[i] U (N[j] + Pi)
16.         end
17.     end
18.     A = A U N[i]
19. end
```

Table 1 Tabular representation of microarray dataset

ORF	alpha 0	alpha 7	alpha14	….	diauf	diaug
YBR166C	0.33	– 0.17	0.04	….	0.14	−0.27
YOR357C	−0.64	−0.38	−0.32	….	−0.97	−1.79
YLR292C	−0.23	0.19	−0.36	….	−0.43	−1.06
YGL112C	−0.69	−0.89	−0.74	….	−0.43	−1.51
YIL118W	0.04	0.01	−0.81	….	−0.51	−1.4
….	….	….	….	….	….	….
YLR160C	0.07	−0.04	0.12	….	0.41	−0.1

3.3 Dataset

Data from DNA microarray which is expressed in genome-wide expressions from [10] is used as a dataset. This dataset was formed by running 79 experiments (Alpha factor arrests and release, Elutriation, cdc15 arrest and release, sporulation, heat shock, DTT shock, cold shock, diauxic shift, etc.) on 2467 annotated yeasts cells. These experiments were conducted and recorded at different time points (Table 1).

In Table 1, 2467 rows are representing different yeast cell genes and 79 columns representing 79 elements of gene expression vectors, which contain the expression levels of a particular yeast cell gene in all the conditions in which data is recorded.

4 Result Analysis

To compare the concept of Clustering and Biclustering or Co–Clustering, the above-explained algorithms are used to cluster the microarray data and analysis is done with the help of visual depictions like Heat Map and Graphs. BICAT [11], i.e., The Biclustering Analysis Toolbox is used to perform the comparative analysis of ORDER-PRESERVING SUBMATRIX (OPSM) biclustering algorithm [6] and K-means clustering algorithm [2].

4.1 Data Preprocessing

Preprocessing includes normalization and discretization. Normalization is a technique that is applied to dataset values in order to bring all the values to a common scale while avoiding range distortion. After normalization, discretization is done in order to bring continuous data to discrete values.

Before applying algorithms on the microarray dataset, its preprocessing is done. In Fig. 5, original data is shown in the form of a heat map where red represents up-

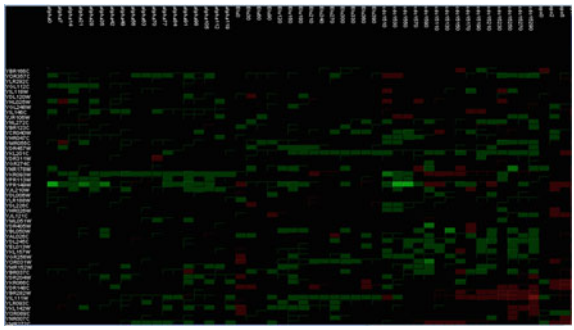


Fig. 5 Original data heat map

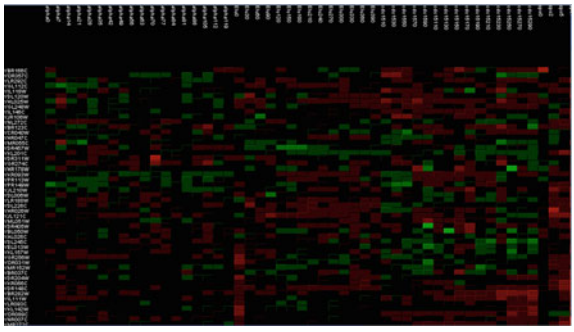


Fig. 6 Normalized data heat map

regulated genes while green represents down-regulated genes and black represents unchanged expression. Then genes are normalized using mean centring (0, 1), its output is shown in Fig. 6.

Finally in order to reduce noise discretized data is formed out of normalized data, i.e. conversion of real values of cells expression vectors into discrete binary values by keeping 10% of 1's using up-regulated patterns discretization scheme. Figure 7 represents its heat map representation.

4.2 K-means Clustering Algorithm

This section provides the results in the form of clusters formed after the implementation of K-means Algorithm [2] using BICAT [9] on the DNA microarray dataset. The random gene method was taken as an initialization method along with Pearson's correlation distance as the distance metric. The number of clusters, i.e. $k = 10$. During the process, any formed empty clusters were dropped.

Figure 8 is a graphical representation of clusters formed by applying K-means on genome data. These graphs represent gene expressions of four different clusters of

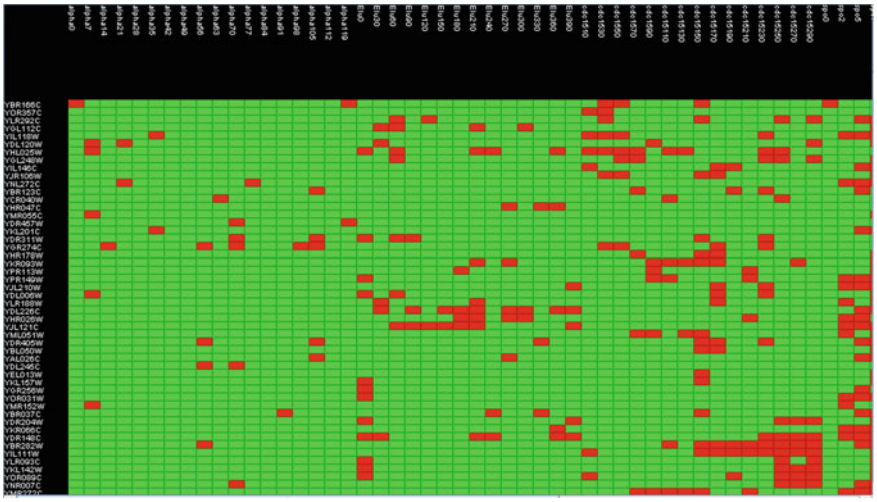
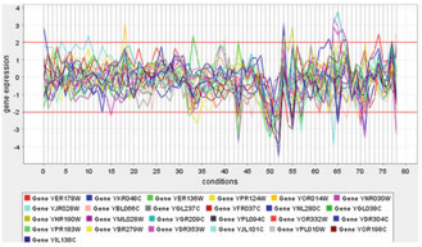
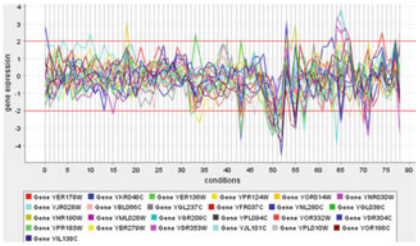


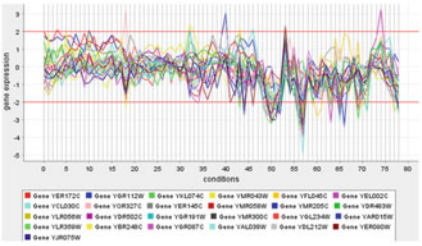
Fig. 7 Discretized data heat map



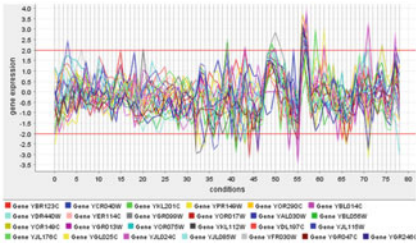
Size: (370, 79)



Size: (282, 79)



Size: (360, 79)



Size: (228, 79)

Fig. 8 Expression profiles of 4 out of 10 clusters formed

genes formed during the clustering process. The graph is plotted by having all the conditions on the x-axis versus gene expressions on the y-axis.

4.3 OPSM Biclustering Algorithm

This section provides information about biclusters formed after the implementation of the Order-Preserving Submatrix Algorithm (OPSM) [6] using BICAT [9] on the DNA microarray dataset [10].

Figure 9 is a graphical representation of biclusters formed by applying OPSM on genome data. These graphs represent gene expressions of four different clusters of genes formed during the biclustering process. The graph is plotted by having all the conditions on the x-axis versus gene expressions on the y-axis. Sizes of biclusters increase with the increase in iterations as it is visible in Fig. 9.

After the implementation of two algorithms, it is observed that K-means is limited to regular shape clusters, i.e., size of all clusters formed are the same, while OPSM is capable of producing irregular shapes. Both algorithms do not handle outliers [12]. The time complexity of OPSM is $O(nm^3l)$, where n is the number of genes and m is the number of samples and l is a number of partial models maintained [8] whereas the time complexity of K-means is O(n) with n as the number of genes. K-means can be categorized as a partitioning clustering algorithm and OPSM falls under the category of the greedy algorithm of biclustering.

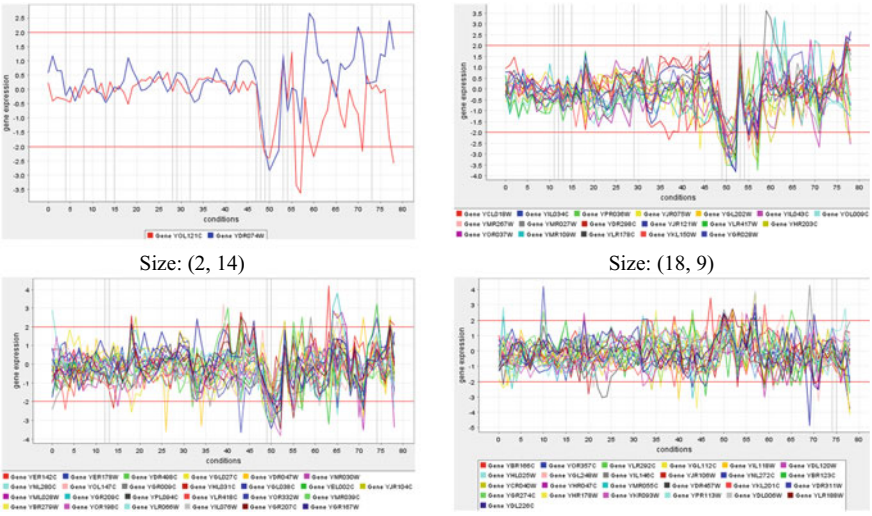


Fig. 9 Expression profiles of 4 out of 10 Biclusters formed

5 Conclusion

We have presented a comparative study of the methods, models, and practicality of a clustering and biclustering algorithm for array mining in gene expression analysis in the field of computational biomedical. K-means, a traditional clustering algorithm, along with OPSM, based on frequent sequential patterns to find OPSMs and deep OPSMs, are presented. Analysis of the result of these algorithms indicates that OPSM obtains more biologically significant clusters as it groups the genes which share expression patterns over different conditions as well and hence, outperforms K-means. However, concerning the higher time complexity in the case of OPSM, K-means is preferred in cases where condition-based clustering is not required due to its simplicity.

References

1. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput. Surv. 31(3):264–323
2. Rohini K, Suseendran G (2016) Aggregated k means clustering and decision tree algorithm for spirometry data. Indian J Sci Technol 9(44):1–6
3. Xu J, Lange K (2019) Power k-means clustering. In: International conference on machine learning, pp 6921–6931
4. Ortega JP, Del M, Rojas RB, Somodevilla MJ (2009) Research issues on k-means algorithm: an experimental trial using matlab. In: CEUR workshop proceedings: semantic web and new technologies, pp 83–96
5. A Tutorial on Clustering Algorithms. http://home.deib.polimi.it/matteucc/Clustering/-tutorial_html/index.html
6. Xue Y, Liao Z, Li M, Luo J, Kuang Q, Hu X, Li T (2015) A new approach for mining order preserving submatrices based on all common subsequences. Comput Math Methods Med
7. Al-Akwaa FM (2012) Analysis of gene expression data using biclustering algorithms. Funct Genomics 51
8. Ben-Dor A, Chor B, Karp R, Yakhini Z (2002) Discovering local structure in gene expression data: the order preserving submatrix problem. In: Proceedings of the sixth annual international conference on Computational biology, pp 49–57
9. Wang H (2007) All common subsequences. In: IJCAI, vol 7, pp 635–640
10. Gene Expression Dataset. http://genome-www.stanford.edu/clustering/Figure2.txt
11. Barkow S, Bleuler S, Preli´c A, Zimmermann P, Zitzler E (2006) Bicat: a biclustering analysis toolbox. Bioinformatics 22(10):1282–1283
12. Bhattacharya A, Chowdhury N, De RK (2012) Comparative analysis of clustering and biclustering algorithms for grouping of genes: Co-function and co-regulation.Current Bioinf 7(1):63–76
13. Joshi A, Sabitha AS, Choudhury T (2017) Crime analysis using K-means clustering. In: 2017 3rd International conference on computational intelligence and networks (CINE), 33–39.
14. Choudhury T, Kumar V, Nigam D (2015) Intelligent classification & clustering of lung & oral cancer through decision tree and genetic algorithm. Int J Adv Res Comput Sci Softw Eng 5(12):501–510
15. Kansal T, Bahuguna S, Singh V, Choudhury T (2018) Customer segmentation using k-means clustering. In: 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS), pp 135–139

16. Gupta P, Sabitha AS, Choudhury T, Bansal A (2018) Terrorist attacks analysis using clustering algorithm. In: Smart innovation, systems and technologies, vol 78. https://doi.org/10.1007/978-981-10-5547-8_33.
17. Kunwar V, Sabitha AS, Choudhury T, Aggarwal A (2019) Chronic kidney disease using fuzzy C-means clustering analysis. Int J Bus Anal 6(3). https://doi.org/10.4018/IJBAN.2019070104
18. Kansal T, Bahuguna S, Singh V, Choudhury T (2018) Customer segmentation using K-means clustering. In: Proceedings of the international conference on computational techniques, electronics and mechanical systems. CTEMS. https://doi.org/10.1109/CTEMS.2018.8769171
19. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of weiler-atherton algorithm to address loose polygons. J Sci Ind Res 78:771–774
20. Kumar V, Choudhury T (2019) Real-time recognition of malignant skin lesions using ensemble modeling. J Sci Indus Res (JSIR) 78:148–153

Fuzzy Logic Method Based MPPT Controller for Solar Energy Generation

Sunny Deol, Shrawan Ram Patel, and Tanupriya Choudhury

Abstract The power demand has been rising gradually because of various industrial developments, population growth. To fulfill the demand of power is a challenging factor for power generation based on fossil fuel alone and the various environmental issues like carbon footprint. To fulfill the demand of power consumption in the world, alternative energy sources can be used. In this paper, the operation of a photovoltaic cell is performed under different conditions of environment. Using fuzzy logic controller, the MPPT controller has been developed based on the performance of PV cell. A fuzzy logic based algorithm is proposed in this paper for tracking optimal power. Modeling and analysis of various subsystems and components has been done in this work. We have tested the models toward validation and connected various models to form an MPPT model that has optimum power.

Keywords MPPT · Fuzzy logic controller · MATLAB/Simulink

1 Introduction

In the present context, looking to the increment in environment pollution and introduction in energy crisis, the renewable energy sources are in high demand because of their characteristics of negligible level of pollution [1]. The Sun is the largest resource of solar energy i.e. easily available over the globe [2]. The increment in energy consumption is observed because of development of industries and human consumption. Due to this, the usage of sustainable and renewable energy sources has been increased exponentially. The researchers are investing technological investments in the field of Energy efficiency optimization. On the other hand, for Power

S. Deol (✉) · S. R. Patel
JIET Jodhpur, Jodhpur, India

S. R. Patel
e-mail: shrawanram.patel@jietjodhpur.ac.in

T. Choudhury
Department of Informatics, School of CS, University of Petroleum and Energy Studies (UPES), Dehradun, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_5

generation purpose, the usage of fossil fuels is decreasing now a day. Depending upon the demand at the load side, the best way to power generation improvement is done by means of combination of more than one energy sources in which the contribution of each load is varied. In the present scenario, there are different methods available to develop a maximum power point tracker in their vast range of applications in terms of cost, speed, efficiency and hardware usability [3].

Photovoltaic solar cell is a hot topic now a days and it will replace non renewable sources like fossil fuels in the upcoming time. To attain this kind of change-over, the power cost per kilowatt-hour for PV must be competitive with the energy sources like fossil fuel. Now, to convert solar irradiance into electricity, the efficiency of PV modules is only 12–26% that is very low in the present scenario [1]. The efficiency of GaAs solar cells has 29% efficiency while the efficiency of Silicon solar cells is only 12–14%. Additionally, Because of decrement in solar insulation, temperature of PV module or load conditions, efficiency can be dropped. For this purpose, there is a requirement of an MPPT controller. The function of MPPT is to perfect tracking of the variation of operating point where delivery of maximum power takes placing to enhance the PV module's efficiency [4].

1.1 Introduction to Fuzzy Logic

Fuzzy Logic deals with imprecision, uncertainty, partial truth, vagueness etc. during the reasoning process. It helps for making decision with incomplete and imprecise information. The theoretical structure of a digital logic can be considered like a basis to formulate the various conceptual definitions without realizing it in real time systems [5]. The value of digital logic is either true or false. The digital logic deal with two states only viz. high and low. In the present scenario, most of the thoughts are of fuzzy in nature rather than digital. So the real world problems cannot be modeled by means of digital logic [6, 15–20] (Fig. 1).

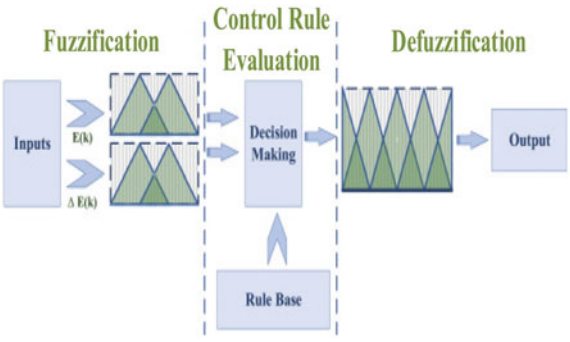


Fig. 1 Fuzzy logic controller block diagram

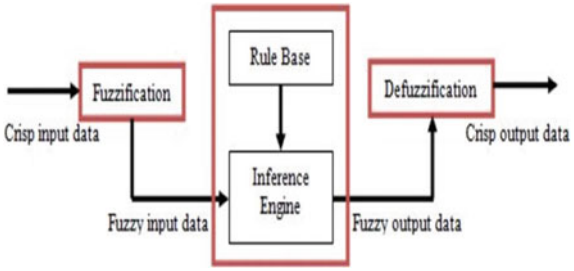


Fig. 2 Fuzzy logic system structure

In a decision making processes or in a specific situation, the nature of real world problems are imprecise and complex [7]. During decision making process, the vagueness problem may arise because of lack of unreliable information. Prof. L. A Zadeh introduced the concept of Fuzzy logic to resolve these situations.

To resolve vagueness issue, real life situations based model can be developed by means of fuzzy logic [8]. In real world, while making any decision, these types of models are intolerant of hesitations and ambiguities. Thus, to form a more realistic model, a more generalized approach is used termed as IFS (Intuitionistic Fuzzy Set).

Fuzzy logics are related with reasoning algorithms belong to the Artificial Intelligence. These logics are used for imitation of human thought patterns as well as abilities of decision making in machine learning [9].

Fuzzy logics perform the various tasks by rectifying the rules that are the combination of inputs by experts and the desired outputs [10] (Fig. 2).

1.2 Fuzzification

The process of conversion of a crisp input date into a fuzzy data by means of knowledge base information is called fuzzification [11]. Various membership functions viz. triangular, gaussian, trapezoidal are used in fuzzification. First of all, the decomposition of a specific linguistic variable takes place in the form of composite linguistic terms. In an air conditioner, temperature is used to define room's temperature that can be further defined in terms of general linguistic like "cold" or "hot". Therefore, the temperature (T) is served have five fuzzy sets in terms of linguistic variable temperature as follows: TOO COLD, COLD, WARM, HOT, and TOO HOT.

The linguistic terms are the individual features of this decomposition that represents a part of total values of temperature. The membership functions are used to convert non-fuzzy input into fuzzy value [21, 22].

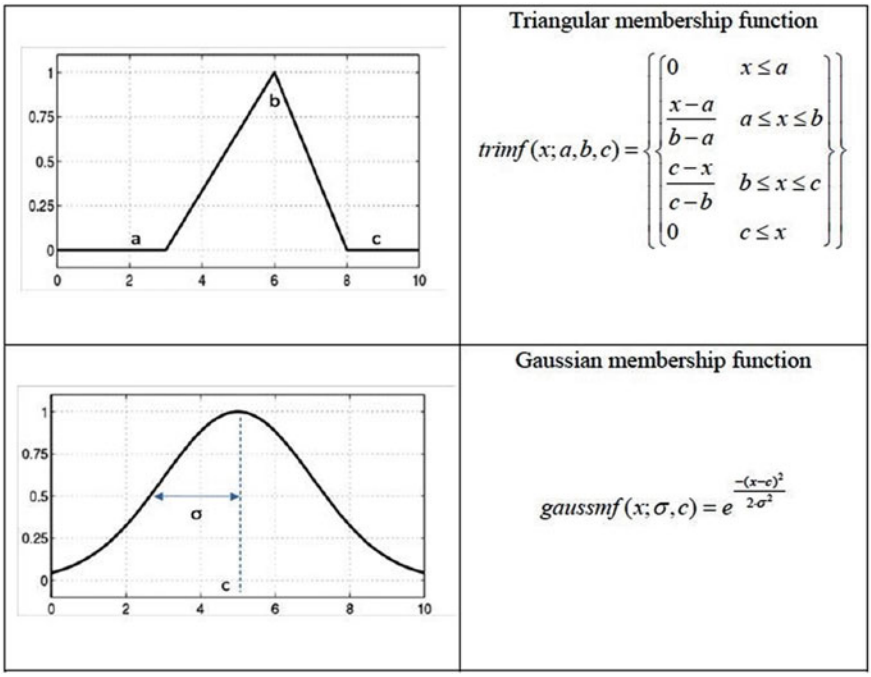


Fig. 3 MF for fuzzy logic system

1.2.1 Membership Functions

The way of expressing the level participation of individual inputs graphically is known as membership function. A particular value is allocated to the inputs by the membership function that also serves as a function of overlap among the inputs. By doing so, the output response also manipulated by membership function. "Configuration" or shape is a critical determining factor of membership function. These shapes are triangular, Gaussian, sigmoid, generalized bell or trapezoidal [12]. The range of degree of membership function lies between 0 and 1. Different types of MF are shown (Fig. 3).

1.2.2 Fuzzy Rules

The output variables are controlled by fuzzy rules. A theoretical background created by fuzzy logic based systems for making logical assumptions among inaccurate terms of reference. This is called approximate reasoning in fuzzy logic technological systems [13]. There are different subjects and verbs for fuzzy logic are used such as fuzzy operators and fuzzy sets or rule statements like "if–then". The "if–then" statements are also known as hallmark of the fuzzy logic.

1.2.3 Interference Engine

The process of deriving novel information from existing information is known as interference [14]. There are two popular methods to obtain interference viz. Takagi–Sugeno- Kang and Mamdani. Ebrahim H. Mamdani developed the Mamdani approach in 1975 which is used to rectifying the behaviour of a boiler and steam engine.

1.2.4 Defuzzification

The process of changing the output of interference engine (i.e. fuzzy in nature) into an exactly defined output (crisp) using membership functions is known as defuzzification [15]. There are different examples of defuzzification process such as Mean of Maximum (MOM), Centroid of Gravity (COG), First of Maxima and Last of Maxima, Weighted Average, Bisector of Area (BOA). The defuzzification process is performed by choosing a convenient defuzzification strategy i.e. based upon the application's properties.

2 Methodology

We have designed the FLC based MPPT for PV cell and simulated the same using MATLAB Simulink software to obtain the output of various power components i.e. used in this report. We have obtained the simulation [23–25] results in detail for the FLC based MPPT system.

2.1 Membership Functions Defined for FLC

We have chosen three subsets for input and output variables, viz. small, medium and high. We have adopted trapezoidal shapes for the membership functions. The range of input membership variables (PV voltage and PV current) has been modified. The range for output of fuzzy logic controller i.e. duty cycle has been kept between zero and one (Figs. 4, 5 and 6).

2.2 Control Rules

According to analysis of system behaviour, fuzzy control rules have been extracted. Different conditions have been assumed.

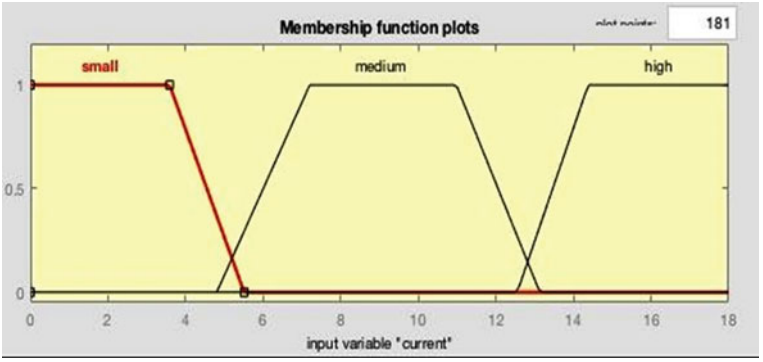


Fig. 4 Input variables 'Current'

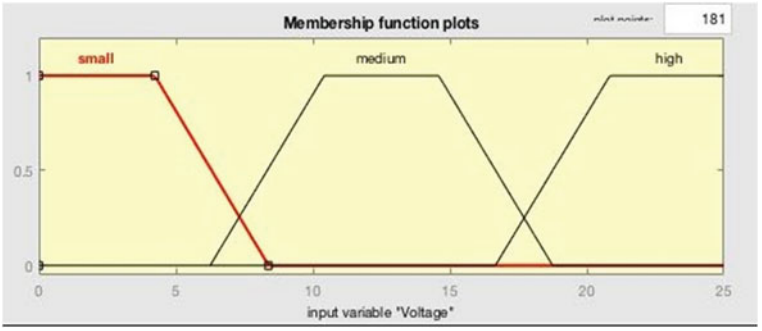


Fig. 5 Input variables 'Voltage'

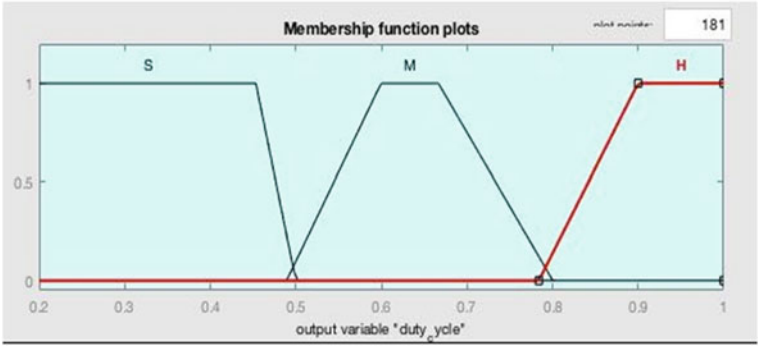


Fig. 6 Output variables 'duty cycle'

Initially, with duty cycle D = 0, the tracking process has been started. We have then measured the converter input voltage (Vm) and current (Im) and observed the duty cycle on which maximum power of the converter can be extracted according to the predicted values of fuzzy sets. This process has been repeated continuously until the system becomes stable. The description of Rules in Fuzzy controller is shown in Fig. 7.

We have designed and simulated the fuzzy logic algorithm based controller using MATLAB Simulink software (Fig. 8).

The graphical representation of the rule base (Fig. 9).

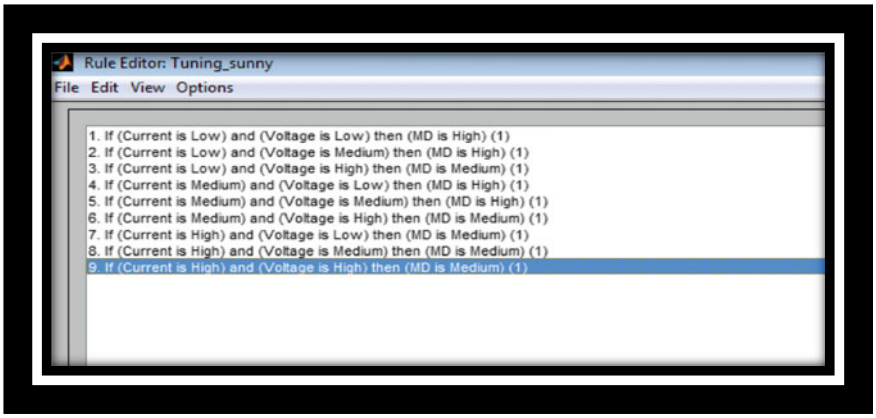


Fig. 7 Rules designed for fuzzy controller

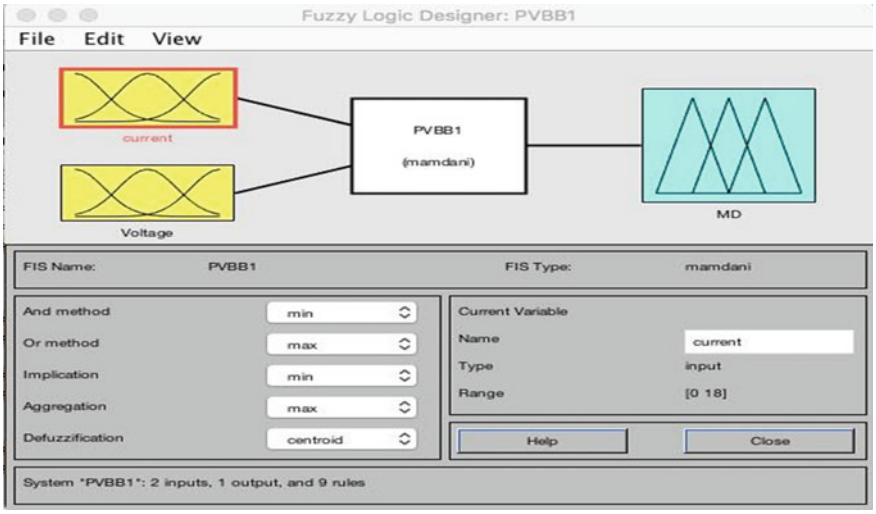


Fig. 8 Fuzzy logic design using Matlab/Simulink software

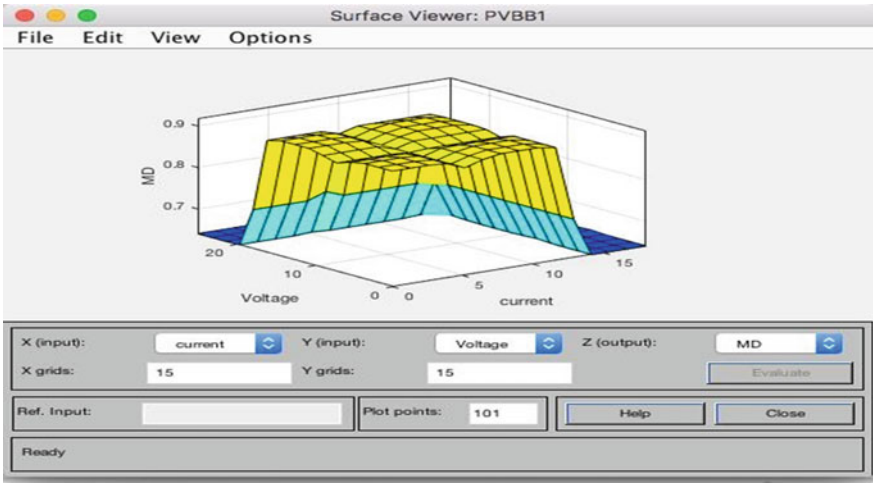


Fig. 9 Graphical representation of rules based on Fuzzy Logic

The rule viewer is shown in Fig. 10 which describes the operation of fuzzy logic.

FIS file was created after the modification of fuzzy controller in MATLAB. The performance of the fuzzy controller has been observed at different values of inputs as shown in Fig. 11.

The overall system has been combined and simulated using Matlab (Fig. 12).

The readings of input power as well as output power have been taken at solar irradiance viz. 1000, 800, 600, 400, 200 W/m^2. And the duty cycle has been observed on the same values of irradiation.

3 Analysis and Result

The simulation results of Fuzzy based MPPT controller [26] is shown (Table 1).

The Pi-Chart representation of Duty Cycle with Irradiance (W/m^2) is shown in Fig. 13.

Representation of Efficiency with Output and Input Power is shown in Fig. 14.

The Fig. 15 shows the relationship between Output Power with Irradiance (W/m^2).

The graphical representation of different parameters is shown in Fig. 16.

The average efficiency of the converter has been observes as 94.49%. Therefore, it is clear that, by using the fuzzy controller based MPPT; the maximum power can be extracted from photovoltaic module.

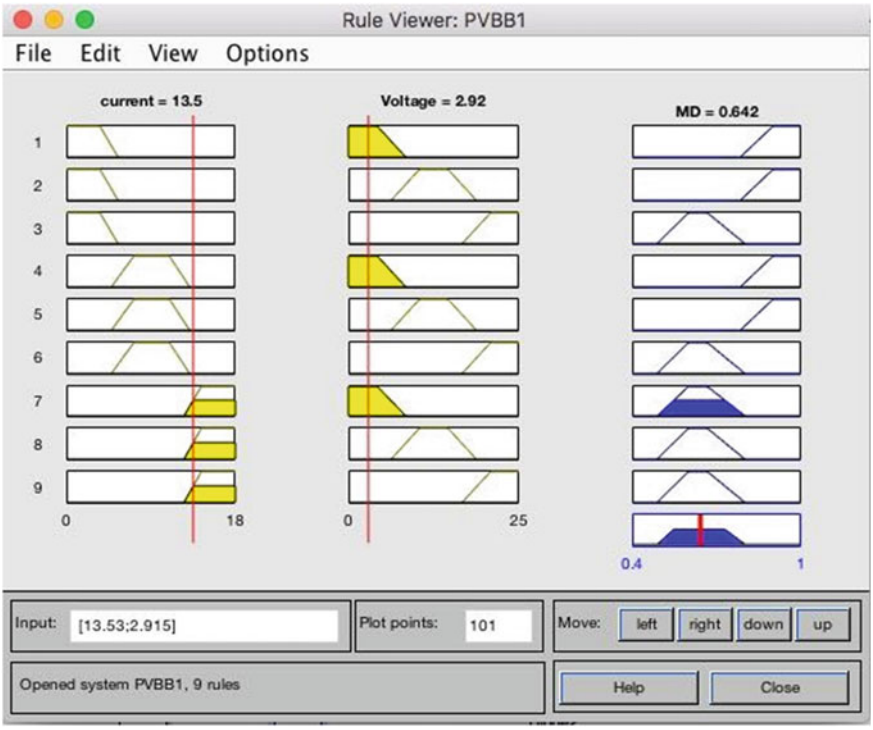


Fig. 10 Rule viewer

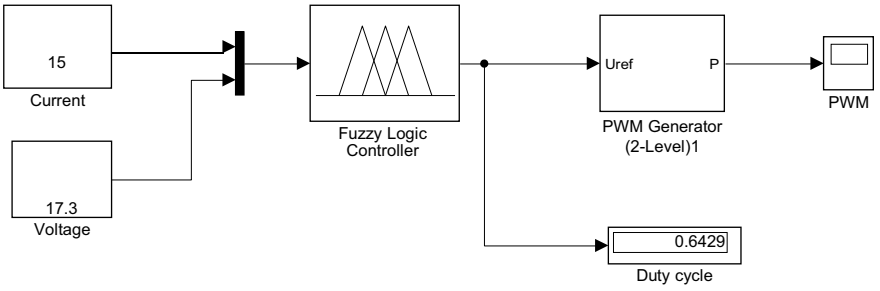


Fig. 11 The fuzzy controller Simulink model

4 Summary and Conclusion

With FLC (Fuzzy Logic Controller), the complete MPPT model has been created and we have used it to match the rules and membership functions of Fuzzy logic controller. As per our work is demonstrated, the Photovoltaic model is practically accurate and it can be used to model solar panel with the help of applying information i.e. available

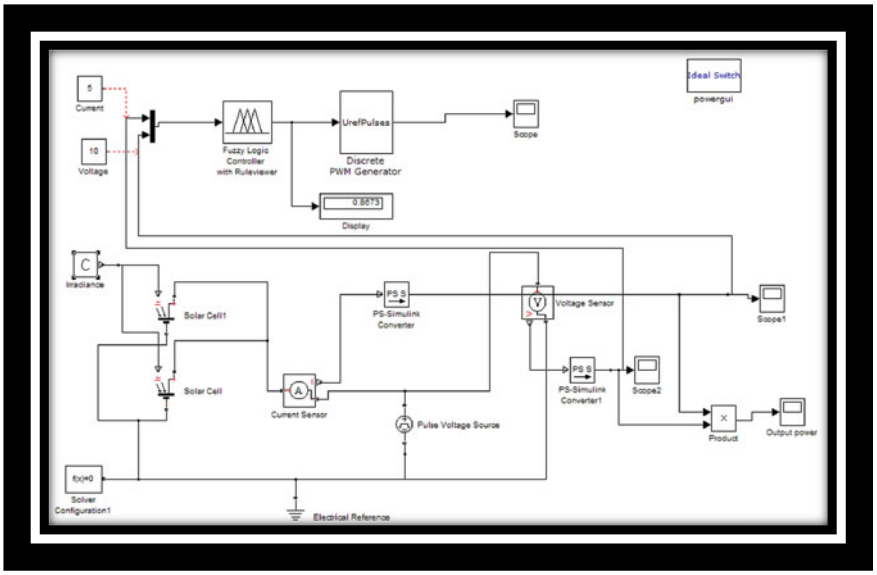


Fig. 12 Simulink model for Fuzzy System

Table 1 Tabular representation of simulation data

Irradiance (W/m^2)	P_{in} (W)	P_{out} (W)	$\eta = \frac{P_{out}}{P_{in}}$ (%)	DC
1000	257.7	252	$\frac{252}{257.7} = 0.98$	0.65
800	209.7	200.9	$\frac{200.9}{209.9} = 0.96$	0.65
600	150.9	142.7	$\frac{142.7}{150.9} = 0.95$	0.59
400	98.98	92.5	$\frac{92.5}{98.98} = 0.92$	0.49
200	48.86	44.4	$\frac{44.4}{48.86} = 0.91$	0.35

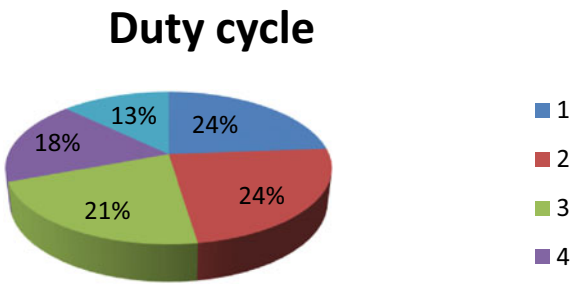


Fig. 13 Pi-chart representation of duty cycle with irradiance (W/m^2)

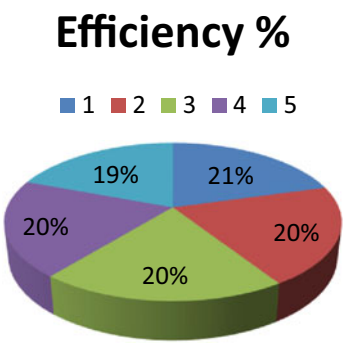


Fig. 14 Representation of efficiency with output and input power

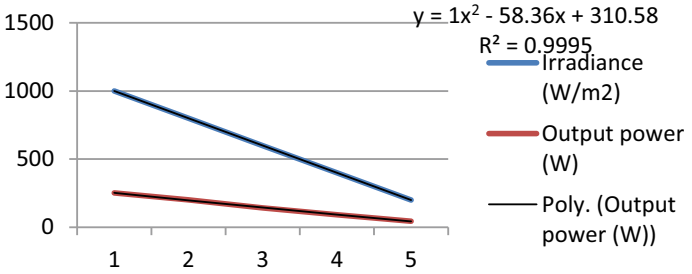


Fig. 15 Relationship between outputs power with irradiance (W/m^2)

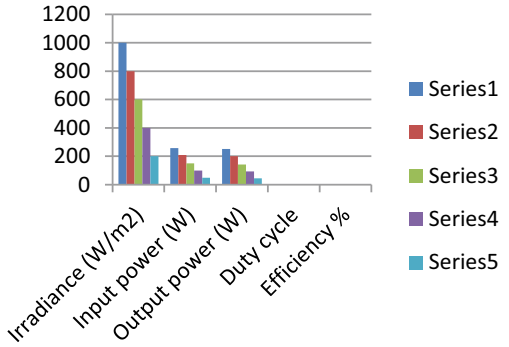


Fig. 16 Representation of different parameters

in the data sheets of manufacturer. Additionally, we have observed that FLC based Controller offers superior performance as compare to other MPPT with reference to speed, MPP oscillations and sensitivity to parameter changes. It is possible due to separately designed and assigned rules of FLC in various operational regions, ensuring in extremely successful large signal as well as small signal operation.

The simulation results also stat that one can extract additional energy in large amount from a PV module using a FLC based MPPT. These results identified a potential improvement in energy efficiency in PV power system. Ultimately, improvement in energy efficiency parameter is a cost saving measure i.e. completely guaranteed. In the near future fuzzy logic based MPPT algorithm could be implementing using a single chip microcontroller. Furthermore, the FLC circuit could be designed that can make use of power control circuit with the help of MPPT chargeable batteries. In general, we conclude that we have implemented FLC based MPPT for a PV power system very well, but still more work can be done by converting lab prototype into a product i.e. available commercially.

References

1. Rezk H, Aly M, Al-Dhaifallah M, Shoyama M (2019) Design and hardware implementation of new adaptive fuzzy logic-based MPPT Control method for photovoltaic applications. IEEE Access 7:106427–106438. https://doi.org/10.1109/ACCESS.2019.2932694
2. Priyadarshi N, Padmanaban S, Holm-Nielsen JB, Blaabjerg F, Bhaskar MS (March 2020) An experimental estimation of hybrid ANFIS–PSO-based MPPT for PV grid integration under fluctuating sun irradiance. IEEE Syst J 14(1):1218–1229. https://doi.org/10.1109/JSYST.2019.2949083
3. Amara K et al (2018) Improved performance of a PV solar panel with adaptive neuro Fuzzy inference system ANFIS based MPPT. In: 2018 7th international conference on renewable energy research and applications (ICRERA). Paris, pp 1098–1101. https://doi.org/10.1109/ICRERA.2018.8566818
4. Vasantharaj S, Vinodhkumar G, Sasikumar M (2012) Development of a fuzzy logic based, photovoltaic maximum power point tracking control system using boost converter. In: IET Chennai 3rd international on sustainable energy and intelligent systems (SEISCON. Tiruchengode, pp 1–6. https://doi.org/10.1049/cp.2012.2220
5. Aurangzeb M, Ai X, Hanan M, Jan MU, Ur Rehman H, Iqbal S (2019) Single algorithm mpso depend solar and wind mppt control and integrated with fuzzy controller for grid integration. In: 2019 IEEE 3rd conference on energy internet and energy system integration (EI2). Changsha, China, pp 583–588. https://doi.org/10.1109/EI247390.2019.9061805
6. Antal B, Hajdu A, Szab-Maros Z, Trk Z, Csutak A, Pet T (2012) A two-phase decision support framework for the automatic screening of digital fundus images. J Comput Sci 3:262–268
7. Gautam AK, Singh SP, Pandey JP, RP Payasi, Verma A (2016) Fuzzy logic based MPPT technique for photo-voltaic energy conversion system. In: 2016 IEEE Uttar Pradesh section international conference on electrical, computer and electronics engineering (UPCON). Varanasi, pp 275–281. https://doi.org/10.1109/UPCON.2016.7894665
8. Narendiran S, Sahoo SK, Das R, Sahoo AK (2016) Fuzzy logic controller based maximum power point tracking for PV system. In: 2016 3rd international conference on electrical energy systems (ICEES). Chennai, pp 29–34. https://doi.org/10.1109/ICEES.2016.7510590
9. Arora A, Gaur P (2015) AI based MPPT methods for grid connected PV systems under non linear changing solar irradiation. In: 2015 international conference on advances in computer engineering and applications. Ghaziabad, pp 542–547. https://doi.org/10.1109/ICACEA.2015.7164752
10. Laagoubi T, Bouzi M, Benchagra M (2015) Analysis and comparison of MPPT nonlinear controllers for PV system. In: 2015 3rd international renewable and sustainable energy conference (IRSEC). Marrakech, pp 1–5. https://doi.org/10.1109/IRSEC.2015.7455012

11. Khan MJ, Mathew L (2017) Artificial intelligence based maximum power point tracking algorithm for photo-voltaic system under variable environmental conditions. In: 2017 recent developments in control, automation and power engineering (RDCAPE). Noida, pp 114–119. https://doi.org/10.1109/RDCAPE.2017.8358251
12. Shi J, Wu Y, Wang H, Li C (2011) A method for MPPT control based on soft-switching circuit. In: 2011 international conference on electronics, communications and control (ICECC). Ningbo, pp 2429–2432. https://doi.org/10.1109/ICECC.2011.6067817
13. Won C-Y, Kim D-H, Kim S-C, Kim W-S, Kim H-S (1994) A new maximum power point tracker of photovoltaic arrays using fuzzy controller. In: Proceedings of 1994 power electronics specialist conference—PESC'94, vol 1. Taipei, Taiwan, pp 396–403. https://doi.org/10.1109/PESC.1994.349703
14. Nayak BP, Shaw A (2017) Design of MPPT controllers and PV cells using MATLAB Simulink and their analysis. In: 2017 international conference on nascent technologies in engineering (ICNTE). Navi Mumbai, pp 1–6. https://doi.org/10.1109/ICNTE.2017.7947932
15. Panchal D, Chatterjee P, Shukla RK, Choudhury T, Tamosaitiene J (2017) Integrated fuzzy AHP-CODAS framework for maintenance decision in urea fertilizer industry. Econ Comput Econ Cybern Stud Res 51(3)
16. Sharma M, Mittal R, Choudhury T, Satapathy SC, Kumar P (2018) Malaria detection using improved fuzzy algorithm. Adv Intell Syst Comput 673. https://doi.org/10.1007/978-981-10-7245-1_64
17. Khanna D, Choudhury T, Sabitha AS, Nhu NG (2019) Microarray gene expression analysis using fuzzy logic (MGA-FL). Adv Intell Syst Comput 755. https://doi.org/10.1007/978-981-13-1951-8_16
18. Kunwar V, Sabitha AS, Choudhury T, Aggarwal A (2019) Chronic kidney disease using fuzzy C-means clustering analysis. Int J Bus Anal 6(3). https://doi.org/10.4018/IJBAN.2019070104
19. Raj R, Mittal R, Gupta SC, Choudhury T (2018) Proposal of software development model using fuzzy logic. In: Proceedings of the 2nd international conference on green computing and internet of things. ICGCIoT. https://doi.org/10.1109/ICGCIoT.2018.8753049
20. Dhamija J, Rawat K, Choudhury T, Kumar P (2018. Enhancing data retrieval using fuzzy keywords and data security using feistal encryption algorithm. In:Proceedings of the 2nd International conference on green computing and internet of things. ICGCIoT. https://doi.org/10.1109/ICGCIoT.2018.8752988
21. Tomar R, Sastry HG, Prateek M (2020) A novel protocol for information dissemination in vehicular networks. Lecture notes in computer science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol 11894 LNCS. Springer, Cham, pp 1–14
22. Tomar R, Prateek M, Sastry HG (2017) A novel approach to multicast in VANET using MQTT. Ada User J 38(4):231–235
23. Tomar R, Sastry H, Prateek M (2020) Establishing parameters for comparative analysis of V2V communication in VANET. J Sci Indus Res (JSIR) 79(01):26–29
24. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of Weiler-Atherton algorithm to address loose polygons. J Sci Ind Res 78:771–774
25. Choudhury T, Aggarwal A, Tomar R (2020) A deep learning approach to helmet detection for road safety. J Sci Ind Res (JSIR) 79(06):509–512
26. Kumar V, Choudhury T (2019) Real-time recognition of malignant skin lesions using ensemble modeling. J Sci Indus Res (JSIR) 78:148–153

A Proposed Approach to DDoS Attacks Detection on SDN Using Machine Learning Technique

Saumitra Chattopadhyay and Ashok Kumar Sahoo

Abstract Software-defined network (SDN) is becoming very popular because of numerous advantages it offers, such as scalability, centralized management, better performance and other parameters. However, it suffers from many security threats including DDoS attacks. These attacks are performed more on the controllers of SDN, which is the most important component in this new network architecture. When a DDoS attack is performed in the controller it becomes over-engaged and its power to communicate with network applications and devices is reduced drastically. Due to this overloading, the switch flow tables become exhausted and the overall network performance suffers and reduces to a bare minimum, and even stops at many times. In the proposed work, an approach to DDoS attacks in a software-defined network is described. About 130,000 feature vectors will be acquired from SDN for experimentation under normal traffic and expected DDoS-based traffic. Using feature selection methods, a proposed data set will be created. The nature of the feature selection method is to shorten the training time and prediction time, simplification of models and facilitate the interpretation of results. The original feature vector set and the data set filtered by the feature selection method will be trained with classifiers, namely naïve Bayes (NB), k-nearest neighbor (k-NN), artificial neural network (ANN) and support vector machine (SVM). The test results will be compared against existing researches. The aim of the experiments will be to obtain efficient results in terms of DDoS detection in SDN and pruning of time and processing loads.

Keywords DDoS · SDN · SVM · k-NN · Artificial neural network · Naïve Bayes

1 Introduction

Existing network infrastructure has been improved several times to solve the challenges of bandwidth, speed, accessibility, cloud computing and virtualization. Software-defined network [1] is one of the latest network architecture that is based on the decoupling of physical plane and control plane [2] which manage the networks

S. Chattopadhyay (✉) · A. K. Sahoo
Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_6

by means of software. The most prime component of this SDN-enabled network is a controller which is also considered to be the brain of SDN. These network architectonics with separated control planes generate many queries regarding network reliability, extendibility and network throughput in contrast to more traditional distributed systems.

The primary objective of the control plane is to direct the transmission devices present in the data plane. The most important component called controller is present here. The controller is considered to be the processing unit of the SDN. All network devices present in the data plane carry the function of packet forwarding based on the set of rules in the controller plane. The application plane interacts with various components present on the network base through the controller.

Network devices present on the particular network are controlled by an application software system. The software applications are either Northbound or Southbound application program interfaces. The SDN controller interacts with the network services which are operating on the application layer with the help of Northbound API and interacts with the network devices on the data layer with the help of Southbound API. There are many protocols used in these interfaces like the OpenFlow protocol [3].

In SDN all decision-making is performed in the controller and not in individual devices it can be optimized [4]. On the other side, any fault or attack in the controller will affect the entire network. SDN has many types of active and passive threats. At least seven popular threats have been identified in SDN [5]. The most important among all is DDoS attacks. The DDoS attack is performed mainly in two places: the first is in the controller and the second in the flow table entries of the OpenFlow switch.

The DDoS attack is performed in the controller in the transmission link between the data plane devices and the controller. This creates a huge amount of packet flow in the direction of the switch on the data plane. When the packets are checked at the OpenFlow switch and if a mismatch (misflow) is identified then the incoming data are stored into the flow buffer. After this, they are taken to the SDN controller with a message containing packet-in data to create a new principle. As a result, the memory, processor and bandwidth of the controller become choked and, as a result, the whole network becomes not available. Also, the overall network performance declines drastically. Moreover, the OpenFlow switch and the network link between the controller and other devices in the data plane are severely affected after the occurrence of the attack [6].

The DDoS attack affects the network devices in the data plane through their flow table. After the attack, data frames are forwarded to the OpenFlow switch with an unknown source identifier. The SDN controller creates a specific regulation for data packets and forwards them to an OpenFlow switch, and the flow table entries are created. As a result, after some time, the flow table entries become full and no new entries are allowed. As of now, no more new rules can be created, so the packets may not be forwarded. Now, any new incoming packets dropped as the buffer in the flow table are exhausted because of the attack [7]. Overcoming a DDoS attack in an SDN network is more complex than in the case of the traditional network.

At this point, we propose that early detection of these attacks can reduce the damage caused and proper countermeasures can be taken. Also, we propose that machine learning [27–32] models provide the best SDN management and security features.

This research paper is divided into five different sections. In the second section, a brief review of the DDoS strikes on an SDN controller and distributed DoS attack on the OpenFlow switch along with the description of the SDN platform, features and security flaws is provided. In the next section, the details regarding the process are taken to study the machine learning [27–32] models in the discovery of the DDoS attacks in a software-defined network, and the various proposed experiments and experiments conducted by earlier researchers in this regard are discussed. In the results and discussion section, a framework and experimentation are proposed for future work. However, in this section, a discussion on existing results in this area will be presented also from existing researchers in this area. In the last section, the concluding remarks were made and future scopes were discussed.

2 Related Works

Here we will look into the various studies taken up related to DDoS attacks on SDN. Data flow decision-making work is now performed by the controller and is no longer performed at the switch level. That is the reason the controller is called the brain of SDN. The controllers in SDN are capable of implementing the top-level rules and also create, modify and add new rules for the data plane devices. Controllers implement these rules by communicating with the devices with the help of secure channels using OpenFlow protocol. These channels play a very crucial role and any problem in this disrupts the connection between the controller and the data plane devices.

As discussed earlier, one of the major attacks in SDN architecture is a DDoS attack. There are three main areas where the attackers try to attack. It tries to deplete the resources of the controller, swallow the transmission capacity of the channel and load flow tables of the switch with wrong entries. Therefore, it becomes very difficult for the controller to distinguish between legal and illegal traffic.

The task of the OpenFlow switch is to find a perfect match in the flow by examining the header of the packet. If it is unable to identify a match, the flow packet is advanced to the controller after wrapping the header of the message packet including the Packet_IN message. After this, the controller returns the FLOW_MOD message [8]. The more is the number of packets dispatched to the controller of SDN, the greater is the controller resource exhausted. As a result, the new genuine packets receiving at the network are not able to be executed and the whole SDN collapses.

The authors [9] found out that many existing solutions to stop DDoS strikes are not much effective. The authors have conducted an investigation to study the effects of misbehavior and also on new flow attacks on a software-defined network. They have collected traffic information from data plane devices at the periodic interval and

then on this data set, they applied machine learning [27–32] for the classification of various algorithms to study the abrupt change in data traffic at the time of the attack. The researchers considered the Packet_IN message packets transferred between SDN controller and data plane devices at the time of the attack as the base. The algorithms to be used for classification are SVM, NB, k-NN and ANN.

The researchers in [10] identified the DDoS attacks on SDN by evaluating the count of the arriving packets at the time of the attack. When some of the packet's rate of arrival at the controller becomes more than the predetermined threshold value, an inspection unit was activated. It uses a support vector machine to ascertain the DDoS attacks.

The authors in [11] proposed a model which uses bidirectional recurrent NN applied at each layer (application, control and data) of the SDN to identify and block DDoS attack. Even though this proposed model has a very high accuracy rate in the case of real-time detection and mitigation of the DDoS attack, but its accuracy is reduced in the case of large networks having multiple controllers' setups. Also, this has other problems like degrading the network performance and disruption of the controller device synchronization.

The DDoS attack most frequently tries to target the OpenFlow switch and the corresponding flow table as it contains many vital network information related to accessing control, administration and transmission parameters of the network [12]. In most cases, the attacker first tries to gain unauthorized access to the target SDN network and then tries to stop the network services so that authorized users cannot access the network. The attacker exploits the fact that the OpenFlow switch has limited storage and cannot store all the rules at a time. So as and when the new packets arrive from a spurious source, new rules have to be created and appended to the OpenFlow switch. The attacker floods the SDN network with many packets in a very short period of time from an unknown source address. The controller creates the new rules for these communicating messages and sends them to the SDN flow table. As the flow table has a very less capacity for storage, it becomes overflow soon after the DDoS attack is initiated. As a result, no place is left available in the flow table for the new regulation entry and the network services stop. There is also another memory called flow cache memory in the OpenFlow switch. Once a packet is received in the input port it is sent to the cache memory [13] and it is matched with the rules in the flow table. If it matches then the packet is sent from the cache memory directly to the exit port of the OpenFlow switch. In the case of a non-matching, it is sent to the SDN controller with a Packet_IN message. After this, the controller sends back the packet to the OpenFlow switch with OFPT_FLOW_MOD message. When a massive number of message packets are sent to the OpenFlow switch by attacker nodes, the message packets are sent to the cache memory and the rules from the controller are also stored for these packets in the flow table [14]. As explained earlier this fills the flow table and also the buffer due to this the new packets from the authorized legitimate users are dropped [15].

The researchers in [16] collected network traffic data from the data plane devices, and the values associated with DDoS attacks were collected from the flow table of the OpenFlow switch. The SVM algorithm [33, 34] was used to ascertain the DDoS

attack. This approach has the advantage of a very high detection accuracy rate and has the disadvantage of a low accuracy rate in the case of an ICMP attack.

Xue et al. [17] in their approach state that we have to take care of many security parameters where multiple switches are managed by a controller. It also states that pre-SDN switches create more threats than the OpenFlow compatible switches. It also conveys that DDoS attacks have to be managed properly if the SDN architecture has to be implemented on a large scale. It uses SVM implemented by a genetic algorithm to detect the DDoS attack.

The authors in [18] suggested the use of a pattern recognition algorithm to be properly trained with the previous network attach data sets. It has used a Bayesian classification network, decision tree, naïve Bayes classifiers and C4.5. It also states that user recognition can help the SDN controllers to write new regulations more effectively and quickly, prevent DDoS attacks and increases network efficiency.

This study mainly emphasized DDoS attacks in SDN architecture and proposed the best solutions to detect attacks and minimize its impacts as quickly as possible. The purpose of this research is to develop DDoS attack detection systems after going through the existing studies on high-accuracy rate artificial intelligence techniques for SDN structure. The early spotting of the DDoS strike in the data plane plays an important role in the SDN controller and the switches for the efficient working of the network and the detection of authorized traffic before an actual attack takes place. If identification of unauthorized traffic on the controller is detected, it will be simpler for the SDN controller to implement a new set of regulations to the flow table of SDN switches situated in the data plane to counter the early strike. This comes up with an ideal solution for preventing the attack.

The researchers suggest and recommend the use of extraction of important features and various feature selection techniques with pattern recognition models to protect against DDoS attacks. According to their experiments, they are very confident about a concrete solution toward the detection of distributed DoS attacks in SDN.

3 Feature Selection Techniques and ML Techniques

As discussed in the introduction, the aim of this paper is to propose an effective machine learning approach to ascertain DDoS attacks. Machine learning uses the theory of probability and statistics to predict something on the basis of training vectors. In the related works in the above section, a number of such methods have been exploited by many researchers in this area. The researchers have used a variety of machine learning approaches to solve this problem. To conclude, it is recommended that the use of SVM, k-NN, naïve Bayes and artificial neural network-based classifiers outperformed the performance. Therefore, in the current study, such classifiers are selected for the experiments.

The heart of any machine learning algorithm is an efficient data set in terms of the number of records and the dimension of each record. These factors also influence the cost factors such as time and processing. It is therefore a requirement to reduce the

number of dimensions of each record for efficient and timely processing of feature vectors.

In addition to the collection of feature points for classification, it is also important for the selection of important features and discarding unimportant features for the experiments. This is called the feature selection process. In the proposed research, wrapper and embedded feature selection methods are considered. The wrapper feature selection technique gives more importance to the feature-based performance of classifiers. The other method focuses to use wrapper and filter techniques to optimize the learning algorithms used. The beauty of each feature selection method is its utility in different algorithms. As it is a pattern recognition problem, there is no scope for missing data. Only two class labels are present here, attack on DDoS or no attack. This is the simplest case of pattern recognition called binary classification. In the filter-based feature selection, the relief filtering technique influenced by instance-based learning is considered.

For the feature subset, the option is a greedy structure. Sequential forward selection search algorithm (SFS) is considered as its capability of selection of excellent feature subset, including another wrapper postulated feature selection techniques. Among others, Lasso or L1 algorithm is considered as it supplements a penalty against complexity for curtailing overfitting degree and drastically improving the effectiveness of the proposed model.

3.1 Filter-Based Feature Selection

This method contributes to achieving optimal results using statistical techniques. By using a relief algorithm, a reduced feature set is to be created. In the data set, the relief algorithm establishes the relation between samples in the data set within instances of its own class and other classes. The distance of sample within the data set and instances of data sets will lead to the creation of a model and enforce the feature selection process [19, 20].

3.2 Wrapper-Postulated Feature Selection

This feature selection algorithm selects the most important ideal attributes on the trial-and-error basis in the classifier. Once the flawless subspace of features is achieved, the task is said to be finished and the moderated data set has been created. This might prove more successful than other statistical methods; however, it is fully dependent on classification algorithms. The primary disadvantage is speed. In considering many fold facts, the sequential forward floating selection algorithm is proposed. The same direction will be used till the subdivision accuracy reaches a higher figure [19, 20].

3.3 Embedded Postulated Feature Selection

This is a feature selection algorithm that is included in the classification process. These are based on the selection of feature set by identification of feature sets that will provide the best recognition accuracy to the model. These algorithms take advantage of both filter and wrapper-based techniques. Here, a learning algorithm is used which takes advantage of the variable selection process and concurrently performs selection and classification [19]. The Lasso algorithm eradicates the weights of unimportant features. In this manner, the selection of a reduced set of features takes place [19, 20].

3.4 The Support Vector Machine (SVM)

SVM [21] is a pattern recognition model based on supervised learning. It uses classification and regression for analyzing the obtained patterns. It is a binary classification model but can be extended to any number of class labels. It is not a linear classifier and uses a hyper-plane that is perfectly separated into two classes.

3.5 Naïve Bayes (NB)

It is one of the earliest but powerful classification techniques and predicts input vector to appropriate class based on the principle of probability [20, 21]. Here, a sufficient number of training data is provided to the system. However, it computes the model every time when one needs to perform the classification task. Therefore, it is slow as compared to other classifiers and also known as a lazy learner. It uses the concept of prior, likelihood and posterior probability concepts for predicting the class labels of the test data set. The performance of this classifier also depends on more number of training data.

3.6 Artificial Neural Network (ANN)

It is an efficient tool for classification problems based on the biological neuron model of a human body. Neurons are a basic building block of this particular classification model. Due to limitations of processing, a limited number of artificial neurons are used for calculation and hence restricts its perdition than actual human beings. In the ANN model [21–23], one input layer depends on the number of dimensions of the input vector from the data set (training or testing), one or many hidden layers depend on the problem statement and one output layer depends on the number of output classes. Between input layer and hidden layer, hidden to hidden layers (if any)

and hidden layer to output layer, fully connectionist weights are randomly provided. When training data is provided to the model, it will stable the connection weights between the layers to provide maximum throughput.

3.7 K-Nearest Neighbors

It is applicable to regression and classification problems. Implementation of this algorithm is quite easy, and it follows a supervised machine learning technique. When a new data vector is appearing in the n-dimensional space of the k-NN model [23], its distance from all k neighbors will be calculated. Its class label will be decided on the basis of distance measurement. It belongs to a class with the lowest distance among all the neighbors.

4 Experimental Setup and Collecting Data from SDN

The following setup is required to carry out the experiments. A number of personal computers (at least six in number) will be acted as virtual machines are required. As most of the network-related research is based on Unix platforms, the PCs should have "Ubuntu" latest version, at least 1 GB primary memory. One CPU should have "VirtualBox-KVM", two VM switches, one Open vSwitch, sFlow, and InfluxDB set up in the same machines. These are required to be well connected to carry out experimentation.

To gather data points (training and testing data sets), all network-related metrics are to be collected from the OVS switch. This can be done by using sFlowDocker image at normal traffic and at the moment of attack traffic. JavaScript can be employed to collect network metrics with a timestamp on an InfluxDB image.

Transmission control protocol (TCP), internet control protocol (ICMP) and user datagram protocol (UDP) are to be used to create an "hping3" packet generator using a DDoS attack data set. The "hping3" toolbox is present on a PC in the network and is defined as a striker, while some other PC is to be selected as the victim. The victim PC should have the IP address 10.0.0.6 with a payload size of 512 bytes and 2000 PPS (packet per second) packet rates for each of the three protocol-oriented flooding attacks. It might influence the figure of flow mitigation from a controller with more than 1304/s, which could be 6996 entries in the flow table. The communication capacity of the controller might be overwhelmed due to the induced DDoS strike packet alongside the processing, controller, CPU capacity, and bandwidth and communication capacity. The controller might create attacking packets; hence the capacity of flow table entries can be exhausted. The experiments will be conducted for all three protocol packets (TCP, ICMP and UDP) for about a quarter of an hour on about 65,000 samples, and all are network-related traffic flow when DDoS attack will be planned. Two personal computers are also to be deployed to record the flow

of normal traffic in the above protocols and also to be consisting of about the same amount of samples as required for the attack. Therefore about 130,000 samples are required for the experiments. The following feature sets are to be considered in order to collect samples.

- If_in_pkt: Packets coming to the system
- If_out_pkt: Packets going out of the system
- Hits: Number of outgoing packets from the system
- Masks: To display the statistics of the mega flow mask.
- Miss: If a packet appears as a transmission device like a switch or a router, a match is sought in the flow table. If the incoming packet is not matching with any flow input, the packet will be forwarded to the controller to a new set of flow rules.
- Open_Flow_From: This is the flow modification message interpreted by the controller.
- CPU_Util: The computer executes and calculates the incoming data with help of mathematical operations. If the value is high, it indicates the computer executes above the normal level.
- Magt_Int_In_Pkt: Input packets at control plane.
- Magt_Int_Out_Pkt: Packets going out of the control plane.
- Open_Flow_To: Primary message sent by the controller.
- OVS_Flow: Total number of flows in data pathway.
- OVS_Lost: The number of dropped packets.
- Class: The traffic class of data. The supervised learning classes are normal, ICMP Flood, UDP Flood and TCP Flood.

The data set as described above will also be reduced by some of the feature selection methods. After proper collection of data set vectors, these data sets will be provided to classifiers such as support vector machine, artificial neural network, naïve Bayes and k-nearest neighbor.

After proper collection of data sets, the next step is classification. The data sets are divided into training and testing sets. The experiment or classification will be held in two phases. In the first phase, testing and training of classifiers will be carried out by all data set features collected so far. The data sets are divided into training and testing by the method of k-fold cross-validation [24, 25]. The feature selection method will be used in the second phase. Some of the following feature selection methods like Lasso algorithm, relief or any other feature selection method will be used on the basis of most effective in the entire data set. The performance of feature selection methods will be examined by all the classifier algorithms as described above.

5 Experimental Results

First, twofold experiments will be carried out: normal traffic movement of data and traffic movement of DDoS attack on SDN architecture. No feature selection method will be used in this study. The parameters of evaluation are accuracy, sensitivity,

specificity, precision and F1_Score (Eqs. 1–5). These parameters will be obtained from the confusion matrix of all classifier performance. The definitions of each of them are as under:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \tag{1}$$

$$\text{Precision} = \frac{\text{Number of true positives}}{\text{Number of true positives plus the number of false positives}} \tag{2}$$

$$\text{Sensitivity (or Recall)} = \frac{\text{Number of correct positive predictions}}{\text{Total number of positives}} \tag{3}$$

$$\text{Specificity} = \frac{\text{Number of correct negative predictions}}{\text{Total number of negatives}} \tag{4}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{5}$$

In the experiments carried out by [26] using all four classifiers, accuracy is reported to be 91–96%, sensitivity is in the range of 87–94%, specificity is about 96–98%, the precision range is 90–97% and, finally, the F1 score is 90–95%. In the proposed research, the range of results is expected to be more than that has been reported in the experiments [26].

Most of the results can be analyzed with the help of a confusion matrix. The results which can be obtained from the experiments are through all the classifiers applied (k-NN, SVM, ANN and naive Bayes) with or without feature selection. Therefore, a large number of results will be obtained and required to be analyzed by a number of parameters like accuracy, precision, sensitivity, specificity, precision and F1 score. Some of the probable combinations of results are expected and described below.

- Confusion matrix of k-NN classifier without feature selection
  Four different types of traffic data will be considered in one of the experiments, namely normal, TCP, UDP and ICMP. The confusion matrix size will be a 4 × 4 matrix, and it will be analyzed with the above five parameters. In the experiments [26], the true-positive rates reported are in the range of 91.00–99.6%, and false-negative results range from 0.4 to 14.4%. In our proposed method, we are expecting more appropriate results.
- Interpretation of filter-based feature selection method results with ML models
  In the proposed experiment we expect the estimation of all five parameters on all four classifiers with varying feature sizes. The experiment carried by [26] used 10, 6, 6, 8 as feature sizes from the data sets for SVM, k-NN, ANN and naive Bayes respectively. They reported the results as shown in Table 1.
- Confusion matrix of k-NN classifier with feature selection
  Four different types of traffic data will be considered in one of the experiments, namely normal, TCP, UDP and ICMP. The confusion matrix size will be a 4 × 4

Table 1 Results obtained by [26] with help of filter-based selection method

Parameter	Result range (%)
Accuracy	92.28–97.15
Precision	91.62–95.07
Sensitivity	89.02–95.88
Specificity	96.99–98.68
F1 score	90.20–96.92

matrix, and it will be analyzed with the above five parameters. In the experiments [26], the true-positive rates reported are in the range of 91.7–99.8% and the false-negative results range is 0.2–8.3%. In our proposed method, we are expecting more appropriate results.

- The wrapper-based and embedded feature selection method performance
  In the proposed experiment we expect the estimation of all five parameters on all four classifiers with varying feature sizes. The experiment carried by [26] used 10, 6, 6, 8 as feature sizes from the data sets for SVM, k-NN, ANN and naive Bayes respectively. They reported the results as shown in Table 2.
- Confusion matrix of the k-NN classifier with wrapper-based method.
  Four different types of traffic data will be considered in one of the experiments, namely normal, TCP, UDP and ICMP. The confusion matrix size will be a 4×4 matrix, and it will be analyzed with the above five parameters. In the experiments [26], the true-positive rates reported are in the range of 94.0–99.6% and the false-negative results range is 0.4–6.0%. In our proposed method, we are expecting more appropriate results.
  In the proposed experiment we expect the estimation of all five parameters on all four classifiers with varying feature sizes. The experiment carried by [26] used 10, 6, 6, 8 as feature sizes from the data sets for SVM, k-NN, ANN and naive Bayes respectively. They reported the results as shown in Table 3.
- Confusion matrix of the k-NN classifier with embedded technique.
  Four different types of traffic data will be considered in one of the experiments, namely, normal, TCP, UDP and ICMP. The confusion matrix size will be a 4×4 matrix, and it will be analyzed with the above five parameters. In the experiments [26], the true-positive rates reported are in the range of 84.1–99.4% and the false-negative results range is 0.6–15.9%. In our proposed method, we are expecting more appropriate results.

Table 2 Results obtained by [26] with help of wrapper/embedded-based selection method

Parameter	Result range (%)
Accuracy	91.44–98.30
Precision	88.11–97.72
Sensitivity	87.82–97.73
Specificity	97.26–99.45
F1 score	87.89–97.70

Table 3 Results obtained by [26] with help of all four classifiers

Parameter	Result range (%)
Accuracy	91.44–98.30
Precision	88.11–97.72
Sensitivity	87.82–97.73
Specificity	97.26–99.45
F1 score	87.89–97.70

In the second last experiment, a comprehensive result analysis will be done with the feature selection methods (normal features, filtered, wrapped, and embedded) with varying number of features and five different parameters (accuracy, precision, sensitivity, specificity, F1 score). From the result table, some conclusions can be drawn for the implementation of the proposed research. However, it is expected to select the highest number of feature selection method as an implementation model.

At the last, one more compressive table will be prepared to compare the results obtained with the results obtained by previous researchers in the field.

6 Conclusions

Machine learning systems can be amalgamated with software-defined networks to detect DDoS attacks. Systems with high accurate results can be designed with the help of machine learning algorithms and these systems can be successfully deployed in practice also. However, the implementation is not easy. There are various difficult ways to collect a huge amount of training and testing data. Experiments are to be set up with varying feature selection techniques and the varying number of features to be selected. After exhaustive experiments, a ray of hope might be obtained to implement the model in real time. There is a requirement to integrate the SDN network with machine learning systems to work together against DDoS attacks on SDN.

References

1. Open Networking Foundation. Software-Defined Networking: The New Norm for Networks. Available online: www.opennetworking.org/images/stories/downloads /sdn-resources/white-papers/wp-sdn- newnorm.pdf. Last accessed 10 April 2020
2. Kreutz D, Ramos FMV, Verissimo PE, Rothenberg CE, Azodolmolky S, Uhlig S (2014) Software-defined networking: a comprehensive survey. Proc IEEE 103(1):14–76
3. Xie J, Guo D, Zhiyao Hu, Ting Qu, Lv P (2015) Control plane of software defined networks: a survey. Comput Commun 67:1–10
4. Suciu G, Vulpe A (2015) An overview study of software defined networking. In: Proceedings of the IE international conference, vol 30. Bucharest, Romania

5. Kreutz D, Ramos FMV, Verissimo P (2013) Towards secure and dependable software-defined networks. In: Proceedings of the second ACM SIGCOMM workshop on hot topics in software defined networking, pp 55–60
6. Shu Z, Wan J, Li Di, Lin J, Vasilakos AV, Imran M (2016) Security in software-defined networking: threats and countermeasures. Mobile Netw Appl 21(5):764–776
7. Ali ST, Sivaraman V, Radford A, Jha S (2015) A survey of securing networks using software defined networking. IEEE Trans Reliabil 64(3):1086–1097
8. Kuerban M, Tian Y, Yang Q, Jia Y, Huebert B, Poss D (2016) FlowSec: DOS attack mitigation strategy on SDN controller. In: 2016 IEEE international conference on networking, architecture and storage (NAS). IEEE, pp 1–2
9. Alshamrani A, Chowdhary A, Pisharody S, Lu D, Huang D (2017) A defense system for defeating DDoS attacks in SDN based networks. In: Proceedings of the 15th ACM international symposium on mobility management and wireless access, pp 83–92
10. Latah M, Toker L (2018) A novel intelligent approach for detecting DoS flooding attacks in software-defined networks. Int J Adv Intell Inf 2018
11. Li C, Wu Y, Yuan X, Sun Z, Wang W, Li X, Gong L (2018) Detection and defense of DDoS attack–based on deep learning in openflow-based SDN. Int J Commun Syst 31(5):e3497
12. Li W, Meng W, Kwok LF (2016) A survey on openflow-based software defined networks: security challenges and countermeasures. J Netw Comput Appl 68(2016):126–139
13. Padmaja S, Vetriselvi V (2016) Mitigation of switch-DoS in software defined network. In: 2016 international conference on information communication and embedded systems (ICICES). IEEE, pp 1–5N
14. Polat H, Polat O (2017) The effects of DoS attacks on ODL and POX SDN controllers. In: 2017 8th international conference on information technology (ICIT). IEEE, pp 554–558
15. Bholebawa IZ, Dalal UD (2016) Design and performance analysis of openflow-enabled network topologies using mininet. Int J Comput Commun Eng 5(6):419
16. Ye J, Cheng X, Zhu J, Feng L, Song L (2018) A DDoS attack detection method based on SVM in software defined network. Secur Commun Netw 2018
17. Li X, Yuan D, Hu H, Ran J, Li S (2015) DDoS detection in SDN switches using support vector machine classifier. In: 2015 joint international mechanical, electronic and information technology conference (JIMET-15). Atlantis Press
18. Nanda S, Zafari F, DeCusatis C, Wedaa E, Yang B (2016) Predicting network attack patterns in SDN using machine learning approach. In: 2016 IEEE conference on network function virtualization and software defined networks (NFV-SDN). IEEE, pp 167–172
19. Hira ZM, Gillies DF (2015) A review of feature selection and feature extraction methods applied on microarray data. Adv Bioinf 2015
20. Venkatesh B, Anuradha J (2019) A review of feature selection and its methods. Cybern Inf Technol 19(1):3–26
21. Sahoo AK, Ravulakollu KK (2014) Vision based indian sign language character recognition. J Theor Appl Inf Technol 67(2):770–780
22. Sahoo AK, Ravulakollu KK (2014) Indian sign language recognition using skin color detection. Int J Appl Eng Res 9(20):7347–7360
23. Sharma M, Pal R, Sahoo AK (2014) Indian sign language recognition using neural networks and kNN classifiers. J Eng Appl Sci 9(8):1255–1259
24. Stamp M (2017) Introduction to machine learning with applications in information security. Chapman and Hall/CRC, Boca Raton, FL, USA
25. Somol P, Novovicová J, Pudil P (2010) Improving sequential feature selection methods performance by means of hybridization. In: Proceedings 6th IASTED international conference on advances in computer science and engineering, vol 2010. ACTA Press
26. Polat H, Polat O, Cetin A (2020) Detecting DDoS attacks in software-defined networks through feature selection methods and machine learning models. Sustainability 12(3):1035
27. Sengupta E, Garg D, Choudhury T, Aggarwal A (2018) Techniques to elimenate human bias in machine learning. In: Proceedings of the 2018 International conference on system modeling and advancement in research trends. SMART. https://doi.org/10.1109/SYSMART.2018.8746946

28. Verma A, Shukla A, Choudhury T, Chauhan A (2020) Recommendation system for prediction of tumour in cells using machine learning approach. In: Communications in Computer and information science, vol 1229. CCIS. https://doi.org/10.1007/978-981-15-5827-6_18
29. Kumari A, Choudhury T, Chitra Rajagopal P (2019) Exploring open source for machine learning problem on diabetic retinopathy. Adv Intell Syst Comput 841. https://doi.org/10.1007/978-981-13-2285-3_67
30. Sharma S, Aggarwal A, Choudhury T (2018) Breast cancer detection using machine learning algorithms. In: Proceedings of the international conference on computational techniques, electronics and mechanical systems. CTEMS. https://doi.org/10.1109/CTEMS.2018.8769187
31. Singh UV, Gupta E, Choudhury T (2019) Detection of rheumatoid arthritis using machine learning. In: Proceedings of 2019 international conference on computational intelligence and knowledge economy. ICCIKE. https://doi.org/10.1109/ICCIKE47802.2019.9004234
32. Choudhury T, Dangi G, Singh TP, Chauhan A, Aggarwal A (2018) An efficient way to detect credit card fraud using machine learning methodologies. In:Proceedings of the 2nd international conference on green computing and internet of things. ICGCIoT. https://doi.org/10.1109/ICGCIoT.2018.8753077
33. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of weiler-atherton algorithm to address loose polygons. J Sci Ind Res 78:771–774
34. Kumar V, Choudhury T (2019) Real-time recognition of malignant skin lesions using ensemble modeling. J Sci Ind Res (JSIR) 78:148–153

Model Analysis of Collaborative-Based Recommender System

Tanuj Pawar, Vaishali Sharma, Parul Kalra, and Deepti Mehrotra

Abstract With the data growing in abundance amount and usage of social media increasing day by day, the features of Youtube videos, video streaming apps, e-commerce sites etc. have been an important concern for the companies. Every user wants the recommendation according to his or her choice. Therefore, it becomes very important for these applications and sites to have an efficient recommender system that not only recommends the result as per the user but also do not overlay excessive burden on the applications. In this paper, we would be analyzing the collaborative-based recommender system algorithms like (Pearson, cosine and matrix factorization) that are used under the different models of collaborative-based recommender systems. Following on the basis of which, we have proposed a generic book recommender system that works under the collaborative-based architecture.

Keywords Collaborative filtering · Recommender system · User-based · Book

1 Introduction

The constantly growing offering of online learning materials to students is making it more difficult to locate specific information from data pools. Personalization systems attempt to reduce this complexity through adaptive e-learning and recommendation systems [1]. Nowadays most e-commerce companies are using recommendation systems to lure buyers to purchase more by offering items that the buyer is likely to prefer [2]. Book recommender systems are used by companies like Kindle, Google Books, Audible etc., to the customers to offer them the desired book they want to read so that they are lured to use the product. Recommender systems utilize user ratings to precisely order the results. For providing accurate results, the system depends upon the number of ratings assigned to each item (as in result). This requires a large group of users to provide ratings to the populace of items. This data is number-crunched by a collaborative filtering engine to associate similar clusters of users and predict the best-next recommendation. It results in an accurate rating; however, the system behavior

T. Pawar · V. Sharma · P. Kalra (✉) · D. Mehrotra
Amity University, Noida, UP, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_7

is static toward learning the latest aggregation after it has reached the concentration level. Collaborative filtering recommender systems evaluate users' ratings in order to give them better recommendations [3]. The data then becomes obsolete and results are less accurate. Artificial intelligence methods developed at the beginning of research are now being applied to information retrieval systems [4]. The recommender systems are basically of three different types: collaborative filtering [5, 6], content-based and hybrid approach [7, 8]. While collaborative filtering is an advanced and proven methodology to predict better recommendations, the system should weigh more on an individual to upscale the accuracy of the existing recommender system. The only disadvantage that lies for collaborative-based recommender systems [9, 10] is that these do not work well for sparse data which means data contains zero values in maximum numbers. But as it is earlier said it all depends on the type of use scenario we are dealing with.

2 Related Works

The use of this recommender system started way back when the internet started to become commercialized, like when the internet started to become a part of life. Ever since the internet websites and applications got commercialized, recommender systems became the key part of the framework. A recommender system tries to suggest the right product to the right person at the right moment. Well, at least the ad-targeting agencies have made their business track and tell if the user is ready for their hundreds of advertisements. The recommendation system's everlasting role is to provide quality. A web-based learning environment has led to the creation of a big number of digital learning materials that offer various topics. Learners spend much time browsing and filtering information that suits their needs. Limited time can prevent learners from finding useful learning materials. One of the most successful techniques to solve the problem is the recommender system which can select items the user is interested in from a large amount of data [11]. Collaborative filtering is considered the most promising method for the recommendation because it recommends items based on the common interests of users [12].

While collaborative filtering is an advanced and proven methodology to predict better recommendations, the system should weigh more on an individual to upscale the accuracy of the existing recommender system. The following report dives into the shortcomings of the existing recommender system: how the new collaborative recommender systems are proving to be a permanent methodology; and how rewarding the recommender system is on each correct or even incorrect recommendation that may help elevate an individual's preferences. The algorithms that are applied under collaborative filtering are Pearson correlation and k-NN-based approach. All these methods are easy to understand and also provide accurate results. The only disadvantage that lies for collaborative-based recommender systems is that these do not work well for sparse data. Model-based algorithms were proposed to tackle a portion of the weaknesses of memory-based strategies [13]. With currently existing systems,

recommendations are based and given through pre-defined rules, and a definitive set of observed patterns is derived through the assessment of user interaction with static content. This provides good recommendations which might not always be relevant, and do not get updated with changing norms and user interests. Other recommendation concepts included are stereotyping, item-centric recommendations and hybrid recommendations [14]. One main goal of the work is to analyze which particular challenges of news recommendation (e.g., short-item lifetimes and recency aspects) have been well explored and which areas still require more work [15]. When this wealth of data is leveraged by recommender systems, the resulting coupling can help address interesting problems related to social engagement, member recruitment and friend recommendations [16]. In this paper, we will analyze the accuracy of collaborative-based recommender systems working on two different models. Following which we will be formulating a generic book recommender system that will be dynamic in nature. Usage of technologies like Django and Python among certain others helps to create a working prototype of a new kind of recommender system. This system is made to provide accurate results to the user and can definitely work very well in the field of recommender systems.

3 Experimental Setup

In this section, we basically will be going through the architecture of a recommender system. In this research, we will be either working on implicit data or a dynamic dataset. While working on an implicit dataset we will need to preprocess the dataset so as to remove all the errors from the dataset. Going on to the next step focuses on user collaborative filtering. It further processes the system by using associate rule mining techniques, and finally, in the result section, it generates the recommendations for the users. Under the two collaborative-based models, we will be analyzing the model-based techniques and memory-based techniques for which accuracies are calculated based on their recommendation approach. Fig 1 lays the framework of the generic recommender system model.

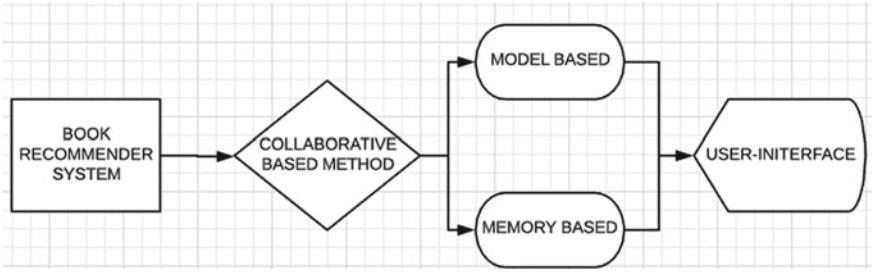


Fig. 1 Outline of book recommender system

title	Correlation
This Year It Will Be Different: And Other Stories	1.000000
Rick Steves' Europe DVD: Germany, the Swiss Alp...	0.748353
Medicina Prehispanica De Mexico/Prehispanic Med...	0.672556
Trigger Points (Quickstudy: Academic)	0.536117
Oncology and Basic Science (Orthopaedic Surgery...	0.377433

title	num of ratings
This Year It Will Be Different: And Other Stories	584
Rick Steves' Europe DVD: Germany, the Swiss Alp...	368
Medicina Prehispanica De Mexico/Prehispanic Med...	507
Trigger Points (Quickstudy: Academic)	420
Oncology and Basic Science (Orthopaedic Surgery...	130

Fig. 2 Results of Pearson correlation with an implicit dataset

The recommender system has been in use now in every field, whether it is social networking websites or any application available in Android or IOS. The growth of a recommender system is tremendous, like every time we visit an Amazon site we see different ads and various types of recommendations. So the answer for where all these things coming from is the recommendation system. But behind these recommendations, there are a great set of algorithms, and machine learning is used to make a system work efficiently. In this project, we have built a book recommender system that is based on various algorithms and machine learning techniques. But before going forward it is very important to understand the workflow of our model so that it is easy to understand each and every step more clearly and precisely. Figure 2 gives us a detailed view of the architecture of this model. Now it's time to understand each step in a more elaborative way so that it is easy to understand. The collaborative-based model basically works on an assumption that it takes for granted that people who liked the item in the past will also like the item in the future. It basically uses mathematical formulas to calculate the closeness between the two items. The collaborative-based model is further divided into model-based and memory-based. The memory-based model basically uses mathematical formulas to calculate the closeness between the two items like cosine similarity and Pearson correlation. On the other hand, the model-based approach uses machine learning algorithms to identify the ratings. Basically, this approach is very much useful when the user has not rated the item at that time this approach comes in handy. In the final step, we have built the user interface of the book recommender system using the k-means algorithms. On the website, the user can see the reviews of a particular book, see different ratings of various books and also can log in to write a review for a particular book. The website can also give suggestions to a particular user after signing into the website. Recommender systems have come a long way from hard-coded systems to

machine learning systems, the times when every user was recommended the same content in the past.

3.1 Collaborative Filtering

Recommendation systems or recommendation engines are used to give suggestions for different types of items. The use of recommendation systems is growing at a tremendous rate. Recommender systems are nowadays used by every other social website or by applications like YouTube, Audible, Kindle etc. It generally takes an input in the form of items and applying different algorithms and machine learning algorithms generates the result. But the items they recommend are basically based on our behavior, which means these sites or applications track our usage behavior and then recommend us the suitable items. But the foremost question arises how these sites collect our usage preference. The answer is simple. They basically with the help of cookies trace our data and our browsing history. Collaborative-based recommender systems have results that are accurate and fast. Also, such type of system is easy to understand. The only disadvantage in these types of systems is that a very large amount of data is needed to recommend the user their accurate choice. Also, it is not much scalable.

3.1.1 Memory-Based

This is the first type of collaborative-based method; it basically finds users based on cosine similarity and Pearson correlation and then takes their average vote ratings. It also generates the results based on two types: user-item filtering and item-item filtering. In user-item filtering, basically, the users rate a particular set of items, and then on the basis of that rating, it finds similar users who had given a similar rating to the user or had a similar kind of history. On the other hand, in item-item filtering, the results are based on the item like the rating given to that item and users who have liked that item. On the basis of that results are generated. Figure 2 shows that the system predicted the top 5 books based on the correlation score, so this basically is the Pearson correlation approach in which it displays the top 5 books similar to the book that has been input by the user. This has been executed with the help of a Python script and here we have provided the output for the user-item-based filtering.

The conclusions for a memory-based model are that no system or machine training is required; no optimization is needed; it is a simple approach and also easy to understand. It is also very useful for those datasets that are implicit, meaning datasets that are already stored as it works on them much faster as compared to the dynamic dataset. But it's not very useful when we have sparse data that actually occurs in real-life world problems, which means it is not so much scalable.

Pearson correlation is a linear correlation between two variables, say (x and y). Figure 3 displays the formula of Pearson correlation formula, which is very important

$$r = \sum(xi - x')(yi - y')/\sqrt{\sum(xi - x')^2 \sum(yi - y')^2}$$

Fig. 3 Pearson correlation formula

Fig. 4 Cosine similarity formula

$$cos\Theta = \vec{a}.\vec{b}/\ ||\vec{a}||\ .\ \left\|\vec{b}\right\|$$

to understand as it is the backbone of the memory-based model. (x) is the values in the first set of data and (y) is the values in the second set of data. Another important similarity for the model-based recommender systems is the cosine-based similarity. Figure 4 displays the cosine-based similarity that we are going to use in the model-based recommender system. Both the formulas, that is the Pearson correlation and cosine-based similarity, are useful for building a model-based recommender system.

3.1.2 Model-Based

The model-based approach also comes under collaborative-based recommender systems or models. This approach comes very much in handy when the data is sparse or the user has not rated a certain number of items. This model basically uses machine learning algorithms to give recommendations. These type of approaches are very useful for datasets which are sparse in nature, which means dataset that has maximum zero values. This model-based approach is further divided into clustering-based, matrix factorization-based and deep learning. All these approaches have various types of algorithms under them. But in this project, we have applied the matrix factorization-based algorithm from which we have applied the SVD algorithm.

Memory-based algorithms explore the user-item rating matrix and make recommendations based on the ratings of item i by a set of users whose rating profiles are most similar to that of user u. Basically, this algorithm is used when the users have not rated the books properly or have not given any ratings. It is very widely used in Netflix or Amazon Prime since there are users who have rated the items, but we want to know what rating they will give to those items which have not been rated by them. In this project, we have taken the sample rating of users and books, and the rating which not has been rated is denoted by zero as shown in Table 1.

The following algorithm is applied in Python language from which the results have been generated (Fig. 5). As we can see that the zero-rating means where the user has not given rating are now filled-with values calculated using the model-based approach. Now according to the result, we will be analyzing the prediction accuracy of matrix factorization.

Table 1 Rating matrix between user and books

Book/User	B1	B2	B3	B4
User1	5	3	0	1
User2	4	0	0	1
User3	1	1	0	5
User4	1	0	0	4
User5	0	1	5	4

```
In [2]: runfile('C:/SPB_Data/.spyder/untitled0.py', wc
[[4.98725844 2.9573206  3.34496042 0.99973323]
 [3.9701758  2.36494963 2.87331574 0.99772604]
 [1.05136568 0.87616207 5.66171166 4.96390124]
 [0.97310036 0.77795934 4.5934374  3.97411742]
 [1.60309814 1.1475449  4.93777393 4.02541083]]
```

Fig. 5 Result of matrix factorization

4 User Interface

This is the final section of this paper where on the basis of the collaborative approach we have created a user interface of our book recommender system. It has been created with the latest technologies like Django Framework, Python and SQLite. The user can create their login, rate a book and then get recommendations on the basis of that. It is pretty much dynamic in nature and interacts with the user dynamically. Figure 6 shows a glimpse of our generic book recommender system.

For code go check: https://github.com/Tanujp7/Book_Recommender_System.git.

5 Results

As per the analysis, both the (Pearson and cosine) algorithms are applied to the dataset and their accuracy is calculated as highlighted in Fig. 7. Therefore, as per the graph, both the Pearson and cosine algorithms have almost the same similarity index. Pearson comes into handy when we want to analyze the similarity of how correlated are two persons among the commonly selected data item. On the other hand, cosine similarity is useful when we want to analyze how two persons have rated a particular item and have the similarity between themselves. In a similar manner, we have also analyzed the accuracy for the results of matrix factorization, which is highlighted in Fig. 8. Here, we can see how proximately the matrix factorization has predicted the ratings for the users for which there were zero ratings provided. Hence, it proves how

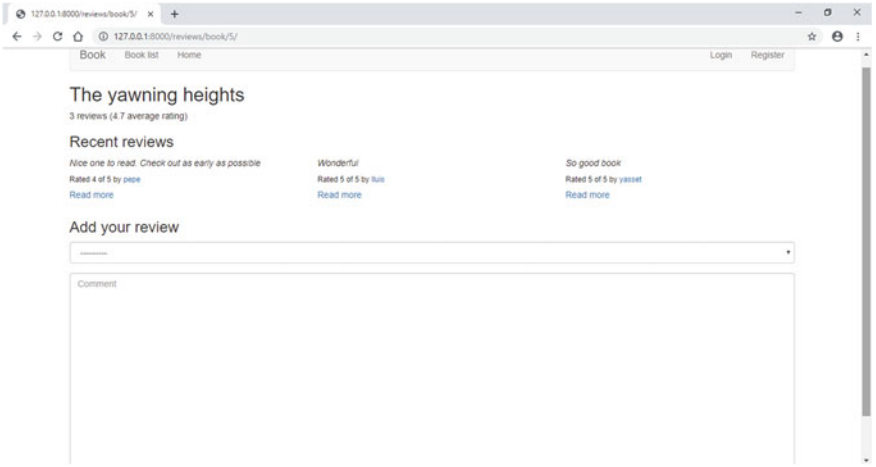


Fig. 6 User interface page

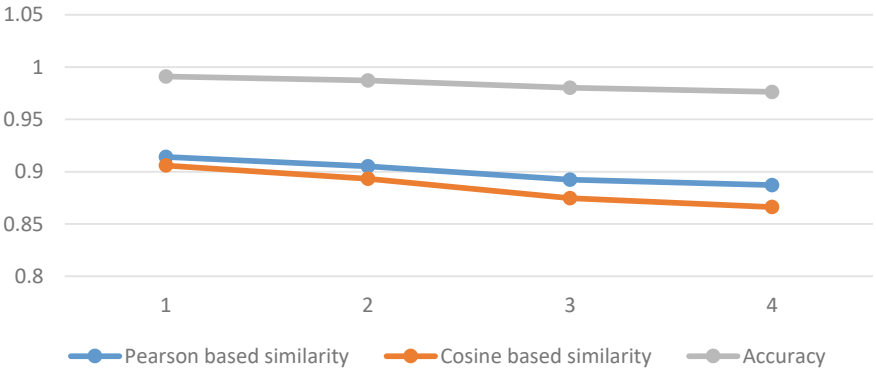


Fig. 7 Accuracy of Pearson and cosine-based similarity

dynamic it is and adapts to the actual environment with accuracy. Concluding this, we have proposed a generic-based recommender system that works on the algorithm of k-means clustering.

6 Conclusion

The usage of recommender systems has increased drastically like it has become an important asset for every business. It plays a pivot role in getting the needs of the user and hence improving the usage dynamically. This paper focuses to analyze the accuracy of the models of the collaborative-based filtering recommender systems.

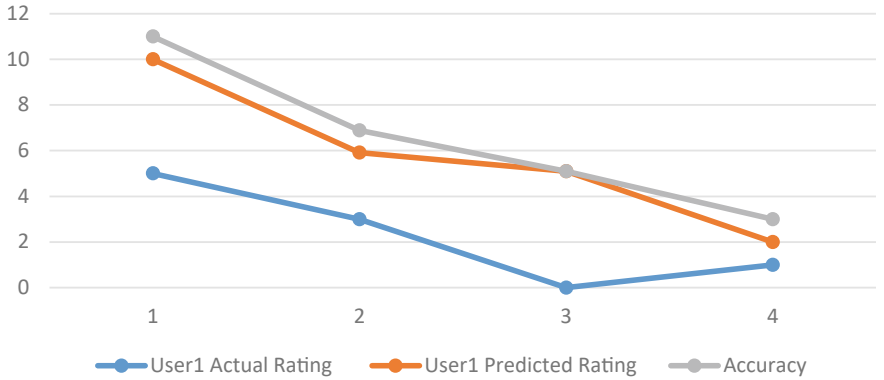


Fig. 8 Accuracy of matrix factorization

Finally, a user interface system is presented, which works on the architecture of collaborative filtering. Also, the user interface that has been proposed interacts with the user dynamically.

References

1. Khanal SS, Prasad P, Alsadoon A et al (2020) A systematic review: machine learning based recommendation systems for e-learning. Educ Inf Technol 25:2635–2664. https://doi.org/10.1007/s10639-019-10063-9
2. Rana A, Deeba K (2019) Online book recommendation system using collaborative filtering (With Jaccard Similarity). J Phys: Conf Ser 1362(1). IOP Publishing
3. Al Hassanieh L et al (2018) Similarity measures for collaborative filtering recommender systems. In: 2018 IEEE Middle East and North Africa communications conference (MENACOMM). IEEE
4. Ibrahim ME et al (2018) Ontology-based personalized course recommendation framework. IEEE Access 7(2018):5180–5199
5. Verma A, Shukla A, Choudhury T, Chauhan A (2020) Recommendation system for prediction of tumour in cells using machine learning approach. In: Communications in computer and information science, vol 1229. CCIS. https://doi.org/10.1007/978-981-15-5827-6_18
6. Rohit SS, Choudhury T (2018) Proposed approach for book recommendation based on user k-NN. Adv Intell Syst Comput 554. https://doi.org/10.1007/978-981-10-3773-3_53
7. Jain KN, Kumar V, Kumar P, Choudhury T (2018) Movie recommendation system: hybrid information filtering system. In: Intelligent Computing and Information and Communication. Springer, Singapore, pp 677–686
8. Bhasin S, Choudhury T, Gupta SC, Kumar P (2017) Smart city implementation model based on IoT. In: Proceedings of the 2017 international conference on big data analytics and computational intelligence. ICBDACI. https://doi.org/10.1109/ICBDACI.2017.8070836
9. Singh N, Sharma T, Thakral A, Choudhury T (2018) Detection of fake profile in online social networks using machine learning. In: Proceedings on 2018 international conference on advances in computing and communication engineering. ICACCE. https://doi.org/10.1109/ICACCE.2018.8441713

10. Gupta SS, Thakral A, Choudhury T (2018) Social media security: analysis of threats and security measures. In: Proceedings on 2018 international conference on advances in computing and communication engineering. ICACCE. https://doi.org/10.1109/ICACCE.2018.8441710
11. Turnip R, Nurjanah D, Kusumo DS (2017) Hybrid recommender system for learning material using content-based filtering and collaborative filtering with good learners' rating. In: 2017 IEEE conference on e-learning, e-management and e-services (IC3e). IEEE
12. Saleem F et al (2019) Using Trust in collaborative filtering for recommendations. In: 2019 IEEE 28th international conference on enabling technologies: infrastructure for collaborative enterprises (WETICE). IEEE
13. Adewole AP, Victor Kawedo C (2017) A model-based collaborative filtering with dimensionality reduction
14. Beel J et al (2016) Paper recommender systems: a literature survey. Int J Digital Lib 17(4):305–338
15. Karimi M, Jannach D, Jugovac M (2018) News recommender systems–Survey and roads ahead. Inf Process Manage 54(6):1203–1227
16. Eirinaki M et al (2018) Recommender systems for large-scale social networks: a review of challenges and solutions, pp 413–418

Low-Cost Obstacle Detection and Avoidance Robots with Swarm Intelligence

Abhijit Das, Divesh Pandey, Aman Sharma, Nitish Jha, Anurag Pandey, and Uma Gautam

Abstract In robotics, it is very important to adapt to the real-time variations taking place in the surrounding for executing a task. It is not only required for the robots designed to be robust but also they must be easy to implement, power-efficient and affordable for wider adaptability and application. Such robots can be used with ease by anybody and anywhere for detecting obstacles where it is difficult for a person to reach. Swarm robots can have a wide variety of uses in agriculture, surveillance, mining, etc. due to their cheap manufacturing. Swarm intelligence can be made use of in the case of multi-robotic practices where coordination among a number of mobile robots is required which perform tasks autonomously without human intervention. In recent years, it has been observed that multi-robotic systems are substituting single robots because of the factors that multi-robotic systems are very time-efficient and have become much more affordable than before. This paper delivers a framework of swarm intelligence implemented into robot clusters for detecting obstacles and avoiding them in a master–slave configuration. Swarm robots are based on a progression strategy that relates to social insects. Swarm intelligence-based robots are a rapidly emerging field. This work highlights the approach of making swarm robotic systems easy to implement, economical and power-efficient, making them a possible objective for further advancement.

Keywords Swarm intelligence · Mobile robots · Multi-robot systems · Master–slave · Wireless communication

1 Introduction

Today, it has become common for engineers working in different fields, varying from material metallurgy to robotics, to take concepts from the environment [1]. Robotics is an area of advancement which surrounds the research, planning and execution of robots, while a robot can be described as an encoding device able to accomplish certain listed duties or procedures which are being allotted by the operator

A. Das (✉) · D. Pandey · A. Sharma · N. Jha · A. Pandey · U. Gautam
Department of Mechanical & Automation, HMRITM, New Delhi, Delhi 110036, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_8

or otherwise the programmer. Swarm robotics can be described as making the use of robots with swarm intelligence to avoid obstacles and complete tasks swiftly. A swarm is a big cluster of people or insects, specifically ants, honey bees etc. Swarm intelligence can be described as AI-based concept comprising cooperative activities of distributed and auto-controlled structures. Swarm intelligence is closely related to swarm robotics and they both happen to be initially inspired by the actions of a group of ants.

Swarm robotics is based on the basic principle of natural inspiration and their prominence is on the dispersed local control and communication among huge numbers of miniature robots [2].

In today's age of modernization, movable robots are seen as one of the main attentions of many researchers. There is a massive field in which robotics has been applied, like automation education, agriculture, medical etc. The major intent of our task is to evade an obstruction, which totally depends upon the input contributed through master bots. The implementation of natural swarm intelligence has applied to both master and slave robots. Whenever the master robot confronts any obstruction then the data of obstruction has been shared with slave robots, and the infrared sensors are used which range from inches to some feet. Any obstruction in the pathway has been sensed by IR sensors and their signal is sent to the controller, established upon the data from the sensor clear path is guided by the controller. Similar data is shared with the slave robot so that it gets through the obstruction as the path followed by the slave robot will be similar to the path of the master robot. This information sharing among the slave robots assists in the reduction of time and as well the hard work needed by slave robots for evading the obstruction; ultimately results in reducing cost and time. Protection of the surrounding is significant for the developers in all areas. The recent report shows the flock of robots used in the place of fireworks for celebration in future Olympic events. It plays a very important role in reducing air pollution, and we can also attain entertainment purposes with it. Else ways, the workload can easily be divided between a flock of small robots instead of a single robot; efficiency and accuracy are also improved. Former studies continue more on the concentration of the practice of trouble-free active robots that completely communicate remotely toward the surrounding. Additionally, in this era, swarm robotics study has advanced far-off from those boundaries to inquire the necessity of insensible robots which can coordinate openly and can also teach them to execute a definite function and to symbolize their surroundings. The popularity of mobile robots is increasing gradually in this era. Connecting robotics with bio-inspired management and grouping to attain insect-like actions is a new research area.

Everybody recognizes robots and their applications today in different areas. There are a lot of different uses regarding swarm robotics, for example, pulling or pushing numerous objects, farming, foraging, mining, cooperative tracking, inspection and saving operation in a tragedy such as property collapse, fire hazards etc. A systematic explanation for these operations is given. First, to make synchronization better between the robots, we use Arduino to build communication. Secondly, to improve our end-result, we are going to carry out a test of obstacle avoidance for server robots via Arduino and IR sensors and finally apply the swarm intelligence (SI) in obstacle avoidance with the help of a wireless communication system.

The primary robot is supposed to be equipped with better sensing as well as actuating devices having exceptional ability to control computing functionality and control. Along with that, it happens to be capable of giving instructions throughout the system while the secondary robots can only trail those instructions because of their restricted sensing capability. That's how, a master–slave flock testing can be produced with an aim, e.g., collision avoidance or line following etc. Our major aim comprises building a dispersed radiocommunication control arrangement in which master–slave coordination techniques be made use of for performing superior robot grouping processes. This study is from the very beginning in the direction of attaining the ultimate swarm-oriented trail succeeding to movable robots consisting of wheels. We offer a structure that comprises mobile robots for chasing the path along with maintaining their arrangement as well as evading hindrances. An easy way is established while supposing ideal communication among movable bots for cooperatively executing the required task which can be controlled depending upon the situation. To exchange information securely, every robot transfers information inside excellent scope for attaining synchronization. The cluster of robots executes a certain duty while rearranging themselves corresponding to their definite aim. This research paper has been planned as follows: Section 2 describes all related discoveries. Section 3 reviews the swarm behavior inspired by nature. Section 4 describes the implementation of the scheme and flowchart, with Sect. 5 emphasizing the result.

2 Related Works

Couceiro et al. [3] established the concept of "A low cost education platform for Swarm Robots" for preserving inexpensive design for robots, which stands for Swarm of economic robots exclusively targeting technical learning along with swarm of robots. Jevtic et al. [4] established the concept of "Swarm Intelligence and its applications in Swarm Robotics" that can presumably be an exciting substitute for traditional access to robotics due to their problem troubleshooting ability. Lee et al. [5] established the concept of "Improving energy efficiency in cooperative foraging swarm robots using behavioral model" that uses behavior representation on the subject of task distribution and search space distribution for developing energy effectiveness. AriSantoso et al. [6] established "A New Obstacle Avoidance Method for

Service Robots in Indoor Environments" in which procedure for avoiding obstacles is followed for assistance robots in internal surroundings while making use of ultrasonic and optical sensors. Hedjar et al. [7] recommended a concept for "Real Time Obstacle Avoidance for Swarm of Autonomous Mobile Robots" providing a conflict-resistant trial preparation by means of indefinite optimizations merged with a controller doing one-step-ahead projections. Ankit et al. [8] established a concept for "Obstacle Avoidance Robotic Vehicle Using Ultrasonic Sensor, Android And Bluetooth For Obstacle Detection". In this, several mechanisms are used to sense an obstruction and evade it. Nurmaini et al. [9] introduced an "Intelligent Low Cost Mobile Robot and Environmental Classification", a self-governing moving robot comprising FKN is planned that is formed for an inexpensive application platform.

Self-governing moving bots comprise the ability of collaborating among further many bots for improving functionality during co-operative execution of different tasks. Besides, on the basis of job characteristics, a distinct club of moving bots which comprise diverse observing and executing abilities are appointed to complete the task. For making it easy, the procedure of synchronism in the supportive action is deserted which has to be accomplished via communication structure otherwise. The sample product arrangement is not only concerned with the implementation stage but is also important in the development stage. The moving robots which are in the implementation stage achieve distinct responsibilities by using shared communication. The stationary robotic agents offer support to moving bots for achieving the objective. While primary formation, every robotic agent provides assistance to the other companion robots for the mutual achievement of a general task. In other formation, each agent robot executes their tasks by determining the problem between all agents to perform their tasks.

Agent robots are classified into two types: one is "Coordinator" while the other is "Cooperator". Cooperative robots which work as a coordinator are responsible for identifying the trouble that isn't achievable for a single robot to implement. Coordinator uses several agent robots for performing tasks co-operatively by compromising with another robotic agent during communication. The work of the coordinator is to determine the helping agent robots from the candidate agent robots which are responsible for executing tasks assigned to cooperators. The task that is executed by the coordinator is the development of path design so as to implement the mutual work of all cooperators through data exchange, in which the path design then gets distributed among all cooperators for executing their role considerately. If in any case the coordinator fails and is not capable of accomplishing the preferred objective, then the coordinator transfers the function to any other robotic agent which then performs a similar function. Communication is a very important aspect of this process.

The routing codes are formed to remove numerous problems linked to the lifespan of battery, for example, sustainability, survivability, flexibility, adaptive character and much more [10]. During the formation of sensor communication channels so as to interact wirelessly, for researchers, many factors exhibit different meanings comprising the fixed energy of scalability, latency, network lifespan, nodes, etc.

3 Swarm Behavior Inspired from Nature

During the process of allotting actions of dispersing over to a category of comparable progressing robots, there is a sequence of complications that are of utmost value, which are recorded/registered. First, the collection of the symmetrical structure of progressing robots in the set. Secondly, alignment and positioning of every single robot while showing commitment to assigned universal objective. Thirdly, interaction amongst robots when they are positioned in a cluster. Fourthly movement perception in the cluster composition and the last is sustaining stability in a cluster while in motion control.

During the past decade, a number of researchers have been attracted toward remote sensor techniques. Specialist professionals are mainly focused on one of the principal issues that are taking into account remote sensor system so as to build a communication algorithm for interacting wirelessly. Revolutionizing these communication algorithms can cause troubles such as—difficulty, survivability, flexibility, and elasticity and battery life in wireless transmission techniques. Path planning algorithms allotted for wireless methods are generated by means of a precise job of solving the problems. For performing constructive interaction amongst mobile robots, controlling protocols are selected where retaining a number of essential emphases in our mind. These are restricted power utilization when transferring data between nodes, to decrease the information damage amongst nodes caused by too much traffic. Data can be assembled in packets. Latency time is defined as the total time required for a bundle of data to transverse from one to another node. Present swarm-based approaches can be classified into three categories.

3.1 Ant-Based Approach

The way in which ants work influences the optimization of the ant colony, which further explains performing difficult tasks together by a cluster of ants, for instance, making a nest and searching. In this approach, programs are created on the basis of individual evaluations which are executed with some certain final objective to fulfill some of the conditions which can be delayed, energy level and velocity. These aspects are perfectly planned with paths being found out. Hence, this designed method has numerous ways to avoid being overcrowded, producing an increase in the data transfer speed and reducing the data damage.

3.2 Bee-Based Approach

There are approaches inspired by the honey bee searching practices. The honey bee has a small number of resemblances with routing in a computer system. This method is equipped with three kinds of honey bee experts; namely spy bees, collector bees and hunter bees. For taking the data packet to a certain node, collector honey bee searches for apt hunter honey bees for that; other than that, for searching the path to other targets, spy honey bees are responsible. There is a substantial amount of data bundles for hunter honey bees to convey to the sink node.

3.3 Slim-Based Approach

Heterotrophic living organisms use a slim mode approach being considered unicellular and are a parasite. There are concrete resemblances amongst such parasitic organisms, ant's colony construction and wireless communication which can be observed easily. Fault tolerance, organizing themselves and flexibility with the surroundings are prerequisites for wireless communication which resembles the working objectives characterized by the clusters of ants. The standard is invigorated by slime-based organisms. Making use of the vector-valued algorithms to coordinate and control information stream by using a technique to overcome information load when sent to a number of sink nodes which depicts fault tolerance along with autonomy.

4 Implementation Scheme and Flowchart

To acknowledge any plan or an idea, implementation plays a major responsibility. The execution of swarm robots is indistinguishable from general robots, apart from the decrease of overall size, time and expense. This research paper is established on the synchronized movement of movable robots, which comprises twice stages: one of them is assembling of the hardware and the next is encoding via Arduino IDE.

The fabrication of hardware includes building chassis, mounting sensors accordingly and utilization of the power source. Our utmost goal is to establish a short-radius transmission system among the master and its slave so as to attain the required results. The master robot drives the slave robot to execute essential activities. Two methods are executed through communication and coordination amongst mobile robots. First method detection of objects and then transferring the protocol of avoidance. Prior to the execution stage, inspecting for the availability of various modules is the primary step. Next is the working status of every module in the perspective of their correct functioning and as per our need is the following stage, for attaining required results at the termination.

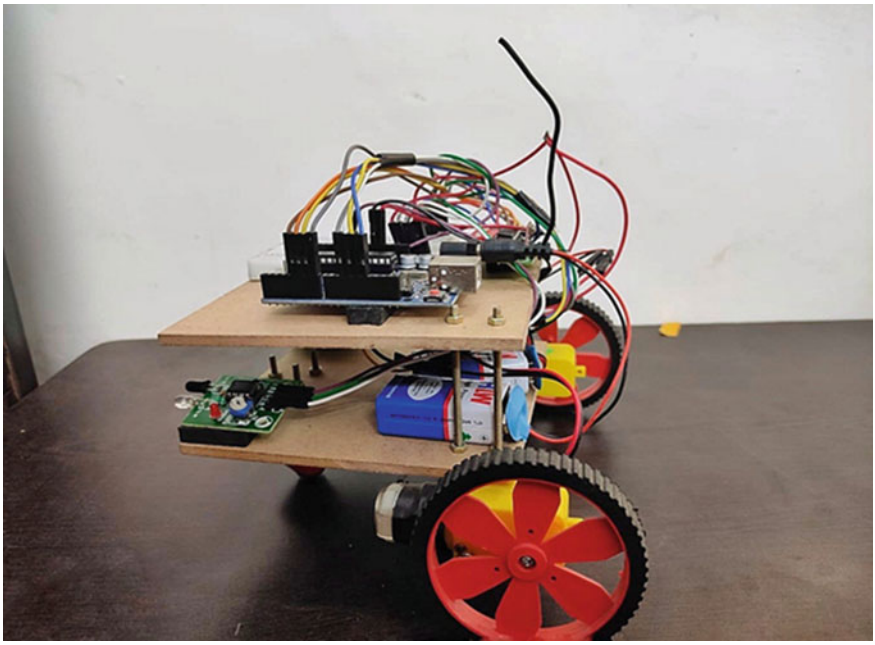


Fig. 1 Master robot

The utmost essential and important process is to carry out the required job effectively by establishing interaction among various robot units. In this we are using two robots, one being the master robot and the other one is the slave robot. It can be seen in Fig. 1, infrared sensors are mounted on the master robot along with Arduino UNO, RF module (433 MHz), Motor driver IC (L293D). For moving the robot, motors are connected with the motor driver which provides the suitable voltage required for the correct functioning of 9v geared motor. Robot receives input from the eyes which are the front-mounted infrared sensors being used by the brain, i.e., Arduino UNO, of our project for making decisions. For receiving and transferring data among the robots, Arduino makes it possible. The slave robot uses the data directly from the master robot making it less of a need to acquire the infrared sensors by the slave robot other than which it consists of all the similar components.

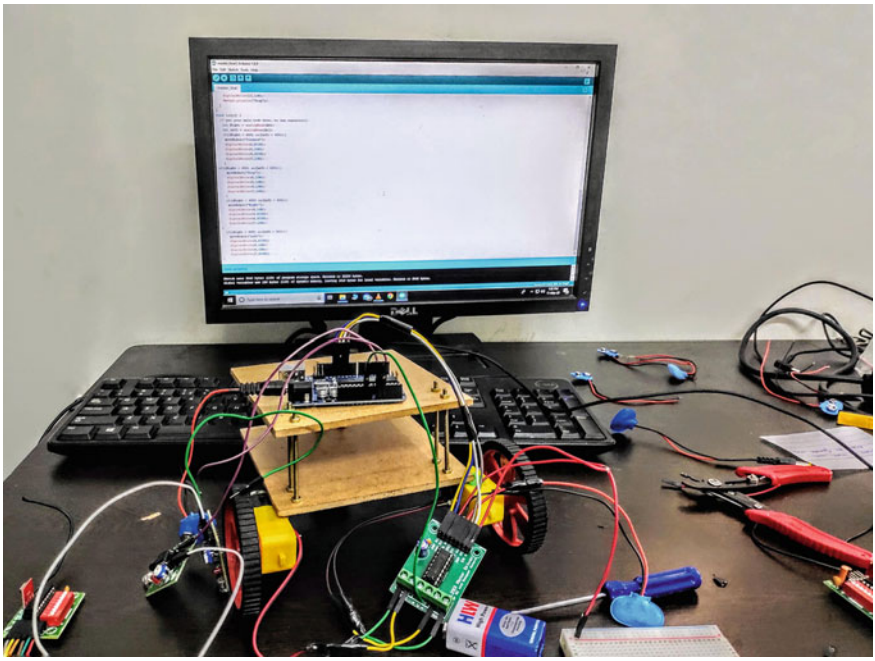


Fig. 2 Swarm robot interfacing

Figure 2 shows the interfacing and programming of the master robot. Essentially, the master robot moves in an area consisting of an obstacle, where obstacles are avoided by the master robot with the help of the input from infrared sensors and the Arduino UNO of the master robot and then transfers the information of the obstacle to another robot. The transferred information makes the slave units re-define their paths and trail the lead of the master to reach their destination. Electric charge is converted into torque with the help of DC motors connected to the motor driver which makes the robot move. For better torque output for movement of the robot chassis, we use DC geared motor which reduces wheel rotation in return. The motor is directed mounted with the gearbox to it. For robotic application, we are using L293D motor driver circuit that is ideal to drive and operate both the DC motors independently effortlessly toping at a peak current of 2A.

A microcontroller integrated development environment is a prerequisite for executing the program. Arduino UNO is an Atmega328 microcontroller that executes the task of establishing a network among master and slave robots as well as controlling the locomotion of the robots as a whole. For transmitting and receiving data, the radio frequency transmitter and receiver are bonded with the encoder and the decoder, respectively, which are interfaced to the Arduino's of the respective master and slave robots. When an obstacle comes in front of the master robot, the infrared sensor gets the reflection to the photodiode from the surface of the obstacle. The robot takes its decision and sends it to the slave. The slave acts upon the received data immediately saving time, power, etc. For this to work, the components are interfaced as given in Fig. 3 followed by the code fed into the Arduino UNO via an IDE.

Code

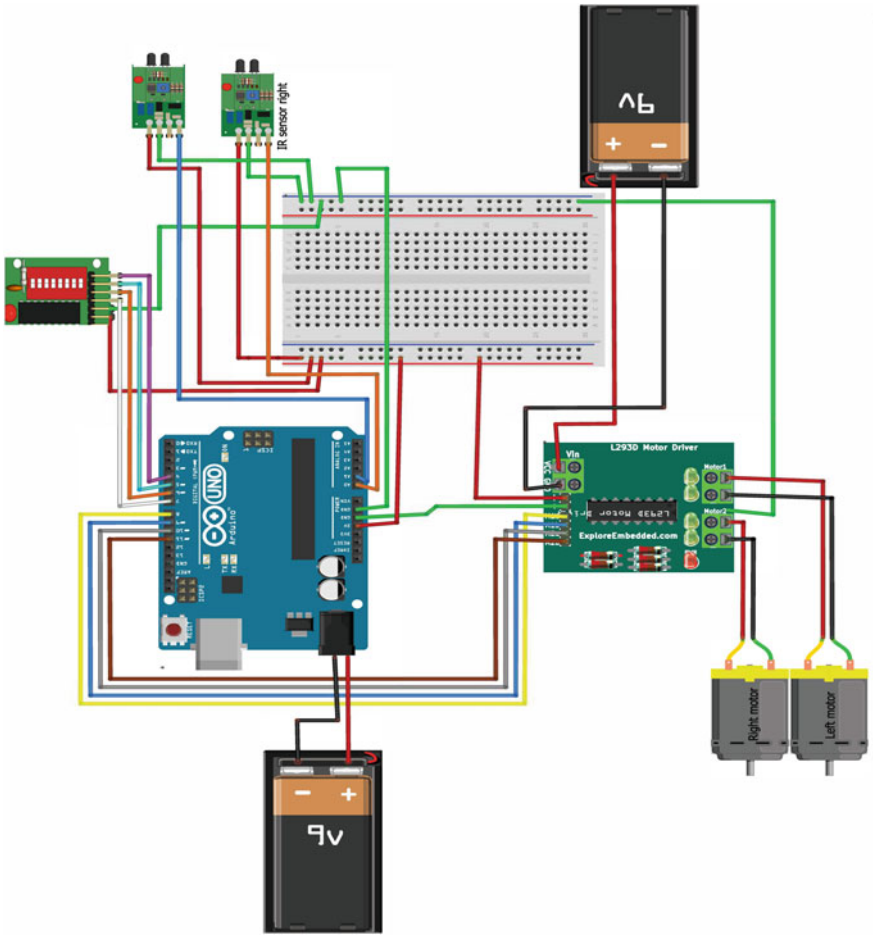


Fig. 3 Connection diagram

```
#define B1 4
#define B2 5
#define B3 6
#define B4 7
int count=0;
void setup() {
  // put your setup code here, to run once:
Serial.begin(9600);//Serial Communication begins.
pinMode(10,OUTPUT);
pinMode(11,OUTPUT);
pinMode(12,OUTPUT);
pinMode(13,OUTPUT);
//For RF Communication with Slave Bot.
pinMode(B1,OUTPUT);
pinMode(B2,OUTPUT);
pinMode(B3,OUTPUT);
pinMode(B4,OUTPUT);
//Analog Pins for reading data from IR sensors.
pinMode(A0,INPUT);
pinMode(A1,INPUT);
}
//Defining functions for movement of Master Bot.
void moveRobot(String motion){
  //For Moving Forward.
  if(motion == "Forward"){
    digitalWrite(10,HIGH);
    digitalWrite(11,LOW);
    digitalWrite(12,HIGH);
    digitalWrite(13,LOW);
    Serial.println("Forward");
  }
  //For Moving Backward.
  if(motion == "Backward"){
    digitalWrite(10,LOW);
    digitalWrite(11,HIGH);
    digitalWrite(12,LOW);
    digitalWrite(13,HIGH);
    Serial.println("Backward");
  }
  //For Turning Left.
  if(motion == "Left"){
    digitalWrite(10,HIGH);
    digitalWrite(11,LOW);
    digitalWrite(12,LOW);
```

```
    digitalWrite(13,HIGH);
    Serial.println("Left");
   }
   //For Turning Right.
   if(motion == "Right"){
    digitalWrite(10,LOW);
    digitalWrite(11,HIGH);
    digitalWrite(12,HIGH);
    digitalWrite(13,LOW);
    Serial.println("Right");
   }
   //To Stop.
    if(motion == "Stop"){
    digitalWrite(10,LOW);
    digitalWrite(11,LOW);
    digitalWrite(12,LOW);
    digitalWrite(13,LOW);
    Serial.println("Stop");
   }
}


void loop() {
  // put your main code here, to run repeatedly:
   /*Storing value of Pin A0 in Right connected to IR 1.
    */
   int Right = analogRead(A0);
   /*Storing value of Pin A1 in Left connected to IR 2.
    */
   int Left = analogRead(A1);
   //Choosing 600 as Threshold Value given by IR.
   if((Right < 600) &&(Left < 600)){
    //No Obstacle detected by sensors.
    moveRobot("Forward");
    //Transmitting data to Slave Bot.
    digitalWrite(4,HIGH);
    digitalWrite(5,LOW);
    digitalWrite(6,HIGH);
    digitalWrite(7,LOW);
   }
   if((Right > 600) &&(Left > 600)){
    //Obstacle detected by Both sensors.
    if(count<5){
    moveRobot("Backward");
    digitalWrite(4,LOW);
```

```
    digitalWrite(5,HIGH);
    digitalWrite(6,LOW);
    digitalWrite(7,HIGH);
    delay(1500);
    moveRobot("Right");
    digitalWrite(4,LOW);
    digitalWrite(5,HIGH);
    digitalWrite(6,HIGH);
    digitalWrite(7,LOW);
    delay(1500);
    moveRobot("Forward");
    digitalWrite(4,HIGH);
    digitalWrite(5,LOW);
    digitalWrite(6,HIGH);
    digitalWrite(7,LOW);
    delay(2500);
    moveRobot("Left");
    digitalWrite(4,HIGH);
    digitalWrite(5,LOW);
    digitalWrite(6,LOW);
    digitalWrite(7,HIGH);
    delay(1500);
    count++;/*Using this value to stop robot if there is no way for Master Bot to move
any further ahead.
            */
    Serial.println(count);
    }
    else{
    moveRobot("Stop");
    digitalWrite(4,LOW);
    digitalWrite(5,LOW);
    digitalWrite(6,LOW);
    digitalWrite(7,LOW);
    }
   }
   if((Right > 600) &&(Left < 600)){
    //Obstacle detected by Left sensor.
    moveRobot("Right");
    digitalWrite(4,LOW);
    digitalWrite(5,HIGH);
    digitalWrite(6,HIGH);
    digitalWrite(7,LOW);
    delay(1500);
    moveRobot("Forward");
    digitalWrite(4,HIGH);
```

```
    digitalWrite(5,LOW);
    digitalWrite(6,HIGH);
    digitalWrite(7,LOW);
    delay(2500);
    moveRobot("Left");
    digitalWrite(4,HIGH);
    digitalWrite(5,LOW);
    digitalWrite(6,LOW);
    digitalWrite(7,HIGH);
    delay(1500);
   }
   if((Right < 600) &&(Left > 600)){
    //Obstacle detected by Right sensor.
    moveRobot("Left");
    digitalWrite(4,HIGH);
    digitalWrite(5,LOW);
    digitalWrite(6,LOW);
    digitalWrite(7,HIGH);
    delay(1500);
    moveRobot("Forward");
    digitalWrite(4,HIGH);
    digitalWrite(5,LOW);
    digitalWrite(6,HIGH);
    digitalWrite(7,LOW);
    delay(2500);
    moveRobot("Right");
    digitalWrite(4,LOW);
    digitalWrite(5,HIGH);
    digitalWrite(6,HIGH);
    digitalWrite(7,LOW);
    delay(1500);
   }
  }
```

5 Results

The foremost challenge in small robots is organizing an economical energy consumption so as to make the most use of a battery for a longer period of time. Our system minimizes energy consumption with respect to other existing systems by using low-power data transmission components.

For most cases, high processing power is essential for swarm robots. Our suggested structure makes use of an Arduino UNO microcontroller that is capable of doing every

Table 1 Cost distribution for master and slave robots

List of components:	Cost
Arduino UNO—2 Nos	1000
RF transmitter and encoder module—1 No	600
L239D motor driver circuit—2 Nos	450
DC motors—4 Nos	400
Wheels for motors—4 Nos	100
Castor wheels—2 Nos	40
Chassis board—4 Nos	100
Wire stripper	50
M3 nuts and bolts	150
Breadboard	100
Connecting wires and miscellaneous	150
Total	3140

essential task which is required from detection and locomotion to operation and data transmission among master and slave robots.

The total cost of the entire assembly is minimized by shortlisting the compulsory components which can perform the required tasks. Cost is further reduced by mounting sensors only on the master robot but no compromises are made with the overall functioning of the robots. This increases the efficiency of the system to work for longer periods and the batteries will last longer. Hence it can be seen that energy, cost and computational power, all have reduced significantly. Upgrades will be highly economical and repairs can be made easily due to less complexity.

Table 1 shows the cost distribution of these low-cost functional robots. These components are readily available in the market.

6 Conclusion and Future

Technology is wide-spreading very swiftly, with technological advancements taking place every day around the world. Without evolving with the growing technology, one will definitely be left behind. A situation has come where it has become impossible to survive without technology. The environment has always been an inspiration that plays a vital role in exploring and inventing the latest technologies. This paper also inspires form a natural phenomenon that is replicating the activities of insects with the help of swarm robots which is quite similar to the implementation of tasks individually in a cluster collectively so as to complete a bigger objective as in the case of a group of ants. The robots make use of the algorithms programmed into them for detecting the obstacle and avoiding it. The master robot implements the operation and further transmits the data to the slave robot wirelessly so as to make the slave robot aware of the position of the obstacle and hence avoiding the obstacle instantly

without delay. It becomes very difficult for an individual to complete a complex task on one's own than being distributed among a group of people who can do it with greater efficiency. So, swarm robotics has a long way to go and will play a major role in the scenarios to come.

Future work will include mapping robot's position in real time in an economical and power-efficient way so as to continue the aim of making robots more affordable for people around. A deeper dive would be taken into making the robots autonomous, and improvement of the overall robot function making it much more result-oriented.

References

1. Herbert-Read JE (2016) Understanding how animal groups achieve coordinated movement. J Exp Biol 219:2971–2983
2. Trenchard H, Perc M (2016) Energy saving mechanisms, collective behavior and the variation range hypothesis in biological systems: a review. BioSystems 147:40–66
3. Couceiro MS, Figueiredo CM, Luz JMA, Ferreira NM (2012) A low-cost educational platform for swarm robotics. Int J Robot Edu Art (IJREA) 2(1)
4. Jevtic A, Andina D, Swarm intelligence and its applications in swarm robotics. In: 6th WSEAS International conference on computational intelligence, man-machine systems and cybernetics
5. Lee JH, Ahn CW (2013) Improving energy efficiency in cooperative foraging swarm robots using behavioral model. In: Sixth international conference on bio-inspired computing: theories and applications, IEEE (2011), pp 39–44; Vaishak NL, Shilpa B (2013) A review of swarm robotics: a different approach to service robot. Int J Sci Eng Technol Res (IJSETR) 2(8)
6. Budiharto W, Santoso A, Purwanto D, Jazidie A (2015) A new obstacle avoidance method for service robots in indoor environments. ITB J 44(2):148–167
7. Hedjar R, Bounkhel M (2014) Suggested the idea of real time obstacle avoidance for swarm of autonomous mobile robots. Int J Adv Robot Syst
8. Ankit V, Jigar P, Savan V (2016) Introduced the idea of obstacle Avoidance robotic vehicle using ultrasonic sensor, android and bluetooth for obstacle detection. Int Res J Eng Technol (IRJET) 3(2)
9. Nurmaini S (2011) Intelligent low-cost mobile robot and environmental classification. Int J Comput Appl (0975– 8887) 35(12)
10. Zengin A, Tuncel S (2010) A survey on swarm intelligence-based routing protocols in wireless sensor networks. Int J Phys Sci 5:2118–2126
11. Desai JP, Ostrowski JP, Kumar V (2001) Modeling and control of formations of nonholonomic mobile robots. IEEE Trans Robot Autom 17:905–908
12. Naveed K, Khan ZH, Hussain A (2014) Adaptive trajectory tracking of wheeled mobile robot with uncertain parameters. In: Computational intelligence for decision support in cyber-physical systems. Springer, Berlin, pp 237–262
13. Naveed K, Khan ZH (2017) Adaptive path tracking control design for a wheeled mobile robot. In: 3rd international conference on control science and systems engineering (ICCSSE 2017). Beijing, China
14. Lubbers S (2014) Behaviour influencing robots. Circuit Cellar 18–29

Quality Evaluation of Fitness Centres Using Gap Analysis

Pushpendra Neniwal and Vikas Kapoor

Abstract The number of gymnasiums is increasing day by day in India due to the laziness of people and advancements in automation industries. Indian Gyms are expanding their facilities in terms of the increasing number of equipments, improved quality of service (QoS), customer satisfaction etc. The expected and perceived QoS has been compared in this paper. The gap analysis has been performed using the SERVQUAL questionnaire in terms of 'Responsiveness' as well as 'Tangibility'. It means 'Actual Service Performance', 'Promptness', 'Always be willing to help' and 'Respond to the Customers' got to be focused upon.

Keywords SERVQUAL · Indian Fitness Centre · Analysis

1 Introduction

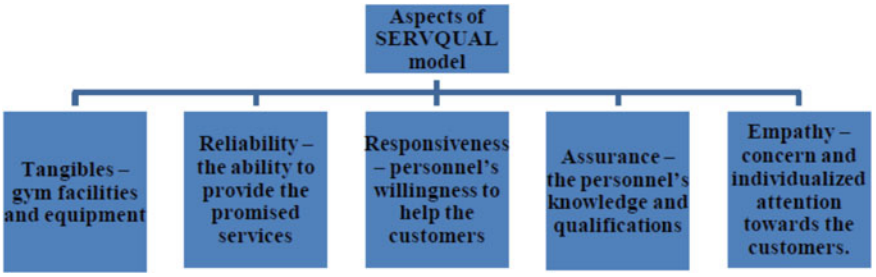
The Indian service sectors are surrounded by different industrial services, viz., hospitality, trade, construction, healthcare etc. Also, the service sectors have played a very important role in raising the Indian economy [1]. Most of the highly qualified Indian workers are working at a very low salary. Therefore, quality-based service sectors are easily available in India [2]. Since IT industries demand long sitting hours, extensive desktop working and a sedentary lifestyle, the massive population of ITES professionals are drawn towards gyms, health and fitness centres [3, 20, 21].

2 SERVQUAL in Fitness Industry

The most important role in various strategies available in the market has been played by QoS evaluation [4]. To estimate the QoS, the SERVQUAL model has been used by different users across the globe [5]. The five aspects of the model are shown as follows.

P. Neniwal (✉) · V. Kapoor
Department of Production and Industrial Engineering, MBM Engineering College, Jodhpur, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_9



3 Instruments of the Study

We have used a questionnaire as a tool for data collection. SERVQUAL instrument has been used to prepare a questionnaire to measure the satisfaction of the customer. The questionnaire has been distributed among 50 members.

4 Dimensional Analysis with Gap Score

The gap analysis [17,18,19] has been used to identify the difference between Perception and Expectation (P–E) [6]. This difference has been obtained with the help of SERVQUAL with 22 attributes and 50 samples. If the difference is found negative, then it clearly states that there is a shortfall in the quality of service.

a. **Tangibles at the service centres**
Figure 1 shows the difference between Perception and Expectation (P–E) for the tangibles dimension.

b. **Reliability level**
Figure 2 shows the difference between Perception and Expectation (P–E) for the reliability dimension.

c. **Responsiveness level**
Figure 3 shows the difference between Perception and Expectation (P–E) for the different levels of responsiveness.

d. **Level of assurance**
Figure 4 shows the difference between Perception and Expectation (P–E) for the different levels of assurance with the largest manifesting in 'Courteous with Customers' (−1.18) and the smallest being 'Feel safe in Transaction' (−1.14) which must be generally addressed (Fig. 5).

e. **Empathy levels**
Figure 4 shows the difference between Perception and Expectation (P–E) for the different levels of empathy. The largest gap was 'Understand the specific

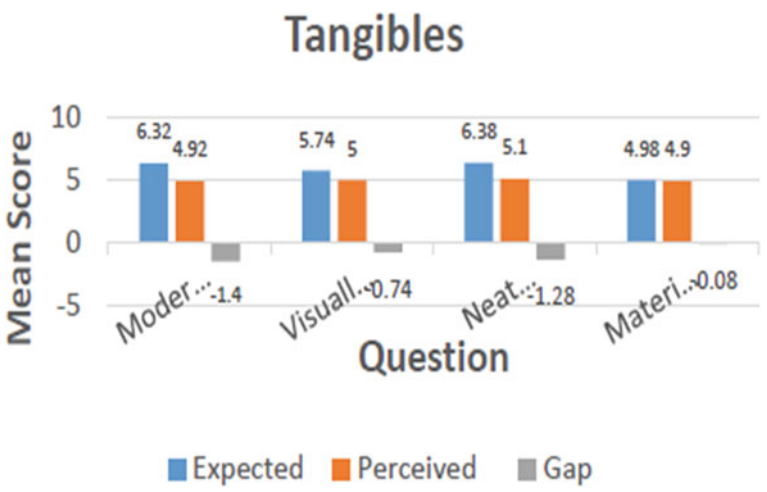


Fig. 1 (P–E) for the tangibles dimension

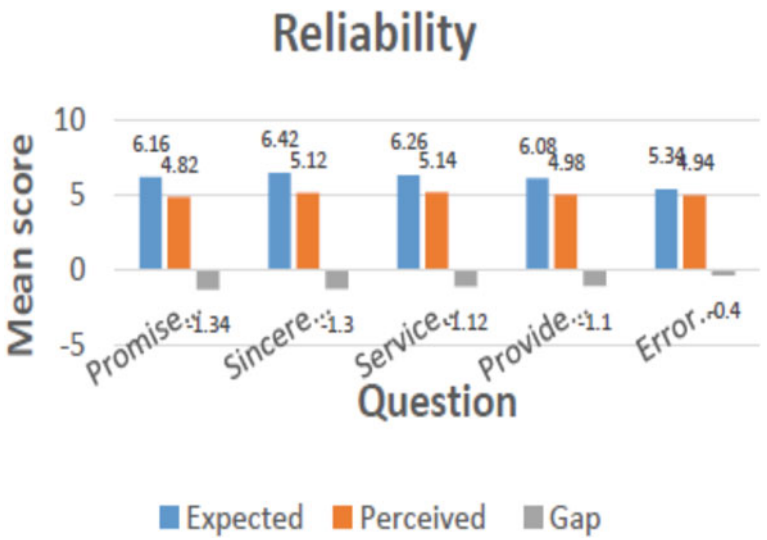


Fig. 2 (P–E) for reliability

needs of customer' (−1.34) while the smallest gap was 'Convenient operating hours' (–0.84).

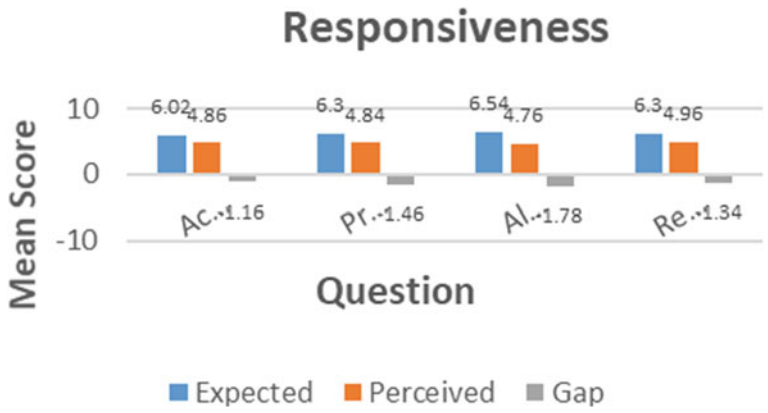


Fig. 3 Responsive levels

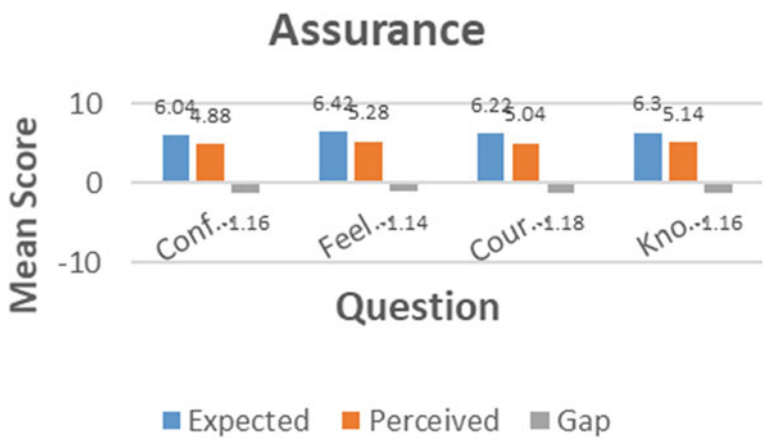


Fig. 4 Different levels of assurance

5 Comparison of Consumer's Perceptions and Expectations of the Health and Fitness Centre

Regarding expectation among the dimensions, 'Responsiveness' has got the highest mean score of 6.290 and is ranked first. This basically indicates how important and crucial attributes under Responsiveness are for maintaining sustainable fitness centre management and development [7]. The remaining four attributes assurance, empathy, reliability and tangibility rank second, third, fourth and fifth, respectively (Table 1).

This is a real experience that the customers [14,15,16] have whenever they use the service, even it can be compared to another fitness centre. Whenever they are satisfied with the way the fitness centre is handled and managed then it means that

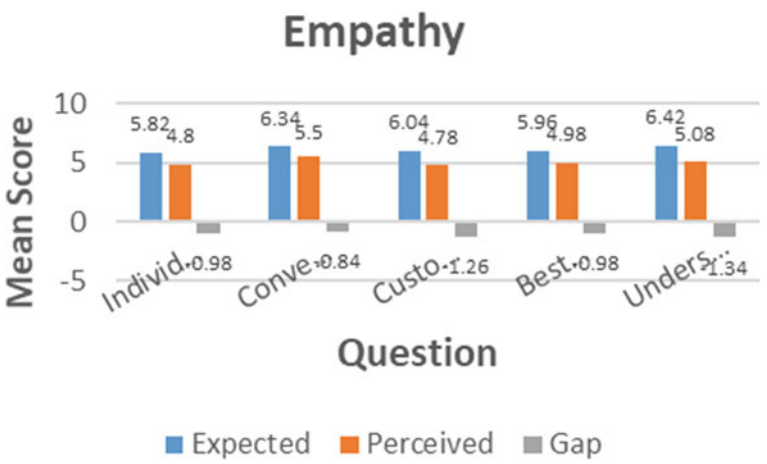


Fig. 5 Different empathy levels

Table 1 Perceptions and expectations comparison of the health and fitness centre

Dimension	Expectation			Perception			(P–E)	
	Mean	SD	Rank	Mean	SD	Rank	Mean	Rank
Tangibility	5.855	0.746	5	4.980	1.155	4	−0.87	5
Reliability	6.052	0.736	4	5.000	1.247	3	−1.052	4
Responsiveness	6.290	0.864	1	4.855	1.210	5	−1.435	1
Assurance	6.245	0.826	2	5.085	1.134	1	−1.160	2
Empathy	6.116	0.848	3	5.036	1.184	2	−1.080	3
Aggregate mean	**6.109**			**4.993**			**−1.115**	

they expect a lot from the fitness centre. The measurement is perfect for determining the best sustainable fitness centre development. If customers are not satisfied it means that they prefer other fitness centres [8]. In this regard 'Assurance' scored the highest mean score of 5.085 and stood first as compared to the remaining four attributes. The finding ranked empathy, reliability, tangibility and responsiveness second, third, fourth and fifth, respectively (Fig. 6).

Table 2 shows the calculation of gap score, which is calculated using the following equation.

$$\text{GAP SOCRE} = \text{PERCEPTION} - \text{EXPECTATION} \quad (1)$$

Table 3 shows the calculation of weighted SERVQUAL scores which is calculated as follows.

$$\text{Weighted SERVQUAL Score} = \text{Average for Each Dimension} \times \text{Importance Weight} \quad (2)$$

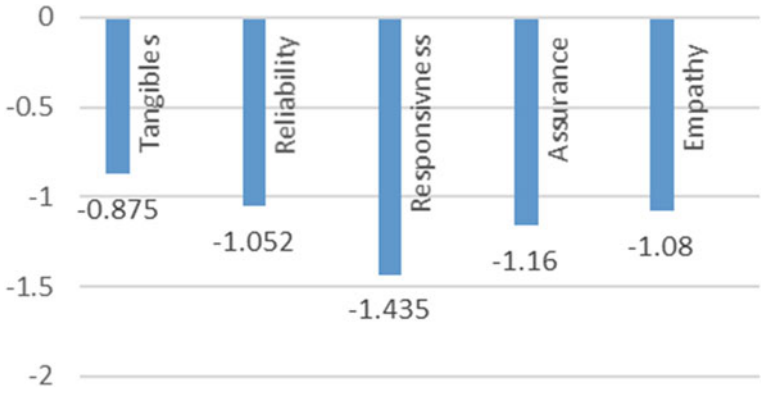


Fig. 6 Gap score for each dimension

Table 2 Gap scores calculation

Dimension	Statement	Expectation score	Perception score	(P-E)
Tangibles	1	6.32	4.92	−1.40
	2	5.74	5.00	−0.74
	3	6.38	5.10	−1.28
	4	4.98	4.90	−0.08
Reliability	5	6.16	4.82	−1.34
	6	6.42	5.12	−1.3
	7	6.26	5.14	−1.12
	8	6.08	4.98	−1.1
	9	5.34	4.94	−0.4
Responsiveness	10	6.02	4.86	−1.16
	11	6.30	4.84	−1.46
	12	6.54	4.76	−1.78
	13	6.30	4.96	−1.34
Assurance	14	6.04	4.88	−1.16
	15	6.42	5.28	−1.14
	16	6.22	5.04	−1.18
	17	6.30	5.14	−1.16
Empathy	18	5.82	4.84	−0.98
	19	6.34	5.50	−0.84
	20	6.04	4.78	−1.26
	21	5.96	4.98	−0.98
	22	6.42	5.08	−1.34

Table 3 Weighted SERVQUAL scores calculation

SERVQUAL dimension	Average for each dimension (Table 3)	X	Importance weight (Table 4) (%)	=	Weighted score
Reliability	−1.05	x	32	=	−0.336
Responsiveness	−1.43	x	22	=	−0.314
Assurance	−1.16	x	19	=	−0.220
Empathy	−1.08	x	16	=	−0.172
Tangibility	−0.87	x	11	=	−0.095
TOTAL					−1.137
AVERAGE (Total/5)					−0.227

Equation (2) shows the weighted SERVQUAL score.

6 Recommendations and Conclusion

6.1 Findings of Talwalkars Health and Fitness Centre

After getting the questionnaires filled up, a follow-up was carried out, wherein the respondents were asked a specific question about how they would like the quality improvement to happen [9]. Some specific findings obtained from that follow-up are.

1. Trainer tends to focus more on female clients as compared to male clients.
2. During the peak time approximately 60 members and 6–8 trainers are available, but still, customers are not fully satisfied [10].
3. Many complaints are there from changing room regarding security purpose. Some members complain that the key of the lockers does not work and sometimes they forget the shoes, clothes etc. but the following day they don't find them. Such thefts or pilferage should be checked.
4. There is no focus of the trainers on the cardio area. Trainers have more focus on bodybuilding [11].
5. Many clients say that trainers make personal relations or become friendlier even beyond the gym with some particular client, so they take personal interest when these clients are doing workouts.
6. In the cardio section of the gym, only 1–2 trainers are available but they do not take a personal interest to train the members.
7. Repairing a machine is another issue in Talwalkars. One member says, his complaint about the cardio cycle machine was not attended and the machine went unrepaired for months [12].
8. Hygiene problem occurs most of the time in this gym. If one member is using the machine for exercise, sometimes he leaves a lot of sweat but to clean the

perspiration from a machine or bench takes 15–20 min and other members suffer.

9. During the joining, they take full body measurements. After that, they promise to schedule monthly measurements but they do not actually do so for the clients [13].
10. There is no follow-up after doing the workout month after month. No one takes interest in how much the client lost or gained in a particular time duration.
11. The employees pressurize the clients to refrain from spreading (bad points) or criticism regarding their health centre with their friends.

7 Conclusion

In this study, the largest gap (Perception–Expectation) is found in the 'Responsiveness' dimension, So a larger focus should be on the responsiveness dimension. It means that 'Actual Service Performance', 'Promptness', 'Always be willing to help' and 'Respond to the Customers' need to be focused upon.

The weighted average score of P–E ≥ 0 will be considered as satisfactory or good service as perceived by the customer. It specifies questions targeted to capture the five dimensions. It is suggested, as appears from the responses arrived by the weighted average P–E score, that compromised perception of response is slightly towards satisfaction. It does not show full satisfaction.

References

1. Gillison FB, Skevington SM, Sato A, Standage M, Evangelidou S (2009) The effects of exercise interventions on quality of life in clinical and healthy populations: a meta-analysis. Soc Sci Med 68(9):1700–1710
2. Serbes S, Cengiz C, Sivri M, Filiz T (2017) Health-related fitness knowledge of middle school students in public and private schools. Montenegrin J Sports Sci Med 6(1):29–35
3. LLOYD, C. (2008) Recruiting for fitness: qualifications and the challenges of an employer-led system. J Educ Work 21(3):175–195
4. Lloyd HC, Reynolds KL (2003) The consequences of dysfunctional customer behaviour. J Serv Res 6(2):144–161
5. Frew M, McGillivray D (2005) Health clubs and body politics: aesthetics and the quest for physical capital. Leis Stud 24(2):161–175
6. Farrell D, Laboissiere MA, Rosenfied J (2005) Sizing the emerging global labour market. McKinsey Q 3:93–103
7. Tansik DA (1990) Managing Human resources issues for High-Contact Service personnel. In: Bowen DE, Chase RB, Cummings TG (eds) Service management effectiveness. Jossey-Bass, Scan Francisco
8. Prasuraman A, Zeithaml VA, Berry LL (1988) SERVQUAL, a multi- item scale for measuring consumer perceptions of the service quality. J Retail 64(1):12–40
9. Tsintsifa E, Sarri K, Zourladani S, Tsiligkiroglou- Fachantidou A, Costa G (1999) Measuring the quality of fitness facilities. In: Proceedings of the 7th congress European association for sport management. Thessaloniki, Hellas, pp 319–321

10. Brady M, Cronin J (2001) Some new thoughts on conceptualizing perceived service quality: a hierarchical approach. J Mark 65:34–49
11. Harris EK (2003) Customer service: a practical approach, 3rd edn. Prentice-Hall, New Jersey
12. Kotler P, Keller KL (2006) Marketing management. Prentice Hall, Upper Saddle River, NJ
13. Armstrong G, Kotler P (2006) Marketing: an introduction, 8th edn. Prentice-Hall, New Jersey
14. Joshi A, Sabitha AS, Choudhury T (2018) Crime Analysis using K-means clustering. In: Proceedings—2017 International conference on computational intelligence and networks. CINE. https://doi.org/10.1109/CINE.2017.23
15. Sharma D, Sai Sabitha A, Choudhury T (2019) Discovering the unknown patterns of crop production using clustering analysis. Adv Intell Syst Comput 841. https://doi.org/10.1007/978-981-13-2285-3_70
16. Raj G, Sharma S, Choudhury T (2018) Load analysis in SDN for distributed topologies. In: Proceedings of the 2nd international conference on green computing and internet of things. ICGCIoT. https://doi.org/10.1109/ICGCIoT.2018.8753099
17. Mehra S, Choudhury T (2018) Sentiment analysis of user entered text. In: Proceedings of the international conference on computational techniques, electronics and mechanical systems. CTEMS. https://doi.org/10.1109/CTEMS.2018.8769136
18. Mittal V, Gupta S, Choudhury T (2018) Comparative analysis of authentication and access control protocols against malicious attacks in wireless sensor networks. In: Smart computing and informatics. Springer, Singapore, pp 255–262
19. Chauhan N, Agarwal R, Garg K, Choudhury T (2020) Redundant iaas cloud selection with consideration of multi criteria decision analysis. Procedia Comput Sci 167:1325–1333
20. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of weiler-atherton algorithm to address loose polygons. J Sci Ind Res 78:771–774
21. Kumar V, Choudhury T (2019) Real-time recognition of malignant skin lesions using ensemble modeling. J Sci Indus Res (JSIR) 78:148–153

Load Balancing Issues in Cloud Computing

Priyanka Agrawal, Subhash Chand Gupta, and Tanupriya Choudhury

Abstract Load balancing is done to distribute the network traffic over various servers. This is done so that one server does not get overloaded and its resources do not run out. This also increases the efficiency of the application and it also increases the availability of the application for users. This is done by a load balancer. It is a device used for load balancing that is for distributing traffic over the server. In today's world cloud computing is growing day by day and the major concern is to increase the efficiency and performance of applications, and load balancing is one way to increase the performance. But there are several issues involved in load balancing. This paper discusses those issues involved in load balancing and how load balancing is done in different types of applications.

Keywords Load balancing · Load balancer · Cloud computing · Performance

1 Introduction

Cloud computing is a technology growing day by day and is widely used by many large-scale organizations to handle their data. It has many features and a considerably reasonable rate in terms of money because of which it has become so popular in such a short span of time.

Cloud computing [4] is basically an on-demand delivery of IT services that are provided over the internet through some providers who own their own cloud, like Amazon Web Services, Microsoft Azure etc. Cloud computing [6] is not just easy to use but also very secure and easy to maintain as compared to the traditional way of storing the data physically.

P. Agrawal · S. C. Gupta
Amity University, Noida, UP, India
e-mail: scgupta@amity.edu

T. Choudhury (✉)
Department of CSE, University of Petroleum and Energy Studies (UPES), Dehradun, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_10

The speed at which cloud computing [33, 26] is getting researched there is not much time left that every organization will use this technology no matter small-scale or big organizations.

Load balancing is one of the most important features of cloud computing [22, 11] and is very important for the successful work of an application as it prevents the application from getting crashed. Load balancing is the process of distributing the network load over different nodes to prevent them from getting overloaded. The device which is used to perform the load balancing process is called a load balancer. There are both hardware load balancers and software load balancers. There are different types of load balancers discussed in this paper. The main objective of the paper is to discuss various issues and challenges faced during the process of load balancing and few techniques for dealing with it. Load balancing is a process of distributing the load, and if the load is not distributed randomly across nodes there are many algorithms used to find how the load will be distributed among the nodes. A few of them are discussed here. So, this paper covers basic knowledge of cloud computing, one of its main features, which is load balancing, and how it is performed through what device and in which ways, and issues and challenges faced during load balancing.

2 Literature Review

Cloud computing is a service basically an on-demand delivery of IT services provided over the internet by a third-party vendor. For example, Amazon Web Services (AWS), Microsoft Azure, Google Cloud, IBM Cloud etc.

It provides services to customers over the internet with various facilities. It allows the user to store their data in the cloud rather than storing the data physically as used to be done in traditional ways. It is not necessary for the user to store all the data in the cloud and there are different models in which the user can store the data.

There are many features of cloud computing:

(1) On-demand, self-service: The user can store data on their own without human help and this will reduce the labor cost for the owner.
(2) High range of network: The data can be accessed anywhere over the internet, thus solving the problem of carrying physical files everywhere.
(3) Resource pooling: Using a multi-tenant model, the resources can be sent to multiple consumers.
(4) Rapid elasticity: The services provided can be scaled based on the demand. It can be both increased and reduced.

There are different models of cloud computing and the model is considered to be most efficient when its resources and services are used in the best possible ways. And this efficient utilization can be done when the resources are employed properly and managed properly.

A. Different Models of [3] Cloud Computing

(1) Private cloud: In this, enterprises have their own infrastructure, and the applications are deployed on their own data centers.
(2) Public cloud: In this, the services used are delivered by a third-party vendor and can be accessed anywhere over the internet.
(3) Hybrid cloud: This type of cloud is a combination of the above-mentioned two types of clouds, which means some things are private and some are public.
(4) Multi-cloud: In this, at one time two or more public clouds can be used or multiple private clouds.

B. Cloud Computing Services
There are three cloud computing services:

(1) Software as a service (SAAS)
These are the vendor-provided applications accessible over the internet. This is directly for the end-users. For example, Gmail.
(2) Platform as a service (PAAS)
In this, the vendors provide the platform on which the user develops the applications. This is for the software developers who can develop software on the platform provided. For example, Google app engine.
(3) Infrastructure as a service (IAAS)
In this, the vendors provide the infrastructure, that is the hardware, which includes servers, resources and the storage space for application development. This is for the infrastructure architects who just need all the hardware and resources for developing the application. For example, Amazon S3, Amazon EC2.

C. Issues in Cloud Computing
Ethical Issues
Cloud [2] computing faces some major ethical issues like security, integrity and privacy because it is the duty of the service provider to secure the customer data stored in their data centers.

(1) Security
This includes confidentiality, integrity and availability of data through secure systems. Security is a major issue in the cloud computing environment because data stored in the data centers are highly prone to unauthorized access. So, important security measures should be taken.

- Confidentiality and Privacy
Confidentiality means that only a person who is authorized can view the data and no one else. In cloud computing, there are very high chances that the data can be accessed by unauthorized as there is sharing of same resources like memory or storage or network and this thing can lead to a breach of data confidentiality.

There have been several issues of unauthorized access which makes the privacy of the customer the biggest issue.
The meaning of privacy exactly is that the information owner is the one who decides who all can view their information. A privacy breach is a major issue in cloud computing as the data is stored at several locations in different regions because of which there are high chances of a privacy breach.

- Availability
  Availability means that the data stored or the software should be accessible in a usable condition when in demand. Availability also means that the system should be available in all conditions even if there are some problems with security, but the providers cannot guarantee this because cloud computing needs the internet to work, and if there is some problem with the internet it will affect the cloud environment. So, this also becomes a major issue.
- Integrity
  This means that only the people who are authorized can alter or make changes in the data and the software or hardware stored in the data centers.
  There are several users who are sharing the same resources so authentication is very important in a cloud environment to maintain security.

(2) Strategy Issues
Cloud computing is a different type of environment and its introduction into the system will bring a major change in the organization's existing IT system.
One of the main change occur in the system is the reduction in the size of IT department as now most of the work is done by cloud service provider. This can also lead to a decrease in jobs as now most of the work is done by cloud and very little work is left, like documenting work or issues resolving.
This is of concern because sometimes the employees do not accept the cloud environment as there are chances of losing their jobs.
There is one more issue that the organization should have, which is skilled labor to handle the cloud environment. The effect of the cloud on an organization and how well the management addresses the issue to its employees tell about the success or failure of cloud computing.

(3) Technological Issues
Implementing a cloud environment requires up-gradation of the existing systems of an organization so as to have a large amount of data and files. This process involves buying new technologies according to the cloud and discarding the old ones which are not needed in the new systems. The main issue with this is that the management has to decide an efficient way of adding this without putting too much cost into it.

This will not be a major issue if the organizations are able to buy all the necessary technologies.

(4) Customization
This is an important issue because every organization likes to modify or customize its products and schemes as it is believed that this makes it different and unique from others. But customizing the products is an issue for the service provider as the resources are shared by everyone and different customers require different services.
Customization is an important issue as uniqueness is very important for an application to win over other applications, and hence it is important for the customer to choose the service provider wisely depending upon who provides customization or not.

D. Load Balancing
Considering large applications one server is not enough to handle all the requests as the server gets overloaded very quickly, so to fulfill this requirement the organization distributes the load over multiple servers and this is called "load balancing". This helps in increasing the efficiency of the application and prevents it from getting crashed. It also makes the application more available to the user and helps in processing more requests quickly.
Various algorithms are used to distribute load through the load balancers to various servers according to the specifications of servers.
For the users, this is not something visible to them. It's just that their requests should be processed quickly even if there is a large number of requests.
For example, consider a shopping site that has a large number of users and during sale time the users increase even more and using the application at the same time. In this case, if there is only one server then it will get overloaded very quickly and also might crash resulting in the failure of the application, So to prevent this, load balancing is used which distributes the traffic over multiple servers so that each request can be processed quickly without harming the application.
So load balancing is generally used for high traffic sites and it is also used for domain name system (DNS) servers and for FTP file transfer protocol sites.
Major tasks performed by load balancing are:

- Distributes traffic over servers
- Minimizes the request-response time
- Increases efficiency of the application.

E. Load Balancer
To perform this load balancing, some device is required and the device used is the load balancer.
This load balancer with the help of different algorithms distributes the traffic over servers and it makes sure that each server handles only the number of requests it can handle and does not get overloaded.
Working of Load Balancer [8, 30, 31]

(1) An application receives a request which has to be passed onto the server.
(2) A load balancer is between a client and a server. So now the load balancer receives the request and based on the algorithm applied it forwards the request to the servers according to their capacities.
(3) The server then receives the request and gives back the response to the client through the load balancer.
(4) When the load balancer receives the response it checks the details like IP and gives a response to the client that matches it.
(5) And this process is repeated again and again until the application is closed.

F. Benefits of Load Balancer

- Required and best for high-traffic applications where a large number of queries has to be processed.
- It increases resource utilization and reduces response time.
- Makes the application accurate and smooth and keeps the users satisfied as there are quick responses and they don't have to wait for results.
- Flexibility: It makes the application flexible, which means the servers can be added or removed according to the requirement of the application and this feature also helps during maintenance of servers as when one is under maintenance the load gets distributed to other servers without affecting the performance of the application.
- Scalability: This is similar to flexibility where servers can be added if the traffic increases, and when servers are added they get included in the process and traffic distributes smoothly.
- Redundancy: While the process is going on and if any server fails, quickly the traffic gets divided among other servers without getting known by the user that any server has failed.

G. Types of [9] Elastic Load Balancers

(1) Application Load Balancer:
This [27] load balancer works at the application layer (HTTP/HTTPS). Application load balancers support dynamic host port mapping. This load balancer also performs health checks and see if any node is experiencing any issues and based on the health check result it distributes traffic.
This type of load balancer uses listeners which are used in applications to react to specific events and the listeners route the requests according to the content of the requests.
(2) Network Load Balancer:
This load balancer works at the transport layer (TCP). It can handle a large number of requests, like millions of requests in one second.
(3) Classic Load Balancer:
This works either at the transport layer (TCP) or application layer (HTTP/HTTPS). Classic load balancers require a fixed connection between the load balancer port and the connection instance port. With

classic load balancers, traffic can be distributed over single or multiple availability zones.

H. Hardware and Software Load Balancing

(1) Hardware Load Balancers
Hardware [12] load balancers consist of physical hardware and distribute the traffic based on the number of connections with each server and the performance of those servers.
These types of load balancers are used in properly developed firms and they need to be maintained and updated regularly as new security patches need to be installed. These are better than software load balancers as they offer more range of features but it requires trained staff for maintenance and management. Because these load balancers are hardware-based they are less flexible and scalable and that is why sometimes they result in over-provisioning of load balancers.
The hardware load balancers are generally implemented on the transport layer (layer4) or the application layer (layer7) of the OSI model of the network.

(2) Software Load Balancers
These are easier to deploy and also cost-effective and more flexible. But being flexible requires more work that needs to be done in setting this up. In this, it is easier to make changes.
Software load balancers are available as solutions that need to be installed. These types of load balancers generally combine multiple scheduling algorithms to distribute the network traffic.

I. Load Balancing Model
There is an architecture model for clouds having varying loads. This model is used for best load distribution among different virtual machines.
The first [1] level of load balancing is performed at the physical machine (PM) and the second level is performed by the virtual machines (VMs) (Fig. 1).
The request generator generates the user request which is basically tasks given by users and these tasks need computation resources for execution. Then the next part is the data center controller which is in charge of task management between machines.
Then the load balancer present at level 1 checks which VM can be allotted to the task given by the user. The load balancer at the first level balances the given workload on the physical machines by further distributing the load among the designated virtual machines. The second level load balancer balances the workload among various virtual machines associated with a specific physical machine.

J. Load Balancing Classification
Figure 2 shows how load balancing algorithms are classified. They are basically divided into two types:

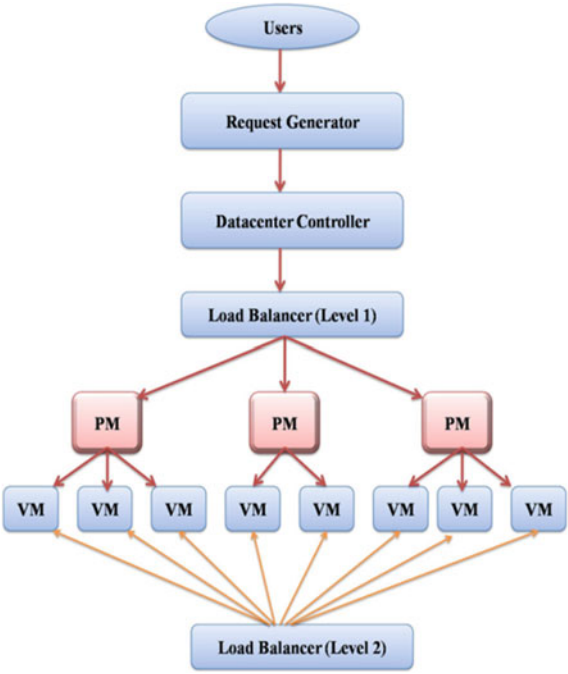


Fig. 1 Load balancing model

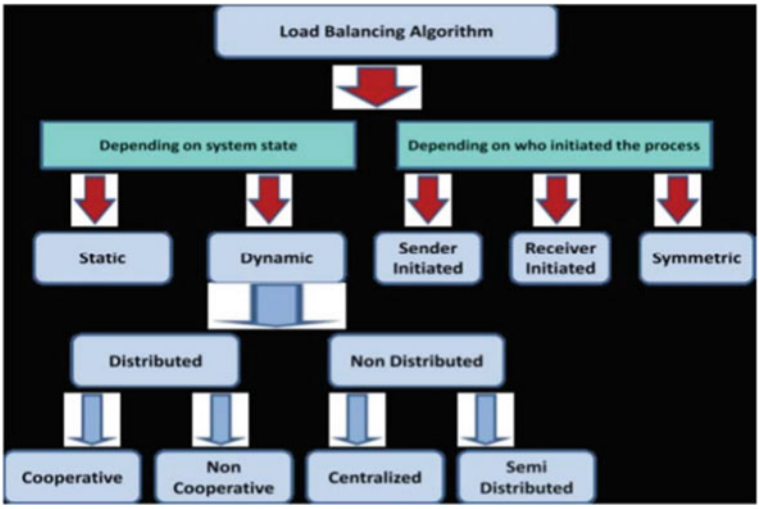


Fig. 2 Load balancing classification

(1) Static Approach
This approach is designed during the design and implementation phase of the system. Load balancing done through this approach divides the network traffic equally between all the servers.

(2) Dynamic Approach
This approach works on the present situation of the system while doing load balancing and it is best suitable for large distributed applications like cloud computing applications.
Dynamic load balancing is further classified into two categories:

- Distributed approach.
- Non-distributed or centralized approach.

3 Load Balancing Issues in Cloud Computing

1. Fault Tolerance
This is one of the most [15] important metrics of load balancing and also one of the issues of load balancing. This means that the load balancing algorithm should be performed properly in case of any node failure, and even if a node fails it should not affect the load balancing process. This issue needs to be resolved and it is expected that load balancing algorithms should have a good fault tolerance approach. A fault-tolerance policy is expected to identify the fault, then solve it, and thus improve the efficiency of an application.
Some types of faults are:
Network faults: These occur over the network when there is a packet loss or problems with the destination node.
Physical faults: These are hardware faults.
Process faults: These faults occur during the process because of a lack of resources.
Service expiry faults: These faults occur when the service time of any resource expires.
There are two types of fault tolerances:

Reactive fault tolerance

Reactive fault tolerance this decreases the effect of failures while an application is running and when the fault starts occurring frequently.

Proactive fault tolerance

This fault tolerance avoids recovery from faults or failures by identifying them before only and then immediately replacing the defected parts with better working parts.

Fault Tolerance Techniques [14, 33]

Fault tolerance [21] techniques are also specific to the type of fault tolerance. Fault tolerance techniques for reactive fault tolerance are:

- Check Pointing
  This technique is generally used for large applications which run for a long time. While updating or creating the applications several marks like checkpoints are made in the program whenever any changes are made so that when the application runs and any failure occurs then the application can be restarted from the previous checkpoint only instead of again and again starting from the beginning.
- Replication:
  In this technique, the tasks performed are copied identical or replicated on different resources so that even if any particular resource fails the task is performed and the expected output is obtained.
- Job Migration:
  In this technique, if any task gets failed due to any reason then that task can be immediately shifted or migrated to any other machine.
- Task Resubmission:
  In this technique when a task gets failed then the job remains undone, so the application is made to run again and the task failed is resubmitted to another resource or maybe the same resource for getting executed.
- S-Guard:
  This technique is based on the rollback recovery process.
- Retry:
  This technique is the most simple and easy technique as in this the failed task is retried again and again on the same resource to get completed.
- Rescue Workflow:
  In this technique the task is performed till the time it can be performed without thinking about failure.
  Fault tolerance techniques for proactive fault tolerance:
- Self-Healing:
  In this technique, the big tasks are divided into several smaller parts which are running on different virtual machines and when many parts of an application are running on a different machine it automatically handles the failure problems.
- Software Rejuvenation:
  In this technique, the system is restarted periodically as set and the system restarts with a fresh start and a clean state.
- Pre-emptive Migration:
  In this technique, the application is under continuous monitoring and analyzed regularly and is based on a feedback control strategy.
  So, using these fault tolerance techniques the application can be made fault-tolerant and one of the issues of load balancing is resolved resulting in increasing efficiency of the application.

2. Virtual Machines Migration
A virtual [28] machine is like an emulator which is created using the software. It uses the physical resources of the system like RAM, CPU but it is separated from the other software of the system. These virtual machines generally called as VMs are also considered as an instance of OS which runs in isolation in the system.
There can be multiple virtual machines running on a single physical system. Each virtual machine consumes its resources and hence making the physical system getting overloaded, and in such cases, the traffic has to be transferred dynamically to another system without getting noticed by the user. And this process of moving one virtual machine over another machine is called migration. In virtualization, the machine can be seen as a set of files or a file. Earlier when migration was performed between two physical systems the VMs had to be stopped and the needed resources are to be allocated to the new physical system and all the files related to that virtual machine are to be moved to the new system. There are two important properties of virtual machines:
Encapsulation
Hardware independence
These properties are responsible for the migration process that is for the movement of a virtual machine from one physical system to another.
This migration process should not be visible to the user and also this process should be quick. If the process takes time due to any reason then the user may feel service unavailable and this is not right. Virtual machine migration allows a continuous transfer of memory, storage and network from one system to another.

Virtual [24] *Machine Migration Requirements*

Load balancing:
The virtual machines are generally migrated from an overloaded host to a less overloaded system to prevent the crashing of applications.

Maintenance:

The virtual machines are generally migrated to maintain the network balance, increase the fault tolerance and properly maintain a system.

Recovery from the host failure:

The virtual machines are migrated to recover from the network failure that occurred due to the overloading of the host system.

Types of Virtual Machines Migration

Cold Migration

This is an old type of migration; in this, the virtual machine is first stopped at one system, and then it is started again only when the migration process is finished and the VM has been transferred.

Live Migration

This is a technique used for load balancing. In this type of migration, the virtual machine need not be stopped while transferring, and it is continued to run. Migration is performed with very a less downtime. The transfer of virtual machines refers to the transfer of resources also like memory, storage etc. Out of all the resources, memory transfer is the most time-consuming.
The parameters [16] considered while doing live migration are:
Downtime: This is the time at which the services of a virtual machine are stopped and are not available to the user.
Migration time: This is the total time which is required to transfer the virtual machine from source to destination system without affecting the availability of virtual machine to the users.
The cost of the live migration process is considerable but it cannot be ignored especially at times when the availability and responsiveness of virtual machines are strictly concerned and governed by the service level agreements (SLA).
Pre-copy: In this type, memory is transferred first, and then the execution is transferred.
Post-copy: In this type, the execution is transferred first and then memory is transferred.
Advantages of virtual machine migration are:

- Server consolidation
- Hotspot and cold spot migration
- Load balancing

3. Scalability
This is the process of either increasing or decreasing the number of nodes in a system according to the requirement.
Cloud computing [32] offers many advantages to organizations but one of the main advantages is scalability.
Before cloud was introduced, adding extra servers was a great problem as new hardware is to be installed and maintained. More space is required for servers and this increases the cost heavily, and not just that the downtime is also increased. In such cases as the time in which more servers are installed the previous servers had to be shut down.
For businesses which are growing slowly, this is not of much concern, but for business growing quickly it increases infrastructure costs heavily for them.
When data is stored physically scalability is always a concern, as a lot of factors have to be considered, but with the cloud, you can start small, and as the business grows data storage is expanded without increasing any hardware at the organization and making scalability easier. And one specialty of the cloud is it follows a pay-as-you-go pricing strategy, which means the money has to be paid only for the storage used.
For example, consider [10] you have a small startup company doing an average business and then you get a chance to publicize your product on a very large scale

and in front of a huge audience. As a result of this, your website has increased the number of users and a large number of people are trying to place orders. But can your existing infrastructure handle this increase in network traffic? If you have a scalable infrastructure there will be no problem as you can increase the number of servers according to the traffic without thinking about the space and pricing but if you don't have a scalable infrastructure then the site will become slower and eventually might stop, thus resulting in knocking off the customers instead of attracting them.

Types of [29] *Scaling in Cloud Computing*

Vertical Scaling (Scaling up)

This method of scaling means adding more power to the existing system. This could be in the form of adding more CPU processors or increasing RAM or more storage. In this, no code is altered and no new infrastructure is added. Vertical scaling is also finite, so after a point limit is reached.
In such cases, the application has to be powered down and resized which results in downtime and extra maintenance, so this requires advance planning.
The disadvantage of this is performance decreases as the compute capacity remains the same as there is no change in the infrastructure even if storage or RAM is increased.

Horizontal Scaling (Scaling out)

This method of scaling means adding more infrastructure to the existing infrastructure and connect them so that they work smoothly.
This requires complex architectural designing which is a very time-consuming and labor-intensive process that requires a lot of people. But this is infinite scaling, which means that an infinite number of instances can be added to increase the performance.

Diagonal Scaling

This is a combination of both horizontal and vertical scaling, which means the benefits of both types of scaling can be used in this.
First, they are scaled vertically till the time it reaches the infrastructure capacity and then new resources are added to a server that is connected to the system and this process repeats continuously.
This type of scaling is most effective in organizations that are handling various workload volumes.
Scalability can be maximized by using load balancing. Load balancers offer a strategy of scaling the system by distributing the load across various nodes to utilize resources. Load balancer takes all the requests and then distributes the request to different nodes based on their capacity.
Load balancers work in a similar way as a scaling group, just the difference is it does not create or delete instances; it distributes the traffic across multiple instances.

4. Load Estimation Policy For Load Balancing

   - The first issue in this is to decide which method is to be used to estimate the workload on each node separately.
   - The workload of a node is determined by various parameters:
     (1) Total number of processes allotted to a node.
     (2) Total resources required by all these processes.
     (3) It also depends on the speed of every processor.
   - Many load balancing algorithms use the total number of processes to increase efficiency.
   - In many cases the actual load of a node varies depending upon the remaining time of that process on the node. And this remaining time can be measured in many ways.
   - A considerable method for load estimation is by measuring the CPU utilization of the nodes.

5. Process Transfer Policy for Load Balancing
   This policy is basically used to transfer the processes from heavily loaded nodes to comparatively less loaded nodes.
   There is a limiting value of workload for each node which is known as a threshold value and there are many algorithms that use this value to determine whether the node is heavily loaded or lightly loaded.
   The threshold value of a node can be determined by two policies:

   - Static policy
     In this policy, there is an already defined threshold value that is based on the processing speed of the node.
   - Dynamic policy
     In this policy the threshold value is calculated based on the average number of processes on a node and a predefined constant.
     Below the threshold value, the node accepts the processes, but above that threshold value, the node does not accept the process; instead, it shifts the process to a less loaded node.

6. Location Policy for Load Balancing
   These policies are used for selecting the destination nodes on which the process will execute.
   These policies are:

   (1) Shortest method
       In this method, nodes are chosen randomly and checked for their load. The process is transferred to the node having the minimum value of the load.
   (2) Threshold method
       In this method, a node is chosen randomly and checks if the node is capable of taking the process and if yes then the process is transferred; if not then another node is chosen.

(3) Biding method
In this method, each node broadcasts requests for bids, and then other nodes respond with bids. The offer best suited for the process is chosen.
(4) Pairing method
In this method each node randomly asks another node to become a pair and share the load, but if the asked load rejects to be in pair and the node asks another node randomly again.

7. Single Point of Failure
There [26] are various algorithms that follow the non-distributed approach and in such algorithms, the decision is taken by the central node. In such cases, if the decision device that is the center device crashes, then it affects the entire cloud computing environment and the load balancing process.
And thus it is required to develop algorithms that do not work on just one node and that one node only controls the entire cloud environment and load balancing process (Fig. 3).
Consider a scenario in which many servers are connected to a single router or switch and then that router is connected to the application server where users can access it.
And if the router fails then no single system can be accessed by the users. There is one more problem if many single servers get attached to the router then there comes the problem of lack of resources.
Amazon's EC2 cloud service has suffered this kind of problem for few days resulting in the loss of data and service due to the problem of a single point of failure.
Hence it becomes very important to remove the issue of a single point of failure to maintain and continue the services provided by the cloud.

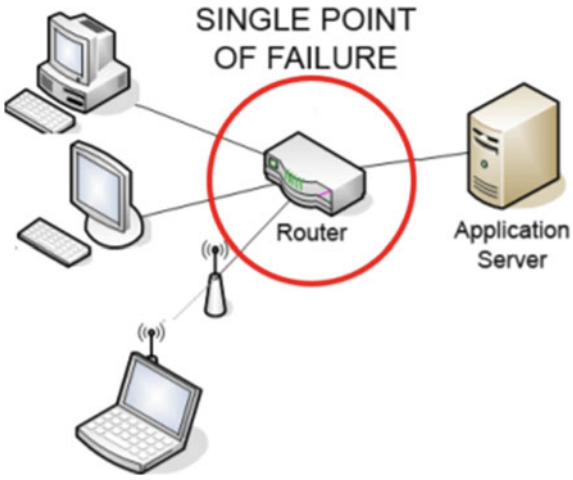


Fig. 3 Single point of failure

Elimination of Single Point of Failure

High availability clusters and different types of redundancies can be used to remove a single point of failure (SPOF). High availability clusters are a group of systems that acts as a single system and has a very low downtime, but adding too many HA clusters is also not good as it complicates the load balancing process. If used in limited numbers they are very useful also.
In a load balancing process if a system fails then immediately the load gets transferred to the other server. Similarly, if there are multiple switches attached to multiple servers then the failing of one switch will not affect the entire application as some other switch will take its place maintaining the functioning and availability of the application.

Types of Redundancy

1. *Logical Redundancy*

   In this multiple application, servers are used instead of one and this helps in removing the problem of a single point of failure from the architecture of applications. There are many software flaws and the service provider should have enough resources to identify, solve and isolate those flaws. Using logical redundancy also increases the chances of cyber-attacks so they should be handled properly by features of cloud-like multi-factor authentication.
2. *Physical Redundancy*
   In big systems, no hardware or software should completely depend on a single hardware system, so in this, the SPOF can be removed by using high availability clusters. Single-way communication between the components of clouds should be avoided in order to implement physical redundancy.
3. *Avoiding Data Loss*
   n cases where a single hardware instance holds all the data of the application and if that single instance crashes or fails this results in loss of data. To prevent this data replication method can be used.
4. *Data Replication*
   This means replicating or copying the data in multiple servers which ensures data availability to the users. While replication there are two types of copies made one primary copy and multiple secondary copies. All the operations are performed on the primary copy but if in any case, it is not available then a secondary copy from a server is treated as a primary copy, and operations are performed.
5. *Heterogeneous Nodes*
   Generally, in the systems, there are homogeneous types of nodes but in cloud computing the user requirements changes rapidly and constantly because of which it is required to use heterogeneous types of nodes for maximum resource utilization, minimum response time and efficient performance.

Hence still a lot of research is going on in developing algorithms and techniques which work on heterogeneous nodes.

6. *Storage Management*
   Storing data in a cloud environment has already resolved the problem of storing data in a traditional way which was not just costly but was very difficult to maintain as well and requires a big amount of labor for maintenance. Cloud allows the user to store data efficiently and easily. The major concern here is how the data can be stored in the cloud efficiently with effective storage and also fast and easy access. For this, the cloud gives an opportunity of keeping replication of data to maintain data consistency. But still, a lot of research is going on full data replication and its policies. In such cases, partial replication can be sufficient and efficient, so still some efficient load balancing techniques need to be developed which works on partial replication.
7. *Automated Service Provisioning*
   The main feature of cloud computing is its elasticity or flexibility. This means that the services are assigned and then delivered automatically. Or in other words, it can be said that services are assigned and released automatically. The main challenge and issue here are how we can use or release the resources by maintaining the same efficiency as traditional systems and also using the resources optimally and effectively.
8. *Energy Management*
   The [18] main advantage of energy management which encourages the use of a cloud environment is economy of scale. Power saving is the most important aspect where many different types of resources are provided by a minimal number of providers instead that everyone possesses their own separate private services.
   The challenge here is how can we use the resources and also keep an economically acceptable performance.

3.1 Issues in Load Balancing Algorithms in Cloud Computing

Static Versus Dynamic Algorithm

An algorithm is based on the state and nature of the system it has to be applied. These algorithms can be static or dynamic.

Static algorithms are the algorithms that work on the previous state of the system and not the current state of the system and that is why these types of algorithms face major problems in case of resource failure or task failure as they don't have any information about the current tasks running and the resources being used.

Dynamic algorithms are the algorithms that work on the current state of the system without considering the previous knowledge of the system. These algorithms are

better than static algorithms as they may be a little complex but they are considered more efficient and have more fault tolerance.

Geographical Distribution

This [20] is used in large applications where applications and users are in different regions separated geographically. To maintain or increase the efficiency of the system, well distribution of resources over the system in cloud computing is required. This kind of geographical distribution is very important in a cloud environment where there are multiple users accessing data from different places.

Algorithm Complexity

The performance of the system can affect the complexity of any load balancing algorithm. Sometimes it is seen that the algorithm is complex but it is efficient and performs well.

On the other hand, sometimes the algorithms which are not complex are not very efficient or may give poor performance and is not fault-tolerant.

So, system requirements should be considered as a major factor in choosing the type of load balancing algorithm. It should be decided whether system requirements are important or the complexity of the algorithm.

Traffic Analysis Over a Different Geographical Area

For load balancing algorithms that work on real-time systems, it is very important to identify first the flow of traffic over different regions and then balance the load accordingly. There are different regions over the globe separated geographically which have different time zones and peak hours because of which network load generally varies. So the load balancing algorithm applied should be capable of handling the traffic in those peak hours to obtain increased throughput and achieve maximum resource utilization.

Load Balancing Execution Algorithm

Load balancing is basically a method to distribute network load over various computers to maximize resource utilization. Recent studies have shown that resource utilization and load balancing are one of the major issues of cloud computing. The algorithms discussed can handle load balancing and provide better resource utilization.

4 Methodology

Algorithms Used for Load Balancing

(1) ***Round-Robin Algorithm***:
This is [31] the most widely used algorithm in load balancing in cloud computing. The reason why it is most widely used is that it is easy to implement

and use. In this let's suppose there are two servers attached to a load balancer and waiting for requests. When the first request has arrived the load balancer allots the request to server1 and then when the second request arrives the load balancer forwards it to server2. And because there are only two servers, so server2 becomes the last server and when request three arrives the load balancer forwards it automatically to server1, then the fourth request to server2 and so on. The requests are forwarded in a cyclic manner.

As it is seen, this method is very simple but still it cannot do well in certain cases and situations (Fig. 4).

For example, consider if server1 has a higher CPU, RAM and other specifications than server2. So server1 should be able to handle more requests than server2. But a load balancer running on round-robin algorithm will not handle the server accordingly, so it will forward the requests in a cyclic manner only even when the two servers have different capacities. As a result, server2 will get overloaded faster and can even go down.

So, this algorithm is best suited in the scenario in which the servers present are of identical specifications.

(2) ***Weighted Round-Robin Algorithm***:

To overcome the above-stated problem, there are many other algorithms out of which the weighted round-robin algorithm is selected.

Considering the same example in which server1 has higher specifications than server2, so this algorithm will assign more requests to the server having higher capabilities. Weighted round-robin algorithm is similar to round-robin algorithm as it forwards the request in a cyclic manner but the difference is that it allots a greater number of requests to the server having higher specifications (Fig. 5).

But the question is how the load balancer will know which server will have higher capacity and the answer is we need to tell it explicitly to the load balancer the weights of each server when we set that up.

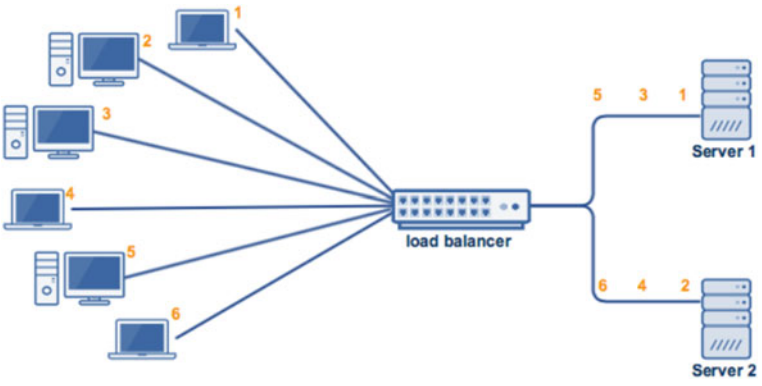


Fig. 4 Round-Robin algorithm

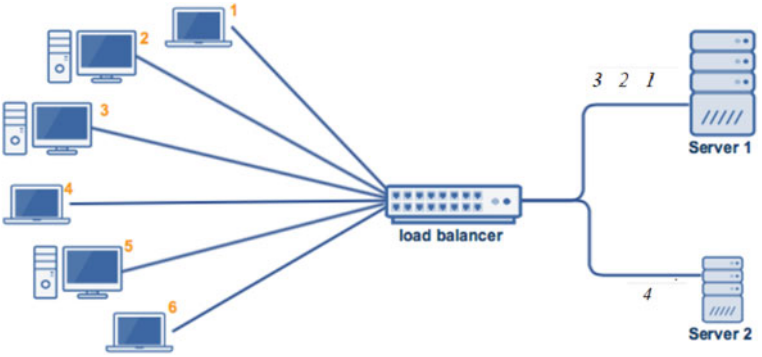


Fig. 5 Weighted Round-Robin algorithm

For example, consider server1's capacity is three times more than server2, so give server1 a weight of 3 and server2 weight of 1.
So when the requests start arriving, the first three requests will be sent to server1 and the fourth request will be sent to server2. Further, if more requests arrive they will be sent in the same manner as in round-robin in a cyclic manner.
Capacity is not the only reason to choose a weighted round-robin algorithm. Sometimes it is used when we want fewer number of connections than an equally capable server because maybe server1 is running important applications and we don't want to overload it easily.

(3) ***Least Connection Algorithm***:
There can be cases where two servers can have exactly the same specifications but still one server gets run out faster than the other.
One reason for that could be clients connecting to server2 stay connected with the server for a much longer time than server1.
This can result in piling up the total connections for server2, whereas the total connection for server1 remains the same as they are getting connected and disconnected over a shorter period of time. And in the end, the resources of server2 run out faster.
It can be seen from the figure that clients 1 and 3 of server1 are already disconnected but the clients of server2 are still connected (Fig. 6).
In such cases, the least connection algorithm is used. So considering the last example when client 6 attempts to connect, the load balancer will check the server having the least number of connections and looking at that clients 1 and 3 are already disconnected, whereas clients 2 and 4 are still connected and the load balancer will forward the client to server1 even though it was supposed to be forwarded to server2 (Fig. 7).

(4) ***Weighted Least Connections Algorithm***:
Like the round-robin algorithm had weighted round-robin, in a similar way least connections algorithm has weighted least connection algorithm. It introduces a weight component on each server which has to be specified before only.

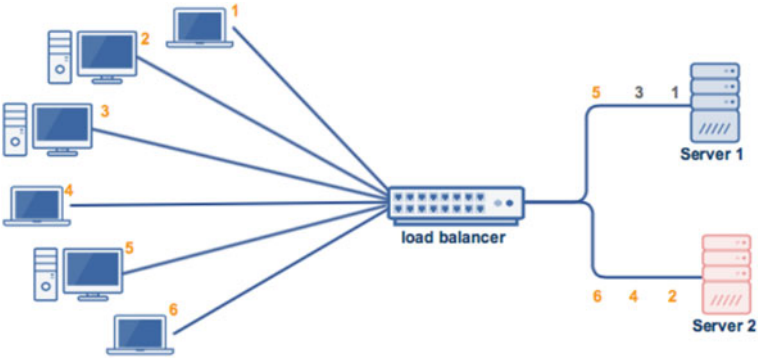


Fig. 6 Least connection algorithm

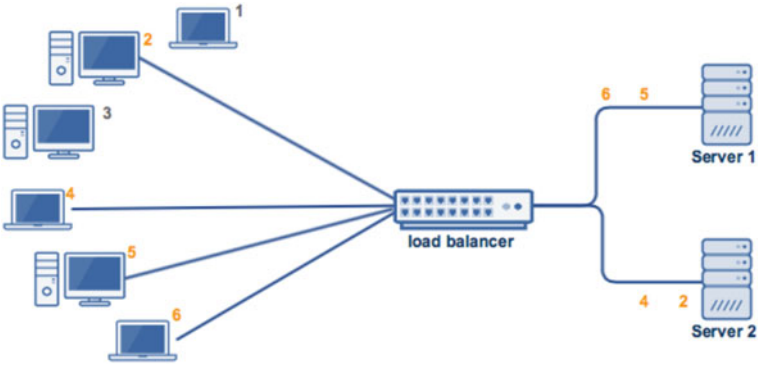


Fig. 7 Least connections algorithm

Now in this, the load balancer takes into consideration two things:

- Weight of the server
- Number of clients currently connected to each server.

(5) ***Random Algorithm***:
Like the name in this algorithm, the load balancer matches the clients and server randomly by using a random number generator.
If the load balancer gets a large number of requests, then it distributes the requests evenly to the servers. So just like round-robin is suitable for servers having identical specifications this is also suitable for servers having similar specifications like CPU, RAM, etc.

(6) ***Active Clustering***:
This [30] is a modified version of the random algorithm. This algorithm works by grouping a similar kind of nodes together and then the working of the application is done on those groups.

In this way, this method increases resource utilization as it uses high resources efficiently and thus increases the throughput and performance.
In this method, a technique is used called a matchmaker, which basically helps in the grouping of nodes. When a process is initiated it starts searching for the next matching node by a matchmaker which satisfies all the criteria and should also be not exactly similar to the previous one.
When a matching node is found out the process gets started and as soon as the process finishes the nodes get separated and with this method load balancing is performed efficiently.

(7) ***Performance Optimized Algorithm***:
This algorithm is also known as the optimized resource time algorithm. In this algorithm, the service provider chooses the data center according to their response time and the data center that has the best response time is chosen. This algorithm is very good and it also increases the performance of the application and makes it more efficient.

(8) ***Throttled Algorithm***:
This algorithm [23] is mainly developed for the load balancing of virtual machines, i.e., VMs.
In this, the throttled load balancer maintains a table that contains the index of all the VMs with their states, that is, if the virtual machine is available or busy. The state of each virtual machine is checked because a new request can be allotted to the virtual machine only when the previous one gets completed.
Now when a new request arrives the load balancer checks the entire table and the first virtual machine having the available state is chosen and its index is sent back to the controller who then assigns that request to the selected virtual machine.
But if there is no suitable virtual machine found then to tell the controller about it a -1 is sent, and in such cases, the request is queued until any virtual machine becomes available.
The notification of the allocation of new virtual machines and de-allocation of old virtual machines after their completion of the request is sent to the load balancers by the data center controller.
Figure 8 shows the process of throttled load balancing.

(9) ***Modified Throttled Algorithm***:
This [25] is a modified version of the existing throttled algorithm which takes into consideration two things:

- The availability of virtual machines for the requests.
- Uniform sharing of load among all the virtual machines for the current requests.

This is considered better than the existing throttled algorithm and also seen as an efficient method for load balancing as it considers both the response time and uniform load sharing of requests.

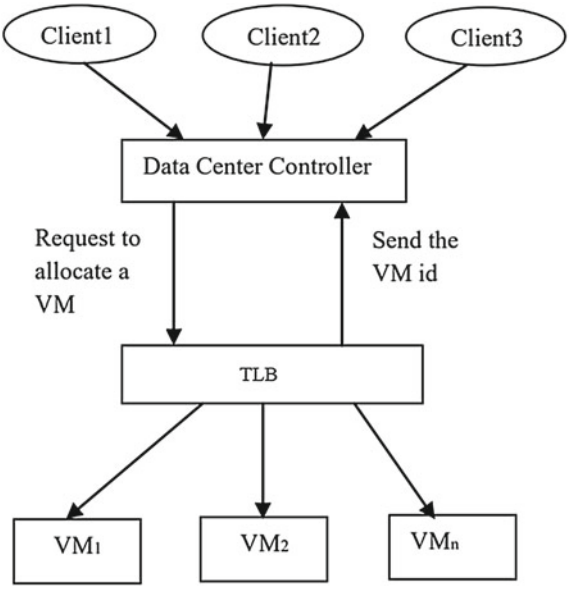


Fig. 8 Throttled algorithm

In terms of response time as compared to the existing round-robin or throttled algorithm, this is considered a better algorithm. In terms of the overall distribution of load, round-robin is considered efficient but this modified algorithm proves to be more efficient in uniform distribution of load and also with better and less response time.

So, this algorithm proves to be more efficient than existing algorithms for load balancing.

(10) ***Active Monitoring Load Balancing Algorithm***:

This [13] algorithm is also known as an equally spread current execution load balancing algorithm.

This algorithm is quite similar to the throttled algorithm with only a little difference in maintaining the table.

In this, the table contains the id of each VM along with the number of requests currently allocated to a particular VM.

Now when a new request arrives the load balancer scans the entire table and the virtual machine which has the least number of requests allocated and which has the lowest load is chosen to allocate the new request and its ID is sent to the controller which in return notifies the load balancer about the new allocation of request.

Now there are also chances that there is more than one virtual machine having the same least allocated load; in that case, the active monitoring load balancer uses the first come first serve (FCFS) method to choose the virtual machine.

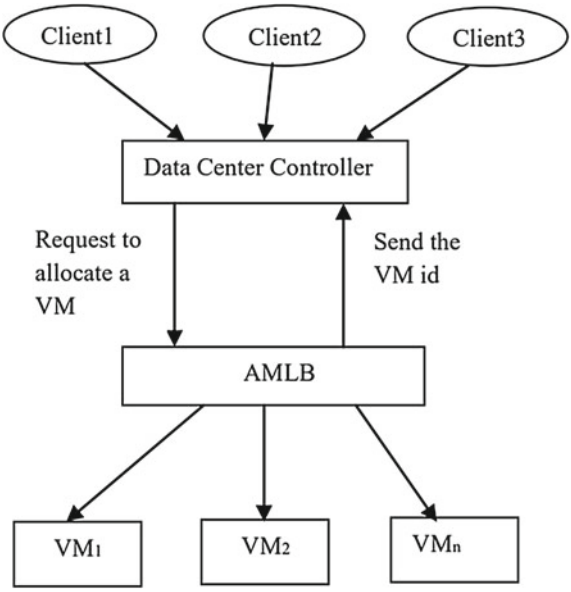


Fig. 9 Active monitoring load balancing algorithm

After this request gets completed, the controller again notifies the load balancer about the de-allocation of VM to reduce its count.
Figure 9 shows the process of active monitoring load balancing.

5 Conclusion

This paper is based on load balancing issues in cloud computing which is a very wide topic and is still unexplored. This topic can be further researched in many different ways. Cloud computing is like a service provided by a third-party vendor which has an end number of facilities provided to users in different forms.

It has various models in which ways cloud services can be used. Cloud services are of three types, which are SAAS, IAAS and PASS.

Cloud computing has various features and its one feature load balancing is discussed in this paper. Load balancing is performed by the load balancer and there are mainly three types of load balancers.

This paper majorly focuses on the issues faced during load balancing and various algorithms used in load balancing to minimize the issues as much as possible.

Major issues faced are fault tolerance, scalability, virtual migration, single point of failure, response time, storage management, energy management and many more discussed.

Few existing algorithms which are discussed are round-robin algorithm, least connections algorithms, throttled algorithms and many others.

This paper focused on the benefits and problems in a cloud environment, various issues faced in load balancing and a few existing load balancing algorithms implemented to resolve the load balancing issues and make the application efficient.

6 Result

The result which can be inferred from this paper is that load balancing which is one of the biggest features of cloud computing has many issues which need to be considered while incorporating cloud computing.

References

1. Afzal S, Kavitha G (2019) Load balancing in cloud computing—a hierarchical taxonomical classification. [Online] J Cloud Comput 8(22) (2019). Available from: https://doi.org/10.1186/s13677-019-0146-7
21. Akande AO, April NA, Van Belle J-P (2013) Management issues with Cloud Computing. [Online] 2013; 120, 121, 122. Available from: https://doi.org/10.1145/2556871.2556899?download=true.
2. Bharti S, Singh N, Researchgate.net/publication. Load balancing issues and its solutions in cloud computing.[Online] Available from: DOI: https://doi.org/10.24297/ijct.v14i6.1919
23. Chauhan N, Agarwal R, Garg K, Choudhury T (2020) Redundant Iaas cloud selection with consideration of multi criteria decision analysis. Procedia Comput Sci 167:1325–1333
31. Dewangan BK, Agarwal A, Venkatadri M, Pasricha A (2018) Autonomic cloud resource management. In: 2018 Fifth international conference on parallel, distributed and grid computing (PDGC). IEEE, pp 138–143
24. Dewangan BK, Agarwal A, Choudhury T, Pasricha A, Chandra Satapathy S (2020) Extensive review of cloud resource management techniques in industry 4.0: Issue and challenges. Software: Practice and Experience.
30. Dewangan BK, Jain A, Choudhury T (2020) GAP: Hybrid task scheduling algorithm for cloud. Revue d'Intelligence Artificielle 34(4):479–485. https://doi.org/10.18280/ria.340413
29. Dewangan BK, Agarwal A, Choudhury T, Pasricha A, Chandra Satapathy S (2020b) Cloudresource optimization system based on time and cost. Int J Math Eng Manag Sci 5(4):758–768
6. AWS Documentation. Load Balancer Types[Online]. Available from: https: //docs.aws.amazon.com /AmazonECS/latest/developerguide /load-balancer-types.html.
14. DSM (2018) Cloud benefits and scalability and load balancing.Weblog[Online] 2018. Available from: https://www.dsm.net/it-solutions-blog/a-beginners-guide-to-cloud-scalability-and-load-balancing
28. Gupta A, Choudhury T, Lal R (2017) An efficient scheme to secure cloud with diversified fortified mechanisms. In: Proceedings of the 2017 international conference on big data analytics and computational intelligence. ICBDACI. https://doi.org/10.1109/ICBDACI.2017.8070829
5. IBM Cloud learn hub (2019) Load Balancing.[Online] 2019. Available from: https://www.ibm.com/cloud/learn/load-balancing#toc-hardware-v-4dnN_P3x
16. Jjena SR, Ahmad Z (2013) Response Time Minimization of Different Load Balancing Algorithms in Cloud Computing Environment. [Online]2013; 69(17):22, 23. Available from: https://pdfs.semanticscholar.org/ece6/d0ae072e54ff02a071000070c8947c0b3be7.pdf

32. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of weiler-atherton algorithm to address loose polygons. J Sci Ind Res 78:771–774
4. Kathalkar P, Deoranker AV (2018) Challenges and issues in load balancing in cloud computing. [Online] 2018; 6(IV). Available from: https://www.ijraset.com/fileserve.php?FID=15746
9. Kaur P, Rani A (2016) Virtual Machine Migration in Cloud Computing.[Online] 2016; 8(5). Available from: https://doi.org/10.14257/ijgdc.2015.8.5.33
26. Kaur A, Raj G, Yadav S, Choudhury T (2018) Performance evaluation of AWS and IBM cloud platforms for security mechanism. In: 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS), pp 516–520
18. Khan RZ, Ahmed MO (2016) Load balancing challenges in cloud computing. [Online] 2016. Available from: https://www.researchgate.net/publication/309172982_Load_Balancing_Challenges_in_Cloud_Computing_A_Survey
33. Kumar V, Choudhury T (2019) Real-time recognition of malignant skin lesions using ensemble modeling. J Sci Indus Res (JSIR) 78:148–153
17. Kumar P, Kumar R (2019) ACM Digital library. Issues and Challenges of Load Balancing Techniques in Cloud Computing. [Online] 2019; 51(6).Available from: DOI: https://doi.org/10.1145/3281010
8. Kumar S, Rana DS, Dimry SC (2016) Int J Adv Res Comput Commun Eng. Fault Tolerance and Load Balancing algorithm in Cloud Computing.[Online] 2016; 4(7). Available from: https://www.ijarcce.com/upload/2015/july-15/IJARCCE%2020.pdf
27. Kumra S, Choudhury T, Nhu NG, Nalwa T (2018) Challenges faced by cloud computing. In: Proceedings of the 2017 3rd international conference on applied and theoretical computing and communication technology. ICATccT. https://doi.org/10.1109/ICATCCT.2017.8389105
15. Kushwaha M, Gupta S (2015) Int J Comput Appl. Response Time Reduction and Performance Analysis of Load Balancing Algorithms at Peak Hours in Cloud Computing. [Online] 2015; 128(17):26, 27. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.734.8004&rep=rep1&type=pdf
11. Patel SR, Panchal KJ (2017) Virtual machine Migration Strategy for load balancing in cloud computing. [Online] 2017; 3(2). Available from: http://ijariie.com/AdminUploadPdf/Virtual_machine_migration_strategy_for_load_balancing_in_cloud_computing_ijariie4997.pdf
20. Priya EAB (2014) Int J Comput Sci Mobile Comput. Advanced load balancing algorithms in cloud computing environment. [Online] 2014; 3(6): 525, 526. Available from: https://www.ijcsmc.com/docs/papers/June2014/V3I6KJ16.pdf
19. Ranjithprabhu K, Sasirega D (2014) Int J Sci Res (IJSR). Eliminating Single point of failure and data loss in Cloud Computing. [Online] 2014; 3(4):335, 336, 337. Available from: https://www.ijsr.net/archive/v3i4/MDIwMTMxNDY1.pdf
7. Rouse M, Application Load Balancer. Weblog[Online]. Available from: https://searchaws.techtarget.com/definition/application-load-balancer .
10. Sankar S, Dath D (2014) Migration of Virtual Machines.[Online] 2014; 2(12). Available from: www.ijarcsms.com
13. Stonefly (2018). What is scalability in cloud. Weblog [Online] 2018. Available from: https://stonefly.com/blog/what-is-scalability-in-the-cloud
22. Uma J, Ramasamy V, Kaleeswaran A (2014) Int J Adv Res Comput Eng Technol (IJARCET). Load Balancing Algorithms in Cloud Computing Environment. [Online] 2014; 3(2):273, 274. Available from: http://ijarcet.org/wp-content/uploads/IJARCET-VOL-3-ISSUE-2-272-275.pdf
3. Villanueva JC, Comparing load balancing algorithms. Weblog. [Online]. Available from: https://www.jscape.com/blog/load-balancing-algorithms
12. Vonnegut S (2017) Scalability in the cloud. Weblog [Online] 2017. Available from: https://www.stratoscale.com/blog/cloud/scalability-cloud-organizations-win-cloud/
25. Wadhwa M, Goel A, Choudhury T, Mishra VP (2019) Green Cloud computing-a greener approach To IT. In: International conference on computational intelligence and knowledge economy (ICCIKE), pp 760–764

Design of Smart Mirror

A. Hameedhul Arshadh, I. Dinesh, S. Ganesan, and J. Friska

Abstract People are spending a considerable amount of time in front of the mirror every morning. We can make this time a lot more productive if we can show the information like appointments and news for the respective day during this time. This is the concept behind the smart mirror. The design of smart mirror is implemented using Raspberry Pi, MagicMirror^2 API, and Google Assistant API. There are many smart speaker devices are available in the market. The main drawback is that it only has audio output which makes it quite difficult to use. To make it more usable, we are adding a visual output with an LCD screen to the existing smart speaker system. This screen is placed behind the glass which is covered with a one-way acrylic sheet. This makes the setup to blend in with the environment. The external frame of the mirror is made out of a wooden chamber. Normally it just looks like a regular mirror but the difference is that it displays details regarding weather forecast, temperature, time, and web applications, calendar, schedules, etc., in the mirror itself. The Raspberry Pi connects to the Internet and retrieves all the necessary data from the MagicMirror^2 platform then it is displayed on the mirror. To make this mirror more functional Google Assistant is added to the mirror. To interact with the mirror a set of speakers and a microphone are added to the mirror. This adds a whole bunch of features like making conversations with mirror, getting smart answers for questions, controlling smart IoT devices over a Wi-Fi network, and many other features. This way we can get all the benefits of a smart speaker while eliminating its biggest problem.

Keywords Smart home · Smart mirror Google assistant · MagicMirror · Internet-of-Things

A. Hameedhul Arshadh (✉) · I. Dinesh · S. Ganesan · J. Friska
Department of Electronics and Communication Engineering, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_11

1 Introduction

1.1 Smart Home

Please Smart home is the common theme that is taking center stage in the modern days. Smart home technology, also known as home automation technology which provides homeowners a whole host of features like security, comfort, convenience, and energy efficiency by allowing them to regulate smart devices, often by a sensible home app on their smartphone or other networked devices. Smart home devices are working with the Internet-of-Things (IoT) to provide the above-mentioned features.

1.2 Origins of the Smart Home

Since the release of X10 in 1975, a communication protocol for home automation devices. This kick-started the whole idea of the smart home. X10 used 120 kHz radio frequency (RF) bursts of digital information onto the existing electric wiring of the home to programmable outlets or switches. These signals contain commands and convey the commands to the corresponding devices. These commands are used for controlling how and when the devices operate. X10 was a one-way technology when it was debut, which made a scenario where the smart devices can actually take commands but cannot reply back to the central network. Later two-way X10 smart home devices became available which allowed a two-way communication between the smart devices and the central network. Then Zigbee and Z-wave protocols made its debut which made it convenient to create smart home devices. Then Google and Amazon came up with their smart home AI-enabled platforms. These platforms made a revolution in smart home devices. These platforms enabled the devices to use the power of AI to manage smart home devices which made it very convenient to use. Since then the number of smart home devices is increasing at a rapid pace. Nowadays, Google Home and Amazon Alexa are the benchmarks of smart home devices.

1.3 Smart Mirror

Smart home products like Google Home and Amazon Echo have been very successful over the past years. But they haven't been updated over the period of these 5 years. Now Lenovo is taking action to create a smart home device with Google assistant with an integrated display that looks like a tablet computer. These tablet-like structures don't blend well with the home environment. In order to solve these issues, smart mirror concept is proposed. Normally we spent somewhere between 20 and 30 mins Infront of the mirror every day. Smart mirrors make these 20–30 mins even more productive by showing important details in the mirror itself. Smart mirrors hide the

display cleverly under a mirror in order to make it blend in with the environment. At normal glace, the setup will look like a normal mirror which can be used for all the things that a person normally does with a normal mirror. But due to the added smart features and display underneath the mirror, it can show various details when needed.

1.4 Applications

Smart mirror has so many applications which make it superior compared to its smart speaker counterparts. This makes the smart mirror more likable than its smart speaker counterparts.

Some of these applications are listed below

- Greeting the person with a message
- Controlling smart devices over a Wi-Fi network
- Giving daily newsfeed in the mirror
- Giving date and time details in the mirror
- Showing reminders, appointments, etc. in the mirror along with Audio and video streaming.

2 Existing System

There are very few smart mirror systems that come close to the feature set offered by our smart mirror. One such system is offered by Lenovo. It has all the features such as Google Assistant integration and display for media streaming purposes. But it falls short because of its form factor. As Lenovo's offering is in the shape of a tabletop tablet computing device. Lenovo does this to ensure the portability of the device but due to form factor, it couldn't quite blend in with the environment that it is placed. Another major disadvantage of this being its price. It comes at a price of 15,000 INR which makes it the most expensive smart home device available in the market. Due to this its sales figures are pretty low.

Most of the other systems offered by top vendors like Google and Amazon come in smart speaker format. They aren't be equipped with a display. They have a processor and a speaker in their setup. Due to their complex AI algorithms, voice recognition from both Google and Amazon is very accurate. They are constantly updating their algorithms to improve their service. They can be used for streaming audio directly from online platforms like play music, Spotify, Apple music, etc. but they can't be used for streaming videos as they don't have a screen. The only way to interact with the speakers is by using the voice. The outputs will also be given in the audio format as well. If the person fails to hear any words from the assistant, once again they have to start the process from the first. This plays as the major disadvantage of these kinds of smart speaker devices. One major benefit with these kinds of smart speaker

devices is that it is relatively cheap and it serves as a good Bluetooth speaker as well. A good smart speaker can be purchased from 4000 INR.

There are some smart mirrors that are made by people as a DIY project. They lack the finesse and functionality compare to other offerings from top vendors. This is due to the less sophisticated proprietary AI algorithms used in these mirrors. Another big disadvantage is that many of these vendors don't update their algorithms as they are not for commercial purposes.

3 Proposed System

In our final design, we use Raspberry Pi, Google assistant API, and MagicMirror^2 API to create the smart mirror. We have included a display mounted behind the mirror. This allows the time, date, and weather to be displayed constantly in the mirror along with a screen that scrolls through the headlines. These details are directly fetched from the MagicMirror^2 API. These details will be constantly updated by the MagicMirror^2 developers. Raspberry Pi is running on Raspbian OS which is a Linux-based OS developed specifically for Raspberry Pi hardware. The integration of Google assistant in the mirror makes it efficient in terms of speech recognition. This also provides many features such as using the regional language for interacting with the mirror, to give reminders based on the events that you have stored in your Google calendar, and so on. Actual physical construction of the mirror has a wooden frame and a two-way mirror. The two-way mirror film-coated glass creates a durable and functional mirror. Speakers and microphones are added to the mirror setup to provide VUI. Speakers are connected using 3.5 mm audio port available in the Raspberry Pi. The USB microphone is plugged into any one of the USB ports available in the Raspberry Pi. The power for the whole setup is taken from the standard wall plug (240 Volts). As this has media streaming facility in it, you can directly stream media from your smartphone to the mirror. Overall, the design that was created serves to create an environment that promotes efficient daily preparation, allows easy access to weather information and news headlines, and provides a user-friendly and enjoyable mirror experience. As we are using open-source hardware and software, the cost of the device is kept low.

The major advantage this has over others systems is that we are using Google assistant API which is an AI-enabled platform and has the most accurate speech detection among all the AI-enabled API platforms available. Similarly, MagicMirror^2 is one of the best smart mirror open source API platforms available. This is pretty evident from the literature survey that we have done. So, we are using MagicMirror^2 API over its competitors.

4 Block Diagram of the Proposed System

The Raspberry Pi acts as the processor that takes care of all the processing required. Raspbian OS is booted in the Raspberry Pi (Fig. 1). To provide the mirror with smart features like asking queries, making conversation with the mirror, etc., we are using Google Assistant API. This Google Assistant API is loaded in the Raspberry Pi by creating and running the API in a Python Virtual Environment. To show the details like time and date, we are using MagicMirror^2 API. This provides a high contrast background which helps greatly in providing the mirror effect. User can interact with the mirror by using voice commands. Speech signal is taken in real time by the microphone and given to the speech-to-text algorithm. This speech-to-text algorithm is a part of the Google Assistant API. This algorithm converts the speech signals into words. Google Assistant matches these words with the words in the database to provide an intelligent answer for the queries. The answers are converted back to speech signals and played through the speaker.

4.1 Design Flow of the Proposed System

The design flow is explained in the above-given diagram (Fig. 2). The whole process starts by installing the operating system on Raspberry Pi. There are many operating systems are available for Raspberry Pi but we are using Raspbian OS for its ease of use and native support for many APIs. To run Google assistant we need to create a python virtual environment within which the Google assistant will run. Once the python virtual environment is created then we can install the Google assistant API in the Python virtual environment that we have created. Then to use the VUI we have to enable access for the Google assistant to make use of the microphone and

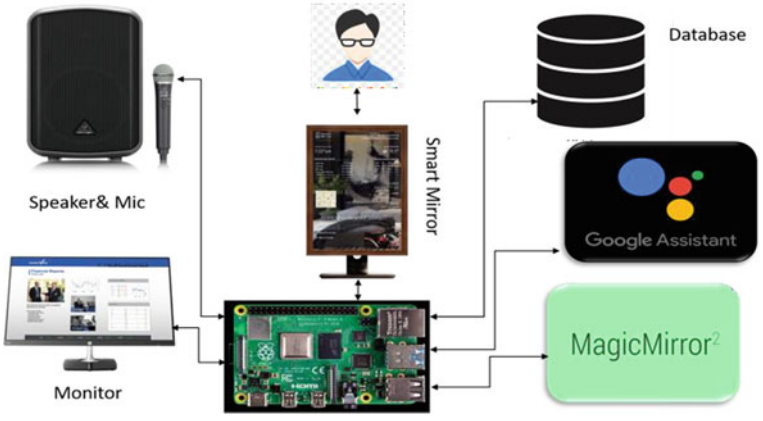


Fig. 1 Block diagram of the proposed system is given above in the picture

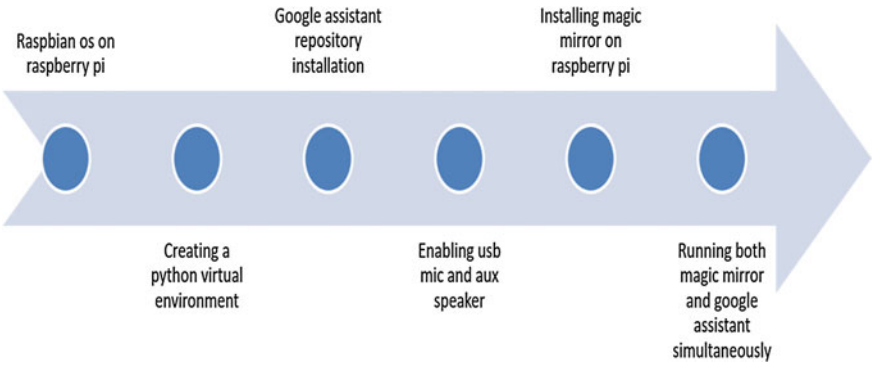


Fig. 2 The above figure shows the design flow of the proposed system of smart mirror

speaker that we connected to the Raspberry Pi. Once the permission is enabled, we have to check whether the microphone and speaker are working correctly. If they are working correctly, we can install MagicMirror[2] API in the Raspberry Pi. Then we have to make it run simultaneously whenever the Raspberry Pi boots up. This can be done by making file with the commands needed to run both these APIs while booting up and invoking the file automatically whenever the Raspberry Pi boots up.

4.2 Flow Chart of the Operation of Google Assistant

Google assistant will not listen to the environment actively (Fig. 3). We must manually trigger the Google assistant. To trigger Google assistant a hot word is used. Generally, this hot word is hey Google or ok Google. Google lets us customize these hot words. But we are going with the default hot words. When the user says the hot word, it triggers the Google assistant to listen to what the user is saying. Then speech to text algorithm comes into play. This converts the speech signals into texts. Right after this conversion is done Google assistant searches its database and web for the query raised by the user.

First, it searches in the database to provide answers. If the answer to the query is not found in the database, it immediately initiates a Google search for the query on the Internet. Once the answer to the query is found, it dictates the smart answer for the query to the user with the help of text to speech algorithm and a speaker. After dictating the whole answer, it waits for the user to say the hot word again. This is how Google assistant provides smart answers to user queries.

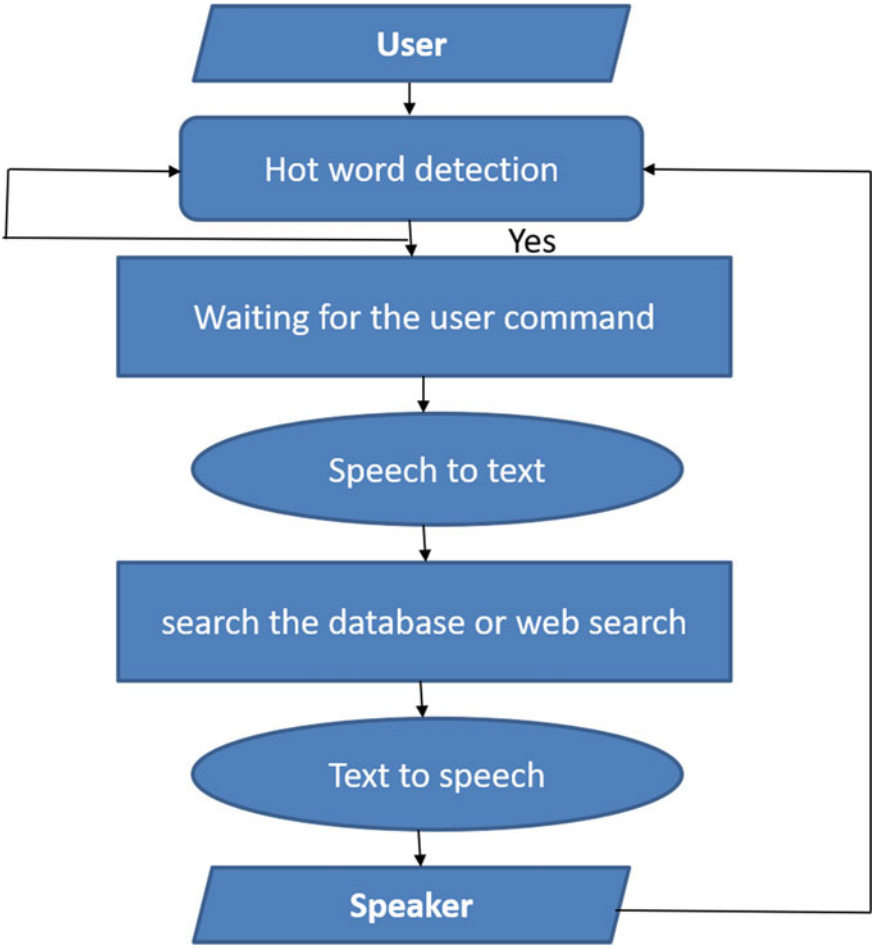


Fig. 3 Above figure shows the way the Google assistant's operation for queries

4.3 Context Awareness of Google Assistant

Google assistant also has a special feature called context-awareness. When the user asked a query about something and he wants to further queries on the same context, Google assistant smartly detects this and instead of starting the process from the first, it can pick it up from where it left off.

For example, if you want to know about Mr. Gordon Ramsay, the user can say "Ok Google, tell me something about Gordon Ramsay". If you want to know the height of him, we don't have to start the query from the first like "How tall Gordon Ramsay is?" instead of that we can directly ask "how tall is he?". Google assistant automatically detects the fact that you are asking about the same person whom the

user have initially asked about. These kinds of smart features makes Google assistant a perfect choice for the smart mirror.

5 Components Used

These hardware and software components are combined together to make a functional smart mirror.

5.1 Raspberry Pi

Raspberry Pi is a card size computer. The entire single-board computer measures just 3.35 × 2.2 × 0.67 inches on its length, breadth, and height, respectively. It has built-in wireless connectivity with Wi-Fi and Bluetooth 4.1. Raspberry Pi can be powered by using a 5 volts 2.5 Amps USB adaptor. The Raspberry Pi 3 B+ has a quad-core Broadcom BCM2837B0 and a Cortex-A53 processor clocked at 1.4 GHz. It has full 64-bit support and offers more power than previous iterations of the pi. It has a built-in 1 GB of ram and can be used with microSD cards varying from 8 GB to 32 GB. It has USB ports to connect peripherals like keyboard and mouse with the Raspberry Pi board. It also has an HDMI port for connecting displays with the Raspberry Pi board. This Raspberry Pi is used as the processing element of the smart mirror.

5.2 LCD Display

We need to show the information directly on the mirror itself. For this purpose, we are using an LCD display. OLED displays are the perfect solution for these kinds of smart mirrors because the unused pixels remain off which saves power and provides deep blacks. But they are very expensive compared to its LCD counterparts hence we are using an LCD display. LCD display is connected with the Raspberry Pi board using the HDMI port available in the Raspberry Pi board.

5.3 Speakers

Smart mirror uses VUI. In a VUI the resultant output for any query is in the form of an audio reply. To playback this audio reply we need speakers. Any normal speaker that use 3.5 mm audio port can be used. The speaker must be small enough to fit in

the mirror while being enough loud so the person near the mirror can be able to hear the playback properly.

5.4 USB Microphone

The input to the mirror is given using speech commands. To get the speech commands from the user we need a microphone. We are using USB microphone for better clarity and ease of use.

5.5 One-Way Mirror and Wooden Frame

The 1-way mirror sits on top of the LCD display. To create a 2-way mirror and 2-way acrylic sheet is stacked on top of the glass. A two-way mirror is a glass that is reflective on one side and clear on the other, giving the appearance of a mirror to those who see the reflection but it allows the people on the clear side to see right through, as if at a window. The wooden frame acts as the housing that keeps all the elements together.

5.6 Google Assistant API

Google Assistant is an AI-enabled virtual assistant developed by Google. It is available on mobile and smart home devices. Google Assistant is available in so many languages from English to most of the Indian regional languages. Google is now letting the developers to create devices that can use the Google Assistant services through Google Assistant API.

5.7 MagicMirror API

It is an open-source modular smart mirror platform. It is free and maintained by a big group of enthusiasts. It also allows third-party developers to build additional modules. There are so many smart mirror APIs are available. But MagicMirror^2 API is regarded as the best for smart mirrors made with Raspberry Pi. It can help to show the details like real-time weather, news feed, calendar, and time.

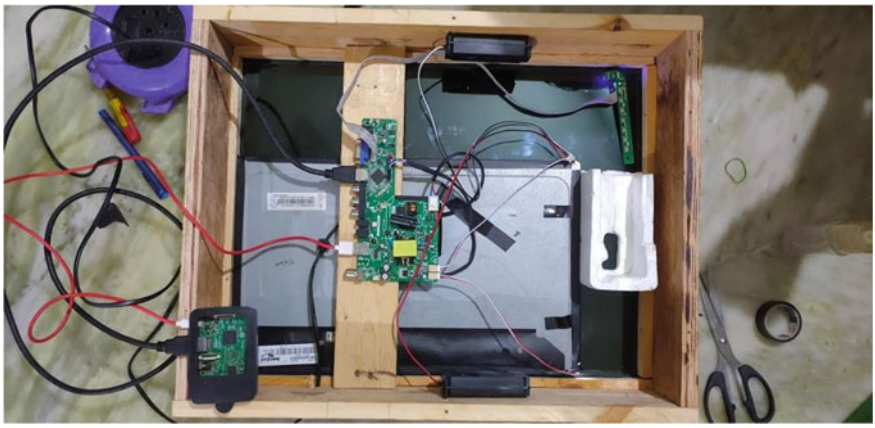


Fig. 4 The above diagram shows how the hardware components are placed inside the frame of the mirror

5.8 Raspbian OS

Raspbian OS is a Linux-based operating system specifically made for Raspberry Pi hardware. It helps to run the Google Assistant API and MagicMirror^2 API to run on the Raspberry Pi hardware.

5.9 VNC Viewer

VNC viewer gives the access to control the Raspberry Pi remotely from other devices like computer or mobile phones. Once the device is placed in the frame, we can't change much in the mirror. In case we want to change or update the software running on the Raspberry Pi we can use this VNC viewer to take manual control over the smart mirror (Fig. 4).

6 Installing the Software Components

6.1 Installing Raspbian OS on Raspberry Pi

Raspbian OS is the operating system on top of which the MagicMirror^2 and Google Assistant API are going to run. Installing Raspbian OS is one of the basic needs to get started with Raspberry Pi. Raspberry Pi uses microSD card as its hard drive.

The entire OS has to be loaded on the microSD card and then inserted into Raspberry Pi.

The microSD card that is selected as the hard drive must have the following minimum requirements.

- The minimum size of the card should be 16 GB
- To have higher read and write speeds class 10 type microSD is preferred as it can provide speeds up to 10 MB/s.
- microSD card should be formatted before loading the OS on the microSD.

Once the microSD card is formatted, it is ready to boot up with Raspbian OS. Initially, we have to download the latest version of the Raspberry Pi Imager and install it on a PC. Then we have to connect the microSD card to the computer through card reader. Once the memory card is detected in the Raspberry Pi Imager, we have to select the Raspbian OS from the given list. Then we can start the loading process by using WRITE option given in the Raspberry Pi Imager. Once the writing process is completed, the microSD card is ready to plug into the Raspberry Pi. Now, we have to insert the microSD card into the Raspberry Pi board. Then the Raspberry Pi will be booted with Raspbian OS.

6.2 Installing Google Assistant API

Installing Google assistant on a Raspberry Pi host is the most important and time-consuming process in building a smart mirror. To successfully boot up Google assistant in a Raspberry Pi host the following 6 step process is used.

- ***Registering for the Google API***

First, we have to register for the access to Google Assistant API. This is done by logging into the Google assistant website and registering the device on the site.

- ***Setting up your Audio for Google Assistant***

We can interact with the Google Assistant using voice commands. Hence this has to be setup before downloading Google assistant on a Raspberry Pi host. In this step, we are setting up the audio input and output through microphone and speakers, respectively.

- ***Downloading and setting up Google Assistant***

Now we have to download the Google assistant repositories into the Raspberry Pi host. Before this, a python environment is created in the Raspberry Pi host for the Google assistant to run on. To access the Google repository each time we are asking queries we have to give the authentication credentials given to us while registering the model.

- ***Authorizing your Raspberry Pi for the Google Assistant***

To access the Google repository each time we are asking queries we have to give the authentication credentials given to us while registering the model.

- ***Using the Google Assistant on the Raspberry Pi***

Then by using the push-to-talk method Google assistant is checked with various queries to check whether the system has any bugs or not. If everything is working fine then we can move on to the final installation process.

- ***Getting the Google Assistant to start on start up***

We want Google assistant to toggle on whenever the device boots up. This is achieved by saving the initiation push to talk command in a file and invoking the file during boot up by including it in the boot up folder.

6.3 Installing MagicMirror² API

Installing MagicMirror² API on a Raspberry Pi host is very easy. Due to this easy installation and feature set, it is preferred over other smart mirror APIs. We have to open the terminal on the Raspberry Pi host. Then the following command is typed.

bash -c "$(curl -sL https://raw.githubusercontent.com/MichMich/MagicMirror//installers/raspberry.sh)".

Once this command ran successfully, we have a system that has magicmirror² API on a Raspberry Pi host. Now we have to create another file with commands to run this API during the bootup just like how we did for Google assistant. Now we have a system that has both Google assistant and MagicMirror API running on it.

6.4 Installing VNC Viewer

Installing VNC viewer on a Raspberry Pi host is a very simple and straightforward task. The executable file to install VNC viewer on a Raspberry Pi host is available on the official website of VNC viewer. All that we have to is to download the VNC viewer in the Raspberry Pi host and click run. To control the Raspberry Pi host through VNC viewer the controller system must also be installed with VNC viewer along with both the devices connected to the same Wi-Fi network.

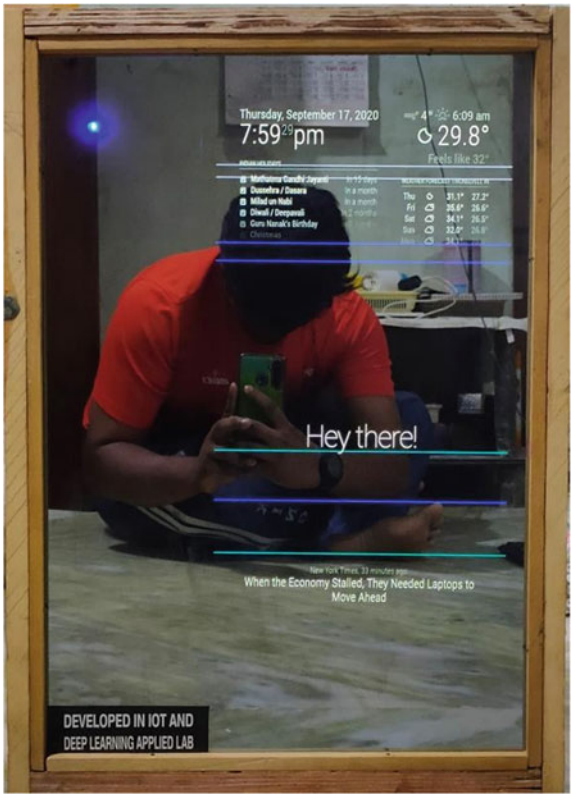


Fig. 5 Above figure shows the smart mirror in operation

7 Result

The mirror screen shows the details such as current date and time (Fig. 5), weather report for the day, weather forecast for the upcoming week, live news, greeting message, and upcoming holidays (Fig. 6).

The operation of the mirror is uploaded in the given YouTube link below: https://youtu.be/VJzVHWJt3Xo.

8 Conclusion

The Smart Mirror designed in this project will provide the user with an enhanced mirror experience. By making use of multiple displays, the user can stay updated on the time, weather, and news headlines while preparing for the day with the fully functional Smart Mirror. Although there are other smart mirror technologies that are

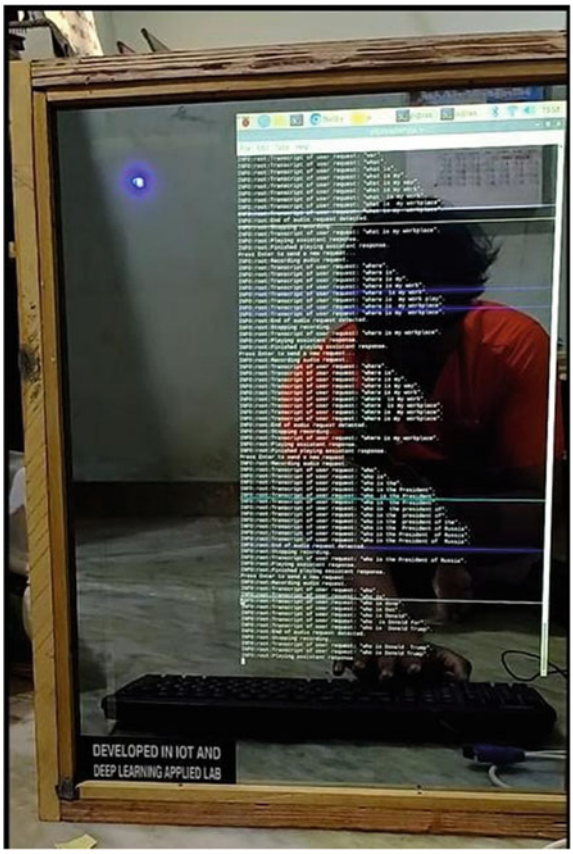


Fig. 6 Above figure shows the operation of Google assistant for the queries we asked

available, the Smart Mirror created in this project stresses saving cost and flexible usage. Through an easy-to-use interface, the mirror can be easily setup to display data that conforms with their desires. The mirror is able to connect to the Internet and parse the proper data to display. The integration of Google Assistant means it can answer various queries with smart answers and also can do Google search whenever the user demands it. While the Smart Mirror will need to be more polished and have a few changes like adding waterproofing before it can be a viable product to be sold, but the Smart Mirror made in this project meets all the design goals set forth before the project and has all the elements that would be needed for a fully functional Smart Mirror product.

Bibliography

1. Balon B, Simic M (2019) Using Raspberry Pi computers in education. In: 2019 42nd International convention on information and communication technology, Electronics and Microelectronics (MIPRO). IEEE
2. Ok F, Can M, Ucgun H, Yuzgecm U (2017) Smart mirror application with raspberry pi. In: 2017 International conference on Computer Science and Engineering (UBMK). IEEE
3. Mukhopadhyay K, Sinha C, Saha HN, Rakshit S, Auddy S (2018) Smart Mirror – a secured application of artificial intelligence recognizing human face and voice. In: 2018 IEEE 9th annual information technology, electronics and mobile communication conference(IEMCON). IEEE
4. Maksimovic M, Vujovic V, Davidovic N, Milosevic V, Perisic B (2014) Raspberry Pi as internet of things hardware: performances and constraints. In: 272175660. ResearchGate
5. Yusri MM, Kasim S, Hassan R, Abdullah Z, Rusla H, Jahidi K, Arshad MS (2017) Smart mirror for smart life. In: 978-1-5090-6255-3/17/$31.00. IEEE
6. Mathivanan P, Anbarasan G, Sakthivel A, Selvam G (2019) Home automation using smart mirror. In: 2019 IEEE international conference on system, computation, automation and networking (ICSCAN). IEEE
7. Jagdish AP, Sadgir Sonal T, Sangaleharshada D, Dokhale A (2018) A review paper design and development of a smart mirror using Raspberry Pi. Int J Eng Sci Invent 7(4) version 1
8. Akshaya R, Niroshma Raj N, Gowri S (2018) Smart Mirror – digital magazine for university implemented using Raspberry Pi. In: Proceedings of 2018 international conference on emerging trends and innovations in engineering and technological research (ICETIETR). IEEE, 978-1-5386-5743-0/18/$31.00
9. Sun Y, Geng L, Ke D (2018) Desing of smart mirror based on Raspberry Pi. In: International conference on intelligent transportation, big data & smart city. IEEE, 0-7695-6368-6/18/$31.00
10. Bai Z, Blackwell AF (2013) See-through window vs. magic mirror: a comparison in supporting visual-motor tasks. In: 2013 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE

Smart Villages: Application of Internet-of-Things in the Development of Rural India

Bushra Sabir, Arsalan Hameed, Mohd Faizan, Vibhav Kumar Rawat, and Mohd Tariq

Abstract In rural development Internet of things have great importance as future technologies depend on the Internet of things (IoT) as it helps in interfacing all communication technology with each other as the population is increasing year by year so, it is becoming a big task to run the functions smoothly, with the help of IoT our tasks turn out to be so simple and efficiently such as its first function is to collect the large data from both sensor and remote devices, the second function is to evaluate the receive data and perform suitable actions accordingly. This paper mainly focuses on the initiative about smart villages, by studying their problems in villages and then discussed the solution for those problems. Moreover, it ensures the improvement in the excellence of life in villages by using smart technologies by providing two-way communications with the help of IoT. This small step can help the peoples of rural areas for development and it is the recognized fact that if our rural areas get established then the whole nation will be truly developed.

Keywords Arduino Uno · GSM module · Web Application · IOT · Strain gauge · Ultrasonic sensors

1 Introduction

The main feature of IoT is to set out the worldwide network of some trillions of objects collected from the worldwide context which can be easily communicated with the end-user where it can be used with smart devices by the internet and can solve many problems regarding farming [1, 2]. Therefore, IoT is defined as it is the network of physical objects or "things" embedded with electronics technologies, software, sensors, and network connectivity that enables these objects to collect and exchange data. A "Thing" in the context of the Internet-of-Things is a physical or

B. Sabir · M. Faizan · V. K. Rawat · M. Tariq (✉)
Department of Electrical Engineering, ZHCET, Aligarh Muslim University, Aligarh, India
e-mail: tariq.ee@zhcet.ac.in

A. Hameed
Department of Computer Engineering, ZHCET, Aligarh Muslim University, Aligarh, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_12

substantial entity that has a distinctive identifier and the ability to transmit data on a large network. The things consist of web-enabled devices that collect, send, and act on data that is getting from sensor, processor, and communication hardware either by Bluetooth module (operates from several meters away from the setup) or by ESP-8266-wifi module (operates from a remote location). These devices are often called smart devices, Users can communicate with these devices to set them up give them instructions or access the information, but the majority of the work is ended by the device on its behalf. The Internet-of-Things life cycle can be understood in four stages which are "collect", "communicate", "Analyze", and "Act", firstly devices and sensors collect data from everywhere such as from our homes, offices and secondly sends information through the network to a respective destination such as to a cloud platform, a private data center, and a home network and after this, the third stage is to examine the information from the data such as visualizing the data, building reports, and the last stages are to perform a suitable action such as communicate with another machine, send a notification (SMS, email, text message, alarm) and communicate with another system. Similarly, IoT connects all the possible devices either hardware or software [1]. India is a developing country wherever the population is jumping up, because of this the prerequisite grown-up for the use of accessible assets in the finest feasible way [3–5].

2 Smart Irrigation System

Agriculture is a major concern in India because more than half of our economy depends on farming and in agriculture water plays a major role and necessitates incredible quality [5]. So, irrigation system is necessary for agriculture especially in summer where water scarcity is more, they do not depend on the rainy season for irrigating the crop many methods are used such as sprinkler system, drip irrigation, ditch irrigation, etc. due to these types of irrigation water wastage is normal. Sometimes surplus water or less water is given to the soil is a big difficulty. Due to these issues, a smart irrigation system is introduced in the existing one, smart means which is very well known familiar with the environmental aspect which affects agriculture [6–9]. In smart irrigation system, wireless sensors are commenced in different modes such as soil moisture sensor (which deal with the quantity of moisture present in the soil), temperature and humidity sensor (which detects the temperature and humidity of the surrounding) along with these sensors a microcontroller Arduino Uno and WIFI module ESP-8266 are used. Arduino Uno continuously check the temperature, soil moisture level from the given threshold value moreover, it adjusts consequently and the WIFI module is interfaced with the microcontroller which provides communication between the system and user by sending a message via SMS or by IOT android app which sends minute to minute live information to the farmer related to their crops with the help of this data, for example, the moisture content is little, the temperature near surrounding goes high microcontroller sends the signal the farmer can switch "ON" the motor accordingly with the help of relay even a remote location

after receiving the signal through the controller. At this moment in a time irrigation system in a village characteristically involves human contribution, primarily the farmers, at each phase of estimation of constraint and process of an electric pump. The anticipated system automates irrigation through conventional motor using sensors and enables distant monitoring. Figure 1a shows the basic idea of the smart irrigation

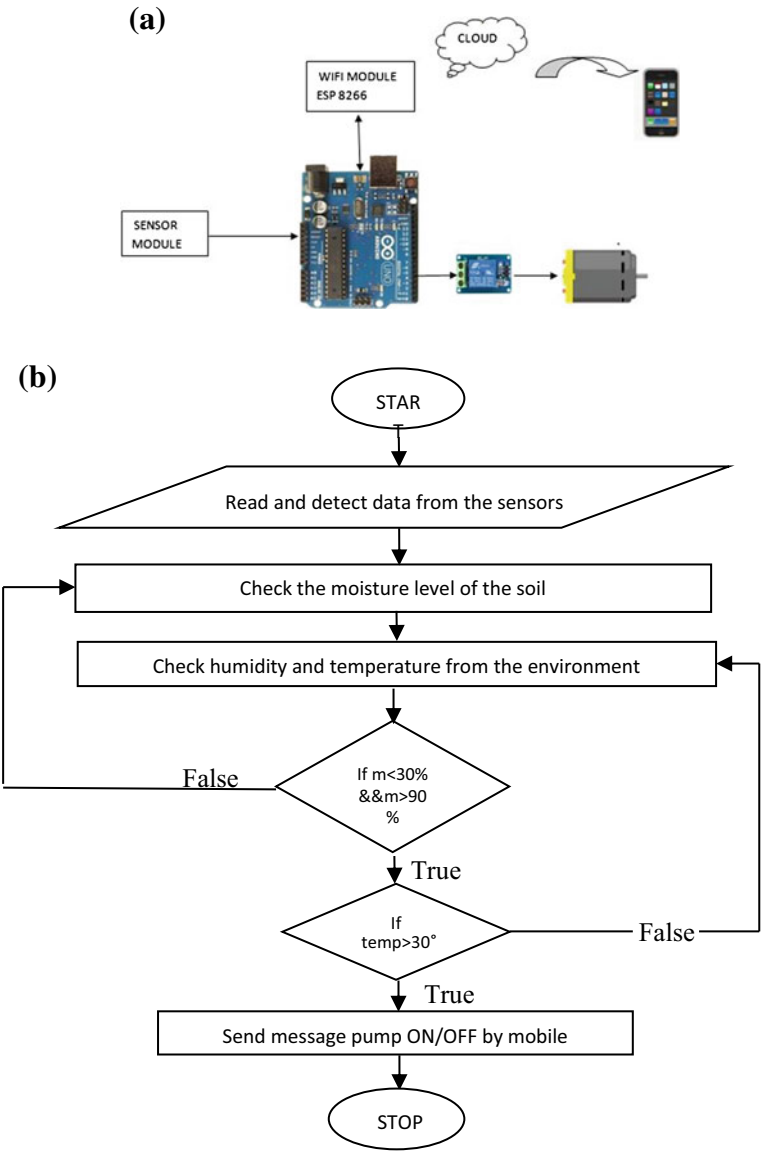


Fig. 1 **a** Smart irrigation system model. **b** Flowchart for smart irrigation system model

System and Fig. 1b shows its flow chart.

3 Solid Waste Management

Due to rise of population among the city's states, people face many environmental related problems with increasing waste day by day, waste disposal is also a very big apprehension for making villages clean as dustbins kept in open places are floating as the waste amount or people dump their waste disposal in open areas without knowing its terrible impact [10, 11]. Because of this disposal in open areas the chances of increasing diverse varieties of diseases and infections may be increased, proper management is required to keep away from some deadly diseases the requirement of a waste management system is required for the villages as it will become a solution which alerts via SMS via GSM module is fitted at dustbin hence the waste collector department gets informed and they will send waste collector vehicle to that respective location. This application helps in remote monitoring of the real-time bin status by using two sensors one is a waste-filled level and another one is a weight sensing sensor. The Waste filling levels inside a bin are determined by the time taken by the ultrasonic pulse to react back to getting reflected from waste material. The weight estimation will be done by placing a conductor material whose resistance will be change when its length changed was it subjected to applied stress and strain, four strain gauges will consist of a wheat stone bridge so when the value of resistance of average strain changes and the bridge becomes unbalanced and corresponding to its output voltage is changed in same proportion and this voltage will be equal to the weight of the waste, Fig. 2a shows the architecture of smart solid waste management and, flowchart is shown in Fig. 2b. This architecture is divided into two parts one is hardware part another one is software which links both hardware and software communicate with the help of GSM module we can also use the WIFI module.

4 Cattle Farming

Dairy Farming is considered to be the subsequent most significant profession of a farmer. The developed nations have more demand for milk in comparison to developing nations. To fulfill this increased demand for milk products improved technologies are required, to get better and healthier yield and this is possible by the use of IoT-based techniques. It can overcome different traditional farming challenges and increases milk. Last few years, milk has become a product with heavy price fluctuations as it is the most demanding product in the market.

IoT facilitates farmers with wearable sensors devices to get a variety of data sources such as data about feeding, calving, nutrition, insemination, and the milk production process. The sensor-based system detects the illness of the cattle correctly before milk production gets affected. The sensor can be placed in neck tail, or leg

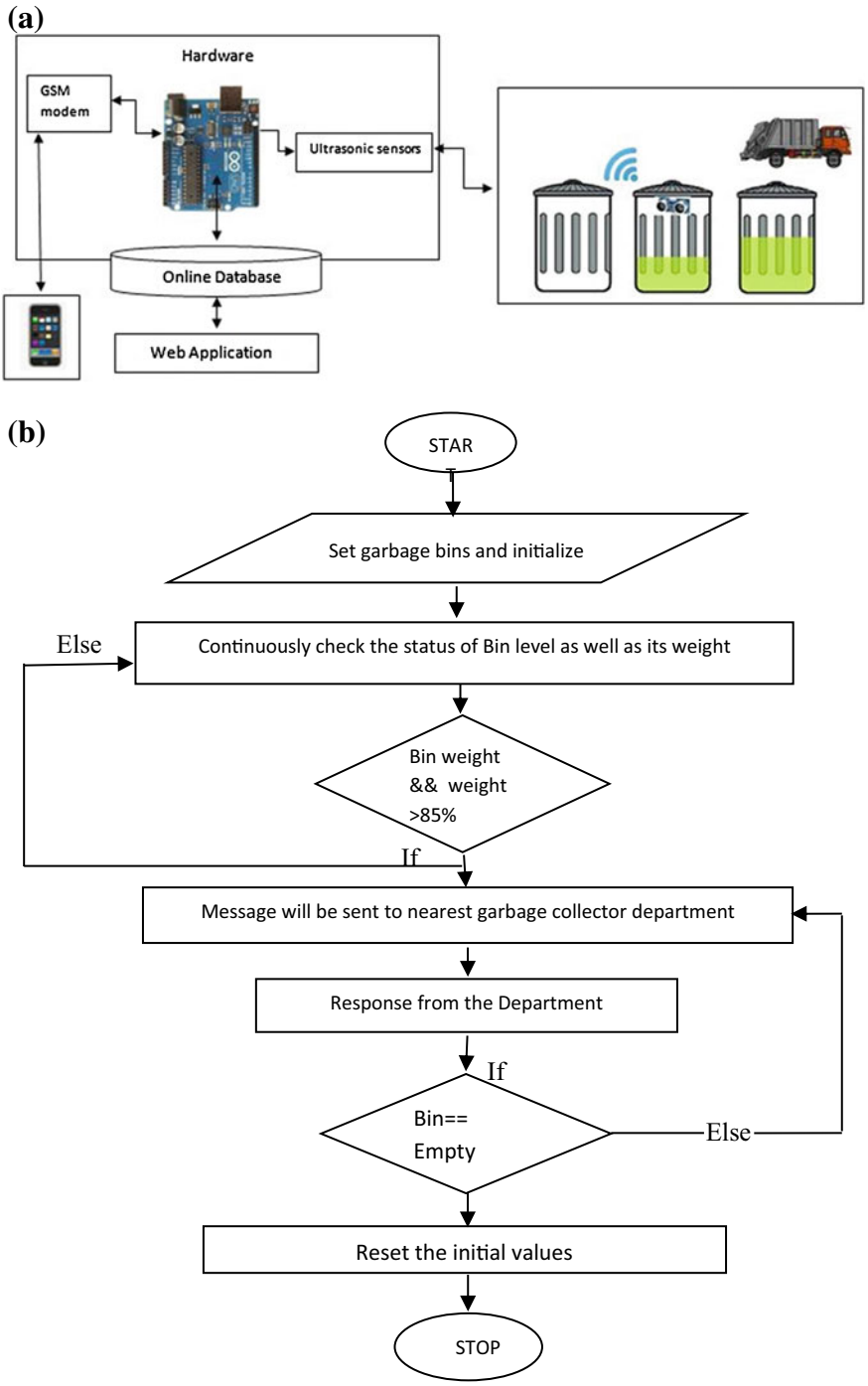


Fig. 2 **a** Architecture of solid waste management. **b** Flowchart of solid waste management

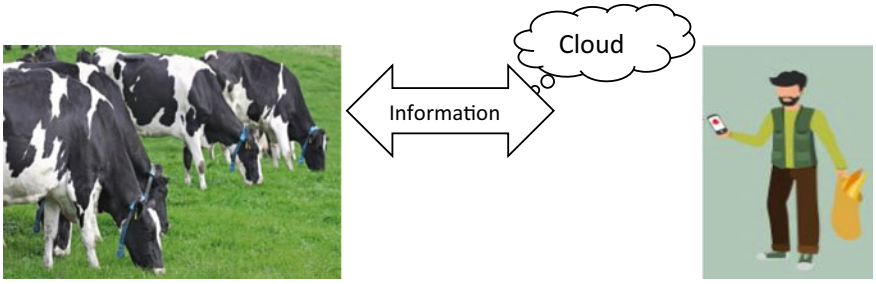


Fig. 3 Smart cattle farming using IoT

of the cattle to examine various activities of cattle such as health, feed consumption, fertility management, and milk production and also wearable sensors also detects the body temperature, illness, and disease of the animal-like salmonella, mycoplasma which can affect the milk production as well as human health when they take it. By using such technologies will formulate it simple for the people as all the things are ended involuntarily by the devices as shown in Fig. 3.

5 Smart Education

India is a developing country that has been facing the need to improve its quality of education in rural areas particularly due to inadequate teaching resources, the requirement of funding for school, lack of study material, school infrastructure is outdated [5, 12]. The idea of smart education provides all these facilities, its scope covers the learning modes, activities inside the school or colleges it improves the literacy rate in villages with the help of IoT-based smart classrooms is a way elegant learning, the availability of connectivity such as WIFI couple with mobiles, tablets, laptops, makes education system and communication easier, online learning, video conferencing, likewise IoT have application in smart desk, **connected sports, smart libraries smart content delivery, student tracking, smart attendance record, provide security of both teachers and students, smart workshops, etc**. By use of smart education, students can study smarter without any trouble, it provides:

(a) Interactive learning

To provide a comprehensive outlook to the children in procedure understanding of new things which better acknowledging and interact with each other, i.e., teachers and students, interactive learning plays a vital role as today's learning process has exceeded the limitations as now it is not only restricted to the amalgamation of images and text but much additional than that.

(b) Educational Apps

These kinds of applications play an important role as educational apps, logical game apps which help to enhance logical thinking as well as enhance the knowledge of the students. They are very helpful in learning as it is very portable to carry your device along with you everywhere and one more advantage by learning from these apps is that children become more familiar with smart devices so being ready for future technology.

(c) Increasing Productivity

IoT-based education's advantage is that An important time of study is waste in schools that don't help in the aim of the study, for example, the presence of the students have to be taken particular times a day, besides, this attendance record has to be sent to the main office, here IOT plays as significant role by putting an end to the efficient system as by using IOT-based devices, this data after collection can be sent to the main office directly through servers automatically so it saves a lot of time.

(d) Security

Monitoring all the student activities is not a simple task but using innovations based on the IoT Position of students can be audit anytime and their existence can be divulged at any instant. There may be an option of emerging alarm in any frightening situation also with a camera. This endeavor can help in avoiding accidents.

6 Health Monitoring

Heath is the measure issue that our country is facing these days and in this pandemic the concern has increased a bit more. Urban areas have good facilities for health and have both government and private hospitals. Rural areas lack good hospital facilities. Villages have some state government hospitals but the facilities over there are not so good and if the condition of the patient becomes serious, then they are referred to the nearest city health centers. But in some cases, due to lack of education of villagers they are not able to track their own health and sometimes a very small disease according to their thinking can take their life and they even can't consult with the doctors or consult at their final stages. Using the application of IOT, track can be made on the temperature, pulse rate; heart rate, etc., [13, 14] on the villagers by their village medical officers and each such sensor can be installed by the government at each residential home. Now if anything goes wrong with them, using the android app hospital staff can know the patient's condition and treat on their level and if they find the case out of their reach then they can also refer them to the city. This will save the time of realization of the patient that he or she is ill and will be able to get good treatment within time. A monthly record can also be maintained by the hospital staff of those who are aged and feel uneasy to come to hospital on a routine basis. A sample health monitoring system is shown in Fig. 4.

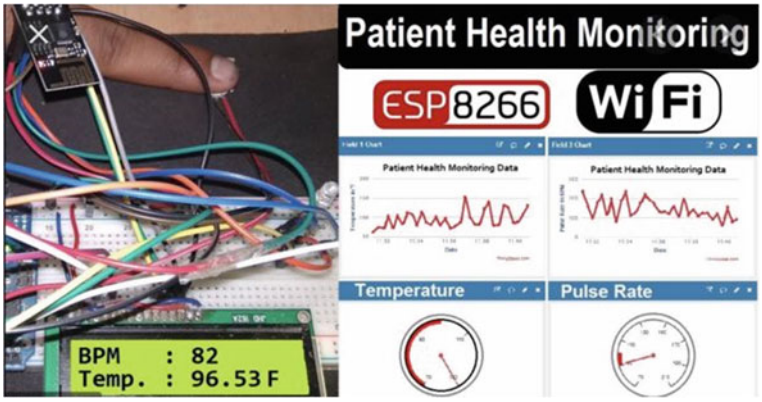


Fig. 4 Health monitoring system

7 Conclusion

A few years back the concept of IoT and Smart villages used to be studious as a future likelihood but it has become realism today, thanks to technical elevations. Many countries have placed the job of revolving their villages into smart villages to many leagues, the excellent use of accessible resources is the requirement of the hour even increasing population has curbed the assets and their usage. IOT combines the interests of various technologies to recognize the idea of smart devices in villages. This idea can be lengthened to cities as well, ameliorate the quality of life of inhabitants. The whole praise of progress is due to the progress in technologies that have been put into the field. We have also research in numerous fields to an internet of things and have given proposals accordingly.

References

1. Farooq MS, Riaz S, Abid A, Abid K, Naeem MA (2019) A Survey on the role of IoT in agriculture for the implementation of smart farming. IEEE Access 7:156237–156271. https://doi.org/10.1109/ACCESS.2019.2949703
2. Chen W et al (2019) AgriTalk: IoT for precision soil farming of turmeric cultivation. IEEE Internet Things J 6(3):5209–5223. https://doi.org/10.1109/JIOT.2019.2899128
3. Tariq M, Yuvarajan S (2013) Modeling and analysis of self excited induction generator with electronic load controller supplying static loads. Can J Electr Electron Eng 4(1):9–13
4. Tariq A, Husain MA, Ahmad M, Tariq M (2011) Simulation and study of a grid connected multilevel converter (MLC) with varying DC input. In: 2011 10th international conference on environment and electrical engineering, Rome, pp 1–4. https://doi.org/10.1109/EEEIC.2011.5874704
5. Tariq M, Shamsi K (2010) Application of RET to develop educational infrastructure in Uttar Pradesh. Int J Recent Trend Eng ACEEE 4:187–190

6. Boursianis AD et al, Smart irrigation system for precision agriculture - The ARETHOU5A IoT Platform. IEEE Sens J. https://doi.org/10.1109/JSEN.2020.3033526
7. Roy SK, Misra S, Raghuwanshi NS, Das SK, AgriSens: IoT-based dynamic irrigation scheduling system for water management of irrigated crops. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2020.3036126
8. Khan R, Zakarya M, Balasubramaniam V, Jan MA, Menon VG, Smart sensing-enabled decision support system for water scheduling in orange orchard. IEEE Sens J. https://doi.org/10.1109/JSEN.2020.3012511
9. Rahman S, et al (2020) Design and implementation of cascaded multilevel qZSI powered single-phase induction motor for isolated grid water pump application. IEEE Trans. Ind. Appl. 56(2): 1907–1917. https://doi.org/10.1109/TIA.2019.2959734
10. Nirde K, Mulay PS, Chaskar UM (2017) IoT based solid waste management system for smart city. In: 2017 international conference on intelligent computing and control systems (ICICCS), Madurai, pp 666–669, https://doi.org/10.1109/ICCONS.2017.8250546
11. Ahmad S, Imran, Jamil F, Iqbal N, Kim D (2020) Optimal route recommendation for waste carrier vehicles for efficient waste collection: a step forward towards sustainable cities. IEEE Access. 8: 77875–77887, https://doi.org/10.1109/ACCESS.2020.2988173
12. Asim M, Tariq M, Mallick MA, Ashraf I (2016) An improved constant voltage based MPPT technique for PMDC motor. Int. J. Power Electr. Drive Syst. 7 (4)
13. Xu G (2020) IoT-Assisted ECG monitoring framework with secure data transmission for health care applications. IEEE Access 8:74586–74594. https://doi.org/10.1109/ACCESS.2020.2988059
14. Pathinarupothi RK, Durga P, Rangan ES (2019) IoT-based smart edge for global health: remote monitoring with severity detection and alerts transmission. IEEE Internet Things J 6(2):2449–2462. https://doi.org/10.1109/JIOT.2018.2870068
15. Sairam AJ, Reddy Induri T, Bagyaveereswaran V (2019) Validation of wearable sensors and RFID for real time monitoring of cattle farming using internet of things. In: 2019 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India, pp 1–4, https://doi.org/10.1109/i-PACT44901.201
16. Andonovic I, Michie C, Cousin P, Janati A, Pham C, Diop M (2018) Precision livestock farming technologies,. In: *2018 Global internet of things summit (GIoTS)*, Bilbao, pp 1-6, https://doi.org/10.1109/GIOTS.2018.8534572
17. Taneja M, Jalodia N, Malone P, Byabazaire J, Davy A, Olariu C (2019) Connected cows: utilizing fog and cloud analytics toward data-driven decisions for smart dairy farming. IEEE Internet Things Mag 2(4):32–37. https://doi.org/10.1109/IOTM.0001.1900045

Online Resale Bike Price Prediction in Indian Market

Bhagirath, Neetu Mittal, and Sushil Kumar

Abstract On the online resale platform customer sell and buy bikes after chatting and physically visiting the bike owner site. The major problem customers face is estimating the realistic price of bike as lots of factors drive the price of resale of bike. Predicting the right price from the historic data is a bit tricky as data of pre-owned bikes is highly complex and noisy in comparison with new product sale data. The existing bike price predicting solutions are based on parameters fixed price range for different parameters of bikes. The proposed work aims at the analysis of resale data of classified platform OLX India and to develop a more accurate pricing model for pre-owned bikes using two machine learning techniques viz. and Decision Tree algorithms and Linear Regression. The evaluated predicted price results have been compared with the actual prices to test the precision of the proposed technique. The change in predicted and the actual prices have been judged by using statistical parameters Mean, Variance and Standard Deviation. The values of these parameters are very low for the Linear Regression technique as compared to the decision tree technique. Thus, a higher degree of closeness of estimated prices with actual prices is ensured.

Keywords Data analytics competency · Bike resale price · Machine learning · Decision Tree Regression · Price prediction

1 Introduction

Online resale market is the new way of selling and buying resale goods online. The online marketplaces such as OLX, Quikr, and others have become the preferred place for selling and buying resale bike online. The price of an old product depends

Bhagirath (✉) · N. Mittal
Amity Institute of Information Technology, Amity University Noida, Noida, India

N. Mittal
e-mail: nmittal1@amity.edu

S. Kumar
UPC Barcelona, Barcelona, Spain

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_13

upon multiple parameters such as age of the product, brand, conditions, and amount of information provided in the portal, etc. Due to multiple attributes influence the price prediction for old goods making it complex work. Even items of the same kind may differ in real conditions. The variable transportation charges, location, seller reliability, quality of product, appearance of listing, brand image, availability of seller/buyer and communication channel, etc., are some of the factors that make the old product price prediction complex. Thus, it is highly complex to quote the best price at which buyer will purchase it. The seller's unsold inventory increases due to the right price estimation for buyers. Having realistic price is win–win condition for both the seller and buyer [1]. The cost of maintaining resale is bit high for all the historical bike sold in past [2]. The competition in seller behaviour is not beneficial for consumers by price rise and product quality reduction [3]. The demand for resource-efficient, energy-efficient, and low-polluting products and with higher resale value has increased in the last few years [4]. Price engine can be made but need to check with good validation technique to prove the efficacy of model [5]. Product manufacturers sufficient efforts in regulating resale price for making monopoly in the market [6]. Price margin can vary with respect to market supply and demand as well as the location of the product [7]. The group pricing discouraging price discrimination [8]. Professional financial advice can help in control efficient plans for selling and buying [9] and also assist to create better-informed financial decisions [10]. The current research into high price item has shown multidimensional which affect the resale price [11]. A large amount of data can be used to build in the distributed environment [12]. Buyer behaviour might be change with respect to gender and social economical factors [13]. The proposed work is to create a resale bike price prediction model for the Indian market by applying machine learning algorithms. For machine learning modelling, it is very important to have sufficient historic data availability. This study utilizes the transactions data of OLX India, one of the major online resale organizations. The multiple models for each bike make have been prepared with high degree of accuracy. The comparison of predicted and actual price provided as output of Linear Regression and Decision Tree algorithm implemented for the research. This paper is organized as follows: Sect. 2 shows the framework of proposed methodology. Resale approaches and data summary have been discussed in Sect. 3 and Sect. 4, respectively. In Sect. 5, algorithms have been discussed and Model implementation in Sect. 6. Results and discussions have been presented in section VII followed by conclusions.

2 Problem Statement and Challenges

India is a big country and bike market is also huge with respect to the percentage of middle-class families. As the primary user of bike for transportation comes from these groups only. Resale of bikes happened in online and offline, both the platforms. There are two types of buyer and seller in both the platforms, business user and end customer. Business users are experience users who buy the bikes in highly negotiated

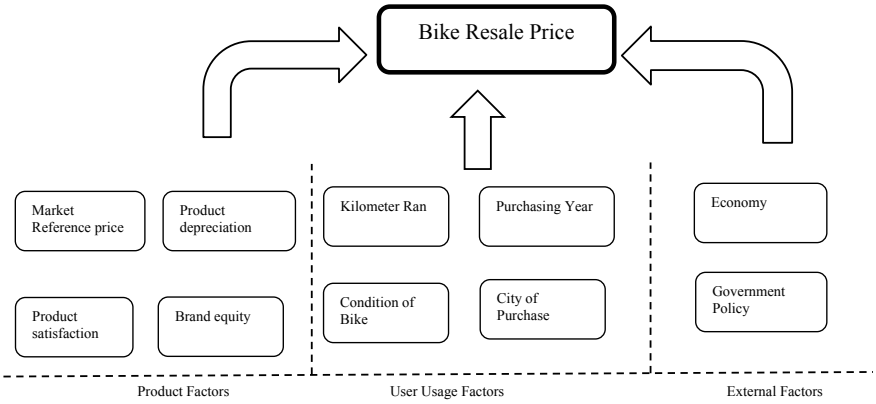


Fig. 1 Block diagram of theoretical model

price and resale it after doing refurbishing it for the sake of earning more margin on sale. End customers are buyers who purchase the bike from other end customer or from business users. Either case end customer could not able to estimate the right price as the resale price of bike influenced by multiple factors. The resale price is influenced by multiple factors such as market reference price, brand equity, product depreciation, product satisfaction, and product logistics. The main source for reference price is an online classified site and refurbished platform such as OLX, Google, etc. When Bike depreciation is high then end-user sets lower resale price. The depreciation is a major factor that influences the bike resale price and condition of the bike. The Bike depreciation can be calculated as per the formula of insurance companies for calculating. We have observed user usage behaviours also play critical part in establishing resale price of the Bike. The condition of the Bike becomes essential before quoting the resale price, it is manual work therefore it also involve time and cost. The brand image also plays a critical role in bike resale price, for good image of the brand it will help in putting high resale by customer. Brand image is driven from the manufactures product quality, value of transaction, after sale service and success in the market (Fig. 1).

3 Data Summary

In the proposed work, the bike listing data of online resale platform has been used. On the online resale platform, there are two types of bike ad data mainly, end customer ad data and dealer ad data (Fig. 2).

Dealer ad data is more authentic and practical as compared to end customer ad listing data. Dealers are businessman who are selling and purchasing Bike on daily basis and they knew from their past experience what would be the ideal transaction closer price after adding their margin. End customers most of the time could not able

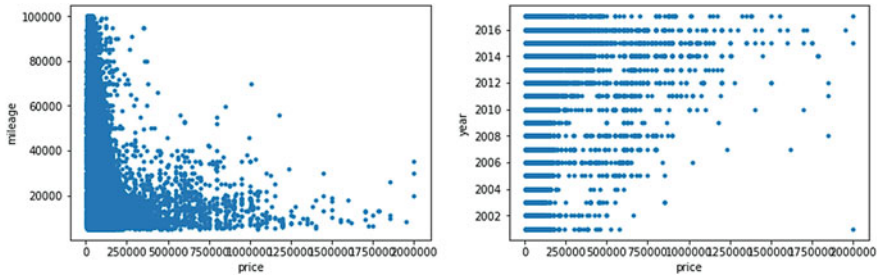


Fig. 2 Bikes price, mileage and year of manufacturing scatter graph

Table 1 Comparison of linear regression and decision tree techniques

	Linear regression technique	Decision tree technique
Number of model created	118	1
Training set	169,285	169,285
Mean absolute error (MAE)	10,717	8751
Root mean square(RMSE)	13,441	11,604
1-MAE/Mean %	80.19	86.91
1-RSME/Mean %	75.15	78.10

to estimate the right resale price while posting their ad and suffer very less queries and reply to their bike ad. Therefore, dealer bike listing data is more appropriate and realist to build any machine learning model for price prediction. The listing data contains key parameters such as make, model, trim, mileage, year of manufacturing, city, locality, fuel type and colours, etc. In Table 1, make, model, year of manufacturing & age of bike are highly correlated, therefore only one variable in the model has been considered as both the variables are providing the same information (Fig. 3).

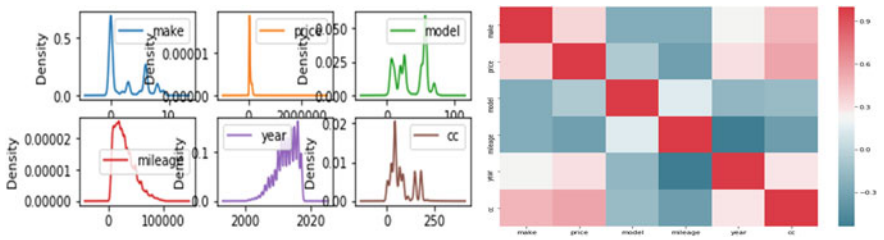


Fig. 3 Variable correlation metrics

4 Machine Learning Algorithms

In this work, 02 machine learning algorithms have been used viz. Linear Regression and Decision Tree Regression algorithms have been used for reselling Bike pricing prediction purposes.

4.1 Linear Regression

The Linear Regression is a basic model for mapping continuous dependant variable and numerous independent variables. In Simple Regression Model, one dependant variable with one independent variable has been used. In case there are several independent variables this is called the Multiple Regression Model. The relationships are modelled using linear predictor function. A modeller should determine the relationship between the independent and dependent variable. Further based on the result, it has been fitted by a linear equation to perceived the data. The least-squares method is the most common method for a tune regression line. This method minimizes the sum of the squares of the vertical deviations from each data point to the line and calculates the best-fitting line for the perceived data. The estimation of new model variable is based on the training data. The least-square method is often used for fitting of Linear Regression models. In Linear Regression model, the data must satisfy certain criteria. If the data does not follow these criteria, then the results could be misleading. The dependant variable has an additive impact in the event of change in the independent variable.

To develop the regression model, the objective is to find a regression line that fits best through the distribution. The line nearest to the points is considered as the best fit. Therefore, the distance between regression line and data points is reduced by adjusting the parameters of the regression line. The original relationship between independent variable x and dependant variable y may be represented by Eq. (1):

$$Y = B_0 + B_1.X + E \tag{1}$$

where B_0 is the constant intercept and B_1 is the coefficient of relationship between X and Y. The differences between actual and predicted values of the dependent variables Y is denoted by E. To reduce the sum of the square of the errors, the Linear Regression line is found by ordinary least square method. By differentiating error with respect to parameters B_0 and B_1 and equating with zero will yield (Fig. 4).

$$\text{Error} = \sum_{i=0}^{n} (Y_i - Y)^2 \tag{2}$$

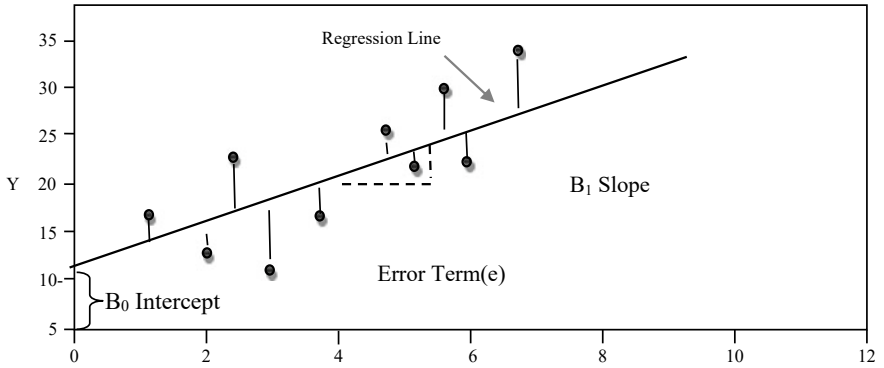


Fig. 4 Regression line, intercept and slope

$$B_0 = Y - B_1.X \text{ and } B_1 = \sum_{i=1}^{n} (Y_i - Y)(X_i - X) / \sum_{i=1}^{n} (X_i - X)^2 \tag{3}$$

4.2 Decision Tree Regression

This is a decision-supporting technique utilizing a model of decisions or tree -ype graph and their possible results. The algorithm sorts in the tree from top to down to classify instances. In the tree, each node corresponds to an attribute. The possible values of attributes are indicated by descending branch. Testing the attributes is specified by this node. The fixed set of attributes describe an instance. It is easiest for the decision tree when each attribute takes on a small number of disjoint possible values. The model can be extended further to real-value problems. The primary task is to select the attribute to be tested at the root. The ability to classify the training data by an attribute determines its performance.

The categorical data is changed into integer data. The model is trained using different hyper parameters to produce the best estimation. The decision tree is made by splitting the source, create the root node into child. The splitting process is based on a set of dividing rules of classification features. This process is recursively repeated for each derived subset. The process is accomplished when splitting provides no value addition to the iteration. The final output is to maximize the information gain at each split is mathematically defined by Eq. (4).

$$IG(D_p, f) = I(D_p) - \left[\frac{N_{\text{left}}}{N_p} I(D_{\text{left}}) + \frac{N_{\text{right}}}{N_p} I(D_{\text{right}}) \right] \tag{4}$$

where

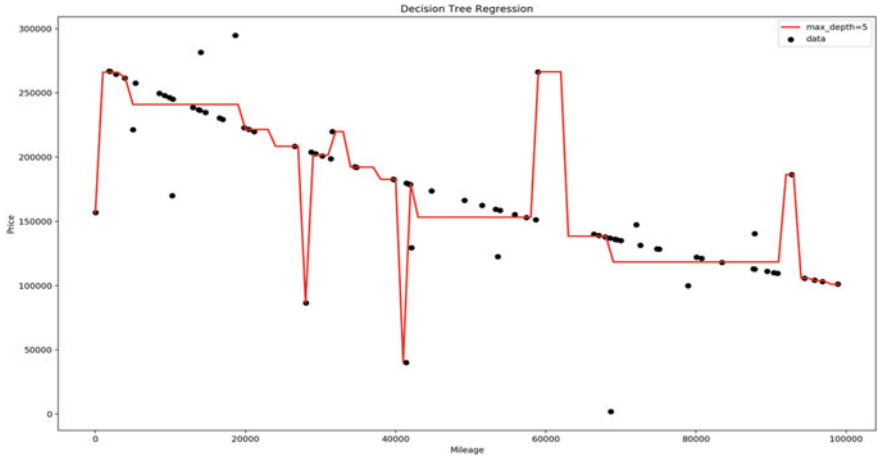


Fig. 5 Decision tree regression

f	feature to perform split,
D_p, D_{left} and D_{right}	data sets of parent and child nodes,
I	impurity measure,
N_p	total number of samples at the parent node,
N_{left} and N_{right}	number of samples in the child nodes.

The statistical test has been conducted to determine the performance of each attribute. For each value of an attribute, the root node descendants are generated. In a particular group, the sorting of training data is performed to a descendent. In a tree, the best node is selected by repeating these steps by utilizing the training example for each individual descendant node. The efficacy of separating the training examples as per target classification is based on amount of information gain (Fig. 5). The error is defined using weighted mean squared error (MSE) of the children nodes as given by Eq. (5).

$$\text{MSE}(t) = \frac{1}{N} \sum_{i \in D_t} (y^{(i)} - \hat{y}_t)^2 \tag{5}$$

Here,

N_t	number of training samples at node t,
D_t	taining subset at node t,
$y^{(i)}$	true dependent value.

The $\hat{y}_t$ is the predicted value mathematically given by Eq. (6).

$$\hat{y}_t = \frac{1}{N} \sum_{i \in D_t} y^{(i)} \tag{6}$$

5 Model Implementation

Two machine learning algorithms have been implemented in the proposed work. Data of classified platform play important role in selecting the mechanism and algorithm. Online classified platform data for bike category has been used. Basic idea chose the source of data very cautiously as end customers put a very high price for the used bike due to lack of knowledge. For our case, we have taken resale listing data of experience user and dealers. The resale bike variables can be divided into two segments, i.e. manufacturer parameters and user usage-driven parameter. Manufacturer parameters are mainly make, model, trim, year, etc., and user usage parameters are kilometre driven, number of owners, condition, etc.

6 Data Preparation

From data correlation Matrix and market experience, critical parameters such as make, model and trim are the most influencing the resale price of the bike. The whole data is divided into data chunks based on make, model and trim of the bike to create a model corresponding to each data chunk. The idea is to create multiple models instead of a single model. By plotting the distribution of price of the resale bikes, this has been found that data consists of noise and outliers. For outlier's detection and removal, the normal distribution approach has been used. All the data points in the training set have been removed whose price < (mean – 2*standard deviation) and price > (mean + 2*standard deviation). For data noise such as wrong value type in column and format issue is tackled by standard typecasting and text mining techniques. The derived variable such as bike age have also been created (Fig. 6).

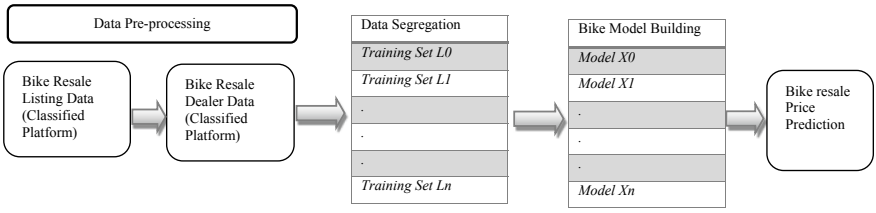


Fig. 6 Model implementation process

7 Performance Evaluation

7.1 Mean Absolute Error (MAE)

In this research mean absolute error (MAE) has been used to measures average quantum of the errors in prediction set. This is the mean of the absolute differences between predictions and actual observations for a sample under test. The MAE is the average distance of identity line and each point. The MAE is one of the well-known techniques to compare prediction with their actual values and mathematically given by Eq. (4)

$$\text{Mean Absolute Error(MAE)} = \frac{\sum_{i=1}^{n} |Y_i - X_i|}{n} \tag{7}$$

where Y and X are predicted and observed values.

7.2 Root Mean Square Error

This method is used to measures the average magnitude of the error and further calculate the quadratic. It is observed as the square root of the average of squared differences between prediction and actual observation.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (Yi - \hat{Y}i)^2}{n}} \tag{8}$$

where, X and μ indicate observed and mean values.

8 Results and Discussion

This paper aims to analyse the resale bike data of online marketplace to develop a relvent pricing model for pre-owned bikes by using two machine learning techniques, i.e. Linear Regression and Decision Tree algorithms. On implementing Linear Regression, 118 multiple models have been created, idea here to build one model w.r.t each make and model. For Decision Tree algorithm, only single model has been created. In the validation process both the models predicted price compared with the listing price quoted by the customer on the classified platform.

The quantitative comparison of Linear Regression and Decision Tree Techniques has been shown in Table 1, this has been observed the value of Mean Absolute Error (MAE) for various models obtained from Linear Regression technique is 10717

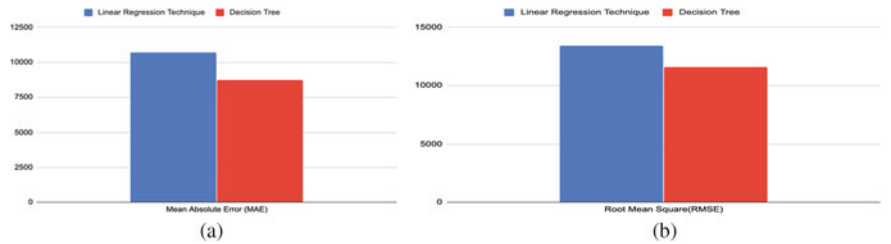


Fig. 7 **a** Comparison of % MEA and **b** RMSE for linear regression and decision tree techniques

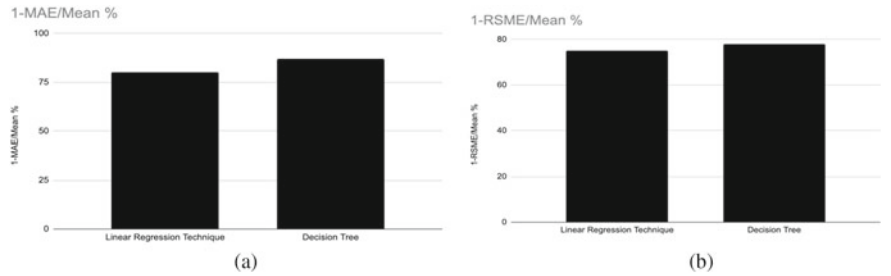


Fig. 8 **a** Comparison of % MEA of Mean and **b** % RSME of mean for linear regression and decision tree techniques

whereas for single decision tree model MAE has decreased to 8751 (Fig. 7a–b). The value of root mean square error from Linear Regression and Decision Tree Techniques are 13,441 and 11,604, respectively. Further, the value of % of MAE w.r.t mean for Linear Regression and Decision Tree Techniques are 80.19 and 86.91, respectively. The % of RMSE w.r.t to mean for both machine learning techniques are 75.15 and 78.10, respectively.

Table 1 shows the comparison between the Mean absolute error and Root Mean Square Error and the same have been depicted as column chart in Fig. 7a–b. Similarly comparative analysis of % Mean absolute error of Mean and % Root Mean Square Error of mean for Linear Regression and Decision Tree techniques are presented as column charts in Fig. 8a–b, respectively.

9 Conclusion

With the rapid expansion of modern technology, the localized market have been replaced by online marketplaces with significant impact on social, environmental and economic, etc. Online marketplaces are highly easy to access, convenient but also include the risk of fraudulent users to show attractive unrealistic price. To attract

online customers and make platform trustworthy, the correct price estimation of pre-owned bikes is a very essential part of business. This paper aims to develop accurate pricing models for pre-owned bikes by using OLX data and machine learning techniques namely Linear Regression and Decision Tree algorithms. The multiple models for various bike manufacturers have been prepared with higher accuracy. To judge the accuracy of proposed technique, the predicted prices have been evaluated and compared with actual price of pre-owned bikes. The MAE and RMSE values obtained with Linear Regression with multiple models and single Decision Tree algorithms have been compared quantitatively and graphically. This has been observed that MAE and RSME values are very high with Linear Regression technique. This indicates higher accuracy with Decision Tree technique is obtained as compare to Linear Regression technique. The correctness of estimated prices may be highly helpful to attract more online customers and making platform more trustworthy.

References

1. Geng X, Wu R, Whinston AB (2007) Profiting from partial allowance of ticket resale. J Market 71:184–195
2. Bowman WS Jr (1995) The prerequisites and effects of resale price maintenance. Yale Law School
3. Elzinga KG, Mills DE, The economics of resale price maintenance, in 3 issues in competition law and policy 1841 (ABA Section of Antitrust Law 2008)
4. Porter ME, van der Linde C (1995) Toward a new conception of the environment- competitiveness relationship. J Econ Perspect 9(4) Fall
5. Park B, Bae JK (2015) Using machine learning algorithms for housing price prediction: the case of Fairfax County. Expert Syst Appl 42(19)
6. Ippolito PM (1991) Resale price maintenance: empirical evidence from litigation. J Law Econ 34(2), Part 1
7. Marvel HP, McCafferty S (1986) The political economy of resale price maintenance. J Political Econ 94(5)
8. Basuchoudhary A, Metcalf C, Pommerenke K, Reiley D, Rojas C, Rostek M, Stodder J (2008) Price discrimination and resale: a classroom experiment. J Econ Educ 39(3)
9. Sunanda Sangwan U, Agarwal P (2019) Effect of consumer self-confidence on information search and dissemination: mediating role of subjective knowledge. Int J Consum Stud 43(1):46–57
10. Liu F, Yilmazer T, Loibl C, Montalto C (2019) Professional financial advice, self-control and saving behaviour. Int J Consum Stud 43(1):23–34
11. Brennan B, Sourdin T, Williams J, Burstyner N, Gill C (2017) Consumer vulnerability and complaint handling: challenges, opportunities and dispute system design. Int J Consum Stud 41(6):638–646
12. Bhagirath, Mittal N, Goel S (2019) Machine Learning computation on multiple GPU's using CUDA and message passing interface. In: IEEE international conference on power energy environment and intelligent control, PEEIC, pp 18–22
13. Mittal N (2019) Female perception towards the buying behavior of Mobile phones. Int J Inf Technol. Springer

COVID-Based Edge-Drone Application Approach

Ramneek Kalra and Vanshdeep Sahni

Abstract The unprecedented outbreak of the 2019 novel coronavirus, termed as COVID-19 has brought new techniques/technologies which are emerging a lot in today's research industry including 5G, Internet-of-Things (IoT), Edge Computing, Blockchain, etc., to more light. This is attracting more researchers across the Globe to enhance the existing models shared through different platforms. In addition to that, many have come across having humanoids/robots deployed into streets for serving and public announcements, delivering medications or emergency support onto doorsteps. These technologies are also helping many organizations across planet to explore COVID-19 related useful solutions which can help to manipulate and process data in real time. In this paper, authors are concentrating on integrating the algorithmic approach of Edge Computing & Drone Technology which can enlighten the challenges of parameters namely latency, connectivity, data processing, onsite computation. etc. In addition to that, proposing an architecture for the design of Edge-based integrated Drone use case for the containment zones detection with offline data processing & computation and deployment of algorithm for thermal sensing. This approach of making UAV bandwidth-independent architecture will create new opportunities for exploring COVID Pandemic smoothness for every aspect of social duties.

Keywords Edge computing · Drone · UAVs · COVID-19 · IoT · Coronavirus · 5G · Blockchain

R. Kalra (✉)
Cloud Support Associate, Amazon Web Services Inc., Hyderabad, India
e-mail: ramneek.kalra@ieee.org

V. Sahni
Project Engineer , Wipro Limited , Bangalore, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_14

1 Introduction

1.1 Edge Computing

Edge Computing deals with the deployment of edge devices near to the application layer to process data on real-time basis using sensor Technology or through the application of Microprocessors/Microcontrollers. This extends the already available solution of Cloud Computing Architecture by sandwiching Edge Layer in between of two other layers namely Cloud & Application Layer as depicted from Fig. 1.

Edge Architecture consists of the following layers:

- *Application Layer*: Application Layer is the physical layer where user/administrator plays their vital role of reporting and analysis outcome. This layer consists of different components including Smartphone, Laptop, System, Website, Application, etc., through which output of data processing/analysis from Edge Layer can be manipulated for Graphical User Interface (GUI). This layer is also independent of any dependency on machines which make key users as human. For example: COVID Related Android Application, Smart AC Application, TV Remote Application, etc.
- *Edge Layer*: The second layer which is sandwiched between other two depicted layer plays the most crucial role in processing the data without any bandwidth-dependency as it is nearby to the Application Layer. This layer helps to sensify or feel the surrounding data and to analyze the same without the use of the Internet. This layer consists of different sensors, microcontrollers, microprocessors, etc., which play like five senses alike of a human being and most often called as Edge Node. This may have less storage to store the real-time data produced by Application Layer but enough to analyze the data to output the required results.

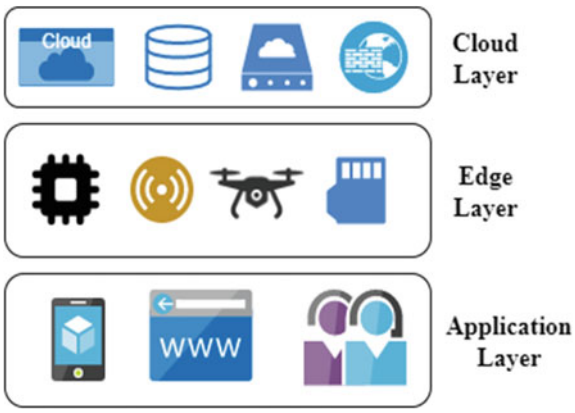


Fig. 1 Edge architecture

For example: Arduino UNO, Raspberry Pi, Water-Level Sensor, Drone Board, etc.

– *Cloud Layer*: The uppermost layer which is Cloud Layer plays the overall role in maintaining the historical data if Edge Layer fails to do so for the dedicated application. This layer plays a role of a Parent Server which is connected to different Edge Nodes distributed from Edge Layer. For example: Water-Level Sensor data per year of a river, bank data of fingerprint sensor per year of a bank accountant, etc.

1.2 Drone Technology

Drone Technology as being promising domain deals out in fully Unmanned Aerial Vehicle (UAV) Operations and Management. During this pandemic situation, drones can be very useful asset for any kind of seller or buyer. With respect to emergency medical delivery as well, drones are helping numerous families for medication from hospitals.

Concentrating on the deployment of hardware and software onto drone system is far complex than operating it. These two aspects can be listed below:

– *Drone Hardware*: When it comes to the hardware aspect of a drone, it may include components including a microprocessor (Raspberry Pi, Beagleboards), sensors (GPS, Accelerometer, Altimeter), brushless motors, propellers, landing gear, flight controller, Camera, etc. The connectivity of all same depends upon the application that can be varied from Emergency First Aid Delivery as reflected by [1], Grocery Delivery, Pizza Delivery, and many more.
– *Drone Software*: Drone has an operating system which is commonly termed as Robot Operating Systems, specially used for deploying different types of software as per application which is called as “Flight Slack or Autopilot”. For software architectural layer, we have three layers namely, Firmware, Middleware, and Operating System (Table 1).

Table 1 Software architectural layer distribution

Firmware	This helps to access the machine code to processor for memory access
Middleware	This includes the sensors functioning including flight control, navigation, radio management, etc.
Operating system	This layer consists of data analysis, decision-making, etc.

2 Literature Survey

With numerous approaches from researchers, aspect of image processing through drone infrastructure, deployment of light-weighted microprocessors/controllers, implication of Edge Computing to achieve the dedicated outcome and many more under the field of proposed work was earlier achieved but with lot of isolation from the aspects of bringing all innovation together.

For the same contrast, [2] discusses a Raspberry Pi with an inbuilt camera which was used to detect the suspicious activity of intruder in the restricted area. A camera was used to take dynamic photographs and videos so that further analysis can be done using tensor flow API. Authors in this proposes model which is good to use if the drone has a flight altitude of 30 meters, and it is illustrated in their research work that it can cover an area of 5 km^2. (if provided a good lighting environment). The major aspects of this model are that it can be used in extreme low light conditions as well, but for that the pre-processing of the image will be required to adjust the contrast and brightness. Another aspect includes the feature of manual control, the Raspberry Pi can even be controlled manually using VNC application viewer which makes the proposed research limited to the range of connectivity.

In the same context, [3] illustrates that there are many techniques to detect humans when they are moving and are not stable, but the most well-founded technique is motion detection with "Kalman filtering" which finds the centroid of the human body and further tracks it in other frames of the video stream. This approach is even proved beneficial in detecting a fast-moving object. The video which is shot with camera is resized to 960×1708 and is further given as an input to MATLAB IDE for processing.

And, with the advancement of Fog Architecture to numerous industries, [4] shares a "Spark oriented Fog architecture (SoFA)" is introduced which has provided an insight to scale and to make system utilization higher. If we talk about embedded systems, the power consumption and high system utilization is still a difficult task to achieve with the usual fog architecture. To overcome this, SoFA has been introduced which can help to overcome these problems as mentioned in their research work. SoFA was proved to provide more scalability by using spark functionalities. Experimental results claimed that for embedded systems the provided architecture (SoFA) can provide scalable solutions with high power efficiency with up to $3.1\times$ energy efficient for word count benchmark when compared to common fog architecture.

With that, to fight with the pandemic, [5] has explained the various steps of COVID19 spread and evaluated the various technologies such as IoT, unmanned aerial vehicles, Artificial Intelligence, 5G, and blockchain which can help to attenuate the adverse effects. The various methods that can help people to prevent themselves were explained and how the government can control the spread in their area with the help of technology that has been illustrated.

In reflection to drones & their challenges in implementation, [6] brings insights on delivering goods to surveillance, the UAV's which are being used among various civil

applications at present. The research has depicted the various classification, usage, challenges, and future insights that can be faced while using UAV's. SAR operations, which is a primary use-case of UAV has been handled in the work and gave a proper insight to path planning process to GPS co-ordination including image detection and searching process which are the most important factors that are considered while using UAV for surveillance.

In the consideration of safe landing of UAV, every aspect needs to be done that [7] focuses. It depicts on the statistics of UAV's including places that are forbidden to fly, potential terror acts, flight, and many other factors that are liked to a safe flight of UAV. Apart from these external vulnerabilities, authors have included the design, manufacturing, operational, and maintenance factors as well with all the safety and pre-planning to overcome the above situations.

On the other hand, there are some factors sharing to the Cloud Challenges which [8] focuses by reflecting on various important factors like power consumption, energy efficiency, and delay in comparison between traditional cloud and fog communication which follows the much advanced architecture. These factors including time, backhaul, power, and coordination have been evaluated to prove the mathematical observations with graphs.

While considering the integration of some prime factors for COVID19, [9] focuses on the major factors like impact, quarantine, social impact, and preventive measures which can help in prevention and to minimize the spread of it. Primary treatments like measuring body temperature has been the most common and successful prevention that can be taken while maintaining social distance. For which, the role of technologies in investigation has been a good help till now and a weight is put on UAV's as well for surveillance to minimize the human–human interaction as much as possible. Drone technology is suggested to spray disinfectant so that minimum human interaction would be there with a higher work yield. Since the technology has reached heights where no possible way is there for human capabilities to put effort. [10] has taken it as a base and suggested methods in a way that UAV can be helpful in monitoring and data analysis in a way that human efforts cannot do for a mobile object. Various AI algorithms have been provided to do the basic tasks of UAV in navigation, mission planning, and detection of moving objects.

Coming up with the integration of all discussed research work, UAV based in Fog Architecture, reflection of Autonomous Vehicular Network (AVN) is also remarkable to understand the reliability of UAV-based approach. Mughal et al [11] discusses methodology to increase the reliability of AVN where it has been depicted that how the decisions can be made locally without sending them at cloud eventually changing the latency. With the help of hypothesis, including the factors such as network usage and scalability, testing is done and has been observed that the latency of proposed fog-based architecture is far less than the latency of cloud-based architecture.

3 Methodology

In this section, authors concentrate on the reflection of the algorithm to use Edge Computing as an integrated technology for Drone Service for the highlighted containment zones detection model.

3.1 Physical View

Through the aspect of using drone (UAV) singly, there are lot of un-avoided parameters including Latency, Flight Time Delay, Battery Life, Data Processing, Decision-Making, etc. To eliminate this problem, Edge Computing plays a crucial role for strengthening the processing without using bandwidth over the Internet. This can only be flexible after the deployment of nearby edge nodes and drone stations connected using 5G spectrum.

Figure 2. depicts the general view of coverage by a single drone into a sub-regional analysis for mapping street data.

Every drone camera is dependent upon the reach and the coverage of number of houses in a street if reflected for the Sub-Regional View. This problem can be sorted by deploying numerous Mini-Drones in different streets and having Edge Nodes deployed for Data Storage and Processing. This eliminates the first crucial factor namely *Latency*. Latency is the time delay due to less bandwidth coverage for data upload and download in a normal network. This delay should be very minimal for fast decision-making and processing of data. This is commonly represented by Round Trip Time (RTT) or the Time to First Byte (TTBT). The mathematical expression for the same can be written as Eq. (1) below:

Fig. 2 Sub-regional view by a drone in the limited image processing

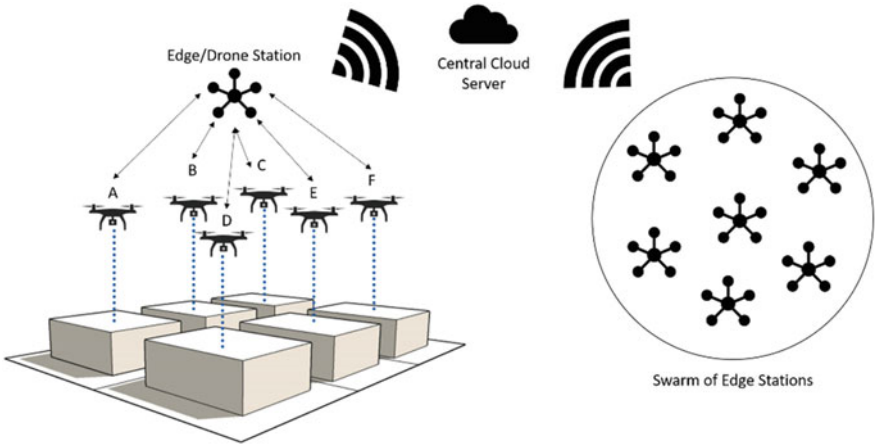


Fig. 3 Physical view of network communication among edge stations

$$\tau = \alpha * \beta / \Delta \tag{1}$$

where

τ represents time of the Flight Time of the drone (hour)
α represents Capacity of Drone's Battery (mAh or Ah)
β represents the Battery Discharge during the Flight Time
Δ represents Average Amp Draw of the drone.

Once the Flight Time is manipulated which is one of the kinds of Latency, but to explore the latency in the data processing, we need to consider the inner network latency which is being discussed in Network View. In addition to that, Fig. 3 depicts the network communication between block-wise distribution over a region.

As depicted, in the similar manner swarm of numerous edge/drone stations can be connected for communication and data storage using a central Cloud Server. This server can be linked to Users Application deployed over that region where detection is being perceived.

3.2 Software View

Before having an overview on Network Communication in deep, establishing the software architecture is more important and concerning. Drones in common, uses Robot Operating System (ROS) which can be Linux, Noobs OS, Debian, etc.

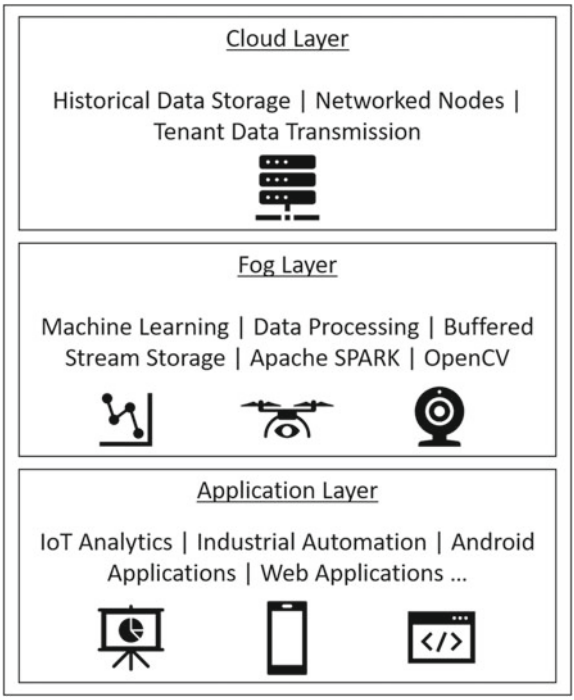


Fig. 4 Technology stack of proposed model

Maleki et al [4] discusses a unique application of Spark-oriented Fog Architecture which leverages spark functionalities to establish higher system utilization, energy efficiency, and scalability. Keeping the data distributed and processing on different layer, model promises the Fog Layer integrated with Application Layer consistently. Under the proposed model, we have used components as depicted by Fig. 4.

As per the different three layers from traditional Fog-Based Approach, different components & functionalities have been included in the proposed model for desired output. The algorithm which has been implemented for the dedicated image-processing onto Fog Layer, has been illustrated in detail below:

Algorithm A.1: Image-Processing for Containment Zones Detection

1. **Calibration**: An input is taken as a camera image-stream which is further considered for calibration in such a way that we can consider everyone on the same plane and plot their actual position on a graph. To estimate the scale, pixel density is considered to calibrate 6 ft distance and, in our camera, captured photo (which can be on a certain angle as well).
2. **Detection**: R-CNN architecture is used to detect the humans in a picture, which further produces image as an output with bounding boxes around the humans detected as being referenced in [12].

Fig. 5 Input test sample image 1

Fig. 6 Input test sample image 2

3. **Measuring**: To measure and to plot the humans detected on a graph, the bottom-central point as reference to plot their positions is considered. Every marked point is then taken to calculate the distance between every other marked points. Points with all the distances higher than 6 ft distance is marked as safe (in green) and the rest as red.

With the mentioned algorithm, execution of test data samples like below is being tested for required depicted results.

For proper parameterization of executed result from the practicality, we executed the same script on Cloud Premises and on Raspberry Pi Microprocessor which will be deployed while proper evaluation on premises.

Figures 5 and 6, illustrate the samples considered for test dataset with required concerns of Social Distancing & containment zones.

4 Results and Conclusions

With the contrast of the primary objective, detecting the containment zones for any violation using the proposed algorithm for image/video processing in real-time offline has been achieved with around 98–99% accuracy as measured by the Mean Squared

Error. The outcome for the same on some sample images can be depicted as below from Figs. 7 and 8, respectively.

As depicted, conclusion points can be made as following:

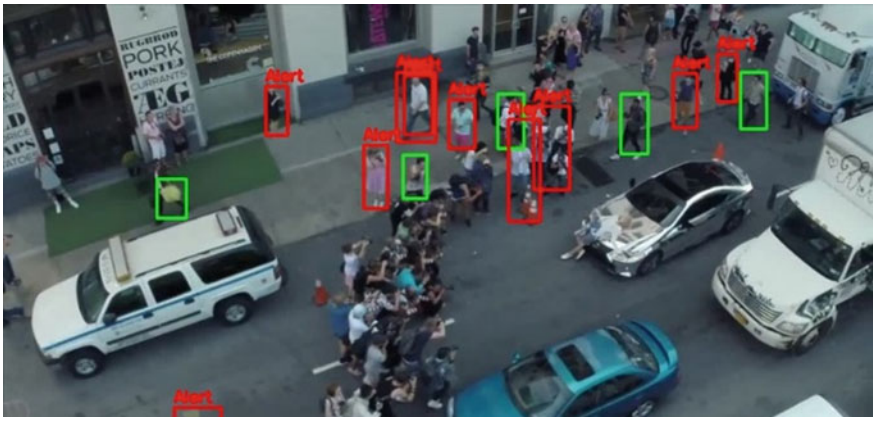


Fig. 7 Sample 1: social distancing violation through captured drone image

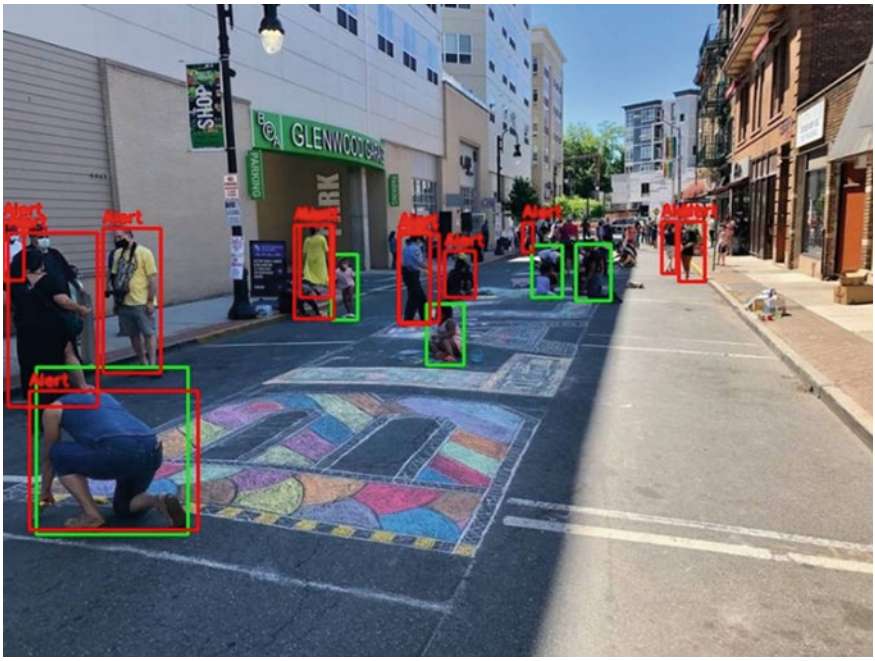


Fig. 8 Sample 2: social distancing violation through horizontal view of drone

– Social Distancing Parameter has been considered by processing real-time image and video streaming onto Drone Infrastructure.
– Alert onto Containment Zone Violation by having the total count of heads and red alerts detected by proposed algorithm.
– Transmission of detected statistics and location through GPS Sensor Data to the nearby authorities for actions against violation.

References

1. Sanjana P, Prathilothamai M (2020) Drone design for first aid kit delivery in emergency situation. In: *2020 6th international conference on advanced computing and communication systems (ICACCS)*. IEEE, Coimbatore, India, pp 215–220
2. Mojib EBS, Haque AKMB, Raihan MdN, et al (2019) A novel approach for border security; surveillance drone with live intrusion monitoring. In: *2019 IEEE international conference on robotics, automation, artificial-intelligence and internet-of-things (RAAICON)*. IEEE, Dhaka, Bangladesh, pp 65–68
3. Krishna SL, Chaitanya GSR, Reddy ASH, et al (2019) autonomous human detection system mounted on a drone. In: *2019 international conference on wireless communications signal processing and networking (WiSPNET)*. IEEE, Chennai, India, pp 335–338
4. Maleki N, Loni M, Daneshtalab M, et al (2019) SoFA: a spark-oriented fog architecture. In: *IECON 2019 - 45th annual conference of the ieee industrial electronics society*. IEEE, Lisbon, Portugal, pp 2792–2799
5. Chamola V, Hassija V, Gupta V, Guizani M (2020) 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. https://doi.org/10.1109/ACCESS.2020.2992341
6. Shakhatreh H, Sawalmeh AH, Al-Fuqaha A et al (2019) Unmanned aerial vehicles (UAVs): a survey on civil applications and key research challenges. IEEE Access 7:48572–48634. https://doi.org/10.1109/ACCESS.2019.2909530
7. Averyanova Y, Blahaja L (2019) A study on unmanned aerial system vulnerabilities for durability enhancement. In: *2019 IEEE 5th international conference actual problems of unmanned aerial vehicles developments (APUAVD)*. IEEE, Kiev, Ukraine, pp 40–43
8. Alhumaima RS (2020) Energy efficiency and latency analysis of fog networks. China Commun 17:66–77. https://doi.org/10.23919/JCC.2020.04.007
9. Kumaravel SK, Subramani RK, Jayaraj Sivakumar TK, et al (2020) Investigation on the impacts of COVID-19 quarantine on society and environment: preventive measures and supportive technologies. 3 Biotech 10:393. https://doi.org/10.1007/s13205-020-02382-3
10. Knyaz V, Zheltov S, Lebedev G, et al (2019) Intelligent mobile object monitoring by unmanned aerial vehicles. In:*IEEE EUROCON 2019 -18th international conference on smart technologies*. IEEE, Novi Sad, Serbia, pp 1–6
11. Mughal S, Awaisi KS, Abbas A, et al (2020) A reliable and efficient fog-based architecture for autonomous vehicular networks. In: Zomaya A, Abbas A, Khan S (eds) Fog computing, 1st edn. Wiley, pp 459–472
12. Ge Z, Jie Z, Huang X, et al (2020) PS-RCNN: detecting secondary human instances in a crowd via primary object suppression. In: *2020 IEEE international conference on multimedia and expo (ICME)*. IEEE, London, United Kingdom, pp 1–6

Challenges and Future Directions for Privacy Dashboard Design of Electronic Health Systems

Param Batavia, Purnima Ahirao, and Deepti Patole

Abstract Every individual has the right to protect and maintain the privacy of the Personal Identifying Information (PII) and Protected Health Information (PHI). But to avail health services, this PHI is essential to be shared with health service professionals. While doing so, the individuals must have a facility to make informed decisions about PII and PHI. The proposed work explores this area of providing better control to people in protecting their data. The paper explains the application of privacy protection policy on electronic health care systems. It compares several different Privacy By Design principles and selects the appropriate one. The features and the protection that should be used to protect such a system are proposed in the paper. The usage of a privacy dashboard is proposed to give the user complete control over their data. Other data protection methods are suggested to cover other aspects like minimization and data hiding.

Keywords Privacy by Design-PBD · Personal Identifying Information (PII) · Protected Health Information (PHI) · Privacy dashboard · Data protection · Health card system

1 Introduction

When people envisioned the future, one-click diagnosis of disease based on history and symptoms is expected. This has to be addressed as people are tired of repeating tests and retelling history to every new doctor. These problems can be solved by a countrywide electronic health card which is the need of the hour (especially during the times of the pandemic). However, with the introduction of such a system, it brings about thousands of privacy concerns in a person's mind. "Will the system ensure proper safeguard of the diseases I have suffered, and prevent leaking of such essential data to anyone other than the doctors treating me". For example, a person visiting a physiatrist may not want the orthopedic to know about him having a mental

P. Batavia (✉) · P. Ahirao · D. Patole
K J Somaiya College of Engineering, Somaiya Vidyavihar University, Mumbai, India
e-mail: param.b@somaiya.edu

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_15

health issue. Hence it is important to safeguard the interest of the user in sharing the relevant data to the required person/doctor. The user needs to procure the right to share parts of his information as and when required. The purpose of this research is therefore to find out the best technologies being used across domains and to bring them together and create a system that not only protects the user's privacy but also gives them control over their data. The current research reveals loopholes in data protection policies and lacks in designing a completed data protection policy with in-depth details and the engineering practices that can be applied to protect the user data. This paper is organized to map functional requirements to each part of the privacy policy. The paper first includes the literature review which decides the privacy policy and philosophy to select them and then go through the engineering practices currently being used. Then the study proposes a privacy dashboard and several other data protection principles and how they can be implemented. In discussions, we compare and contrast different methods that could have been used.

2 Literature Review

This research realizes that many software developers are not well equipped with ways to implement privacy safeguards and thus the privacy concerns are lost in fulfilling functional requirements. The main decisions made during the survey were which data protection guidelines to follow, features of a typical dashboard, and features in a typical health card system. The following papers by Danezis et al. [1], Cavoukian [2], and Langheinrich [3] discuss various data protection guidelines through Privacy by Design. Danezis et al. mentions 8 data protection policies and goes on to give a basic description of the design patterns for each of the policies, the paper discusses 8 privacy design strategies: "MINIMIZE", "HIDE"," SEPARATE", "CONTROL", "ENFORCE", "AGGREGATE", "INFORM", and "DEMONSTRATE" [1]. Cavoukian (2011) provides a framework that can be used for reference to create PBD software, the paper provides 7 core principles for PBD and their interpretations and possible implementations. The paper [2] is one of the first papers which states the principles of PBD software. According to Cavoukian (2011), the 7 fundamental principles of PBD are Proactive, not Reactive, Privacy as the default, Privacy Embedded into Design Privacy, Full Functionality, End-to-End Security, Lifecycle Protection, and Respect for User Privacy above all [2]. This paper by Langheinrich (2001) introduces the concept of Privacy, talks about the legal implications, and then goes on to state about the properties a PBD should contain. The paper [1] majorly deals with how to apply the principles of PBD to actual engineering practices but gives major importance to data protection design strategies than the overall strategies. The paper [2] deals with only the philosophies and 7 principles of PBD. The major pitfall in Cavoukian (2011) is it discusses the possible meaning of philosophies that can be open to interpretation differently by different people and due to this there is a lack of mapping of philosophies to design strategies [2]. Langheinrich (2001) deals with the legal aspects as well as the founding philosophies of PBD but

lacks the suggestions of possible design patterns [3]. Considering all the factors the privacy strategies described by Danezis et al. seem most apt to build a base on and create a software environment for electronic health card systems. Martin and Kung discuss and cover a broad overview of how to achieve privacy protection compliance using various methods like PBD, PET, Risk Management, Automated Privacy Management, Model-driven Design, and Requirement Engineering but the paper just provides a brief overview of the mentioned touchpoints [4]. The paper [5] by Schaar covers the implementation of the health card system in Germany in brief, it covers several points like Voluntary Basis, Extends of Data, Right to Information, and Ability to check. These topics are taken into consideration while creating data protection methods for the proposed methodology. The aim is to expand on the Privacy by Design Concept, software engineers use software security practices to:

- Prevent the introduction of vulnerabilities at its source into a product by better coding
- Detect vulnerabilities from the system created
- Respond to the discovery of vulnerabilities by clients or hackers [6].

The ideal aim of this system is to prevent the creation of vulnerabilities using better coding practices. Langheinrich et al. introduces the topic of privacy and segregates the topic into 5 specific topics: personal behavior, territory, person, personal communication, and personal data [7]. van Rest et al. provide an in-depth definition of Privacy By Design (complete and extended definition), Design Patterns, Data Abstraction Layers, and the Law [8]. Several engineering practices are described by Hoepman which deal with the implementation of PBD design patterns, in reality, some key concepts are selected and improved upon in this study to suit the health card system [9]. The paper [10] delves deeper by identifying the type of legal prerequisites, rights of the user, and Data Taxonomy. The major actors in privacy invasion activities are the data subject(victim) and data holders or data processors. The proposed system introduces the concept of privacy dashboards for websites. The features of the methodology proposed are to cover the legal requirements and rights of the users as discussed in the paper and suggest ideas for functional implementation.

3 Proposed Methodology

The methodology will be broken down into 2 sections. (1) Privacy Dashboard (2) Data Protection Rules. In the privacy dashboard, we try to cover as many Privacy Design Strategies we can via the implementation of Privacy Dashboards. The internal data which is being stored are solved using the implementation of Data Protection Practices for each policy. Figure 1 gives a brief idea about the Data Protection Policies or Principles [11].

The system demonstrates the application of Privacy By Design Principle and Data Protection Policies through the application of the Electronic Health System.

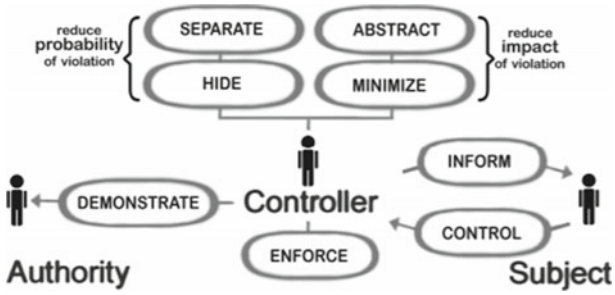


Fig. 1 Data protection policies and actors [11]

3.1 Privacy Dashboard Methodology

Privacy dashboards are meant to provide access to personal data in a structured and interactive manner [14]. The implementation of the dashboard covers 4 out of 8 data privacy protection principles. The principles covered are INFORM, CONTROL, ENFORCE, AND DEMONSTRATE.

Privacy Dashboard is meant to accomplish the following tasks:

- Give informed consent to the user. It is meant to (INFORM) give an informative overview of the privacy policy of the website. The dashboard should contain more informative and easier to understand privacy policies instead of just NOTICE. Some examples [12]:
  - Visual iconography that can flag if a certain issue or practice is present
  - Staging a walk-through the terms, step-by-step
  - Gamification of the terms, being quizzed about what the terms are
  - Storification of the privacy policy and character-based explanations of the terms, through stories.
- Enable the user to view what data and how the data is being processed. The dashboard should be able to give a standard overview of the processes being carried out in a graphical format. The data subject should be given the option of different visualizations to view the stored data. They should be given a full overview of the different entities of user data being used. There should be complete transparency at the user's end.
- Obtain information on the data controllers being used by the company.
- The right to complete erasure of the user data stored by the company
- The right to grant or withdraw consent on the privacy policy.
- The dashboard aims to provide access to DEMONSTRATE the actual compliance with privacy policy and all the legal requirements that are applicable. The right to share only a specific amount of data on case to case basis.

3.2 Data Protection Techniques

As shown in Fig. 2, the data privacy protection policy chart can provide a guideline for providing a better data sharing option for the user. It can play a very important role in providing better data-sharing features to the users. Below is an in-depth explanation of the four types of privacy policies.

3.2.1 Data Minimization

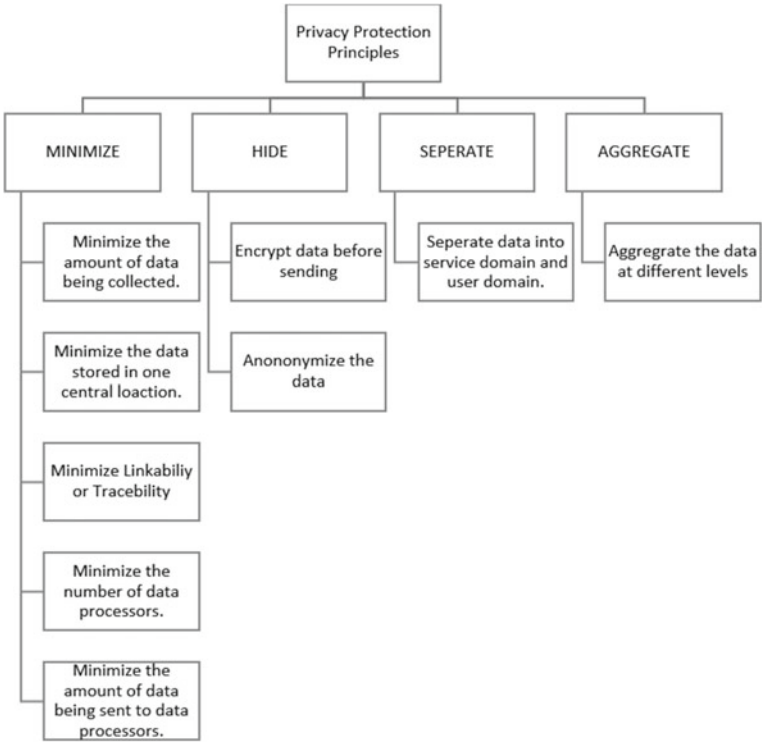


Fig. 2 Data privacy protection policies/principles chart

The most common design strategy is MINIMISE, which states that the size of personal data that's processed should be limited to the minimal. This strategy is commonly used in Design Patterns like "select before you collect" and "anonymization and use pseudonyms".

Techniques used in Data Minimization

- Minimize the amount of data of the data holder collected. Any irrelevant data that is not needed for the proper functioning of the system should not be collected. Strategies like "select-before-collect" can be used.
- Minimize the amount of data being stored in one central location. This makes it vulnerable to attacks which can lead to loss of all data or denial of service.
- Minimize/Remove Linkability or Traceability: It means that the data stored should not directly lead to the real identity of the data subject. Measures like anonymity and pseudo-anonymity should be carried out to prevent such situations.
- Minimize the amount of data being sent to data processors/holders. The data collected should be divided into the user domain and the service domain. Only data relevant and needed for operation at the data processor needs to be sent there. The rest of the user data should be stored in their local device in some encrypted manner.
- Minimize the number of data processors. A limit on the number of data processors allowed to access the data.

3.2.2 Data Hiding

HIDE, states that any personal data, and relationships with other entities, should not be visible in plain sight. The HIDE strategy is quite vital but frequently overlooked. Within the recent past, many systems are designed using seemingly innocuous identifiers that later clothed to be privacy nightmares. Such identifiers include but not limited to RFID tags, wireless network identifiers, and IP Addresses.

The design patterns that are commonly used in the HIDE strategy are a mixed bag and different techniques used on the basis of use case. One frequently used technique is of encryption of PII or any another essential data, both when stored, or when in transit. Other used methods are mixed networks to hide traffic patterns or techniques to unlink certain inter-related events like attribute-based credentials, anonymization, and pseudo-anonymization.

Techniques that can be used:

- Encrypt the data before sending it to the service domain/data processors. The encrypted data will be sent to the data processors and the key will be stored in the local device.
- Anonymize the data. Removing identifications that can uniquely identify the user wherever not needed or make change them to hide the identity of the user [13].

3.2.3 Data Separation

SEPARATE, states that the processing of data should include a mixture of different processors and must be done in separate components. Using multiple different processors or storage for the data of a singular person avoids the creation of a complete user profile even if the data is leaked. No established and ubiquitous Design Patterns.

Techniques that can be used:

- Separate the data collected into the service domain and user domain. Service Domain consists of entities that are to be sent to the data processors. User Domain consists of entities that are stored in the local device.

3.2.4 Data Aggregation

AGGREGATE, states that non-public data should be processed at the absolute best level of aggregation and with the littlest possible detail during which it's (still) useful. Coarse-grained data means the info items are general enough that the knowledge stored is valid for several individuals hence little information is often attributed to one person, thus protecting its privacy.

Commonly used design patterns in the industry that belong to the present strategy are the following: k-anonymity, differential privacy, aggregation over time (used in smart metering), and dynamic location granularity (used in location-based services).

Techniques that can be used:

- Aggregate the data at different levels of abstraction and then send the data to the processors.

3.3 Proposed Electronic Health Record System

Electronic Health Record Systems is truly the future of medical visits but there is a dearth in the research carried out in the features and responsibilities of actors of the system in case of privacy dashboards. The main contribution of the proposed study wants to bring all the data protection policies together for the privacy protection of Electronic Health Systems. There is a lack of techniques to be stated that can be used to satisfy each part of a PDB system. In discussions, we have gone through the adoption of EHS in countries and people's confidence in such systems. Through the techniques suggested above, we want to alleviate people's fears.

Through the literature survey, it can be recognized that there is a lot of research on the concepts of Privacy Dashboard and Privacy Presentation Methods individually but there is still not much work done to understand how a generic privacy dashboard can be adapted to suit the needs of an Electronic Health System. This study aims to set a standard for the basic features that should be included in all EHS, while most

of the features might seem intuitive but the legal protections mentioned under the Health Insurance Portability and Accountability Act (HIPAA) should be covered. The authors aim to cover all these aspects in this study and the list of features mentioned is not exhaustive and limiting to them.

We can implement the techniques studied and will demonstrate potential usage in an electronic health care system. A sample dashboard creation will have to protect the following PII information.

- Biometric of ID
- User Identification Number
- Personal Details like Gender, Age, Address, Allergies, Blood Group
- Personal Health Records.

The Biometric of Id should be protected and un-hackable. There should be no link between Identification Number, Personal Details, and the health records of the person.

3.3.1 Privacy Dashboard Features for Electronic Health Record System

- A center display control to show all the essential information like Identification Number and Personal Details.
- An informative way to show privacy policy and how information is being shared with third parties like insurance companies, physicians.
- Different visualization and documents in human-readable form currently linked to the card. Visualization can be used to extract essential information like prognosis, medicines from doctor reports, and hospital charts
- A complete informational panel about the data processing activities being carried out by third-party companies. The data processing activates here can be whether calculating the premium or the risk index of any disease.
- The option to permanently delete and complete erasure of some or all medical records from the system at the preference of data subject,
- An option to grant or revoke the privacy policy at the discretion of the data subject.
- The dashboard can be used to show in a noble way how their personal data is being secured and how every legal obligation is being followed. An additional feature to privileged users like government officials can be to check whether the data controllers are not misusing the data or breaking any law.
- The data holders should have complete control on whom to share the data with and which physician to share how much data with on a case to case basis.
- The privacy dashboard system should have an OCR (Optical Character Recognition) system that will recognize all the PII data of the patient from scans like X-Rays, CT-Scans, etc. The OCR will then carry out the anonymization process by either blacking or coloring them out. To recognize such data, we can use processes like keyword filtering or Regular Expressions [20]. We also propose

standardization of medical image reports to easily recognize data and thus can be easily censored.

Figure 3 shows the mapping of the functionality of the dashboard to each component of the privacy policy and how we plan to satisfy them using the privacy dashboard.

Figure 4 is a use case scenario that maps the responsibility of the data processors and the services they must provide through the dashboard.

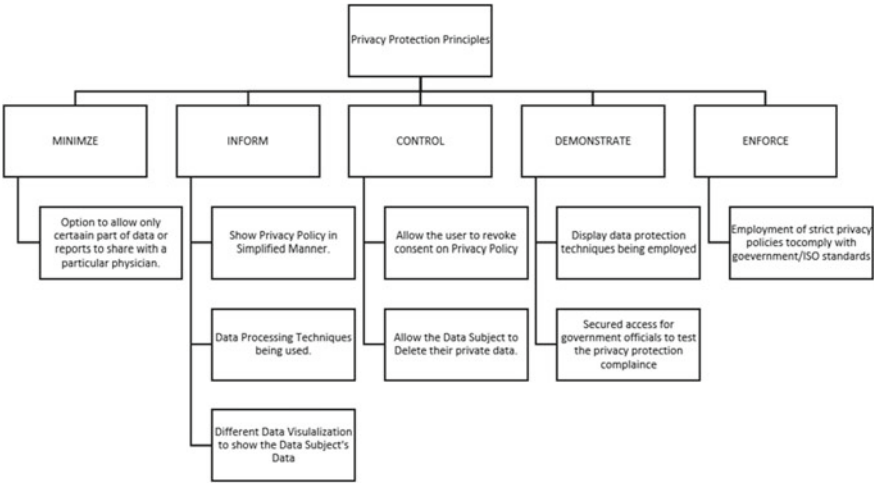


Fig. 3 Mapping of functionality to policy

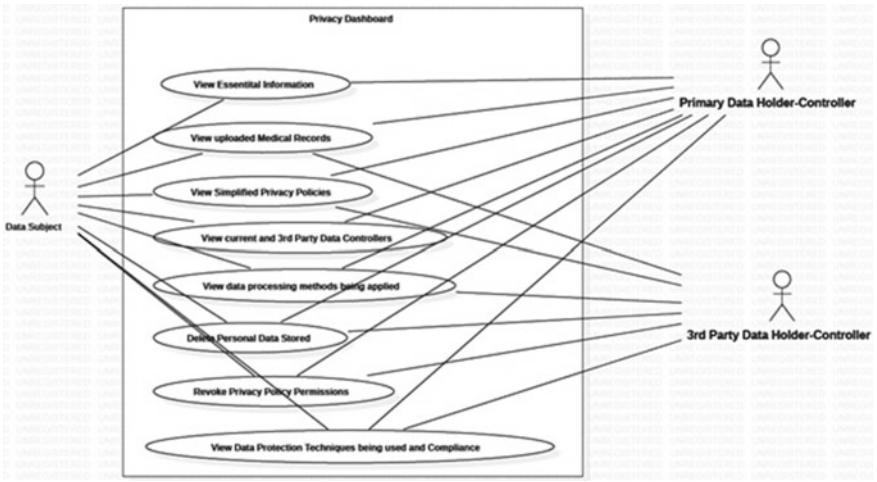


Fig. 4 Use case for privacy dashboard

3.3.2 Data Protection Policies That Can Be Applied to Electronic Health Record

- Divide the entities in the service domain and user domain. The user domain includes user and health-card. Service Domain consists of Government, Physicians, Insurance Companies.
- Data sent to the Service Domain must be limited using "select-before-collect" or "only-collect-necessary-data" policies should be used.
- Data like person identification number, important details like height, weight, and medical records should only be collected and stored. Any other information such as location or other items for targeted advertising must be avoided.
- Identification of data objects must also take place into consideration whether a data is PII or non-PII. The user should have the option to control all the PII data and such data should be stored locally only and be de-identified without sending it to 3rd party controllers. Identification of PII data can be done using Keyword Filtering and Regular Expression. We also propose the use of NER (Named Entity Recognition) to identify such data as they provide better recognition rates in case of typos over the 2 methods stated before, NER should be used to identify any such PII data in medical records and removed before being sent. NER is one of the things not being currently used in such systems and can greatly improve the protection of PII data.
- The data holders should have complete control on whom to share the data with and which physician to share how much data with on a case to case basis.
- Reduce the amount of data stored in a central repository. There should be more than one data center where data is stored based on the proximity to the data holder. This allows for faster access and not the complete leak of data if one center is compromised.
- Use of encryption for "User Identification Number", this should not be shared with the data controllers. It should be encrypted and sent to the concerned third-party apps. Encryption should store the key for the encryption in the local device.
- Private health records must also be stored in an encrypted manner and can only be accessed when the subject grants access to a certain data provider. This can be achieved by storing the data at the local device and encrypting it and sending it to the data processor. The key to be used can be provided, this data cannot be accessed by unauthorized parties.
- If the data needed by an insurance company for a particular city for analytics, then data should be stripped of all identification parameters and the aggregated data should be directly sent to the processor instead of individual records. This can be interpolated when the overall information of a user is required, a summary of all the available or records for which access is granted is sent to the processor instead of individual health records.

Figure 5 shows the mapping of the functionality that needs to be added to the Privacy Protection Policies.

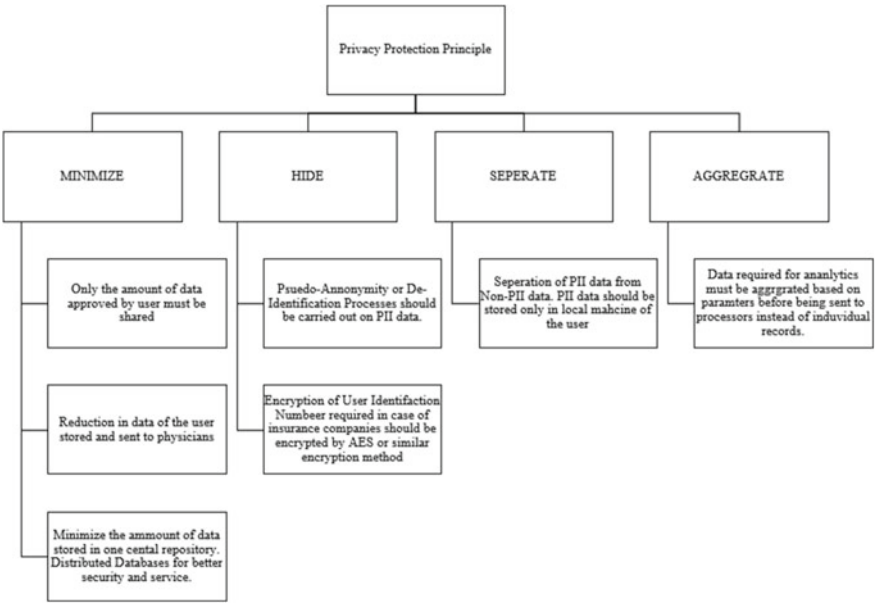


Fig. 5 Mapping of approaches to privacy protection policies

4 Discussions

This section discusses the various concerns and aspects being presented and projected around the globe for data privacy and data protection.

Figure 6 shows that while there is an upward trend of people being confident about the safeguards put in place, there is still a large number of people who are

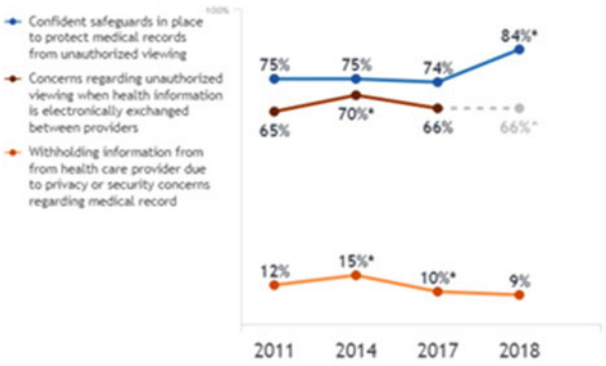


Fig. 6 Security trend among users downloaded from https://dashboard.healthit.gov/quickstats/pages/consumers-privacy-security-medical-record-information-exchange.php [7]

concerned that their data is not secured. This study aims to decrease this trend by improving the privacy policy by making it more understandable about the security measures which are being undertaken. The steps included in the privacy dashboard will alleviate people's concerns.

Figure 7 shows the adoption of electronic health care systems in countries based on their incomes. This clearly shows there are still a lot of countries where such systems have not been implemented and thus have tremendous potential for growth, If the system is made more secure, foolproof, and ubiquitous. More than 83% of healthcare professionals have admitted that EHR has helped them manage prescription order entry, patient records, clinical notes, medications, problem lists, and view lab results [16].

This goes on to show that Electronic Health Systems[EMR] are quite useful and reduce the load of the physician on the administrative side and allows them to focus on diagnosing and treating the patient. The same is also proved by Fig. 8 [17].

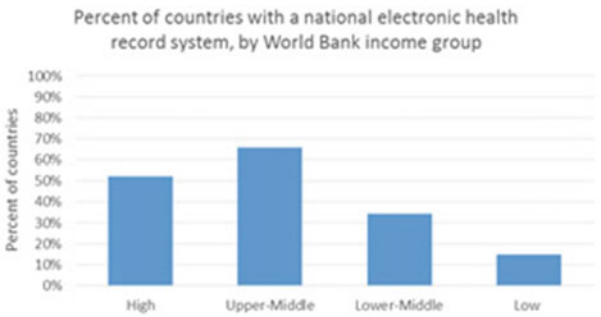


Fig. 7 Usage of EHS based on income downloaded from https://www.who.int/gho/goe/electronic_health_records/en/ [15]

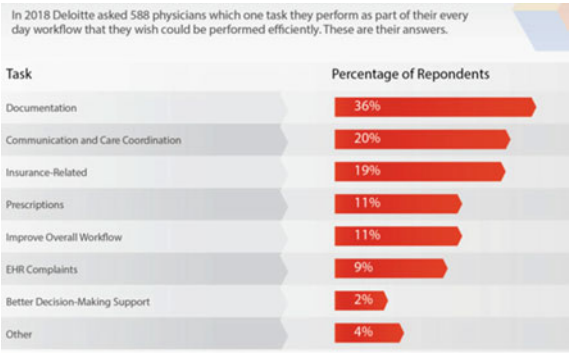


Fig. 8 Tasks that will be reduced on the implementation of EHS downloaded from https://www.selecthub.com/medical-software/emr/electronic-medical-records-future-emr-trends/ [17]

Jeff Riggins, a health IT expert at Drury University, accurately states the problem with current systems: Most EMR software is insufficient. They were built on obsolete platforms with little budget and no thought dedicated to user experience. Most EMR packages were initially billing systems and then the addition of additional features. The planning phase for the development of features was largely overlooked because the functionality of the system was all the government had created guidelines for. We want to avoid this trend by making the EMR more user-centric as we have proposed in our solution.

In 2019, healthcare firms have faced continued attacks from hackers causing large number of data breaches and ransomware attacks. The financial cost of these breaches is following a bad trend with data breaches expected to cost US$ 4 billion by the top of the year [18]. One among the main samples of this was the AMCA Data breach during which 25 million patient records were compromised, there are several large scale data breaches, making data protection a serious concern. The anonymization strategy and other strategies suggested are often wont to make the info safer albeit there's a breach, it can still stop very critical information like identification numbers falling into the hands of wrongdoers.

The various techniques that are being currently used by hackers to illegally access data are masquerading, unauthorized use, alteration and disclosure of services and information, and unauthorized denial of service [19].

5 Conclusions

The Study mainly tries to propose an electronic health system that makes the user feel secure and assure that the data is being properly safeguarded and in the right hands. The privacy dashboard will provide an easy to understand and interactive way in which a person can control their own data. The major drawbacks which we have identified are the lack of research in the area of privacy dashboards in combination with the sector of electronic health card systems. There is a deep lack of standards to be set in this area while there are several laws for data protection in EHS, each interpretation can have a different implementation, we have tried to establish bare minimum features an EHS should contain. We have tried to import the concepts of NER and OCR and fit them in places that can help in the protection of privacy. Also, Privacy by design being the norm in this privacy concerning generation and legal compliances there is still a lack of depth and ideas regarding the application of PBD in different domains. The Proposed system tries to provide a list of possible functional requirements of an electronic health system. Privacy dashboard provides an "all in one" "drill into detail" and mobile-friendly properties appeal to everyone this will give a sense of security to people. We also want to focus that more research work should take place on the subject matter of improving the way privacy policy is presented and the methods mentioned should be worked upon and made mandatory. With the many benefits of the electronic health systems, it will continue to grow and the most important factor of such systems is "Trust" between the provider and

the user, making privacy dashboards and other data protection principles proposed a necessity. In the future electronic health card systems, will replace manual paperwork and will be used by hospitals, private doctors, hospital staff, and patients themselves. In discussions, we have shown the potential EHS systems can achieve and the need of certain standards or bare minimum features that a system should ensure for data protection. Some essential backup system should be designed to prevent complete loss of access and services in the case the system is hacked or is undergoing a DOS attack, this is essential to ensure that the critical care to patient is not denied.

Acknowledgements Author would like to acknowledge the management of K J Somaiya College of Engineering, for supporting this Research Work.

References

1. Danezis G, Domingo-Ferrer J, Hansen M, Hoepman JH, Métayer D Tirtea R Schiffner S (2014) Privacy and data protection by design - from policy to engineering. https://doi.org/10.2824/38623.
2. Cavoukian A (2011) Privacy by design — the 7 foundational principles. Technical report, Information and Privacy Commissioner of Ontario, (revised version)
3. Langheinrich M (2001) Privacy by design — principles of privacy-aware ubiquitous systems. In: Abowd GD, Brumitt B, Shafer S (eds) Ubicomp 2001: Ubiquitous computing. UbiComp 2001. Lecture Notes in Computer Science, vol 2201. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45427-6_23
4. Martin Y, Kung A (2018) Methods and tools for GDPR compliance through privacy and data protection engineering. In: *2018 IEEE European symposium on security and privacy workshops (EuroS&PW)*, London, pp 108–111, https://doi.org/10.1109/EuroSPW.2018.00021
5. Schaar P (2010) Privacy by design. Ident Inf Soc 3:267–274. https://doi.org/10.1007/s12394-010-0055-x
6. Williams L, McGraw G, Migues S (2018) Engineering security vulnerability prevention, detection, and response. IEEE Softw, 1–1, https://doi.org/10.1109/ms.2018.290110854
7. Individuals' perceptions of the privacy and security of medical records and health information exchange, Dashboard.healthit.gov, 2020. Available: https://dashboard.healthit.gov/quickstats/pages/consumers-privacy-security-medical-record-information-exchange.php
8. van Rest J, Boonstra D, Everts M, van Rijn M, van Paassen R (2014) Designing privacy-by-design. In: Preneel B, Ikonomou D. (eds) Privacy technologies and policy. APF 2012. Lecture Notes in Computer Science, vol 8319. Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-54069-1_4
9. Hoepman J-H (2014) Privacy design strategies. IFIP advances in information and communication technology, 446–459. https://doi.org/10.1007/978-3-642-55415-5_38
10. Raschke P, Küpper A, Drozd O, Kirrane S (2018) Designing a GDPR-compliant and usable privacy dashboard. Privacy and identity management. The smart revolution, 221–236. https://doi.org/10.1007/978-3-319-92925-5_14
11. Colesky M, Hoepman J, Hillen C (2016) A critical analysis of privacy design strategies. In: *2016 IEEE security and privacy workshops (SPW)*, San Jose, CA, pp 33–40, https://doi.org/10.1109/SPW.2016.23
12. Haapio H, Hagan M, Palmirani M, Rossi A (2018) Legal design patterns for privacy.
13. Gurses S, Troncoso C, Diaz C (2015) Engineering privacy by design reloaded. In: *Proceedings of Amsterdam privacy conference*, p 22

14. Bierr C, Kühne K, Beyerer J (2016) PrivacyInsight: the next generation privacy dashboard. In: Schiffner S, Serna J, Ikonomou D, Rannenberg K (eds) Privacy technologies and policy. APF 2016. Lecture Notes in Computer Science, vol 9857. Springer, Cham. https://doi.org/10.1007/978-3-319-44760-5_9
15. Electronic health records (2020) World Health Organization. Available: https://www.who.int/gho/goe/electronic_health_records/en/
16. Coreadmin (2019) 15 latest statistics on EHR to prove its Importance. Available: https://physicianxpress.com/15-latest-statistics-ehr-prove-importance/
17. Adair B (2020) Future of electronic medical records | EMR Trends For 2020, Selecthub.com. Available: https://www.selecthub.com/medical-software/emr/electronic-medical-records-future-emr-trends/
18. Srinavas R (2020) Ransomware attacks that affected healthcare providers in 2019, CISO MAG | Cyber Security Magazine. Available: https://cisomag.eccouncil.org/7-times-ransomware-became-a-major-healthcare-hazard/
19. Win K Susilo W, Mu Y (2006) Personal health record systems and their security protection. J Med Syst 30:309–315. https://doi.org/10.1007/s10916-006-9019-y
20. Huang L-C, Chu H-C, Lien C-Y, Hsiao C-H, Kao T (2009) Privacy preservation and information security protection for patients' portable electronic health records. Comput Biol Med 39(9):743–750. https://doi.org/10.1016/j.compbiomed.2009.06.004

A Novel Approach for Morphological Analysis and Inflections Based on POS Tagging

Kanu Goel

Abstract Word processing with tokenization and part-of-speech (POS) tagging is an integral component of computer linguistics. During the process of POS tagging, morphological analysis is crucial for applications like information retrieval, language modelling, machine translation, etc. Generating inflections that relate various forms of a single word is an important step in morphological analysis. In this paper, a novel approach of analysing word structure and generating inflections of nouns and adjectives has been proposed. In this approach, an efficient method using xml-based short notation has been implemented to store the generated inflected forms. A regular expression-based numeral sequencing has been used.

Keywords NLP · POS tagging · Morphology · Inflections · Nouns · Adjectives

1 Introduction

In case of any language words form an integral component of literature. Nouns and adjectives are foremost parts of speech that derive the basic essence of flow and meaning that a text conveys. Inflections associated with any word are basically its different forms as given in Table 1. They can be used to convey agreement, number, person, tense or any other information [1]. For any thorough linguistic study, it is crucial to identify correct part-of-speech (POS) tags and have knowledge of word inflections. The lexical information and the contextual information related to a word are depicted by these two aspects [2]. In this paper, we mainly discuss the importance of inflected forms of various part-of-speech (POS) tags and propose a novel approach to generate and store inflections of nouns and adjectives.

K. Goel (✉)
Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_16

Table 1 Inflections for various categories of grammar

Part-of-speech (POS)	Category of grammar	Inflection	Examples of inflected forms
Noun	Number	-s'	Car-> Cars, Card-> Cards
Noun	Number	-es'	Glass-> Glasses
Noun	Case(genetive)	-'s, -'	Tim-> Tim's, Frankis-> Frankis'
Pronoun	Case(genetive)	-s'	It-> Its
Pronoun	Case (reflexive)	-self'	Your -> Yourself
Pronoun	Case (reflexive)	-selves'	Them -> Themselves
Adjective	Degree of comparison (comparative)	-er	Sharp -> Sharper, Strong-> Stronger
Adjective	Degree of comparison (superlative)	-est	Sharp -> Sharpest, Strong-> Strongest
Verb	Aspect (progressive)	-ing	Sleep -> Sleeping
Verb	Tense (past)	-ed	Walk -> Walked
Verb	Tense (present)	-s	Laugh -> Laughs

Morphemes which are the minimal meaningful elements of the words are extracted and the inflected forms are written in numeral notations in regular expressions. Various approaches have been built in the past which focus on generating the correct part-of-speech tags for given words as stated in Table 2. This study is believed to add useful way of representing the language in precise notation. The purpose of this work is to analyse infections in nouns and adjectives. This paper is organized as follows. Section 2 describes the literature review of significant work in this domain. Section 3 presents the proposed approach. The implementation details of the proposed work are given in Sect. 4 followed by conclusion and future scope in Sect. 5.

2 Related Work

In existing works, in [3, 4] the concept of part-of-speech has been described listing various grammatical properties. Further morphological and grammatical rules are detailed for rule-based POS tagging system [2]. The structure of words with their morphological variations in English language form an integral part of the linguist study [5, 6]. Various kinds of inflections related to adjectives along with their uninflected forms are given in the literature [7]. There have been other studies that have worked to generate inflections of nouns namely nominative, genitive, accusative. In the work [8], bottom-up procedure of characterizing inflected words is given. They have also worked to generate inflections of nouns namely nominative, genitive, accusative. Their feature-based approach for morphological analysis is quite significant.

Table 2 Predefined tags for parts of speech forms

Tags	Parts of speech form
CC	Coordinating conjunction
JJ	Adjective
JJR	Comparative adjective
JJS	Superlative adjective
NN	Singular noun
NNS	Plural noun
NNP	Singular or proper noun
NNPS	Plural or proper noun
PDT	Predeterminer
PRP	Personal pronoun
RB	Adverb
RBR	Comparative adverb
RBS	Superlative adverb
TO	To
UH	Interjection
VB	Base verb form
VBD	Past verb tense
VBG	Gerund verb
VBN	Past participle verb
WDT	Determiner-Wh

Various application-based works as in [9, 10] the authors have presented an approach suitable to handle suffix-based sequences while increasing performance of morphological dictionary for Assamese text corpus. Another work by Dasgupta et al. [11], the authors have parsed compound words retaining inflections and also resolved word ambiguities. They have specifically worked for Bangla language using feature unification-based parser.

3 Proposed Approach

In the proposed work the input text is processed to generate the inflections of the words. Currently, the nouns and adjectives are undertaken into processing in this work. The input words are tokenized and part-of-speech tag is found for the current word. Singular and Plural are the two categories considered for nouns which further labelled as proper nouns or not. On the other hand, in case of adjectives, they are classified according to their degree of comparison which can be positive, comparative or superlative. All these steps can be mapped to Algorithm 1. In this paper three different functions are described in separate algorithms for finding plural nouns,

Fig. 1 Output inflections in xml format

```
<?xml version="1.0"?>
- <data>
  - <brush type="Noun">
      <Form>Regular</Form>
      <Plural>%5es</Plural>
    </brush>
  - <potato type="Noun">
      <Form>Regular</Form>
      <Plural>%6es</Plural>
    </potato>
  - <half type="Noun">
      <Form>Regular</Form>
      <Plural>%3ves</Plural>
    </half>
  - <cat type="Noun">
      <Form>Regular</Form>
      <Plural>%3s</Plural>
    </cat>
  - <copy type="Noun">
      <Form>Regular</Form>
      <Plural>%3ies</Plural>
    </copy>
  - <big type="Adjective">
      <Form>Regular</Form>
      <Comparitive>%3ger</Comparitive>
      <Superlative>%3gest</Superlative>
    </big>
  - <large type="Adjective">
      <Form>Regular</Form>
      <Comparitive>%4er</Comparitive>
      <Superlative>%4est</Superlative>
    </large>
</data>
```

comparative adjective and superlative adjective. Algorithm 2 depicts the process of formulating root word and its plural form both. Here, four different cases of writing inflections of plural nouns, i.e. nouns ending with 'ies', 'ves', 's' and 'es' are extracted and written in numeral short notation in XML output files. Next, in Fig. 1: Output inflections in xml format case of comparative adjectives words ending with two forms namely 'ier' and 'er' are extracted according to Algorithm 3 and written in XML file. Lastly, superlative degrees of adjectives are found using adjectives ending with 'iest', 'est' or 'st'. Algorithm 4 states the steps of writing short numbered regular expressions for superlative adjectives.

Results: The output obtained by processing a random input text string and generating inflected forms in XML notation is depicted in Fig. 1. The results are described in the xml notation for all the morphological forms of the root words. Results for nouns, adjectives with comparative and superlative degrees are shown in Fig. 1.

Algorithm 1 extractPOSTag(TxtStr)

```
Input text string in TxtStr
SplitText = TxtStr.split()
for each word in SplitText do
    struct = pos_tag(word_tokenize(word))
    [structList] = list(struct[0])
    structWordForm = structList[1]
    if structWordForm[0] == 'N' then
        if structWordForm=='NN' then
            print(word+" :It is a singular noun")
        else if structWordForm=='NNS' then
            self.pluralNoun(structWord)
        else if structWordForm=='NNP' then
            print(structWord,":Proper Noun Singular")
        else if structWordForm=='NNPS' then
            print(structWord,":Proper Noun Plural")
        else if structWordForm[0]=='J' then
            if structWordForm=='JJ' then
                print(word+" :It is a positive adjective")
            else if structWordForm=='JJR' then
                self.comparitiveAdjective(structWord)
            else if structWordForm=='JJS' then
                self.superLativeAdjective(structWord)
            end if
        else
            print(structWord,":Neither a noun nor an adjective!")
        end if
    end if
end for
```

Algorithm 2 findPluralNoun(self, word)

```
1: reWord = re.findall("(.*?)(ies|es|s|ves)$",word)
2: reWordList =list(reWord)
3: wordPart=reWordList[0]
4: rootWordPart=wordPart[0]
5: find suffixPart=wordPart[1]
6: if suffixPart=='ies' then
7:     rootWordPart=rootWordPart+'y'
8:     lenRootWord=len(rootWordPart)
9:     pluralForm= '%'+str(lenRootWord-1)+'ies'
10: else if suffixPart=='ves' then
11:     rootWordPart=rootWordPart+'f'
12:     lenRootWord=len(rootWordPart)
13:     pluralForm= '%'+str(lenRootWord-1)+'ves'
14: else if suffixPart=='s' then
15:     lenRootWord=len(rootWordPart)
16:     pluralForm='%'+str(lenRootWord)+'s'
17: else
18:     lenRootWord=len(rootWordPart)
19:     pluralForm= '%'+str(lenRootWord)+'es'
20: end if
21: if self.xmlObject.searchWord(rootWordPart) is None then
22:     wordObject=self.xmlObject.insertWord(rootWordPart,'Noun')
23:     self.xmlObject.insertTag(wordObject,'Form','Regular')
24:     self.xmlObject.insertTag(wordObject,'Plural',pluralForm)
25:     self.xmlObject.prettify(self.xmlObject.root)
26: end if
```

Algorithm 3 findComparitiveAdjective(self, word)

```
reWord=re.findall("(.*?)(ier|er)",word)
reWordList=list(reWord)
wordPart=reWordList[0]
rootWordPart=wordPart[0]
suffixPart=wordPart[1]
comparitive=None
superLative=None
if suffixPart=='ier' then
    rootWordPart=rootWordPart+'y'
    lenRootWord=len(rootWordPart)
    comparitive= '%'+str(lenRootWord-1)+'ier'
    superLative= '%'+str(lenRootWord-1)+'iest'
else
    lenRootWord=len(rootWordPart)
    if (rootWordPart[lenRootWord-1]==rootWordPart[lenRootWord-2])
    then
        rootWordPart = rootWordPart[:-1]
        rootWordPartOrg=rootWordPart
        lenRootWord=lenRootWord-1
        comparitive='%'+str(lenRootWord)+(rootWordPart[lenRootWord-
        1])+'er'
        superLative='%'+str(lenRootWord)+(rootWordPart[lenRootWord-
        1])+'est'
    else
        rootWordPart=rootWordPart+'e'
        comparitive='%'+str(lenRootWord)+'er'
        superLative='%'+str(lenRootWord)+'est'
    end if
    if (self.xmlObject.searchWord(rootWordPart) is None) then
        wordObject=self.xmlObject.insertWord(rootWordPart,'Adjective')
        self.xmlObject.insertTag(wordObject,'Form','Regular')
        self.xmlObject.insertTag(wordObject,'Comparitive',comparitive)
        self.xmlObject.insertTag(wordObject,'Superlative',superLative)
        self.xmlObject.prettify(self.xmlObject.root)
    end if
end if
```

Algorithm 4 findSuperlativeAdjective(self, word)

```
reWord=re.findall("(.*?)(iest—est—st)",word)
if suffixPart=='iest' then
    rootWordPart=rootWordPart+'y'
    lenRootWord=len(rootWordPart)
    comparitive='
    superLative='
else
    lenRootWord=len(rootWordPart)
end if
if rootWordPart[lenRootWord-1]==rootWordPart[lenRootWord-2] then
    rootWordPartOrg=rootWordPart
    rootWordPart = rootWordPart[:-1]
    lenRootWord=lenRootWord-1
    comparitive='
    superLative='
else
    rootWordPart=rootWordPart+'e'
    comparitive='
    superLative='
end if
if self.xmlObject.searchWord(rootWordPart) is None then
    wordObject=self.xmlObject.insertWord(rootWordPart,'Adjective')
    self.xmlObject.insertTag(wordObject,'Form','Regular')
    self.xmlObject.insertTag(wordObject,'Comparitive',comparitive)
    self.xmlObject.insertTag(wordObject,'Superlative',superLative)
    self.xmlObject.prettify(self.xmlObject.root)
end if
```

4 Implementation Details

The proposed approach has been implemented using Python framework with NLTK library. The output inflections generated from the root words are stored in the xml format using the XMLHelper class. The xml form is generated so as to obtain a presentable format of results. Various tagged notation is used to showcase the morphological forms of the root words. The implementation can also be extended to other words and various other POS tags.

5 Conclusion and Future Work

This paper focuses on analysing the inflections of mainly two parts of speech namely nouns and adjectives. As morphological analysis forms an important component of linguistic modelling, it becomes crucial to effectively store inflections of words. The highlight of this paper is a novel approach to output the given root words along with

their inflected forms in short numeral notation. In future, the proposed work can be extended to find inflections of other parts of speech like verbs, adverbs, etc.

References

1. Yordchim S (2012) Inflections in English nouns, verbs, and adjectives. J Royal Inst Thailand
2. Saharia N, Sharma U, Kalita J (2010) A suffix based noun and verb classifier for an inflectional language. In: *2010 International conference on asian language processing*, pp 19–22
3. Kosti´c A, Katz L (1987) Processing differences between nouns, adjectives, and verbs. Psychol Res 49(4):229–236
4. Nugues P, Words, parts of speech, and morphology. In: Language processing with perl and prolog, Springer, pp 169–203
5. Carstairs-McCarthy A (2017) Introduction to English Morphology: words and their structure. Edinburgh University press
6. Z´aˇcek M, Homola D (2017) Analysis of the English morphology by semantic networks. In: *AIP conference proceedings*,1906, AIP Publishing LLC, pp 080006
7. Backhouse AE (2016) Inflected and uninflected adjectives. Adjective Classes: A cross-linguistic typology 2(5):99–110
8. Ralli A (2000) A feature-based analysis of greek nominal inflection
9. Dincer T, Karaoglan B, Kisla T (2008) A suffix based part-of-speech tagger for Turkish. In: *Fifth international conference on information technology: new generations*, pp 680–685
10. Sharma U, Kalita JK, Das R (2008) Acquisition of morphology of an indic language from text corpus. ACM Trans Asian Lang Inf Process 7(3):1–33
11. Dasgupta S, Khan N, Sarkar A, Pavel D, Khan M (2005) Morphological analysis of inflecting compound words in bangla

Forecasting Cryptocurrency Time Series Using Adaboost-Based Ensemble Learning Techniques

Haritima Manchanda and Swati Aggarwal

Abstract In the domain of time series, cryptocurrency data is one of the most complex data available with its inherent volatility and non-linearity, thus making the task of analysis and prediction exceptionally challenging. The econometric models such as ARIMA and ARMA fail to capture the non-linearity of data putting forth the need to adopt other models for forecasting. In this paper, data of five cryptocurrencies-Bitcoin, Ethereum, Litecoin, XRP, and Stellar is analyzed in an effort to predict the next day closing prices of the respective cryptocurrencies. This goal is achieved by applying the ensemble learning technique of AdaBoost to boost the weak learners namely MLP, ELM, SVR, and LSTM all of which individually suffer from the problem of overfitting. Comparing the results produced by these combinations with the individual techniques, it can be proved that boosting gives significantly better performance accuracy as compared to the individual learning methods. Adaboost-LSTM gives the minimum MAPE of 3.6569, 6.1932, 13.1040, 11.0626, and 5.1058 for Bitcoin, Ethereum, Litecoin, XRP, and Stellar, respectively. Also, the consistency of the results produced was tested by applying the same techniques on the opening prices of cryptocurrencies. Adaboost-LSTM combination produced the minimum MAPE of 2.0087, 2.5135, 8.8680, 2.0351, and 3.8633 for the five cryptocurrencies, respectively.

Keywords SVR · LSTM · ELM · MLP · MAPE · Adaboost · Cryptocurrencies

1 Introduction

With each and every passing day, data from different information sources is organized as time series to make limitless choices based upon the study of past data. In this paper, an attempt is made to explore one such application of time series that is the cryptocurrency time series data. As an alternative to the traditional centralized financial ecosystem, cryptocurrencies have emerged as an interesting phenomenon because of their rapid acceleration in market capitalization mounting to a revenue of about

H. Manchanda (✉) · S. Aggarwal
Netaji Subhas University of Technology, Dwarka Sector-3, Delhi 110078, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_17

$100 billion in 2017 [1]. As of February 2020, there are about 2417 decentralized currencies circulating in the market [2].

A considerable amount of work has been done to test the accuracy of the common statistical models such as the Auto Regressive Integrated Moving Average (ARIMA) Model on the cryptocurrency time series [3, 4]. It has been proved that ARIMA fails to capture the complexity of the data and assumes that the circumstances in which the required data is collected remain the same throughout [5]. Therefore, to model the non-linearity of the data other machine learning methods such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), and deep learning solutions such as Long Short-Term Memory (LSTM)-Recurrent Neural Networks (RNN) have been used [6]. Research has also been done to predict the accuracy of cryptocurrency time series using Extreme Learning Machine (ELM) [7]. These models generally prove to be better than the traditional econometric models yet they suffer from the problem of overfitting. Therefore, it necessitates a need to introduce some methods to solve the underlying problems and to obtain better results.

The objective of the research is to analyze cryptocurrency time series and devise strategies to predict that with what accuracy the price of various cryptocurrencies can be predicted. Over the past decades, a lot of research effort has been made to explore and model the financial time series including the foreign exchange. But, not much significant work has been done to study the time series of cryptocurrencies. Therefore, the study and analysis of the cryptocurrency time-series data becomes extremely important. Here, the effect of most commonly used models for time series analysis such as ARIMA, SVR, MLP, and LSTM is tested on data of five cryptocurrencies. Due to the volatile and unpredictable nature of cryptocurrencies, these models overfit the data. Adaboost is used to integrate the forecasting results of the individual models to improve the performance of the models. Boosting would help to provide stability by reducing the variance in results.

2 Literature Review

As the cryptocurrency time series are extremely non-stationary with a reasonable amount of noise present, it becomes necessary to include their non-linearity into the prediction models while forecasting. Therefore, studies have been conducted to capture this non-linearity and complexity with varying degrees of accuracy. Linear Regression has been used to explore the forecasting capacity of the blockchain features on Bitcoin time series giving an accuracy of 55% for predicting one hour ahead in the future [8].

The results generated by Bayesian regression in the study by Shah and Zhang while being run on real trading data with an aim to form a successful trading approach based on Bitcoin time series provided a framework to almost double the investment in a time period limited to 60 days [9]. Though many conventional models such as the ARIMA model have been widely used and experimented, the neural networks have proved to exhibit better predictive capabilities [3]. Nakano applied a deep learning

approach with seven layer neural network structure on a data with technical indicators to forecast the future return trend of bitcoin [10]. McNally used Bayesian optimized RNN and LSTM to predict the Bitcoin price in USD and compared the results with the standard ARIMA model. It was proved that with an accuracy of 52% LSTM performed much better than the ARIMA model because of its ability to remember the behavior of the series over a long period of time [6]. LSTM has the ability to memorize long and short-term temporal information simultaneously [11]. In another paper ELM has been used over the bitcoin time series data collected for a period of three months giving the smallest Mean Absolute Percentage Error (MAPE) to be 2.657% [7]. Other approaches have been to be Support Vector Machine (SVM) [12] and random forests [13] for forecasting.

Boosting refers to a machine learning mechanism of creating a precise forecasting system from a comparatively weak machine learning model. The first empirical and feasible boosting algorithm was developed by Freund and Schapire in 1997 [14]. Adaboost has been applied to several machine learning algorithms to boost the performance. Adaboost based on multi-layer neural network has been used in a paper to predict the accuracy of respiratory signal used further in certain medical researches giving reasonable accuracy [15]. The Adaboost-LSTM ensemble learning technique has been tested on a time-series data to predict the future Sea Surface Temperatures (SSTs) proving that the combination of Adaboost-LSTM performed better than LSTM and Adaboost applied separately. It even outperformed other machine learning techniques used such as SVR and the optimized Back Propagation Neural Network model (BPNN). Another research has been done to forecast the internet traffic where Adaboost-LSTM has given better results than LSTM applied separately [16].

3 Research Gap

Most of the work mentioned in the last section has been done on only one cryptocurrency that is bitcoin, thus the consistency of the applied models on the plethora of cryptocurrencies available in the market today cannot be established. Substantial amount of research has been done to map ARIMA model and LSTM model on Bitcoin data, even the comparison between the two has been drawn, but minimal research has been done to test SVR and ELM models on cryptocurrency data. It has been proved that ARIMA fails to capture the seasonality property accurately which is an important component inherent in data, in turn producing inaccurate results. An alternative to ARIMA, the Seasonal ARIMA (SARIMA) model has not been widely tested on cryptocurrencies. Papers that examine SVR and ELM on bitcoin data use a small sample size, thus reducing the influence and impact of study. Adaboost can be used to prevent overfitting in the models, but it has not been used on cryptocurrency time series before.

In this paper, data of five cryptocurrencies namely Bitcoin, Ethereum, Litecoin, XRP, and Stellar is used instead of one to give a fair idea about the performance of different models on a variety of cryptocurrencies available in the domain. Also, to

test the consistency of the results, each model is applied and tested on the closing as well as the opening prices of data. An in-depth study has been done to evaluate the effect of SVR and ELM models on the five cryptocurrencies with a large sample size of 1493 entries, thus validating their impact. Here, SARIMA model is used instead of the ARIMA model to include the seasonality while measuring results. Therefore, SARIMA, MLP, LSTM, SVR, and ELM are tested on the data of five cryptocurrencies and their results are compared. Further to reduce the effect of overfitting and to stabilize the results, Adaboost is used treating the other models as weak learners. The outcome of applying Adaboost with MLP, LSTM, SVR, and ELM is examined and the results are compared.

4 Methodology

4.1 Data Pre-Processing and Analysis

Time series data of these five cryptocurrencies (Bitcoin, Ethereum, XRP, Litecoin, and Stellar) is collected from a period of December 26, 2015 to January 25, 2020 from coinmarketcap.com [2] giving 1493 entries. The data has seven features namely: Date, Open, Close, High, Low, Volume, and Market Cap which were further normalized. Figure 1 depicts the time series plot after decomposing the closing price feature of the five cryptocurrencies into trend and seasonality components. It clearly shows the presence of a seasonality component in data. Therefore, the machine learning techniques that model this seasonality need to be used.

4.2 Prediction Techniques

SARIMA Approach. The ARIMA (p, d, q) model is unable to process the seasonality component inherent in the data assuming that the seasonality if any in the time series data has already been dealt with. Therefore, SARIMA or Seasonal ARIMA approach which combines seasonality with ARIMA has been applied after converting the data into stationary time series by one level of differencing. SARIMA $(3,1,2)^{*}(1,0,1,12)$ is modelled on the closing price of five cryptocurrencies to predict the next day price.

MLP Approach. As the name suggests, Multi-Layer Perceptron consists of multiple layers of perceptrons in a feedforward network architecture. Each layer of the model consists of an activation function. Using 'ReLU' as the activation function for hidden layers and 'sigmoid' for dense layer, the MLP was trained over 1044 samples and tested for 447 samples. To improve the computational accuracy, Adaptive Model Estimation (ADAM) method was used. Introduced and developed by Kingma and Ba in the year 2014, ADAM works on adaptive estimates of lower order moments resulting in gradient-based optimization and performs best on problem statements

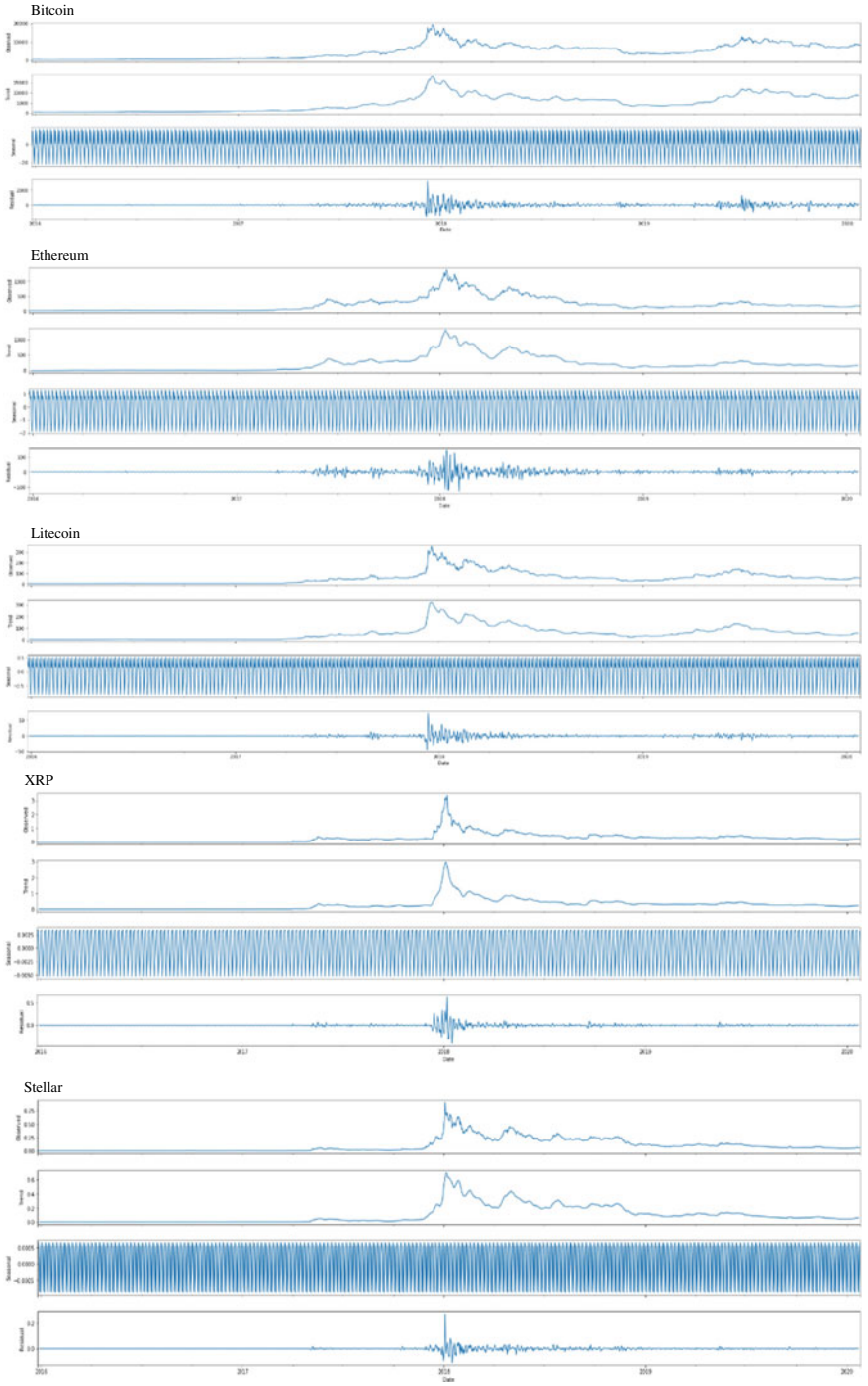


Fig. 1 Decomposition of cryptocurrency data

that require the study of large data or data with a number of parameters [17]. Figure 2 depicts the graphs of predicted cryptocurrency closing price vs the actual closing prices.

LSTM Approach. RNN suffers from the problem of vanishing gradient when dealing with very large time intervals. LSTM is a specific type of RNN which was developed to model long term dependencies accurately. Therefore, they are an effective machine learning technique for dealing with long range time series data. After scaling and reframing the series into supervised data, LSTM is applied on the five cryptocurrencies to predict next day closing price. Along with ADAM optimizer dropout regularization can be used to optimize the results by minimizing overfitting. Figure 3 depicts the graphs of predicted closing prices vs the actual closing prices after applying LSTM network with two hidden layers and 50 percent dropout rate.

SVR Approach. SVR primarily deals with fitting the error value in an acceptable range, adjusting it to a certain degree rather than minimizing the total error. Figure 4 depicts the graphs of predicted closing prices vs the actual closing prices after applying SVR model.

ELM Approach. ELM represents a single layer feedforward network (SLFN) which overcomes certain disadvantages of the traditional neural networks including the problem of overfitting and slow training speed of the network. Proposed by Huang et al., ELM is capable of adaptively setting up the number of hidden nodes in single hidden layer and randomly adjusting the weights of input to the network, weights of output from the network, and the hidden layer biases [18].

When the training set consists of N samples of the form (x, t), where 'x' represents the input and 't' represents the target value at a particular time instance, a SLFN with 'm' hidden neurons can be represented mathematically as in Eq. 1.

$$\sum_{i=1}^{m} a_i \phi(w_j x_j + b_i) \quad \text{where } i = 1, \ldots, N \tag{1}$$

Here,

- w_i: weight vector to be assigned randomly, connecting the hidden node i to the input nodes
- b_i: threshold of each hidden node i
- α_i: weight vector to be assigned randomly, connecting the hidden node i to the output node
- ϕ: activation function

Forecasting using Adaboost. The main aim of boosting is to improve the forecasting accuracy of a training sequence of weak learners where each component of the sequence compensates for the weakness characteristics of its predecessors.

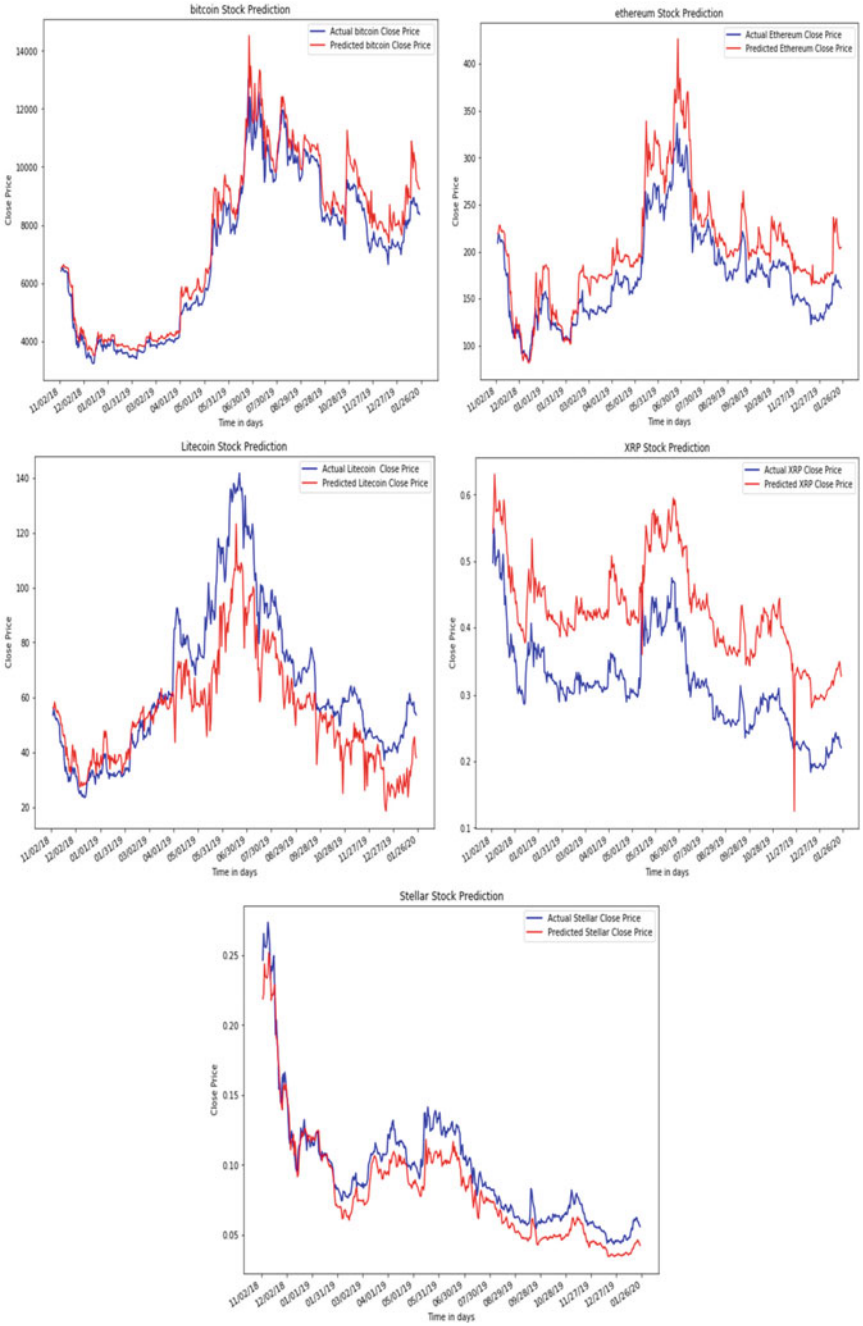


Fig. 2 Actual versus predicted price of cryptocurrencies after applying MLP

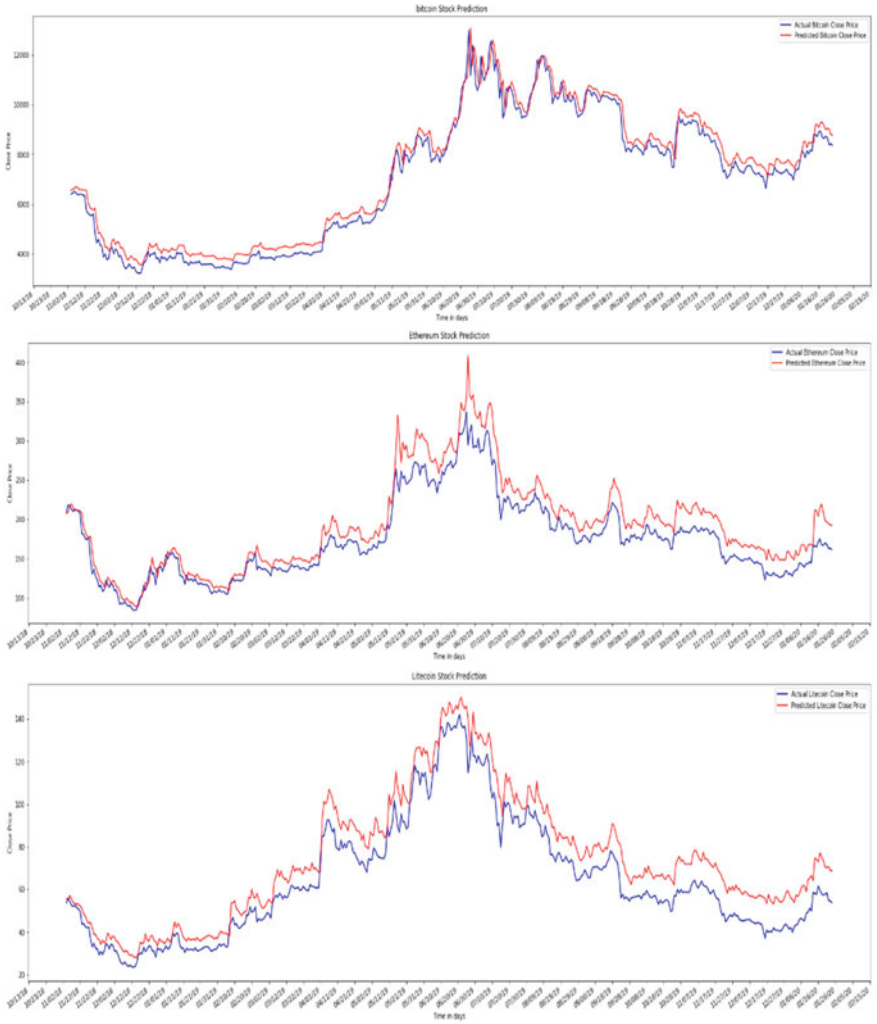


Fig. 3 Actual versus predicted price of cryptocurrencies after applying LSTM

Adaboost is one such boosting ensemble technique where the adaptive data resampling technique used helps to create a strong regressor by combining several weak regressors.

Each training sample is designated a weight that determines its probability of being selected for the training set for a weak regressor. The idea here is, the weights of samples that are accurately estimated by the weak regressor are reduced and the weights of samples for which the estimated values are deviated from the observed values are increased. Therefore, this helps in selecting the data samples which are poorly estimated by the previous regressor more often than the accurately estimated

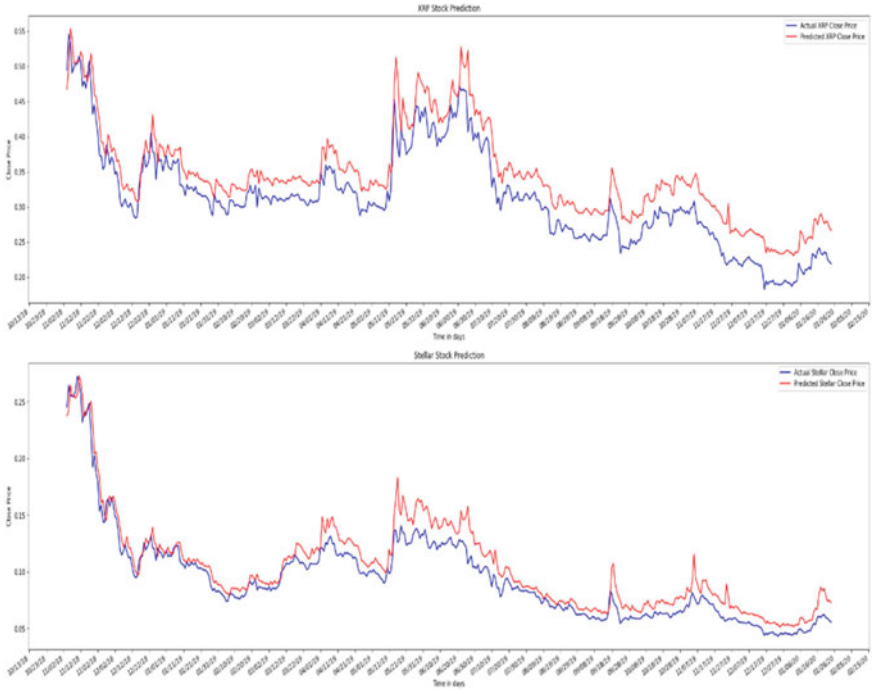


Fig. 3 (continued)

samples. These data samples are now fed to another weak regressor and their weights are updated. Successive regressors focus on poorly estimated data samples from the previous regressors. The whole process of data sampling and weight updation stops when either the error rate of the regressor reaches certain threshold (defined by the user) or maximum number of regressors are exhausted in this process.

5 Evaluation and Results

Table 1 depicts the cumulative results of MAPE calculated after applying SARIMA, MLP, SVR, LSTM, and ELM on the cryptocurrencies data without Adaboost and Table 2 depicts the cumulative results of MAPE calculated after applying the models with Adaboost.

Comparing the results of Table 1 and Table 2 it can be seen that Adaboost—LSTM gives the best results for all the cryptocurrencies. It signifies that the ensemble learning combination of Adaboost—LSTM is an efficient learning approach for cryptocurrency financial time series giving appropriate accuracy in terms of the mean absolute percentage error. Further, testing over five popular cryptocurrencies gives an idea of its applicability on most of the cryptocurrencies present today. The results

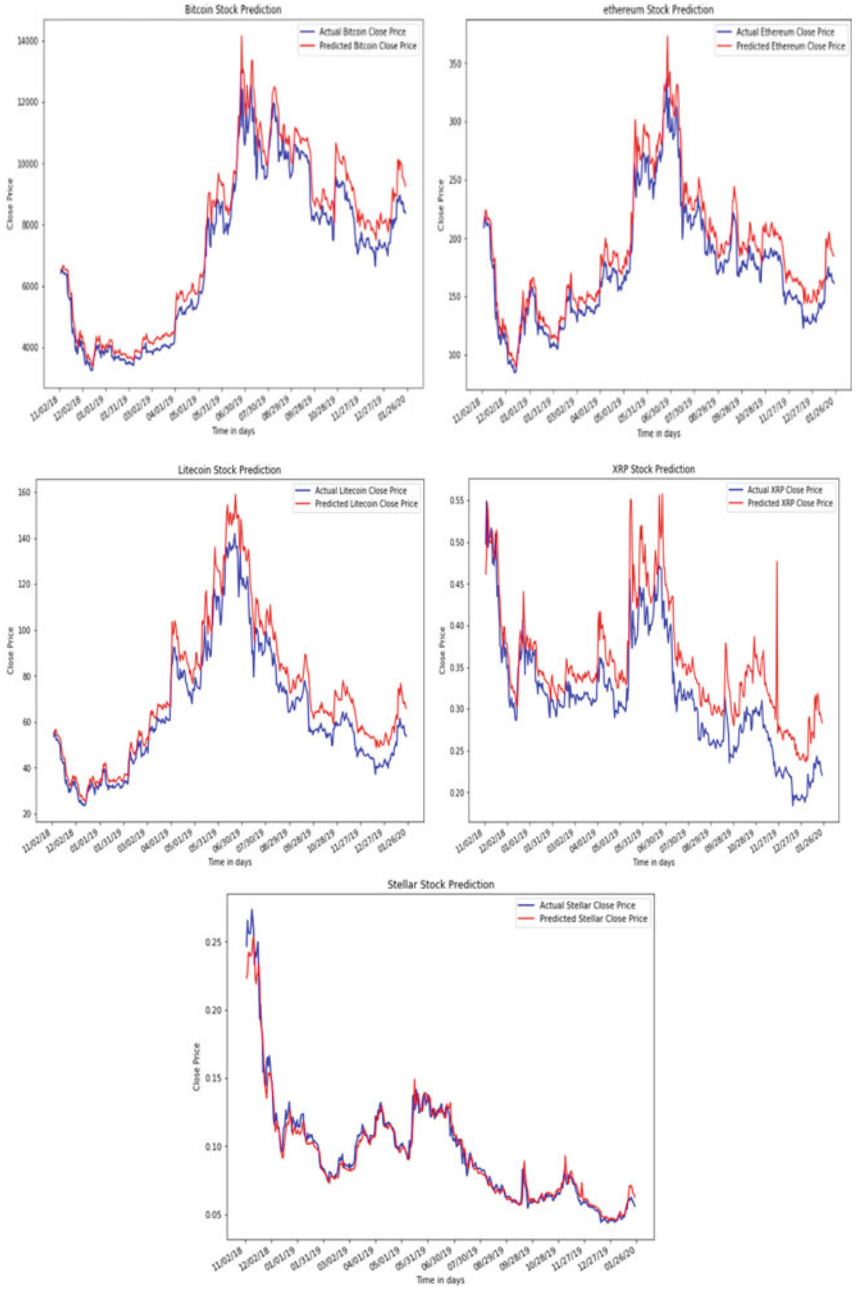


Fig. 4 Actual versus predicted price of cryptocurrencies after applying SVR

Table 1 Cumulative results of closing price without Adaboost (MAPE) (In %)

	Bitcoin	Ethereum	Litecoin	XRP	Stellar
SARIMA	38.1139	31.4681	33.4517	52.4274	87.4781
MLP	7.3182	16.4967	19.3749	23.2190	15.2266
LSTM	5.9584	10.6413	16.9251	12.2095	12.6672
SVR	5.7981	9.1852	13.6909	14.0875	4.5556
ELM	6.7801	7.0337	10.5909	25.3091	12.4287

Table 2 Cumulative results of closing price with Adaboost (MAPE) (In %)

	Bitcoin	Ethereum	Litecoin	XRP	Stellar
MLP	4.0778	7.9467	14.6250	15.09878	7.7557
LSTM	3.6569	6.1932	13.1040	11.0626	5.1058
SVR	5.5964	8.2199	10.2963	14.0591	4.5487
ELM	5.4442	6.4935	9.3397	16.5874	10.0849

also prove that forecasting performance of the machine learning techniques with Adaboost is better than without. One of the reasons for this effect is that boosting a learning method reduces the possibility of overfitting.

To test the consistency of the results produced, the same operations are also performed taking opening price of the cryptocurrencies as the target variable and 'high', 'close', 'low', 'volume', and 'market cap' as the input variables. The results of the experimentation with and without Adaboost are given in Tables 3 and 4, respectively. Testing with opening price proves the stability and the consistency

Table 3 Cumulative results of opening price without Adaboost (MAPE) (In %)

	Bitcoin	Ethereum	Litecoin	XRP	Stellar
SARIMA	38.1383	32.0086	33.5098	53.7114	85.1831
MLP	5.4124	19.8444	12.5775	12.7797	7.2244
LSTM	2.1492	8.5326	11.0472	7.0694	6.9542
SVR	5.2373	12.4899	10.7637	14.5104	5.4497
ELM	2.3407	5.4757	16.3137	4.5074	5.5557

Table 4 Cumulative results of opening price with Adaboost (MAPE) (In %)

	Bitcoin	Ethereum	Litecoin	XRP	Stellar
MLP	3.6332	8.9294	11.6858	A3.0195	4.6532
LSTM	2.0087	2.5135	8.8680	2.0351	3.8633
SVR	4.8109	12.0062	9.3057	7.5490	5.2989
ELM	2.1247	3.4555	10.5843	2.5653	4.0334

of the model. Here, also Adaboost—LSTM ensemble learning gives the minimum MAPE and for all weak learners used. Hence, Adaboost significantly improves the performance.

6 Conclusion and Future Work

The main purpose of the research was to analyze and perceive the underlying complexities of the cryptocurrency financial time series data. Cryptocurrency data was inspected and decomposed into its various components to gain a better understanding of which machine learning approaches would produce best results. The mathematics involved in regards to time series forecasting was explored and the functioning of different machine learning techniques was examined. Stochastic models like ARMA and ARIMA models are considered the best candidate solutions for analyzing any available time series data, but they work only for univariate time series and do not capture the trend and the seasonality components of data. A modified stochastic model called the SARIMA model designed to include the seasonality was applied on the data but it failed giving large error in the form of MAPE. Machine learning models such as SVR and ELM were tested on the data and the results generated prove that these models handle the forecasting fairly. Further, the deep learning techniques were explored starting from a basic MLP and dealing with more advance approaches such as LSTM. After testing all these models individually on the data, they were then fed as weak learners to Adaboost boosting algorithm.

Overall results show that using Adaboost reduces the error percentage of forecasting. The combination of Adaboost-LSTM gives better results than all the other combinations and even LSTM applied individually standing at 3.6569%, 6.1932%, 13.1040%, 11.0626%, and 5.1058% for Bitcoin, Ethereum, Litecoin, XRP, and Stellar, respectively. Adaboost provided stability to the model by understanding the complexity in the data and reducing the overall variance in results. Next, all the approaches with and without boosting were also applied on the opening price of the data. The results for predicting the next day closing prices and for predicting the next day opening prices give similar results proving the consistency of the approaches on the given data.

The algorithm for now predicts only one step ahead in future, but changes could be made to the model to allow multi-step future prediction. This multi-step prediction if implemented effectively would provide for potential gains in the cryptocurrency financial market. However, this form of prediction suffers from a complication of accumulating errors at each step. This can be solved by using the prediction result of each step as an input to the next [19].

The entire process involved in the research could be made faster to a degree by applying some sort of GPU optimization in the approaches used. The TensorFlow library used in this research applies GPU optimization to some extent but more changes can be put in place. A hybrid of different machine learning architectures can be derived and applied on the data. For instance, the ARIMA model can be combined with LSTM and the results can be examined.

References

1. ElBahrawy A, Alessandretti L, Kandler A, Pastor-Satorras R, Baronchelli A (2017) Evolutionary dynamics of the cryptocurrency market. R Soc Open Sci 4(11):170623
2. Coinmarketcap (2020) https://coinmarketcap.com/, last accessed 2020/01/01
3. NA Bakar, S Rosbi (2017) Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: a new insight of bitcoin transaction. Int J Adv Eng Res Sci 4(11)
4. Rebane J, Karlsson I, Denic S, Papapetrou P (2018) Seq2Seq RNNs and ARIMA models for cryptocurrency prediction: a comparative study. SIGKDD Fintech 18
5. Montgomery DC, Jennings CL, Kulahci M (2015) Introduction to time series analysis and forecasting. John Wiley & Sons
6. McNally S, Roche J, Caton S (2018) Predicting the price of bitcoin using machine learning. In: *2018 26th Euromicro international conference on parallel, distributed and network-based processing (PDP)*. IEEE, pp 339–343
7. Faizal R, Setiawan BD, Cholissodin I (2019) Prediksi Nilai Cryptocurrency Bitcoin menggunakan Algoritme Extreme Learning Machine (ELM). Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN 2548:964X
8. Greaves A, Au B (2015) Using the bitcoin transaction graph to predict the price of bitcoin. No data
9. Shah D, Zhang K (2014) Bayesian regression and bitcoin. In: *2014 52nd annual Allerton conference on communication, control, and computing (Allerton)*. IEEE, pp 409–414
10. Nakano M, Takahashi A, Takahashi S (2018) Bitcoin technical trading with artificial neural network. Phys A 510:587–609
11. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
12. Karasu S, Altan A, Saraç Z, Hacioğlu R (2018) Prediction of bitcoin prices with machine learning methods using time series data. In: *2018 26th signal processing and communications applications conference (SIU)*. IEEE, pp 1–4
13. Madan I, Saluja S, Zhao A (2015) Automated bitcoin trading via machine learning algorithms. http://cs229.stanford.edu/proj2014/Isaac% 20Madan, 20
14. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55(1):119–139
15. Sun W, Jiang M, Chen G, You T (2019) A comparison of adaptive boosting algorithms for the respiratory signal prediction. In: *2019 12th international congress on image and signal processing, biomedical engineering and informatics (CISP-BMEI)*. IEEE, pp 1–5
16. Xiao C, Chen N, Hu C, Wang K, Gong J, Chen Z (2019) Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. Remote Sens Environ 233:111358
17. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980
18. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501
19. Ouyang Y, Yin H (2018) Multi-step time series forecasting with an ensemble of varied length mixture models. Int J Neural Syst 28(04):1750053

A Fingerprint Recognition System Using Raspberry Pi

Brajesh Singh, Urvashi Dhand, Paras Bhardwaj, Supriya Mishra, and Gaurav Arora

Abstract Fingerprint matching is a way to identify an individual's identity or can be used as verification depending on the requirements. There are many techniques involved in fingerprint matching. We know that a fingerprint consists of lots of unique points which differentiate one fingerprint from another they are called minutiae. In this paper, we will define various techniques that are used for Minutiae extraction and matching like crossing number, minutiae position, minutiae orientation and how we have implemented fingerprint recognition with Raspberry Pi. It is a small size computer which is based on Debian Linux.

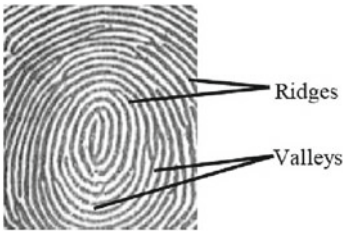
Keywords Fingerprint matching · Crossing number · Minutiae extraction · Minutiae matching · Raspberry Pi

1 Introduction

Biometric system depends on behavioural and physiological features to recognize a person. The behavioural data consists of signature, speech and keystroke which changes with respect to age and surroundings. The Physiological features include face, fingerprint, palm print, iris these features do not change throughout the lifetime of the person. The system works on verification mode or identification mode or any other mode depending on requirements of an application. Out of all these parameters we choose that parameter which must be suited upon a distinguishable human attribute such as fingerprint. We have been using fingerprint for the past many years in biometrics system because of its feasibility, distinctiveness, permanence, accuracy, reliability and acceptability. A fingerprint is the replication of the fingertip epidermis, produced when a finger is pressured against a plain surface. A fingerprint consists of pattern of ridges and valleys ridges are dark and valleys are bright.

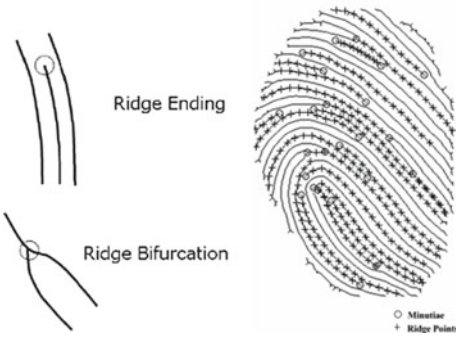
B. Singh · U. Dhand (✉) · P. Bhardwaj · S. Mishra · G. Arora
HMR Institute of Technology and Management, New Delhi, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_18



Fingerprint is also considered as a form of lines and minutiae which are drawn out using inked impression or sensors. The minutiae are unique for every fingerprint, its position and orientation are different for every person so for identifying or verifying the fingerprints we can extract the minutiae from the input and compare it with the template fingerprint already stored in the training dataset, From there we can calculate the matching score and determine the results.

There are various ways for extracting minutiae mostly people detect the ridge ending and bifurcation (bifurcation means where the ridges part their ways).



After the extraction of the minutiae we can now store this information in the training dataset. Now comes the matching where we take the input fingerprint and pre-process it like binarization and thinning we extract its minutiae and match it with already stored template fingerprint. Matching can be done by many ways.

(1) We compare the minutiae type then,
(2) For every minutiae in fingerprint p we search a minutiae in fingerprint q by distance and direction thresholding.

After this the fingerprint is matched with one of those present in the database, this is mostly used to provide access but nowadays can be used for time clock or payroll access.

2 Minutiae Extraction

Minutiae is the pattern on the fingerprint which can be extracted from using the inked impression or when a finger is pressed against a sensor.

The input fingerprint image is gray scale image, in this as we know ridges are dark lines and valleys are the whitespaces between the ridges. Minutiae point is where ridges either terminates or bifurcates (i.e. splitting into two ridges).

We need to pre-process the images to reduce the noise in them so that we can avoid the detection of spurious minutiae. By pre-processing we can get a good quality image from which we can identify the useful minutiae's which is used to identify a person.

Preprocessing

In one of the ways, the fingerprint image is pre-processed by first passing it to median filter which may be a nonlinear smoothening methodology that may reduce the blurring of edges, Then the image is binarized after that orientation field map is obtained then region of interest is got by cropping the particular region. Then this binary image is thinned, then the removal of spikes was done after the minutiae got extracted.

In one of the another way we came across they follow two steps that is binarization and block filter.

Binarization

In this we convert the image's pixel value into binary value by setting a threshold. The pixel having value above the threshold are set to '255' and others are set to'0'.

Block Filter (Thinning)

The images are thinned using this filter, this reduces the width (thickness) of the ridges up to a single-pixel width to extract minutiae point effectively Thinning is done mostly to preserve the outermost pixel so to do that we put white pixels at the outside (boundary) of the image, so the output will be that the first five rows and last five rows the first five columns and the last five columns are assigned the value one. The neighbouring 8 pixels values are accessed based on which a pixel is changed Dilation and erosion are used to thin the ridges.

Minutiae Extraction

After the minutiae is extracted its location and angle (direction) is calculated crossing number tells us the type of minutiae point in the fingerprint image.

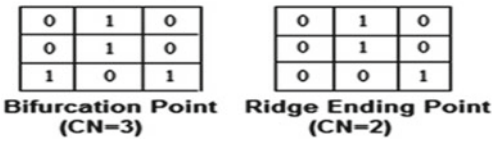
Crossing number is defined as

$$\mathrm{CN} = \frac{1}{2}\sum_{i=1}^{8} |P_i - P_{i+1}|$$

It is defined as sum of the difference between the adjacent pixels.

To identify that which if the pixel we are currently working with is ridge ending or bifurcation we take into account the 8 neighbouring pixels and then find the crossing number.

CN	Property
0	Isolated Point
1	Ending Point
2	Connective Point
3	Bifurcation Point
4	Crossing Point



If the crossing number is calculated to be 0 then that point is called isolation point if it comes out to be 1 then it is ending if it is 2 it's a normal ridge point if the crossing number is 3 then it is ridge bifurcation if it is 4 then it is crossing point.

Now to calculate the bifurcation angle, we have a well known fact which is that ending and bifurcation are having dual nature. The picture of ending agree to the bifurcation in its negative picture so by putting the identical set of rules we can get the bifurcation angle.

3 Minutiae Matching

It is used to differentiate the input fingerprint and the fingerprint which is already stored in the data base.

Each minutiae is expressed as a tuple of three parameters, i.e.:- the X and Y coordinate, direction or orientation and type.

Let m be a minutiae point, then m is stored in this form

$$m = ((X_i, Y_i), \theta_i, T_i)$$

where X and Y are the position coordinates of the minutiae point θ describes the direction and T describes the type of minutiae.

We should also remove the false minutiae before the matching if two minutiae's are having less distance between them (i.e. distance between an ending and bifurcation or two points on bifurcation) they might be considered as the same minutiae point and hence be terminated.

We also remove the two minutiae points if they are located in a short ridge with length less than threshold. After removing the false minutiae we can proceed for the matching.

Now we take a minutiae point from the test image and do the matching we search all the minutiae points for the best match for this we do two calculations

$$\text{Position} = 1 \text{ if } \sqrt{(X_j - Y_j)^2 + (X_i - Y_i)^2} <= R_o$$

R_o = threshold value

Here 1 means that the point is the perfect match for the given point in the test image

$$\text{Orientation} = 1 \text{ if } \min\left(\lfloor \theta_j - \theta_i \rfloor, 360 - \lfloor \theta_j - \theta_i \rfloor\right) <= \theta_o$$

We count the number of one in this and then match if they are more than the threshold and if they are then the fingerprint is matched otherwise not matched.

We take a point from the test image and match with all the minutiae's available and then if find the suitable minutiae then start with the next point and goes on until all the points are searched.

This is the way we have implemented to find matching fingerprint.

There are many other ways for minutiae matching here is the overview.

Now let us look at another way in which the minutiae matching is done by using both the local and global structures the local structure consists of rotation and translation invariant feature of the minutiae in its neighborhood. It is used to find the correspondence in two minutiae sets and increase the reliability of the global matching. The global structure determines the uniqueness of the fingerprint so the combination of both gives a strong basis for the a powerful minutiae matching.

The feature map in the first method is very prone to changes due to rotation and translation which cannot make up a robust minutiae matching. On the other hand, only using the local structure (distances ridge count and directions between two nearest minutiae) is also not advisable because two fingerprint from different finger may have same local structure which gives rise to false acceptance.

However when local and global structures are used for the minutiae matching then it creates a robust model as according to the local structures the other fingerprint database images can be aligned and the global structure which defines the uniqueness will be used to identify the person so the drawback of global structure is fulfilled by local structure and the drawback of the local structure is fulfilled by the global structure.

The next method is where we developed a scheme for minutiae score and use it in fingerprint matching. The scoring scheme is developed on three aspects.

One can be the distance between the minutiae point and the center of the fingerprint for a good score minutiae should be closer to the center, next can be orientation coherence in which the fingerprint image is divided into many blocks and the current orientation of the point is checked with the main orientation of the block which gives us the score if the orientation is difference the value is 0 otherwise 1. The third and last is that the local factors should be taken into account as in the other two methods the bad minutiae may have high score in this we take into account the pseudo and the

genuine minutiae points and calculate the score as the pseudo minutiae have more influence so it should be given less weightage, so the larger the score the better the minutiae quality.

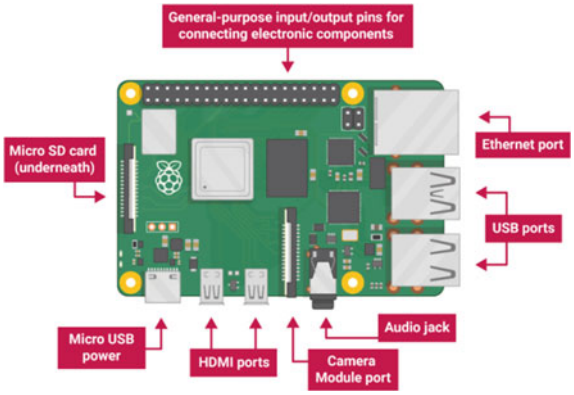
Finally, a total score is generated by combining all the factors using product rule and linear weightage, this scheme of scoring can be applied to the fingerprint matching technique.

4 Hardware Implementation

Raspberry Pi is a small card size computer that operates on Debian Linux. It is easy to use and learn. The operating system of Raspberry Pi can be installed easily by just adding a simple Sd card. The Sd card will contain the OS and it will be added into the Pi board and to power this board a simple USB mobile charger will be required, After this just connect the keyboard and mouse and a TV or computer will be connected by an HDMI cable and it is ready for use. If we need to work on a windows platform we might need Putty and VNC Viewer application to open the Pi desktop on the windows platform. Now as the Pi desktop is opened it can do multiple things you can send emails or create spreadsheets. Raspberry Pi provides us with an opportunity to code with scratch or python. As we are here working with an optical fingerprint sensorR307 so we need to import multiple libraries like Opencv or others.

Raspberry Pi contains a command prompt from which we can import the libraries, It contains Ide's so that we can code easily and run to see the function happening properly.

So here we can start our working, the fingerprint sensor will capture the image of the fingerprint and store it in the PI and with the help of Opencv which will fetch the image and from where we can apply the image pre-processing like binarization and thinning and extract features and then we can match it with the already stored features.

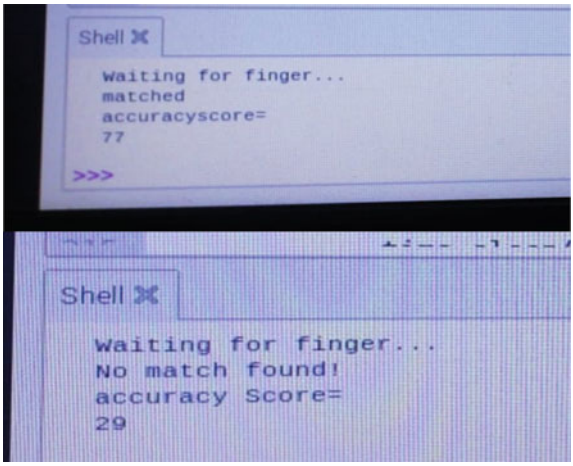


5 Dataset

To implement this Minutiae matching technique we have collected 5 images per person from which we have collectively taken out unique features (minutiae) we have taken a total dataset of 50 images.

6 Results

This fingerprint matching is implemented using Raspberry Pi3 here we have built the dataset of 50 images (5 images per fingerprint), minutiae's are extracted from them and stored, It is a real-time application where user when places the finger on the sensor the image is taken and its minutiae's are calculated and matched with the stored database. Matching is done on the type of minutiae, the position (x and y coordinate) and the orientation of the input minutiae and the stored minutiae, after that matching score is calculated and if it is more than the threshold we say that the fingerprint is matched. There is also a time delay which means the time taken by the system starting from placing the finger and displaying the result is near about 1 min. Here we have shown that the matched fingerprint has accuracy more than 60 so 60 is considered as our threshold value, Which is calculated by observing the scores of the matched fingerprint.



7 Conclusion and Future Scope

Biometric is nowadays the most useful technique in security or any other field like taking attendance or maintain records of the person's identity to do this we need that biometric feature which is easily available and remains permanent throughout the person's lifespan which is fingerprint in this paper we have given a review about various techniques used in minutiae (Pattern present in the fingerprint) matching like local-global structures, distance and direction thresholding, or data matrix (which contains minutiae position, type and direction).

Bibliography

1. Ravi J, Raja KB, Venugopal KR (2009) Fingerprint recognition using minutia score matching. Int J Eng Sci Technol 1(2):35–42
2. Jiang X, Yau WY, Fingerprint minutiae matching based on the local and global structures
3. Krishnamoorthi M, Lavanya K, A scalable approach for fingerprint matchine system based on minutiae extraction. In:IEEE sponsored 2nd international conference on innovations in information embedded and communication systems ICIIECS'15
4. Tong XF, Li PF, Han YX (2012) An algorithm for fingerprint basic minutiae feature matching. In: Proceedings of the 2012 international conference on machine learning and cybernetics, Xian
5. Adhiyaman M, Ezhilmaran D (2015) Fingerprint matching and similarity checking system using minutiae based technique. In: 2015 IEEE international conference on engineering and technology (ICETECH), Coimbatore, TN, India
6. Zheng X, Wang Y, Zhao X, Wei Z (2008) A scheme for minutiae scoring and its application to fingerprint matching. In: Proceedings of the 7th world Congress on intelligent control and automation. Chongqing, China

7. Mohsen SM, Zamshed Farhan SM, Hashem MMA (2004) Automated fingerprint recognition: using minutiae matching technique for the large fingerprint database. In: 3rd international conference on electrical & computer engineering ICECE 2004, Dhaka, Bangladesh
8. Afsar FA, Arif M, Hussain M (2004) Fingerprint identification and verification system using minutiae matching. In: National conference on emerging technologies
9. Kaur M, Singh M, Girdhar A, and Parvinder S. Sandhu (2008) Fingerprint Verification System using Minutiae Extraction Technique. World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering 2(10)
10. Bhargava N, Bhargava R, Narooka P, Cotia M (2012) Fingerprint recognition using minutia matching. Int J Comput Trends Technol 3(4)
11. Sivaranjani S, Sumathi S, Implementation of Fingerprint and Newborn Footprint Feature Extraction on Raspberry Pi. In: IEEE sponsored 2nd international conference on innovations in information embedded and communication systems ICIIECS'15

Impact of Underlap/Overlap of Germanium Source Dopingless Tunnel Field Effect Transistor (Ge-S-DLTFET)

Kumari Nibha Priyadarshani and Sangeeta Singh

Abstract This work illustrates the impact of underlap/overlap of Germanium source doping less tunnel field effect transistor (Ge-S-DLTFET) variation on the tunneling length and the device performance. Germanium region length is considered up to 2 nm inwards of gate contact (Str I), till gate contact (Str II) and 2 nm away from gate contact (Str III). This germanium region variation has potential to alter the electron concentrations, energy band and thus the tunneling length at source-channel junction. The tunneling length is reported to be minimum for Str III, facilitating its highest drive current. Further, due to higher tunneling length for Str I the leakage current is decreased and thus Str I gives maximum I_{ON}/I_{OFF} ratio. Thus, Str I is expected to demonstrate the minimum static power dissipation. The in-depth dc parameters analysis has also been done for the three structures and compared.

Keywords Germanium source · Tunnel FET · Dopingless

1 Introduction

TFETs have gained the interest of device researcher in recent era because of its applicability in ultra-low power designs, immunity toward short channel effects (SCEs) and its potential to prevail over the 60 mV/decade bottleneck of MOS device subthreshold swing [1–4]. TFET gives lesser SS and thus fast switching due to the difference in carrier transport mechanism. Transport of carrier in MOS device is due to drift-diffusion and thus its SS is limited by thermal voltage, but the transport of carrier in TFET is due to band-to-band tunneling [2–5]. The problem of random dopant fluctuation and trap-assisted tunneling which is dominant in physically doped TFET is eradicated in dopingless TFET [6, 7]. DL-TFET also exhibits an ease of fabrication with lower thermal budget, because of the absence of expensive diffusion or ion implantation processes to form ultra-precise doping profiles for drain and source region in nanoscale devices [8–11].

K. N. Priyadarshani · S. Singh (✉)
National Institute of Technology, Patna, India
e-mail: sangeeta.singh@nitp.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_19

The major disadvantage of the TFET is its low ON current. Use of low band gap material in source region, double gate, dual metal gate etc. [2–5], has been reported to improve the drive current of device.

In this work, we present the impact of underlap/overlap of Germanium source dopingless tunnel field effect transistor (Ge-S-DLTFET) variation on the tunneling length and the device performance, by studying its effect on electron concentration and on tunneling length. There are three different device structures are considered, for Str I the germanium region is considered up to 2 nm inwards of gate contact, for Str II germanium source is till the gate contact and finally, for Str III the germanium source is considered till 2 nm away from gate contact (Str III).

The manuscript has been organized as follows: Sect. 2 discusses the device geometry and simulation models incorporated. Section 3 covers the dc performance parameters. Section 4 concludes this investigation.

2 Device Structure and Simulation Models

2D device structure of Ge source DLTFET with variation in germanium source length with structure nomenclature as Str I, Str II and Str III of Ge-S-DLTFET is shown in Fig. 1a–c, respectively. To keep the induced doping concentration uniform the body thickness of device should be kept less than Debye length (L_D) [12–17], i.e., $L_D = ((\varepsilon V_T)/qN)^{1/2}$, where, V_T is the thermal voltage, ε is the dielectric constant

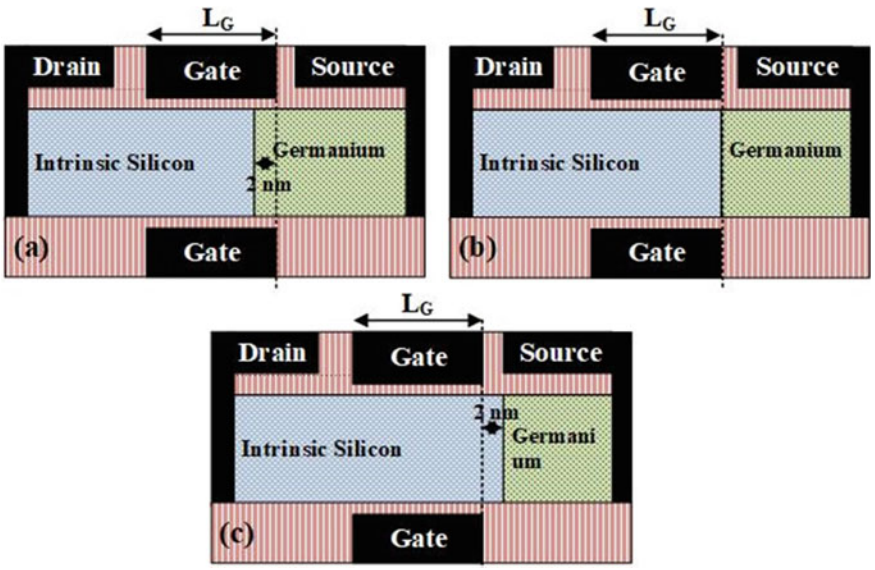


Fig. 1. 2D device structure of **a** Str I, **b** Str II and **c** Str III of Ge-S-DLTFET

of the body material (Si or Ge) and N is the carrier concentration in the body. Hence, the film thickness of the device is taken as 10 nm. For dopingless concept to function properly the work-function of source and drain metal contacts are taken as 5.93 eV and 3.9 eV for p^+ and n^+ doping, respectively. Suitable metal for source contact can be Pt and drain contact can be Hf. The gap between source and gate electrode is taken as 2 nm and gate and drain electrode is taken as 15 nm. The dielectric used is Al_2O_3 (dielectric constant = 9.3). The gate length taken is 50 nm with work-function of gate metal contact as 4.5 eV. The thickness of oxide under gate contact is taken as 1 nm, whereas the thickness of oxide under drain and source contact is taken as 3 nm. All the above parameter is same for the three structures namely Str I, Str II and Str III. The only difference is the germanium source length, which is 2 nm inwards of gate contact for Str I, till gate contact for Str II and 2 nm away from gate contact for Str III.

The device structure simulation has been done on 2D TCAD Silvaco ATLAS device simulator [18]. It solves Poisson's equation along with current continuity equation. Non-local band to band tunneling model has been considered along with trap assisted tunneling, SRH recombination and auger recombination. Mobility models dependent on carrier concentration and electric field have also been considered. Further, the quantum mechanical effect has not been considered as the film thickness of device is 10 nm.

3 Results and Discussion

Figure 2a and b show the electron concentration below oxide–semiconductor interface for Str I, Str II and Str III in OFF state and ON state, respectively. As the germanium region for Str I is 2 nm inwards of gate contact, for Str II it is till the gate

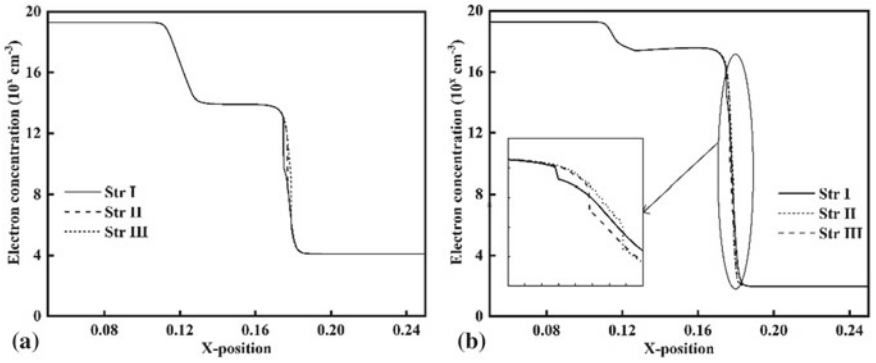


Fig. 2 Electron concentration at **a** OFF state ($V_D = 0.5$ V and $V_G = 0$ V) and at **b** ON state $V_D = 0.5$ V and $V_G = 0.5$ V

contact and for Str III it is 2 nm away from the gate contact. The electron concentration varies with shift in germanium region. The variation in electron concentration is due the electron affinity of germanium ($X_{Ge} = 4$ eV) which is less as compared to silicon's electron affinity ($X_{Si} = 4.17$ eV). The metal work-function needed for p^+ and n^+ induced doping is $\varphi_{m,p^+} > \chi_s + E_g/2$ and $\varphi_{m,n^+} > \chi_s - E_g/2$, where, E_g is energy band gap of semiconductor, χ_s is the electron affinity of bulk semiconductor. whereas, φ_{m,n^+} and φ_{m,p^+} are the workfunction of metals needed to induce n+ doping and p+ doping, respectively.

The decrease in electron affinity of germanium in comparison to silicon leads to increase in hole concentration and increase in electron concentration in germanium region. This causes the change in electron concentration with shift in germanium region for Str I, Str II and Str III as shown in Fig. 2a and b.

Figures 3a and b show the OFF state and ON state energy band diagram, respectively for Str I, Str II and Str III. The decrease in electron concentration with the introduction of germanium source results in upward shift of valance and conduction energy levels as shown in Fig. 3a and b. It can be seen in Fig. 3a and b that the tunneling length increases with shift of germanium region toward gate, with tunneling length for Str I > Str II > Str III for both OFF and ON state. The variation in tunneling length is high for OFF state and lower for ON state.

Figure 4 shows the I_D-V_G graph for Str I, Str II and Str III at drain voltage 0.5 V. Str I has highest tunneling length, thus, lower current and Str III has least tunneling length resulting in maximum current. The variation in OFF state current is higher as compared to ON state current because of higher variation of tunneling length for OFF state as compared to ON state. The maximum ON state current is reported for Str I and minimum OFF state current is demonstrated for Str III. Whereas, the ambipolar current is same for all the three structures.

Further, the dc parameter of three structures has been analyzed. Figure 5a–f show the threshold voltage, ON current, OFF current, I_{ON}/I_{OFF} ratio, transconductance and SS, respectively.

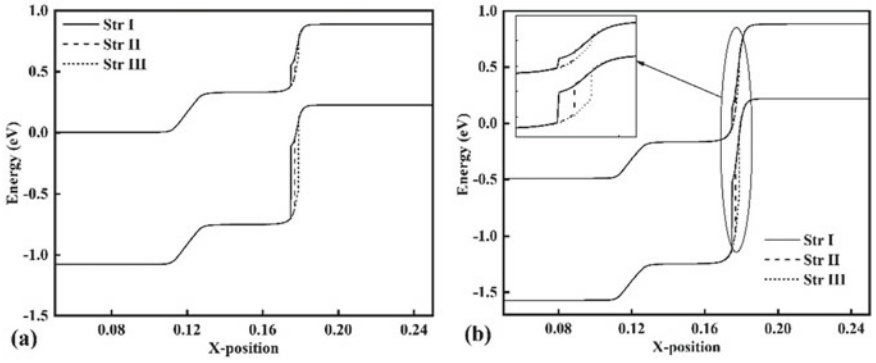


Fig. 3 Energy band diagram at **a** OFF state ($V_D = 0.5$ V and $V_G = 0$ V) and at **b** ON state ($V_D = 0.5$ V and $V_G = 0.5$ V)

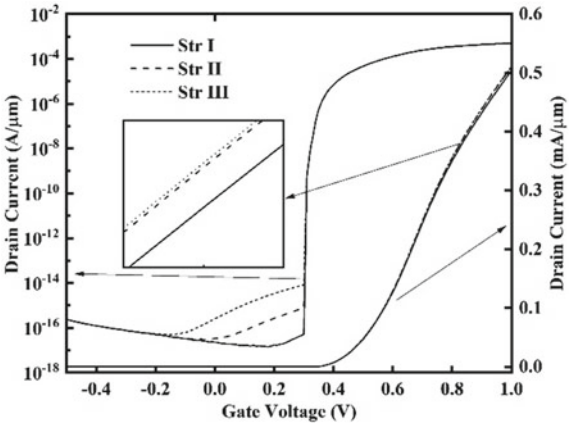


Fig. 4 I_D-V_G of Str I, Str II and Str III of Ge-S-DLTFET at $V_D = 0.5$ V

The threshold voltage is lowest for Str I (germanium region 2 nm inwards of gate contact) then increases for Str II (germanium region till gate contact) and again decreases for Str III (germanium region 2 nm away of gate contact). ON current increases with decrease in tunneling length with maximum ON current for Str III and minimum for Str I.

The OFF current also increases with decrease in tunneling length with minimum OFF current for Str I and maximum for Str III. The increase in OFF current is higher than increase in ON current. The I_{ON}/I_{OFF} ratio decreases with the shift of germanium region away from gate contact, with maximum value for Str I and minimum value for Str III. Increase in drain current leads to increase in transconductance with maximum transconductance for Str III and minimum for Str I. Further, increase in OFF state current with germanium region shifts away from gate leads to increase in SS.

The comparison of device characteristic parameter of proposed structure (Str I) with Ge source DLTFET [19] is shown in Table 1. It shows better performance of proposed structure in terms of drive current, I_{ON}/I_{OFF} current ratio, transconductance and the subthreshold swing.

4 Conclusion

In this work, the effect of underlap/overlap of Germanium source dopingless tunnel field effect transistor (Ge-S-DLTFET) variation has been analyzed. The germanium region length variation is done for 2 nm inwards of gate contact (Str I), till gate contact (Str II) and 2 nm away from gate contact (Str III). This variation in germanium region varies the concentration of electron at source-channel junction and it alters the energy band and tunneling length. The variation in tunneling length results in variation in device performance, with maximum drive current for Str III (Germanium region 2 nm

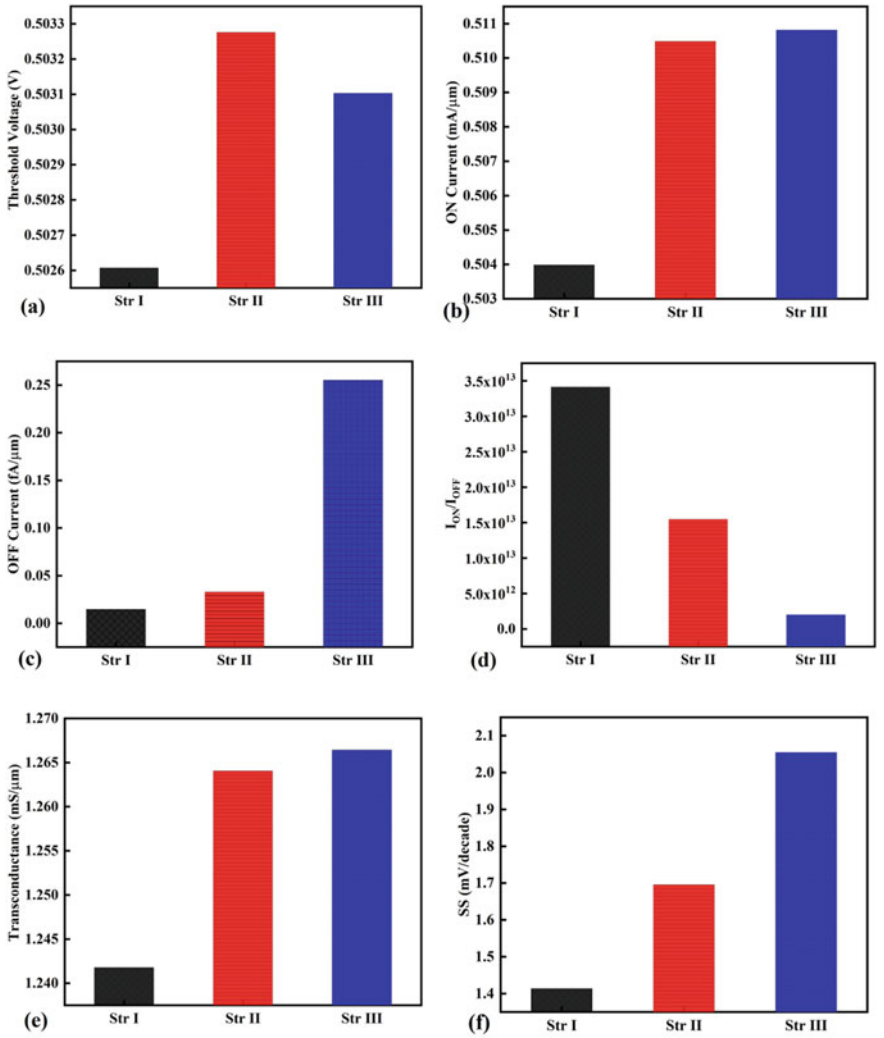


Fig. 5 **a** Threshold voltage, **b** I_{ON}, **c** I_{OFF}, **d** I_{ON}/I_{OFF}, **e** g_m and **f** SS for Str I, Str II and Str III of Ge-S-DLTFET at $V_D = 0.5$ V

Table 1 Device characteristic comparison of proposed structure with existing Ge source TFET

Structure	@V_D (V), @V_{Gmax} (V)	I_{ON} (μA/μm)	I_{OFF} (pA/μm)	I_{ON}/I_{OFF}	g_m (mS/μm)	SS (mV/decade)
Str I	0.5,1	504	1.47×10^{-5}	3.4×10^{13}	1.24	1.4
Ge source DLTFET [19]	1,1	116	8.81×10^{-6}	1.3×10^{13}	0.73	11

away of gate contact) but maximum I_{ON}/I_{OFF} ratio for Str I (germanium region 2 nm inwards of gate contact). Str I demonstrates the minimum static power dissipation.

Acknowledgements The authors are grateful to the National Institute of Technology, Jamshedpur, India for providing the computational resources.

References

1. Koswatta SO, Lundstrom MS, Nikonov DE (2009) Performance comparison between pin tunneling transistors and conventional MOSFETs. IEEE Trans Electron Devices 56(3):456–465
2. Boucart K, Riess W, Ionescu AM (2009) Lateral strain profile as key technology booster for all-silicon tunnel FETs. IEEE Electron Device Lett 30(6):656–658
3. Boucart K, Ionescu AM (2007) Double-gate tunnel FET with high-k gate dielectric. IEEE Trans Electron Devices 54(7):1725–1733
4. Choi WY, Park BG, Lee JD, Liu TJK (2007) Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec. IEEE Electron Device Lett 28(8):743–745
5. Jhaveri R, Nagavarapu V, Woo JC (2010) Effect of pocket doping and annealing schemes on the source-pocket tunnel field-effect transistor. IEEE Trans Electron Devices 58(1):80–86
6. Damrongplasit N, Shin C, Kim SH, Vega RA, Liu TJK (2011) Study of random dopant fluctuation effects in germanium-source tunnel FETs. IEEE Trans Electron Devices 58(10):3541–3548
7. Damrongplasit N, Kim SH, Liu TJK (2013) Study of random dopant fluctuation induced variability in the raised-Ge-source TFET. IEEE Electron Device Lett 34(2):184–186
8. Kumar MJ, Nadda K (2012) Bipolar charge-plasma transistor: a novel three terminal device. IEEE Trans Electron Devices 59(4):962–967
9. Ramaswamy S, Kumar MJ (2014) Junctionless impact ionization MOS: proposal and investigation. IEEE Trans Electron Devices 61(12):4295–4298
10. Amin SI, Sarin RK (2015) Charge-plasma based dual-material and gate-stacked architecture of junctionless transistor for enhanced analog performance. Superlattices Microstruct 88:582–590
11. Hueting RJ, Rajasekharan B, Salm C, Schmitz J (2008) The charge plasma PN diode. IEEE Electron Device Lett 29(12):1367–1369
12. Singh S, Kondekar PN (2017) A novel electrostatically doped ferroelectric Schottky barrier tunnel FET: process resilient design. J Comput Electron 16(3):685–695
13. Hur J, Moon DI, Han JW, Kim GH, Jeon CH, Choi YK (2017) Tunneling effects in a charge-plasma dopingless transistor. IEEE Trans Nanotechnol 16(2):315–320
14. Singh S, Singh AP, Kondekar PN (2017) A novel self-aligned charge plasma Schottky barrier tunnel FET using work function engineering. Microelectron Eng 168:67–75
15. Kumar MJ, Janardhanan S (2013) Doping-less tunnel field effect transistor: design and investigation. IEEE Trans Electron Devices 60(10):3285–3290
16. Nadda K, Kumar MJ (2015) Vertical bipolar charge plasma transistor with buried metal layer. Sci Rep 5:7860
17. Bashir F, Loan SA, Rafat M, Alamoud ARM, Abbasi SA (2015) A high-performance source engineered charge plasma-based Schottky MOSFET on SOI. IEEE Trans Electron Devices 62(10):3357–3364
18. ATLAS User Manual, Silvaco International, Santa Clara, CA (2015).
19. Cecil K, Singh J (2017) Influence of germanium source on dopingless tunnel-FET for improved analog/RF performance. Superlattices Microstruct 101:244–252

Meddoc Drive

Akshay Kumar Goel, Jayesh Gaur, Ajay Rose, and Deepika Rawat

Abstract Changing a doctor because of lack of satisfaction is a common phenomenon and this brings with it a plethora of problems for the patient. Every doctor would want all the tests to be conducted again and this is where the patient is left with no option but to conduct all the tests again. This is where a potential Medical Document Drive will come into the picture in which the patient will be allotted a unique number, i.e., a P.O number and all the tests and report of the patient will be saved online on a portal. While visiting a new doctor the doctor can check the reports, look at the symptoms and detect the disease without conducting the same tests again. This secure online application will be divided into three modules and those are the patient module, the admin module and the doctor module. The ultimate goal of this research is to create a secure portal that will have all the information of the patients and the admin will control the whole process.

Keywords Symptoms · Overlap-S Algorithm · Foreign key · Additive-F Algorithm

1 Introduction

Medical Document Drive will be a secure online portal through which the users will have all the required details about their health online. Users or patients will have usernames and passwords through which they can check the information needed by them anytime anywhere. This document drive will be divided into three broad modules and those are the patient module, the admin module and the doctor module. In the patient module, two algorithms will be used Overlap-S Algorithm and Additive-F Algorithm. The major role of the overlap S algorithm is to predict the disease by analyzing the different symptoms entered by the user/patient. Overlap-S Algorithm stands for Overlap Symptom Algorithm. The second algorithm is the Additive-F Algorithm and it is used to calculate the dosage of medicine analyzing the different

A. K. Goel (✉) · J. Gaur · A. Rose · D. Rawat
Department of Computer Science and Engineering, HMR Institute of Technology and Management, New Delhi, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_20

factors selected by the user/patient like age, habits. Additive-F Algorithm stands for additive factors algorithm. Another feature of this module is to add the feedback of the doctor he has consulted with and book the appointment of a doctor according to his time and date. The next module is the admin module. An Admin or the Administrator would be responsible for four important things. Firstly, the administrator would be responsible for registering a doctor. Secondly, as the admin would be having the authority of registering a new doctor, he/she definitely would be having the authority of removing a doctor. Thirdly, an administrator would easily be able to view the details of all doctors that are registered by him or her. Lastly, the most important part on which the administrator would be having his or her control is the feedback details part or the feedback details section. The last module is the doctor module. In the doctor module, the doctor will be able to view his schedule or previous appointments by retrieving it from the appointment table. In the next feature, the doctor will be able to add the reports of the patients who have consulted him before.

The remainder of this paper is organized into into five sections. Section 2 discusses the Patient Module and the algorithms. Section 3 discusses the appointment process of the doctor. Section 4 highlights the Admin Module. Section 5 discusses the Doctor module. Finally this paper concludes in Sect. 6 giving a brief conclusion on the discussions made in the paper and highlighting several future scopes.

2 Patient Module

In today's world, patients are facing a common problem all over the country. The problem is that they have to undergo their diagnostic tests many times within the same period of time just because they change their hospitals/doctors because of dissatisfaction. The new hospitals/doctors ask them for their reports which they have to either carry to all the places (if the old reports are accepted by the hospital) or have to undergo the same diagnostic tests again (if the hospital doesn't count on the previous reports) [1]. This causes problems and a state of unrest within the patients. In the proposed idea of the medical documents drive, the initial module is the patient module which generates a unique I.D known as the P.O number of a patient. This P.O number will be universal and inside this P.O number, all the reports of patients will be stored online so that by this unique P.O number, the doctors can access his old diagnostic reports anytime and anywhere.

Another feature of this module is that it helps the patients to get their disease information on base of their prima facie symptoms. This module has the Symptom checker system which detects the diseases on the basis of the symptoms entered by the user/patient and also suggest patients with the medicine. This Symptom Checker System predicts the disease on the basis of the "Overlap-S Algorithm". This algorithm looks over the overlapped symptoms and uses the Foreign Key concept which will be discussed later in this paper.

The next feature of this module is the Dosage Advisor feature which advises the dosage according to the various factors entered by the user according to the disease

that is predicted by the Symptom Checker System. The dosage is advised using another algorithm named as "Additive-F Algorithm". In this algorithm, the dosage is advised according to the factors entered such as age, smoking habits, drinking habits, etc. [2]. The algorithm is explained later in this paper.

In the first part of the module, the user is able to search the doctors according to his requirements (dentist, general, etc.). The list of doctors is displayed according to the backend database that is maintained by the administrator of this drive. Now let us look at the algorithms of this module.

There are basically 2 algorithms in this module. These are:

1. Overlap-S Algorithm
2. Additive-F Algorithm

Now, let us study each of the two algorithms.

2.1 Overlap-S Algorithm

The major role of this algorithm is to predict the disease by analyzing the different symptoms entered by the user/patient. Overlap-S Algorithm stands for Overlap Symptom Algorithm.

Now, let us understand the working of this algorithm.

Step 1. A disease has been given a unique Disease Code.

E.g.—Zika Virus has been given code '104'.

Step 2. The symptoms of that disease (one or many) have also been assigned the same code of that disease, such that the disease code works as a foreign key for both the tables 'Disease' and 'Symptoms'.

E.g.—The Disease code of Zika Virus = 104.

It's Symptoms –1. Eye Redness

2. Fever

1. and 2. have been assigned the same code '104'.

Step 3. Repeat Step 1 and Step 2 for all the given diseases and symptoms.

Step 4. When a user selects various symptoms, its codes are extracted from the table and the disease code that has most no. of symptoms, i.e., the disease in which most symptoms overlap is predicted and the output shows that disease to the user [3].

2.2 Additive-F Algorithm

The major role of this algorithm is to calculate the dosage of medicine analyzing the different factors selected by the user/patient like age, habits, etc., Additive-F Algorithm stands for Additive Factors Algorithm.

Now, let us understand the working of this algorithm.

Step 1. The factors such as BP, gender, age, smoking & drinking habits, previous disease, current living area are selected by the users on the interface.

Step 2. These factors are further assigned a value in 0/1 or 0/5 to every choice a user selects.

E.g—A person having drinking habits is given '5' value to that option and '0' if he doesn't have drinking habits.

Here, 5 indicates that he has a higher requirement of medicine dose as compared to the person with no drinking habits, i.e., value'0'.

Step 3. After we get all the values, the values of all the factors are added together [4]. The result after addition is the required dosage amount to the patient.

In an example, suppose the user enter these values:

In this, BP = High = 1; Smoking = Yes = 5; Drinking = Yes = 5; Previous Disease = Yes = 5; Age = 22; Gender = Male = 1; Living Area = Village = 1;

Adding these, we get 1+5+5+5+22+1+1 = 40. Hence, these were the two algorithms of this module.

3 Appointment of the Doctor

Another feature of this module is to add the feedback of the doctor he has consulted with and book the appointment of a doctor according to his time and date. The appointment of the doctor is followed by the generation of the P.O Number which is the unique ID assigned to him for future purposes so as to store the reports of his diagnostic tests and to use this P.O Number as the mobile report documents or a medical drive. By using this P.O Number, the patient can also search his medical reports that are entered by the doctor from the other side on the same P.O number. The P.O number is generated by using a Randomize() function.

A Randomize() function is a function that generates random unique values according to the lengths of the required string desired. In this module, the length of the values of the randomize function varies from 4 digits to 7 digits. Now finally, after this the user logs out of his account [5]. The rest of the modules are Doctor module and the Admin module. These modules are also discussed in the paper further. Now let us look at the Doctor module followed by the User module.

4 Admin Module

An Admin or the Administrator would be responsible for four important things. Firstly, the administrator would be responsible for registering a doctor who has not been registered or the admin would be responsible for registering a new doctor. Secondly, as the admin would be having the authority of registering a new doctor, he/she definitely would be having the authority of removing a doctor, the admin would be easily able to remove any specific doctor that he/she wants to, it would be solely

the admin's decision that which doctor he/she wants to remove, which the admin can decide by considering several aspects of measuring the doctor's performance, one of the most important aspects from these is the feedback which the doctor receives from various of his or her patients. Thirdly, an administrator would easily be able to view the details of all doctors that are registered by him or her, this will help the administrator in viewing the number of doctors that are present and this detail would allow the administrator to realize how many more doctors are there that he should register to increase the doctor's numbers, by adding more doctors or by registering new doctors he would be giving a lot of options to the patients so that they can choose the doctor of their choice. Lastly, the most important part on which the administrator would be having his or her control is the feedback details part or the feedback details section [6].

The feedback details section would be having very crucial and confidential information which only the administrator can view and this information is very important for the administrator, by looking and understanding this information he or she would easily be able to figure out that which doctor should be removed and which doctor should not be removed and this is one of the core responsibilities of an administrator, if he or she fails to follow this responsibility, he or she is not eligible for having the responsible post of the administrator. This was all a brief overview of each section on which the administrator would be having his or control, now each of these sections will be discussed in detail. Starting with the first section which is nothing but the section of registering a new doctor, in this section the administrator would be able to register new doctors which the administrator thinks are eligible and has a good experience in his or her field of expertise. This means that all eligible and experienced doctors would be easily added by the administrator in the register doctor section.

An Administrator can look for good and experienced doctors in his or her locality, after locating and finding such doctors, the administrator can ask the doctor if he or she wants to get registered, if the doctor agrees for getting registered then he or she would be successfully registered by the administrator, if in any case the doctor denies getting registered, then the administrator would for other doctors that are ready for getting registered and hence would be successfully registered by the administrator. In the register new doctor section the administrator has to give a username to the doctor, this username should not be existing previously, if it is a name and that name exists previously then in such case username can be created by adding special characters such as underscore or numbers at the beginning of the username or at the end of the username. After successfully creating a username the administrator would have to create a password for the user, this password can contain alphabets, numbers or special characters. After giving a new username to the doctor and creating a reliable password for the respective doctor, the administrator would be entering complete address of the doctor which will act as a collateral security for the administrator.

Another detail of the doctor which will be kept by the administrator is the email ID of the doctor which can be used by the administrator or the patients to contact the doctor or the patients can also send some of the x-ray or other types of reports to the doctor for his or her review. In addition to the email ID, for contacting the doctor the phone number of the doctor will also be taken by the administrator, this phone

number like the email ID can be used by the administrator and the patients to contact the doctor in any emergency circumstances. Apart from this, the administrator would be saving the gender of the doctor and his or her qualification. This qualification that the doctor gives would be made available or visible to all the patients so that they can choose their own doctors. In addition to the qualification, another aspect that will help the patients in choosing their doctors is the experience aspect, by looking at the experience of the doctor also, the patients can easily decide if he or she is the right doctor for them. Lastly, the age of the doctor would also be saved by the administrator along with the field in which the doctor specializes. This field of specialization of the doctor would also be made visible to all the patients so that they can choose the doctor according to their requirements. Now, as the administrator has to make a decision as to which should be registered and which doctor should not be registered, admin should also have the capability to remove any specific doctor that he wants to remove. For removing any doctor, the administrator just has to select the doctor whom he or she wants to remove, after selecting the doctor, the administrator just has to click on the button remove and that doctor will be removed successfully and instantly, this is the section where feedback of the patients about the doctor is used, after understanding the feedback, the administrator makes the decision whether any specific doctor has a good feedback or bad feedback and based on this, the administrator will choose the doctor that has to be removed. Now, understanding the next section in detail or the next responsibility of the administrator in detail, this section allows the administrator to keep an eye on the number of doctors, keep an eye on their contact details and other basic details like name, gender, age, etc. In any case, any patient or the administrator itself would want to contact the doctor, the doctor details section would be of great use, if any patient wants to contact the doctor, he or she can just contact the administrator and ask for the required details which the patients are asking. The email ID can be used in a great deal by the patients, it can be used to send specific reports of the patients to the specific doctor, by doing so, the patients are allowing the doctor to examine the reports before any formal meeting is scheduled between the doctor and the patient [7]. Thus, this email allows the doctor to work from home and allows the patient to know about his or her reports by just sitting at their homes and thus can also realize that is there a need for any formal meeting between him or her and the doctor. Now, as the email ID would be used to send various documents or reports, the phone number of the doctor would be used to contact him or her whenever the patient wants to, there are many cases when the patient wants to call the doctor at home, in such cases the patient would have the ability to contact the specific doctors by understanding their specialization, looking at their experience and qualification, and thus deciding as to which doctor has to be chosen by them and contacted by them in any type of circumstances, whether the circumstances are normal or whether the circumstances are of emergency [8].

After understanding the importance of doctor details section, another most important section would be looked into detail, this section is most important because it allows the administrator to understand as to which doctor has to be removed and which doctor should not be removed, this section is known as the feedback details section and it would allow the administrator to look at the feedback which the patients

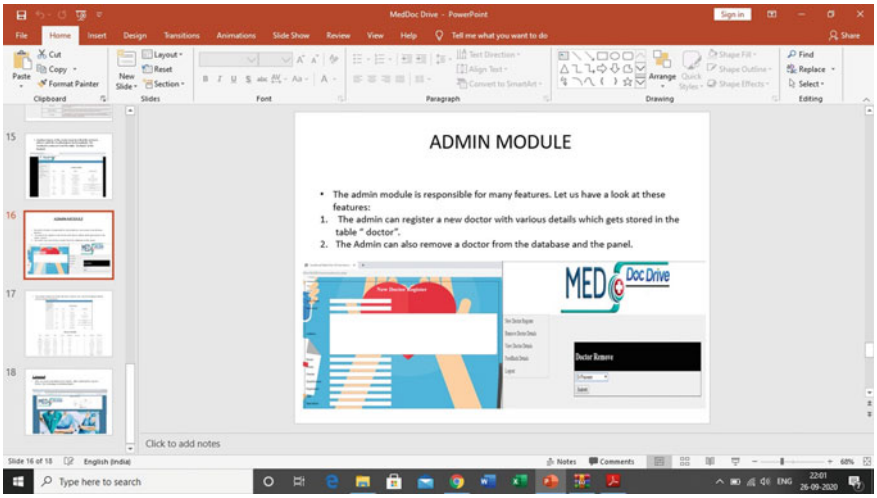


Fig. 1 Admin module

have entered, and after understanding the feedback it would become easy for the administrator to make a decision as to which doctor is not able to perform his or her duties effectively and thus would be immediately removed by the administrator, the feedback details section also allows the patients to give a feedback to the doctors and thus help the administrator in serving the patients in a better way possible. After the administrator finishes his or her work, whether it was about adding any new doctor details or registering a new doctor, or whether it was about removing a specific doctor based on the feedback of the patients, or whether it was looking at the doctor details section for getting the contact information which is demanded by the patients to contact their doctors, or whether it was the task of analyzing the feedback given by the patients to the doctor and understanding which doctor is not performing up to the mark or not performing effectively and thus removing the doctor who is not able to follow his or her duties in a good way. After performing all his or her responsibilities and doing any work that he or she wants to do by visiting any of the four sections discussed above in detail, the administrator can successfully log out and can visit again anytime any of the four sections that the administrator wants to visit and perform his or her responsibilities and duties effectively (Fig. 1).

5 Doctor Module

In the doctor module, the doctor will be able to view his schedule or previous appointments by retrieving it from the appointment table. In the next feature, the doctor will be able to add the reports of the patients who have consulted him before. This report will be available to the patients as well. The patients will have a unique P.O number

and once they give that to the doctor the doctor will be able to access any medical report of the patient. The medical report will have information about all the tests conducted in the past. Another feature of this module would be that the patients would have the facility to give feedback to the doctor about his or her experience with him. The doctor will be able to view the feedback as well. The feedback can be retrieved from the table "feedback" at the backend.

6 Conclusion

The implementation of this idea will be easy but the security and authenticity of the information should be dealt with caution. Our current goal is to minimize the cost of the application and also reduce the work done by the administrator. Future scope of this project could be to add machine learning algorithms in this and predict the disease also. Security in this idea will be of utmost importance and will play a huge role in its performance. This idea will save a lot of hassle for the users and also help the user to choose from many doctors according to their disease.

References

1. The Open Web Application Security Project (OW ASP), Available: hnps:l/www.owasp.orgiindex.php/Top 10 2013-Top 10
2. Health Information Technology (Health IT), Availabe:hnp:llwww.healthit.gov/providersprofessionals/fags/what-patient-portal
3. Healthcare IT News, Available: hnp:llwww.healthcareitnews.com/directory/patient-portals
4. The Electronic Health Reporter, Available: hnp:lleiectronichealthreporter.comipatient-portalssecurityconcem-oreffective-tool!
5. Goel AK, Gaur J, Rose A (2019) Security issues, vulnerabilities and countermeasures in information communication technology and signal processing. In: 2019 2nd international conference on signal processing and communication (ICSPC), Coimbatore, India, pp 300–304, https://doi.org/10.1109/ICSPC46172.2019.8976734
6. Goel AK, Rose A, Gaur J, Bhushan B (2019) Attacks, countermeasures and security paradigms in IoT. In: 2019 2nd international conference on intelligent computing, instrumentation and control technologies (ICICICT), Kannur, Kerala, India, pp 875–880, https://doi.org/10.1109/ICICICT46008.2019.8993338
7. Gaur J, Goel AK, Rose A (2019) Applications and design principles of information communication technology. In: 2019 2nd international conference on signal processing and communication (ICSPC), Coimbatore, India, pp 305–309, https://doi.org/10.1109/ICSPC46172.2019.8976640
8. Gaur J, Goel AK, Rose A, Bhushan B (2019) Emerging trends in machine learning. In 2019 2nd international conference on intelligent computing, instrumentation and control technologies (ICICICT), Kannur, Kerala, India, pp 881–885, https://doi.org/10.1109/ICICICT46008.2019.8993192

Weather Monitoring System Using IoT

Tanya Singh and Mohammed Asim

Abstract A weather monitoring system can be described as system which is capable of providing weather parameters and weather predictions in real-time. It can provide information about temperature, humidity, pressure, and rainfall, etc. This device senses temperature and relative humidity of a small area using sensor DHT11. By using relation between relative humidity and rainfall prediction, device will also predict the probability of rain in that area. Wi-Fi model used is NodeMCU ESP8266. Further, this data is sent in real time to a mobile application Blynk app using the cloud server. If temperature and relative humidity ranges cross a certain point a notification is sent to the mobile app and using actuators relay any device is automatically turned off/on.

Keywords Internet-of-Things (IoT) · Weather monitoring · Temperature · Humidity

1 Introduction

Internet-of-Things (IoT) is one of the most rapidly growing technologies. It's basically communication between devices and machine which are connected to Internet through which data can be accumulated and shared over Internet using various network protocols [1, 2]. The main objective of IoT is to introduce smart objects, make lifestyle easier, removing heterogeneity and develop interoperability of devices all over. IoT system broadly consists of four main components 1. Sensors 2. Connectivity 3. Data transferring 4. User Interface. Main applications of IoT are smart monitoring of devices in agricultural, industries such as the industrial internet of things (IIoT). The objective of this paper is to visualize one such application of IoT in monitoring weather parameters, temperature and humidity of a small area using IoT devices and sending this information real time over the Internet to cloud server using NodeMCU. Cloud server will send data to a mobile app named BYLNK which will display data

T. Singh (✉) · M. Asim
Department of Electrical Engineering, Integral University, Lucknow, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_21

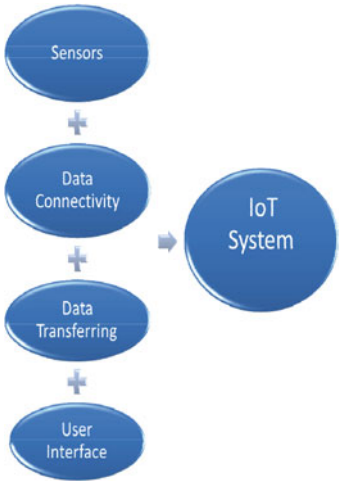


Fig. 1 Diagram of the basic IoT system

real-time on user's mobile device and will perform interaction between user and weather monitoring system [3, 4].

1.1 Weather Monitoring System

Weather monitoring system is a system that provides various parameters such as temperature, relative humidity, rain and pressure, etc. [5, 6]. Weather monitoring plays important role in graphing down weather patterns of an area. Traditional weather stations are placed at a large distance and their real-time forecast can often mismatched the actual status. To have a more accurate analysis of weather conditions now IoT devices are being fabricated for this purpose. The diagram of the basic IoT system is shown in Fig. 1 [7, 8].

2 Proposed Design

This IoT system consists of very minimal devices [9, 10]. NodeMCU is the main board, DHT11 sensor is the sensing device. Two Relays are also used in this setup along with 7805 voltage regulator which is used to give 3.3 V to NodeMCU. All these components are cost-efficient and the software component used is also freeware, hence further making it more economical.

2.1 Devices Used

1. NodeMCU ESP8266 is the main development board used in the design. It comes with an ESP8266 chip having RISC microprocessor. NodeMCU has 128 Kb RAM and 4 MB of flash memory.ESP8266 is low cost Wi-Fi chip developed by Espress if System. It is powered by 3.3 V and consists of 16 general purpose input /output pins.
2. DHT11 is a temperature and humidity sensor. Power supply ranges from 3.5 V to 5.5 V. The range of temperature being from 0 ℃ to 50 ℃. Humidity measured by this device is in between range of 20–90%. It consists of 4 pins. Pin 1 is Ground, Pin 2 is Data pin, Pin 3 is 3.3 V power and pin 4 is null.

2.2 Software Components

Blynk App is a freeware mobile application used as a software tool for IoT projects. It controls hardware remotely, capable of displaying sensor data on a screen in a very interactive manner by the usage of widgets. This application also allows user to have virtual on/off switches which can actuate the same to IoT system. Blynk app consists of three major components. Blynk App, Blynk server which is responsible for data transfer from cloud to device then IoT system and vice versa. Lastly, Blynk libraries are predefined libraries to support for various IoT devices like Arduino, Raspberry Pi, and NodeMCU.

3 Setup

DHT11 sensor data pin (pin 2) is connected to D1. GPIO pin of microprocessor NodeMCU senses the temperature and humidity. A 5 V relay is connected to this IoT system which will turn on and off any external device connected to it when data will be sent from Blynk app. Further, NodeMCU will communicate with Blynk app via Blynk app cloud server. Figure 2 shows the block diagram of a weather monitoring system.

4 Working and Implementation

Figures 3 and 4 show the flowchart for humidity and temperature conditions, respectively. Firstly, collection of weather parameter namely temperature and relative humidity at interval of 10 seconds is done and then this data is sent in real-time to cloud server. Interface of sensor module takes place with NodeMCU device, data parameters are then sent to Blynk app using Blynk app server. First, condition for

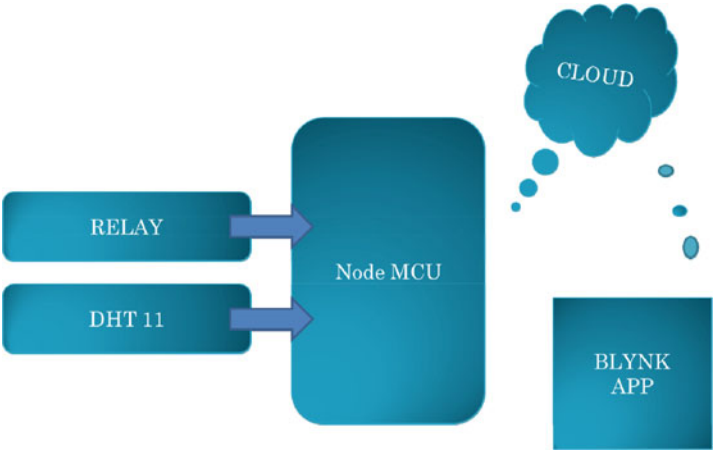
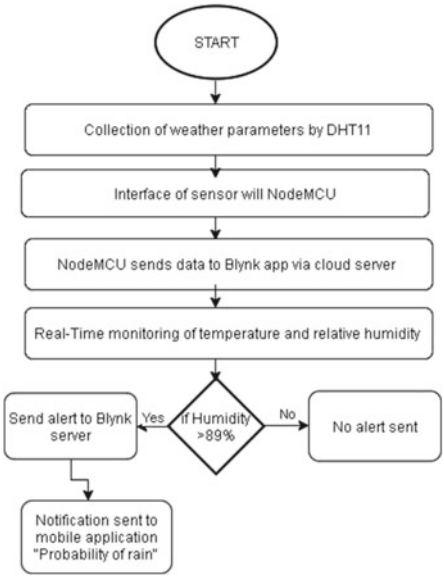


Fig. 2 Block diagram of weather monitoring system

Fig. 3 Flowchart for humidity conditions



relative humidity is checked, if humidity if greater than 90% then there is a probability of rain. If humidity sensed by the sensor is more than 89% considering limit of DHT11 is 90% so 89% is taken. If humidity greater than 89% Blynk app sends notification to its mobile app "Probability of rain" so that user will be alerted about this and be prepared accordingly.

Latter part of Flowchart shows the conditions for temperature sensed by the sensor. If the temperature exceeds 49 ℃, the external device which is controlled by this

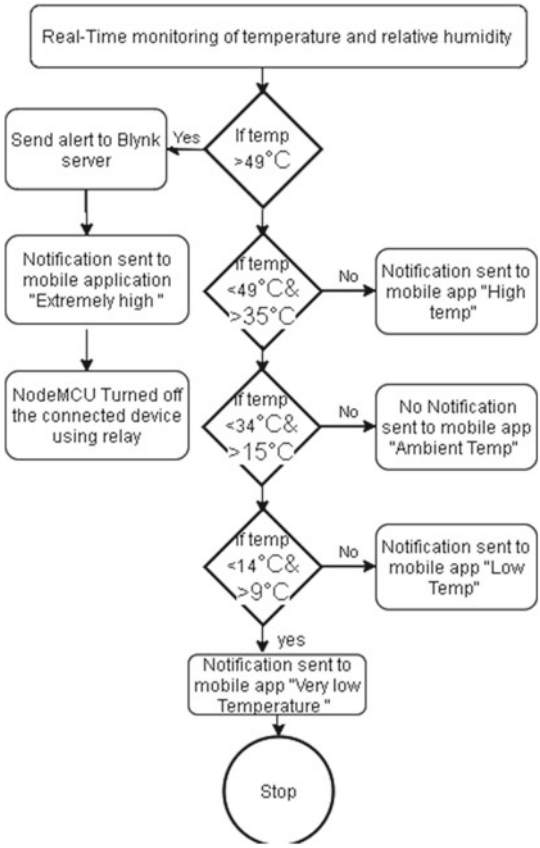


Fig. 4 Flowchart for temperature conditions

setup will be turned off using actuator relay. In actual scenarios, this will help in overheating of the functioning device which might lead to burning of insulations and damage to device. A notification will also be sent to mobile app "Extremely high Temperature". Temperature ranging between 49 ºC and 35 ºC will send a notification to the app "High Temperature", range 34–15 ºC sends no notification as its ambient temperature. Range of temperature from 14 ºC to 9 ºC sends notification as "Low Temperature ". Lastly, if the temperature is below that range notification will be sent "Very low temperature".

This IoT system can be helpful in areas that are not fit for humans, such as nuclear sector, wildlife sanctuaries, etc. This system allows accurate collection of data and also cumulative records can be stored. The widgets of Blynk app allow user-friendly interface to this system and hence can be learnt to use with minimum efforts.

5 Result

After sensing data from DHT11 sensor the values of temperature and relative humidity have been calculated which are refreshed at an interval of 10 sec. Figure 5 shows the hardware setup of IoT-based weather monitoring condition. The sensed data was sent to Blynk app server via cloud, when proper connection is established IoT System and user will contact remotely. This IoT system is Low-Power and efficient system. This device is also very economical as only one sensor is used for predicting three weather parameters.

A Blynk app interface on the mobile end is used to send temperature and relative humidity in real time. The Blynk app widgets consisted of two gauges namely temperature where maximum limit is set to 50 °C, IoT device recorded temperature as 27.8 °C and relative humidity is measured in percentage as 43.1%, so no probability of rain was observed in reading which was observed for a span of 15 minutes, weather parameters value changed after 2–3 min. Recorded temperature of the device was more accurate than the weather forecast of the main city making device more reliable. A virtual on/off switch is also added for manual relay operation, which can also be turned off automatically if the temperature limit exceeds the threshold value. Further, if required labeled displays can also be used. One notification widget is also added which will be sending a pop-up notification if weather parameters come under threshold values, which have already been set up before.

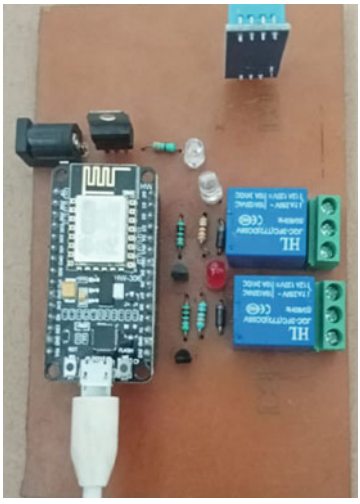


Fig. 5 Hardware setup of IoT-based weather monitoring condition

6 Conclusion and Future Scope

The monitoring of weather parameters by IoT smart devices for a small location can be measured more accurately and hence, enabling reliability of this system. The system proposed can be used to measure temperature and relative humidity in different applications where human intervention is not easy. It can also be used in agriculture and other areas where constant monitoring can be done, more number of sensors can be added for larger area coverage so as to create a system of wireless sensor network which can be of use in agricultural areas for better crop yield. Also, proximity sensors can be added to detect intervention of wild animals in the farm area, this can prevent crops from damage and enhance security.

In areas where is it hazardous for humans to go like radioactive sites such IoT smart devices can come in use. Being low-power and cost-efficient, it is also economical.

These IoT weather monitoring system can be coupled along with other IoT smart systems such as smart energy meters, smart car controls, smart homes to all together create smart cities.

References

1. Ahmad R, Asim MA, Khan SZ, Singh B (2019) Green IoT—issues and challenges. In: Proceedings of international conference on advanced computer and software engineering (ICACSE), pp 1–5
2. Kodali RK, Mandal S (2016) IoT based weather station. In: 2016 international conference on control, instrumentation, communication and computational technologies (ICCICCT)
3. El-Bendary N, Fouad MMM, Ramadan RA, Banerjee S, Hassanien AE (2013) Smart environmental monitoring using wireless sensor networks. K15146_C025.indd, 2013
4. Gopavanitha K, Nagaraju S (2017) A low cost system for real time water quality monitoring and controlling using IoT. In: 2017 International conference on energy, communication, data analytics and soft computing (ICECDS). IEEE
5. Betts A, Desjardins R, Worth D, Beckage B (2014) Climate coupling between temperature, humidity, precipitation and cloud cover over the Canadian Prairies. J Geophys Res 119. https://doi.org/10.1002/2014JD022511.
6. Tariq M, Bakhsh FI, Khan SA, Asim M (2010) Solar based space cooling system in building-a case study of Maulana Azad library, AMU. Int J Recent Trends Eng Technol 4(4)
7. Rao BS, Rao KS, Ome N (2016). Internet of things (IoT) based weather monitoring system. Int J Adv Res Comput Commun Eng 5(9):312–319
8. Shinde SR, Karode AH, Suralkar DS (2017) Review on IOT based environment monitoring system. Int J Electron Commun Eng Technol 8(2)
9. Rahut Y, Afreen R, Kamini D, Gnanamalar SS (2018) Smart weather monitoring and real time alert system using IoT. Int Res J Eng Technol
10. Kamble SB, Rao PRP, Pingalkar AS, Chayal GS (2017) IoT based weather monitoring system. Int J Adv Res Innov Ideas Educ 3(2):2886–2991

Optimal Bayesian Estimation Framework for Reduction of Speckle Noise from Breast Ultrasound Images

Pradeep Kumar, Subodh Srivastava, Y. Padma Sai, and Santosh Choudhary

Abstract A new methodology is proposed in this paper, for the elimination of Speckle noise from the breast Ultrasound Images. Ultrasound is used in radiology to scan the anatomy and other physiological processes of the body and it is an important technique for clinical diagnosis as it gives the live view of internal structures of tissues in the targeted regions and safe also. The existence of Speckle noise in the breast Ultrasound images creates hindrance in getting a clear breast image and making the diagnosis strenuous. In this paper, Optimal Bayesian Estimation Framework intended for the Reduction of Speckle Noise from Breast Ultrasound Images is proposed. The suggested method is compared with that of existing a guided filter and speckle reducing anisotropic diffusion (SRAD) which is the state of art method for reducing the speckle noise with respect to four performance metrics, i.e., Root-mean-square error (RMSE), Structural similarity index metric (SSIM), Peak signal-to-noise ratio (PSNR) and Correlation parameter (CP). Simulation results depict that our present suggested technique shows better outcome with regard to performance metrics.

Keywords Speckle noise · Breast ultrasound · Root-mean-square error · Correlation parameter · Peak signal-to-noise ratio

1 Introduction

In women, breast cancer is the prevailing form of cancer and it is the second leading reason for deaths due to cancer in women. And the rate is higher for women over 40–50 years. A woman has 12% chance to develop breast cancer in her life [1], i.e., a 1 in every 8 have a chance of being affected by breast cancer. Breast cancer is essentially

P. Kumar (✉) · S. Srivastava
Department of Electronics and Communication Engineering, National Institute of Technology Patna, Patna, India
e-mail: pradeepk.phd18.ec@nitp.ac.in

Y. Padma Sai · S. Choudhary
Department of Electronics and Communication Engineering, VNR VJIET Hyderabad, Hyderabad, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_22

the rapid growth of unwanted cells that forms into a lump or tumor overtime. Catching breast cancer in its early stages is very hard. X-Ray, mammography, ultrasound, and Magnetic Resonance Imaging (MRI) are the four major techniques, the doctors used to take a closer look at the lump [1]. In imaging techniques, the scanned images also dealt with some type of noise in them. The removal of such noises is necessary to observe the cancerous region properly. For this purpose, preprocessing of image is required, i.e., the removal of noises. The most common kind of noise in any imaging technique is additive white Gaussian noise (AWGN) [2, 3]. When an X-ray image is acquired, white Gaussian noise is found as well as quantum noise which follows Poisson distribution dominates during the formation of X-rays images [4, 5], in MRI images the Rician noise is found [6]. In ultrasound images, the dominant noise is the Speckle noise [2, 7]. The characteristics of multiplicative noise is different from that of general noise and the removal of the noise becomes very tiring.

In the literature survey, most of the works tired this field were associated with the removal of speckle noise. There square measure 3 elementary strategies: nonlinear filtering, linear filtering, and hybrid filtering techniques. The Linear filtering ways, as an example, mathematician channel and Mean channel, will viably eliminate noise from the picture; but exhibit the difficulty of obscuring edges. In order to defeat this difficulty, nonlinear methods of filtering are planned. These procedures, as an example, non-local means that (NLM) filter [8], anisotropic diffusion filter (AD) [9], and two-sided filter [10] are planned, and are created to get rid of speckle noise whereas reducing degradation of the sting region. Despite limiting decay in edge degradation, these methods aren't cheap for eliminating speckle noise since they don't think about the qualities of noise of the ultrasound image. So as to beat the constraints, a brand-new filter is planned, i.e., Optimal Bayesian Estimation Framework.

Section 1 of this paper provides a concise introduction to breast cancer along with noise removal techniques in breast ultrasound images. Sections 2 and 3 explain the methodology and evaluation metric, respectively. Section 4 illustrates the result and discussion. Finally, we concluded this paper in Sect. 5.

2 Methodology

One medical imaging technique is ultrasound and is a non-invasive diagnosis that doctors utilize for the purpose of treatment. It has been observed that ultrasound follows gamma distribution. It is given in below equation [11–13].

$$N(g) = \frac{g^{\alpha-1} e^{\frac{-g}{a}}}{\alpha - 1! a^{\alpha}} \tag{1}$$

Let Z (i, j) is the degraded pixel of an observed breast ultrasound image and U (i, j) be the free from noise breast ultrasound image pixel which is to be recovered. With the multiplicative noise model,

$$Z(i,\ j)\ =\ U(i,\ j)\ *N(i,j) \tag{2}$$

Mentioned above, N(i, j) is the noise that has gamma distribution. The info dependency nature of speckle noise and conjointly the noise distribution will vary relying upon the look of the imaging system. It's terribly difficult to separate noise-free image U(i, j) from speckle noise N(i, j). We suggest to address these problems related to minimizing of speckle noise by estimating the noise-free image within the exponent area mistreatment associate best theorem statistical method estimation approach supported conditional posterior sampling. To tackle the data-dependent behavior of speckle noise, the measured data Z(i, j) is projected into the logarithm space for decoupling, the free from noise breast ultrasound image U(i, j), and the speckle noise N(i, j),

$$\log(Z(i,\ j))\ =\ \log(U(i,\ j))\ +\log(N(i,j)) \tag{3}$$

$$Zl(i,\ j)\ =\ Ul(i,\ j)\ +Nl(i,j) \tag{4}$$

Zl(i, j), Ul(i, j) and Nl(i, j) represent the logarithmic of Z(i, j), U(i, j) and N(i, j), respectively. In the logarithmic space, optimal Bayesian least squares prediction of Ul(i, j) can be expressed as follows,

$$\widehat{U_l}(i,\ j) = \min_{\widehat{U}_{l(i,\ j)}} \left\{ E\left[\left(\widehat{U}_l(i,\ j) - \widehat{U_l}(i,\ j) \right)^2 \mid Z_l(i,\ j) \right] \right\} \tag{5}$$

Above equation reduces the average squared error of the free from noise data estimated on the basis of the measured data. The simplified version of the above equation is given below,

$$\widehat{U}_l(i,\ j) = \int p(U_l(i,\ j) \mid Z_l(i,\ j))U_l(i,\ j)dU_l(i,\ j) \tag{6}$$

The Eq. (6) shows the finest Bayesian estimation of the noise-free data built on the basis of the measured data. The posterior distribution $p(Ul(i, j)|\ Zl(i, j))$, is highly complex and nonlinear in behavior and making it hard to solve for noise-free data. Mostly, general Bayesian least square estimators [14] and specific parametric posterior distribution methods [6] are used to reduce speckle noise. But these methods are poor to remove noise and it results in significant amount of loss of structure detail. To overcome the mentioned drawbacks and for estimating the posterior distribution, we used optimal conditional posterior sampling. To measure the data in order to estimate the posterior distribution, usage of optimal conditional posterior sampling serves as an added advantage. This method allows preservation of the boundary characteristics while balancing out the speckle noise.

In Markov-Chain Monte Carlo density estimation [15, 16], by initially sampling from a known earliest probability distribution, an unknown target distribution is estimated in an indirect manner. Similarly, the posterior distribution is estimated based on Gaussian distribution which is given by

$$f_g\Big((i,\ j)^{'}/(i,\ j)\Big) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{(i,j)^{'}-(i,j)^2}{2\sigma^2}\right)} \tag{7}$$

σ is the variance and we used the Euclidean norm. The initial distribution generated sites that tend to be closer proximity to the site (i, j). We should calculate the local mean of the neighborhood in a 4 X 4 region centered at (i, j). It provides us with information sufficient for getting good statistics not depending on the resolution of breast ultrasound images. The inclusion of (i, j)′ as a realization of posterior distribution is built on the assumption that the local mean is a reasonable early prediction of noise-free data. This method is replicated until the greatest number of sites are used for the estimation of the original signal. The weight associated with each pixel (i, j)′ in estimating noise-free data is denoted and calculated by equation

$$W((i,\ j)^{'} \mid (i,\ j)) = e^{\left[\frac{\mu(i,j)^{'}-\mu(i,\ j)}{2\sigma^2}\right]} \tag{8}$$

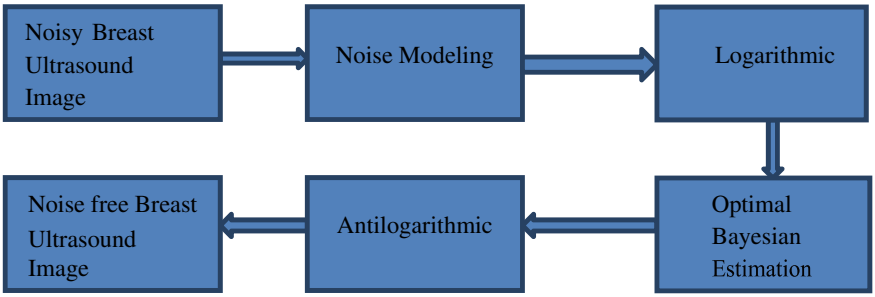
Equation (9) shows the likelihood function of weight associated with each pixel and it helps to estimate the posterior distribution. $\mu(i, j)$ is centered at (i, j) and is the local mean of the neighborhood and σ is the image noise variance as estimated. So, the estimation of posterior distribution is given by

$$\widehat{p}\,(U_l(i,\ j) \mid Z_l(i,\ j)) = \frac{\sum_{k=n} w((i,\quad j)_k^{'} \mid (i,\quad j)\delta\Big(U_l - Z_l(i,\quad j)_k^{'}\Big))}{M} \tag{9}$$

where M is the normalization term so that the summation of posterior distribution becomes one. Finally, we can find the estimation of noise-free data by back-projecting the Bayesian estimate of $Ul(i, j)$determined using the estimated $p(Ul(i, j)|\ Zl(i, j))$. To find the actual value of the image data we have to calculate the antilogarithmic of the estimated value and it's given by equation

$$\widehat{U}(i,\ j) = e^{\widehat{U}_l(i,j)}. \tag{10}$$

The flow chart of despeckling of breast ultrasound image is shown in below diagram.



3 Evaluation Metric

The comparison of the performance of suggested method with earlier method is given by PSNR, MSE, SSIM, and CP. They are defined as below for breast Ultrasound.

Root-mean-square error:

$$\text{MSE} = \frac{1}{mxn}\sum_{i=1}^{m}\sum_{j=1}^{n}[I^{'}(i,j) - I(i,j)]^2 \tag{11}$$

$$\text{RMSE} = \sqrt{MSE} \tag{12}$$

where I and I' denote original and filtered image, respectively.

Peak signal-to-noise ratio:

$$\text{PSNR} = 20\log\left[\frac{255}{\sqrt{MSE}}\right] \tag{13}$$

Structure similarity index map and Correlation parameter:

For the comparison of the luminance, contrast, and structure of two separate images SSIM is used. The SSIM of two different images X and Y is given by:

$$\text{SSIM (X, Y)} = \frac{(2\mu_x\mu_y + C_1) \times (\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1) \times (\sigma_x^2 + \sigma_y^2 + C_2)} \tag{14}$$

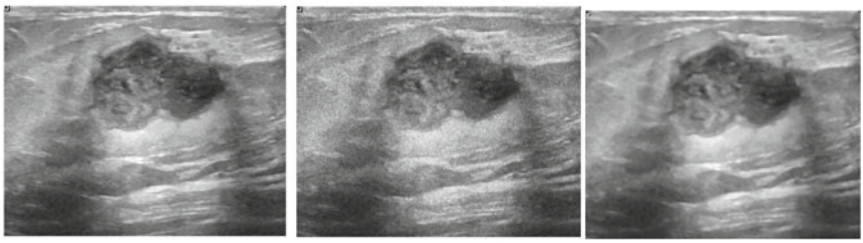
where μ and σ is the mean and standard deviation (i = X or Y), respectively. C (i = 1 and 2) is the constant to avoid instability.

4 Results and Discussion

In the current section, the method that has been suggested is compared with a guided filter and a speckle reducing anisotropic diffusion (SRAD), which is the state of art method for reducing the speckle [17] noise with respect to RMSE, PSNR, CP, and SSIM. Breast Ultrasound Web dataset is used for simulation for comparison of the effectiveness of the proposed method. For the denoising of breast ultrasound. The ground truth is that breast ultrasound data are corrupted with a noise variance having a range from 5% to 30% artificially for the evaluation of the quantitative metrics (Fig. 1). The performance analysis of proposed and other method is shown in Table 1.

Basing on the idea of quantitative and circumpolar results, it's clear that the advised procedure has created correct results. Such as preservation of edges, and better noise removing ability and structural information, at all the possible levels of Speckle noise.

Figure 2 depicts the graph of RMSE vs noise ratio. The results in Fig. 2 depict that the suggested method reduces the value of RMSE at all noise ratios. Figure 3 depicts the graph of PSNR vs noise ratio. Figure 3 shows that the PSNR value is higher for proposed method for all values of noise ratio. Figure 4 depicts the graph of SSIM



(a) Original Image (b) Noisy Image (c) After removal of noise with proposed method

Fig. 1 Simulated breast ultrasound with speckle noise

Table 1 Quantitative comparison

Noise ratio	SRAD				Proposed method			
	RMSE	PSNR	SSIM	CP	RMSE	PSNR	SSIM	CP
0.05	5.3	25.98	0.80	0.96	4.6	27.40	0.82	0.97
0.10	5.4	25.80	0.79	0.96	4.9	26.88	0.81	0.97
0.15	5.5	25.40	0.78	0.96	5.3	26.40	0.80	0.97
0.20	5.6	24.91	0.75	0.96	5.4	26.00	0.79	0.96
0.25	5.7	24.50	0.74	0.95	5.5	25.60	0.78	0.96
0.30	5.8	24.30	0.72	0.95	5.7	25.50	0.76	0.95

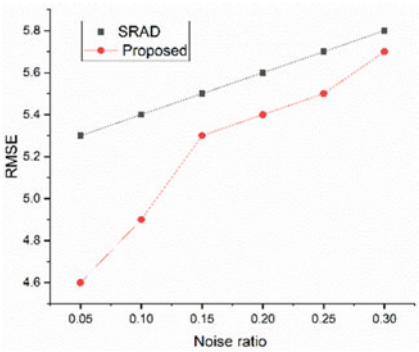


Fig. 2 RMSE w.r.t. noise

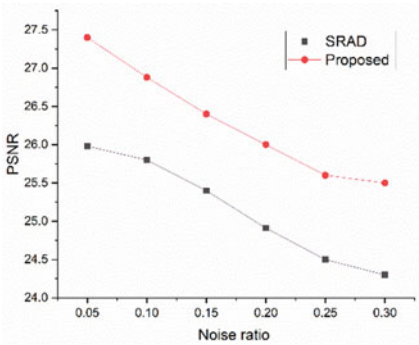


Fig. 3 PSNR w.r.t. noise

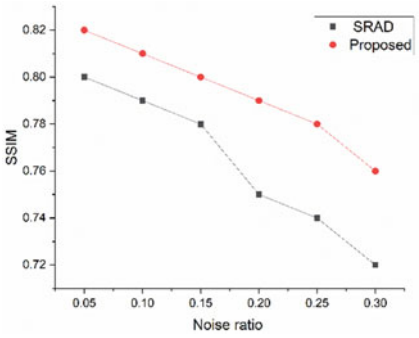


Fig. 4 SSIM w.r.t. noise ratio

versus noise ratio. Figure 4 shows that the SSIM value is greater when compared to a guided filter and the speckle reducing anisotropic diffusion (SRAD). Figure 5 depicts the graph of CP vs noise ratio. Figure 5 shows that the CP value is better than a guided filter and the speckle reducing anisotropic diffusion (SRAD).

From experimental results, we observed that proposed method is better than the existing one. From the obtained result Optimal Bayesian Estimation Framework is a better result than the existing one because all parameters values are optimized.

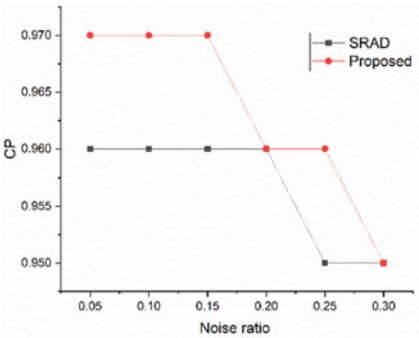


Fig. 5 CP w.r.t. noise ratio

5 Conclusion

Breast cancer is a malignant tumor that forms in the cells of the breast. It contributes to 25–32% of cancers in women in all cities across India. In India, one woman is diagnosed with breast cancer every 4 minutes. The ultrasound is used in radiology to scan the anatomy and other physiological processes of the body and it has become important for clinical diagnosis. The existence of Speckle noise in the ultrasound image of breast causes hindrance for getting a clear breast image and making the diagnosis strenuous. So, in order to combat the effects of speckle noise and for ensuring high performance, an Optimal Bayesian Estimation Framework for removal of speckle in breast ultrasound was proposed which performs superior than the existing methods in terms of RMSE, SSIM, PSNR, and CP.

References

1. Gupta A, Shridhar K, Dhillon PK (2015) A review of breast cancer awareness among women in India: Cancer literate or awareness deficit? Eur J Cancer 51:2058–2066
2. Yadav RB, Srivastava S, Srivastava R (2017) Identification and removal of different categories of noises from magnetic resonance image using hybrid partial differential equation-based filter. Int J Digit Signals Smart Syst 1(2):87–98
3. Srivastav R, Srivastava S (2013) Restoration of Poission noise corrupted digital images with nonlinear PRE vased filters along with the choice of regularization parameter estimation. Pattern Recogn Lett 34:1175–1185
4. Yadav RB, Srivastava S, Srivastava R (2016) Identification and removal of different noise patterns by measuring SNR value in magnetic resonance images. In: 2016 ninth international conference on contemporary computing (IC3). IEEE
5. Pal C, Das P, Chakrabarti A, Ghosh R (2017) Rician noise removal in magnitude MRI images using efficient anisotropic diffusion filtering. Int J Imaging Syst Technol 27:248–264. https://doi.org/10.1002/ima.22230
6. Yadav RB, Srivastava S, Srivastava R (2017) Modified complex diffusion based nonlinear filter for restoration and enhancement of magnetic resonance. Int J Biomed Eng Technol 23(1)
7. Abbott JG, Thurstone FL (1979) Acoustic speckle: theory and experimental analysis. Ultrason Imaging 1(4):303–324

8. Buades A, Coill B, Morel J-M (2005) A non-local algorithm for image denoising. In: Proceedings of computer society conference on computer vision and pattern recognition, vol 2, pp 60–65
9. Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell 12(7):629–639
10. Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images. In: Proceedings of the IEEE international conference on computer vision, pp 839–846
11. Benzarti F, Amiri H (2013) Speckle noise reduction in medical ultrasound images. In: Signal, image and pattern recognition laboratory, engineering school of Tunis (ENIT)
12. Kaur T, Sandhu M, Goel P (2012) Performance comparison of transform domain for speckle reduction in ultrasound image. Int J Eng Res Appl 2(1):184–188
13. Salivahanan S, Vallavaraj A, Gnanapriya C (2008) Digital signal processing. Tata Mcgraw Hill, vol 23. New Delhi
14. Wong A, Mishra A, Bizheva K, Clausi D (2010) General Bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery. Opt Express 18:8338–8352
15. Salivahanan S, Vallavaraj A, Gnanapriya C (2008) Digital signal processing. Tata McgrawHill, vol 23. New Delhi
16. Hastings W (1970) Monte carlo sampling methods using Markov chains and their applications. Biometrika 57(1):97–109
17. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A (2020) Dataset of breast ultrasound images. Data Brief 28:104863. https://doi.org/10.1016/j.dib.2019.104863

Morphological Transformation in Color Space-Based Edge Detection of Skin Lesion Images

Sudhriti Sengupta, Neetu Mittal, and Megha Modi

Abstract Detection of lesion border from skin image is very essential for developing efficient non-invasive computer-aided diagnosis. For creation of an automatized system, edge detection is an essential step. An improved color space-based edge detection method with linear structure element is introduced in this paper. At first, the input image is converted to red color space and the Sobel operator is applied to detect the border. After this, linear structuring element is introduced, followed by filling up of holes to get appropriate segmented image. The performance of this method is measured by comparing its result with the results obtained by using a standard Sobel edge detection method.

Keywords Skin lesion · Color space · Dilation · Erosion · Edge detection

1 Introduction

Analysis and proper detection of lesions in skin images is very important for efficient and timely diagnosis of different types of diseases pertaining to skin [1]. This helps the doctors and other medical practitioners to provide better treatment to the patients. A well-developed CAD for skin lesion analysis will be beneficial to patients in rural and far-flung areas where specialists are not available. The patients in such areas can provide the images to the specialists for their timely intervention in diagnosis and treatment [2]. Edge detection is an important step for developing CAD for detection and identification of skin lesion images. It finds the boundaries of areas of interest in the images. Accurate skin lesion edge detection from background skin is essential for diagnosis which deals with the shape and size of the lesions. This is very essential

S. Sengupta (✉) · N. Mittal
Amity Institute of Information Technology, Amity University, Noida, Uttar Pradesh, India
e-mail: ssgupta@amity.edu

N. Mittal
e-mail: nmittal1@amity.edu

M. Modi
Yashoda Specialty Hospital, Ghaziabad, Uttar Pradesh, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_23

as its accuracy effects the succeeding steps of classification of skin lesion images. However, skin lesion edge detection is difficult due to variety of shape, texture, size, color, etc., along with different pathology of human. There are many techniques developed for determining the border or edge in skin lesion images. Some of the most popular techniques in this category are threshold [3], edge detection operator [4], and region-based method [5]. A combination of median filtering, bi-level threshold, and Laplacian of Gaussian was introduced by Denton et al. [6] for determining the skin lesion edges [7]. For automatized detection of melanoma, Support Vector Machine was introduced in the year 2014 [8]. To analyze, Contour analysis of melanoma and nevus tissues was done by applying various image processing techniques [9].

Classification of benign and malignant melanoma lesion was done by Artificial Neural Network and Support Vector Machine [10]. A novel hierarchical KNN classifier was introduced for classification of color skin lesion images. Feature extraction from skin lesion images for detection of Basal cell carcinoma was done by Random Forest classifier [11]. A method of detection of skin lesion border by using color space-based threshold technique was proposed by Sengupta et al. [12]. ACO-based optimization of edge detection technique was done in the year 2019 [13]. Quality of skin lesion image was improved by using median filtering and further Sobel edge detection operator was used for segmentation [14]. Enhancement of skin lesion image by applying linear structure of specific dimension was done in 2018 [15]. In 2019, fudge factor was introduced to tune the threshold value for improving the result of segmentation [16]. It is found that image processing technique is very popular in medical image analysis like Breast cancer image, MRI images, etc. [17–19].

In this paper, the test skin lesion image is first spited in the red color space, then Sobel edge detector is applied. In the image map obtained, line of dimension 3 at 90 degrees is introduced by process of dilation. After this, fill up of the edge gaps is done by holes using the process of erosion. The proposed method is described in Sect. 2. In Sect. 3, implementation details are discussed. Section 4 deals with result analysis and discussion.

2 Proposed Methodology

The proposed system has four principal steps apart from image acquisition. A test image is obtained for experiment and result analysis. The input image is converted into gray-scale image for further processing. The image is separated into an image in red color space. The Sobel edge detection method is applied in the image in red color space. Linear structuring element having dimension 3 at 90 degrees is introduced by the process of dilation. Then, the edge map is refined by filling up with holes. The flow chart of proposed methodology is given as in Fig. 1.

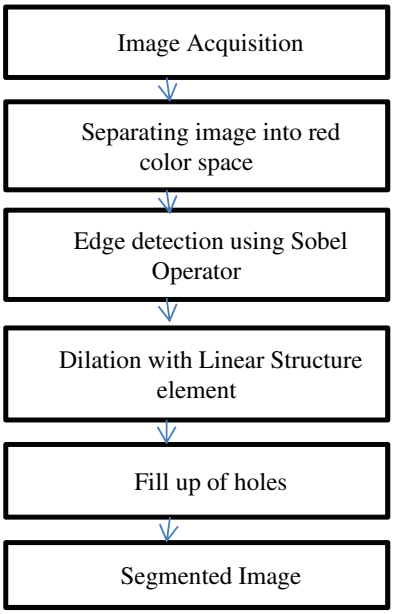


Fig. 1 Flow chart of the proposed system

2.1 Image Input

An image of skin containing lesions occurring due to neurofibromatosis is taken from https://www.dermnetnz.org/image-library. Neurofibromatosis causes brown tumor-like lesions on skin surface caused by genetic disorder of nervous system. Occurrence of brown dome-shaped lesions larger than 15 mm in diameter is typically symptoms for neurofibromatosis diagnosis in adult human (Fig. 2).

Fig. 2 Input image

Fig. 3 Image in red color space

2.2 Red Color Space

A true color (RGB) image is stored as an m * n * 3 array that specifies the red, green, and blue color part of each pixel. Each pixel color is calculated from a combination of red, green, and blue intensity stored in each color plane at its location. A monochrome image in red color space is obtained and used for further processing. The average intensity of the red channel is taken. In red color space, the white represents the highest concentration of pure red color and black represents no red color. The red color space is the first image plane in the true color image. The image in the red color space is separated from the green and blue color space (Fig. 3).

2.3 Edge Detection Using Sobel Operator

An edge map of an image can be created by Sobel Edge detection operator. This operator was developed in 1968 by Sobel and Feldman [20]. Sobel operator uses two kernels to mask the original image and calculate derivatives. These kernels are of size 3*3. Each kernel is present in horizontal and vertical derivative approximately. The Sobel operator provides a smoothing effect to the natural edges of the input images and this creates lines or border in output image. This method is implemented on the test image present in red color space to acquire the edges of lesions in the skin image as depicted in Fig. 4.

2.4 Dilation and Fill Up

Dilation is the process of addition of pixels to the edges of the object in an image. The total number of pixels connected is based on the shape and dimension of the

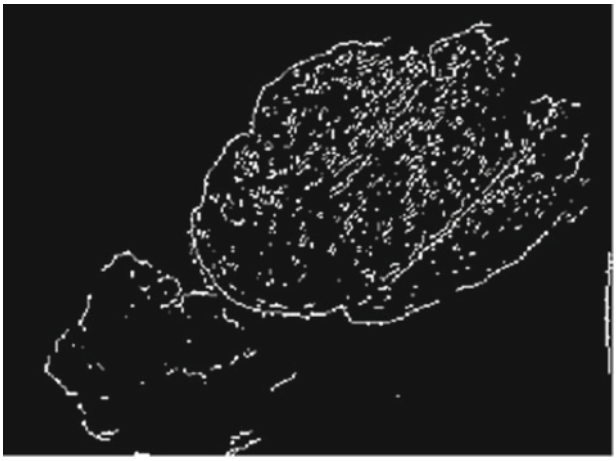


Fig. 4 Edge detection by Sobel operator

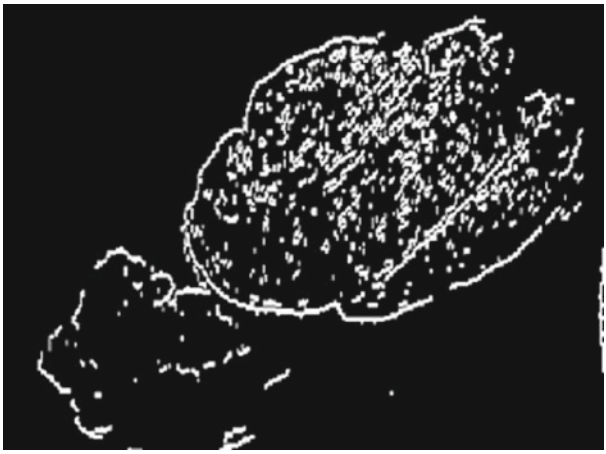
structuring elements applied in processing of images. The maximum value of all the pixel is considered as the output pixel value in this dilation. Structural element of dimension 3 at 90° is introduced in edge map of the gray-scale image of the red color space. The resultant image is shown in Fig. 5a. The final step is to fill holes in the test image depicted in Fig. 5. A hole is a group of pixels in a neighborhood that cannot be filled by surrounding pixels.

3 Implementation

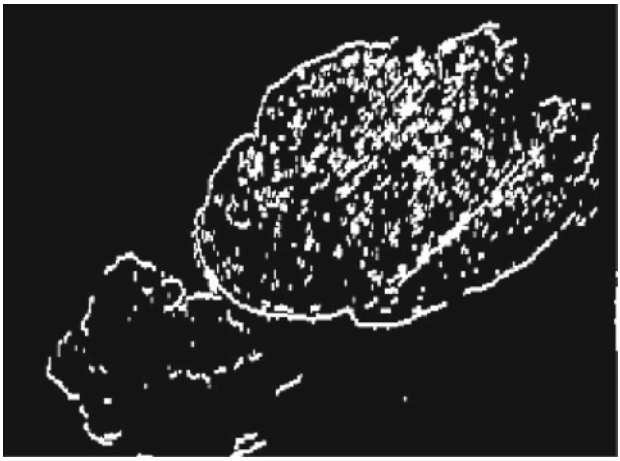
In this paper, an image of neurofibromatosis is obtained from web portal for experimenting. The methodology mentioned in the proposed system is tested by experiments conducted using Image Processing Toolbox of GNU Octave 4.2.1. The configuration of the system used is Windows 7, 64 bit OS with Intel® Core ™, 1.90 Ghz CPU, and 4.00 GB RAM.

4 Results and Discussion

To measure the efficiency of the proposed system, we have used Entropy of the image. This specifies the randomness of data in an image. If the image is of good quality, then the Entropy of the image is high [21]. Here, the comparison is done in two cases. In the first case, we have compared the entropy of the edge map obtained by the Standard Sobel edge detection method with the result obtained by the proposed method. In the second case, we have compared the entropy of the edge map acquired



(a)



(b)

Fig. 5 **a** Image with linear structural element. **b** Image filled up with holes

from the image in RGB color space with the entropy obtained from the image by the proposed system.

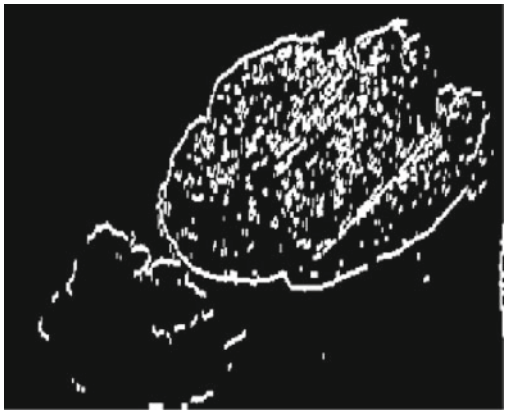
Case 1:

Figure 4 shows the image obtained by the standard Sobel operator. The entropy of this image is 1.0912. The entropy of the final resultant image as depicted in Fig. 5b is 1.9894. Thus, we can say that the proposed system shows better result than the Standard Sobel edge detection operator.

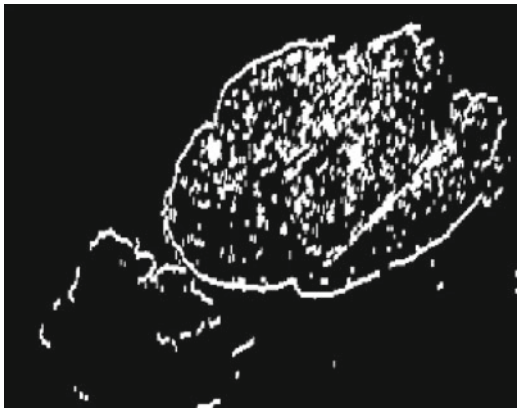
Case 2:

In this case, we discussed the resultant obtained image in RGB color space instead of the image in red color space. We have used Sobel operator in the image in RGB color space to obtain the edge map. In this edge map, we have introduced a linear structuring element of size 3 at 90 degrees by dilation. The holes in the final images are filled up to get the final output image for result analysis.

Figure 6 is used for performance comparison of the image in standard true color space with the image in red color space. This image depicts the edge obtained from the image in RGB color space. Figure 6b is the output image. We have found that the entropy of Fig. 6b is 1.7032.



(a)



(b)

Fig. 6 **a** Image with linear structural element on image in RGB color space. **b** Image filled up with holes in **a**

By comparing the entropy obtained by results of Sobel edge detection operator with the entropy of the final image obtained by proposed system, we can say that the proposed system is better. We have also comprehended that the proposed method of introduction of linear structuring element works better in images present in red color space.

5 Conclusion

By this study, it is found that the performance of the standard Sobel edge detection operator can be improved if we consider the image in red color space. The next step is to add the pixels in the edge boundaries in the form of linear structure. Finally, the holes are filled up in the edge map. We have verified the result of the proposed methodology in two ways. First, with the result obtained by the standard Sobel edge detection operator only, and secondly, by considering the image to be in RGB color space instead of red color space. In both cases, the entropy of the proposed system is significantly larger; hence, it shows an improvement in edge detection.

References

1. Pennisi A, Bloisi DD, Nardi D et al (2016) Skin lesion image segmentation using Delaunay Triangulation for melanoma detection. Comput Med Imaging Graph 52:89–103
2. Al-Masni MA, Al-Antari MA et al (2018) Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. Comput Methods Programs Biomed 162:221–231
3. Ganster H, Pinz P, Rohrer R et al (2001) Automated melanoma recognition. IEEE Trans Med Imag 20:233–239
4. Rubegni P, Ferrari A, Cevenini G et al (2001) Differentiation between pigmented Spitz naevus and melanoma by digital dermoscopy and stepwise logistic discriminant analysis. Melanoma Res 11(1):37–44
5. Celebi M, Kingravi H, Lee J (2008) Border detection in dermoscopy images using statistical region merging. Skin Res Technol 14(3):347–353
6. Denton W, Duller A, Fish P (1995) Robust boundary detection for skin lesions. In: Proceedings of IEEE annual conference engineering in medicine and biology society, vol 1, pp 407–408
7. Caslellini P, Scalise A, Scalise L (2000) A 3-D measurement system for the extraction of diagnostic parameters in suspected skin nevoid lesions. IEEE Trans Instrum Meas 49:924–928
8. Lu J, Kazmierczak E, Manton JH, Sinclair R (2013) Automatic segmentation of scaling in 2-D psoriasis skin images. IEEE Trans Med Imaging 32(4):719–730
9. Tanaka T, Yamada R, Tanaka M, Shimizu K, Oka H (2004) A study on the image diagnosis of melanoma. In: International conference of the IEEE engineering in medicine and biology society IEMBS, vol 26, pp 1597–1600
10. Lucia B, Robert B, Fisher B et al (2012) Non-melanoma skin lesion classification using colour image data in a hierarchical K-NN classifier. IEEE Int Symp Biomed Imaging 9:358–361
11. Fengying X, Haidi F, Yang L, Zhiguo J, Rusong M, Alan B (2017) Melanoma classification on dermoscopy images using a neural network ensemble model. IEEE Trans Med Imaging 36(3):849–858
12. Sengupta S, Mittal N, Modi M (2020) Improved skin lesions detection using color space and artificial intelligence techniques. J Dermatol Treat 31(5):511–518

13. Sengupta S, Mittal N, Modi M (2019) Improved skin lesion edge detection method using ant colony optimization. Skin Res Technol 25:846–856
14. Mittal N, Tanwar S, Khatri S (2017) Identification & enhancement of different skin lesion images by segmentation techniques. In: IEEE international conference on reliability, infocom technologies, vol 9, pp 609–614
15. Sengupta S, Mittal N, Modi M (2018) Edge detection in dermascopic images by linear structuring element. In: International conference on reliability, infocom technologies and optimization, vol 7, pp 419–424
16. Sengupta S, Mittal N, Modi M (2019) Segmentation of skin lesion images using fudge factor based techniques. Lect Notes Mech Eng 1:837–846
17. Sahni P, Mittal N (2019) Breast cancer detection using image processing techniques. Adv Interdiscip Eng 1:813–823
18. Sengupta S, Dubey A, Mittal N (2020) Analysis of MRI images using image processing technique. In: Pandian A, Palanisamy R, Ntalianis K (eds) Proceeding of the international conference on computer networks, big data and IoT (ICCBI - 2019). ICCBI 2019. Lecture notes on data engineering and communications technologies, vol 49. Springer, Cham. https://doi.org/10.1007/978-3-030-43192-1_30
19. Mittal N, Singh H, Gupta R (2015) Decomposition & reconstruction of medical images in MATLAB using different Wavelet parameters. IEEE Int Conf Futur Trends Comput Anal Knowl Manag 1:647–653
20. History and Definition of the Sobel Operator. https://www.researchgate.net/publication/239398674_An_Isotropic_3x3_Image_Gradient_Operator,last accessed 2020/11/21.
21. Du-Yih T, Yongbum L, Eri M (2008) Information entropy measure for evaluation of image quality. J Digit Imaging 3:338–347

A Review on 6th Generation Wireless Communication Networks Based on Artificial Intelligence

Zerksis Mistry, Alok Kumar Yadav, and Mayank Kothari

Abstract Mobile communication technology is evolving day by day and year after year. This paper discusses the requirements of future wireless communication and how the upcoming technologies help in achieving required communication to go beyond 5th generation (5G) and 6th generation (6G) technologies along with its roles in different types of applications. 6G has three main advancement features, namely, cell phone ultra-broadband, Internet of Things (IOT), and Artificial Intelligence (AI). Basically, Tera hertz (THz) communication can be used to achieve super IOT, and machine learning techniques are promising candidates for AI. 6G will have frequencies ranging from 100 GHz to 3THz because wide band unutilized spectrum are unexplored to some extent and can hence be used for high frequency transmission of data. Initial research on this frequency range presents some of the discoveries that are quite promising, unconventional techniques, and latest effects, so one can use the resourceful data to improve and implement the 6G technology of wireless networks. In the current case, 6G networks will suffice the user needs of data rate that is sufficiently faster without any network interrupts. In this paper, comparison between 5G and 6G is discussed and how AI, Intelligent radio will help in achieving smart communication. 6G will be a quickest and quite effective wireless communication technology due to full duplex radio wave transmission capability. The most important purpose of these studies is to implement 6G networks throughout the world that can assist numerous emerging communication technologies to achieve very high speed data transfer at a very less cost with reliability.

Keywords Artificial intelligent · Intelligent radio · 5G · 6G · And deep neural network

Z. Mistry · A. Kumar Yadav · M. Kothari (✉)
SVKM's NMIMS (Deemed To Be University Shirpur), Shirpur, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_24

1 Introduction

The wireless communication enterprise is an industrial sector that has been consistent in a quick developing trend with innovative functions for many years since the development of previous generations. The terrific fulfillment of cell Internet has been the basis behind the massive evolution of wireless technology. The current use of fifth generation (5G) cell structures is constantly revealing obstacles of the machine, in comparison to the unique Internet of Everything (IOE) packages. These drawbacks of 5G focuses on defining the sixth generation (6G) network which are wireless and which could definitely integrate ways of achieving application ranging from self-sufficient systems to extended reality [1]. As this change in 5G is shifting to the new section of 6G networks, it has resulted in an increasing growth through commercial, instructional, and other groups to begin analyzing and designing the 6G networks. 6G radio fiber era utilizes the radio waves with around 5.8 GHz Bandwidth.

The imaginative and prescient will then make new network, i.e., 6G offerings and provide a stable study road map and pointers to shift from present day 5G networks toward 6G networks. 5G is still at the beginning stage since there are still many more researches going on for the same [4]. Hence, to maintain more sustainability and a strong competition in the field of wireless communication system, 6G will require three basic technologies, i.e., the COC, CAeC, and EDuRLLC offerings as shown in Fig. 1. The provider constitutes emerging 6G applications [2]. They can hardly ever be supplied by means of 5G, no longer in hand, due to their crucial necessities for larger and better data rates, decreased lag, dense connections, but additionally because of its severe demand for new and advanced performance metrics which have never been considered in 5G technology, for instance, put off jitter, conditions cognizance, unmanned aerial vehicle (UAV) or satellite television for personal computer system compatibility, and so on. 6G is a data device that is intelligent which is both driven by and is also a driver of the recent artificial intelligence (AI) technology [5, 6].

An agenda of 6G is based on the judicious strategies of many types of required parts, and also, it is likewise projected based on the 5G technology conditions. We envisage that AI shall considerably improve the focus of network operators, and also, additionally allow expansion to beneficial resources. To increase the network capacity, wireless networks are normally driven for higher rates, while the ultimatum for network capacity shows steady increment. The basic fundamentals of IOE systems of uniting billions of peoples and machineries is making a breakthrough from rate centric advanced cellular wideband to ultra-reliable low latency communication [5].

All the problems which are affecting the IOE systems to shift could not be solved by the 5G network. The plans for making smart cities and automation are facing a huge problem, especially, in industries and smart homes, which are not able to connect properly due to lack of network capacity because of its congestion in the network. To overcome this problem in making this world an IOE system, new generation networks, i.e., 6G wireless communication systems came into picture, whose designs are tailored in such a way that they meet the performance requirement of IOE applications [7] (Table 1).

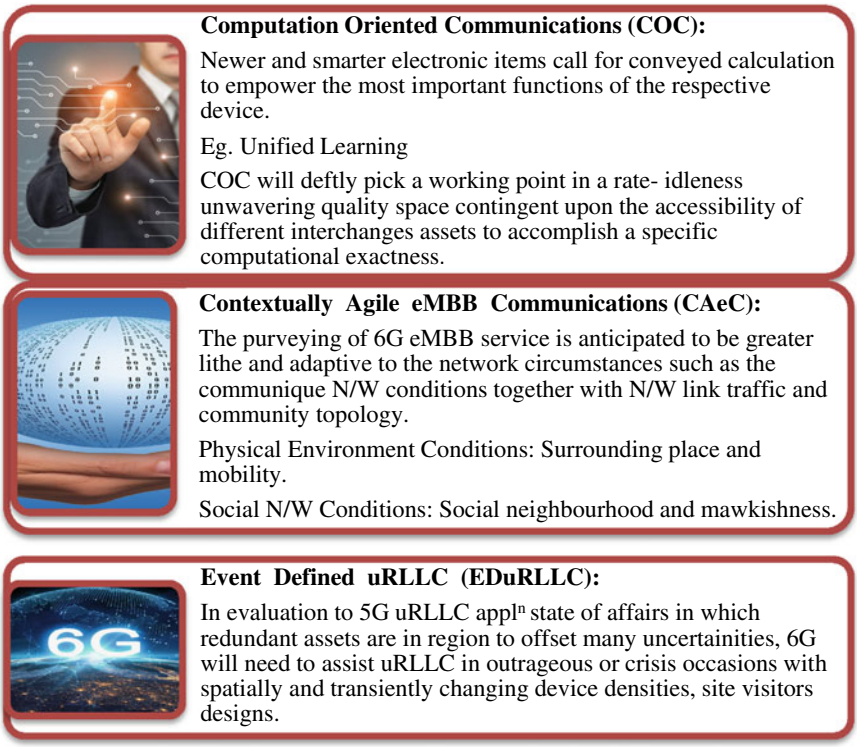


Fig. 1 Technology requirement for 6G [2, 3]

Table 1 Performance comparison between 5G and 6G networks

Features	5G	6G
Speed	50 Mbit/s–Gbit/s	About 100 times faster than 5G Network
Efficiency	Less efficient in comparison to 6G networks	More efficient than 5G networks
Smartness	Less Smart in comparison to 6G	Smartest among the previous generation Networks
Technology	4G and WWWW	5G and satellite
Core Networks	Internet	Internet
Standards	WiMAX LAS CDMA, OFDM, MC-CDMA, UWB, Network-LMDS, IPv6	GPS, COMPASS, GLONASS, Galileo systems
Handoff	Horizontal and vertical	Horizontal and vertical

2 Literature Survey

2.1 Communication Technologies

In this recent era, the technologies that are wireless in nature are quite useful. It helps the users to transfer the information from one place to the other, to be more precise, from one base-station to another base-station, wirelessly. These technologies help the users to communicate short as well as long distances. Also, the mode of communication is being improved gradually to create a customer-friendly environment. Especially, wireless mode of communication has made the lives of the people quite convenient. There are many wireless technologies which have been already made available to the people, i.e., from 1 to 5G networks in which the later have provided high data rates, good network coverage, and incredibly high performance. 6G enables the devices to connect Internet with wireless devices like Wi-Max, Wi-Fi, Bluetooth in the most efficient and a faster way [8, 9].

2.2 Impact on Economy

The mode of data transmission is through the incorporation of radio waves in 6G. The overall cost of the data will be reduced in comparison to earlier plans because 6G allows the user of the respective network to use the connectivity that is very effective. This connectivity is offered from Wi-Fi router systems or even mobile data. This will in turn help the consumers from paying tons of money for subscription or Internet plans to the Internet Service Provider (ISP) [10].

There are a lot of people who are suffering from low loading speeds of the videos or movies that they are viewing. Hence, many prefer to download the movies from torrent or other websites where the respective movies or TV series are available. This emerges as a huge threat nowadays to both the theaters along with film industries. Since the data rate of 6G is extremely high, the users will be able to stream movies, web series, and other videos even when the user is traveling with very less or no buffering at all. The film industries are also quite profitable using this methodology for telecasting the movies, i.e., streaming online movies. The cost spent to view the movies can also be reduced if this methodology is implicated as this is quite convenient for the people, and hence, more and more people will go for this technology as they may stream the same movies with same quality at their home or even while traveling with very less or no buffering, which in turn will definitely increase profits of the producers of the movie [9]. Also, because of such high data speeds of 6G, privacy of movies will be reduced as people would rather stream movies on such online platforms such as Netflix, Amazon Prime, JIO Cinema to name a few. Hence, 6G plays a vital role in such scenarios and emerges itself as a pervasive solution for avoiding the ever-augmenting threat on movie piracy [10, 11].

Table 2 Difference between 1G, 2G, 3G, 4G, and 5G networks [13–15]

Features	1G	2G	3G	4G	5G
Speed	2.4 Kbps	64 Kbps	2 Mbps	Up to 100 Mbps	Up to 10 Gbps
System	Analog	Digital	Digital	Digital	Digital
Voice quality	Poor	Moderate	Good	Superior	Superior
Capacity	Low	Low but better than 1G	High	High	Very High
Video processing	Poor	Poor	Good	Better than 3G	Best as compared to earlier generations
Security	No security	Very Low	Good	High	Very high
Latency	–	300–1000 ms	100–500 ms	10 ms (radio)	< 1 ms (radio)
Additional features	–	Introduce services such as text messages, picture messages, and MMS	Provides faster communication, Send/receive large Email messages, Services like video calling, GPS, mobile television, etc	High quality video streaming, Low cost/bit, Uses Packet Switching for voice and video calls instead of circuit switching	Capable of supporting WWWW, Fastest data transmission, Low cost with high connectivity More effective and attractive

2.3 Gradual Up-Gradation

Nowadays, 4G mobile network technologies are available on almost each and every device, while 5G has recently come into existence and is evolving gradually. Still, the coverage area of the network and the ever-increasing demand of high data rate are continuously increasing and still remains as a challenge [12]. 6G technology has the ability to provide data speeds up to the rate of 1000 Mb/sec which is more than the expectation of the users. Currently, 4G technology is offering 100 Mb/sec internet speeds. Moreover, security is the prime aspect for wireless data communication which is quite advanced in 6G network technology [3] (Table 2).

2.4 Improved Security, Speed, and Storage

Since 6G networks have the ability to deliver very high Internet speed and have the ability to store a massive amount of data, more memory is occupied. As security is

of utmost importance to each and every one, 6G will allow the users to use the data for different applications with high security and encryption of messages [16].

3 6G Network Architecture

Figure 2 shows the potential architecture of 6G networks which includes intellectual network, sub-network evolution, and smart radio.

3.1 Intelligent Network with AI

6G will boost software-based networks to a next level, particularly in the direction of intelligence of the network. Networking and network functions, virtualization, that have moved cutting-edge communication networks towards virtual networks is basically based on software. They additionally permit reduced networking, which can provide an effective virtualization functionality to permit more than one digital network to be created at a shared infrastructure [2]. Nonetheless, the complexity and heterogeneity of the network is increasing day by day, softwarization by itself would not persuade 6G networks. Specifically, for assisting AI-based programs, the network organizations have to aid numerous capabilities, along with data warehousing, communications, computing, or even wireless power switch. Moreover, 6G networks will include new wireless entry to interfaces consisting of Tera Hertz (THz) communications and surfaces that are itself intelligent. It may also support advance Internet of Things (IoT) functions along with monitoring and sensing, collection of data, analytics, and data storage. All of the aforementioned demanding situations require a structure that is biddable, amenable, and quite importantly smarter [17, 18].

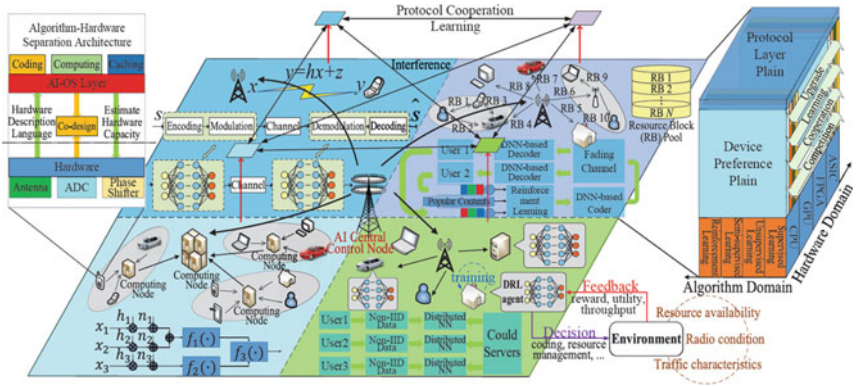


Fig. 2 Architecture of 6G network

Primitive methods of wireless community expansion cannot be used in 6G networks because the networks might not be extraordinarily zestful, and also, the overall complexity may be quite high, due to its dimensions, density, and its heterogeneity [19]. Hence, the other fundamental software of AI in 6G wireless systems is self-driven [20].

Massive statistical analytics is the first basic utility of AI. In all, there are four sorts of data analytics which can be implemented for 6G, namely, analytics that are quite descriptive, analytics that can be diagnosed, analytics which can be predicted, and prescriptive analytics. It substantially advances the cognizance of community operators and carrier providers. Diagnostic analytics can detect network errors or faults and provider impairments, pick out reasons of network peculiarity, and also in the end, improve the community safety and make it more reliable. It is important to know that harvesting and analyzing massive quantity of facts increases issues about security of data and may also temper privacy. Since security, privacy, and integrity are the key requirements of every individual, 6G structure and the respective protocols are to be designed by considering the basic essential factors mentioned above [21, 22].

3.2 A Criss-Cross of Subassemblies

It predicted extreme heterogeneousness which is one of the main functions of 6G networks, i.e., it will be its own potentiality to take advantage of a dynamic subnet-broad development to efficaciously adapt to the surroundings and consumer needs, thereby ensuing in a community of sub-network. The neighborhood development may take place in neighboring cells to do some effort to flexibly practice current trends on new wavelets or complete waveforms, code techniques, and multiple access protocols in subnets without tremendous time-ingesting. Each sub-network gathers and examines its neighborhood records, which may additionally include the environments that are wireless in nature, individual requests, mobility patterns, and so on, which makes the most AI methods to upgrade itself [23].

3.3 Intelligent Radio (IR)

This is in disparity to the structures from1G network to 5G network in which the devices and transceiver algorithms are collectively designed. Based on the hardware statistics and techniques based on AI, the operating system should be capable of designing its own algorithms of trans-receivers through interfacing. In comparison to the gaining knowledge of primarily based intelligent physical layer mentioned ahead, IR basically counts on the algorithm hardware separation architecture. The traditional modulation or coding factor in special Cognitive Radio would be replaced with a way of Deep Neural networks (DNN) in IR that would be adapted in smart manner to the surroundings with the respective hardware. An AI chip is able to

Table 3 Difference between SDR, CR, and IR [8, 25]

Parameter	Software defined radio (SDR)	Cognitive radio (CR)	Intelligent radio (IR)
Frequency band	Invariable	Environment adaptable	Adaptable to the environment and hardware
Hardware upgradability	No	No	Yes
Multiple access techniques	Prearranged	Based on sensing	Distributed ML based
Sharing of spectrum	Invariable	Based on opportunity	AI enabled
Capability of hardware	Claimed earlier	Claimed earlier	Estimated online

impose DNN in low electricity, which in turn benefits the DNNs primarily based on IR and leading to a fundamental change in hardware structure of 6G transceivers [12, 19]. The applications of IR shall assist 6G revel in miles decreased enactment time with a massive bargain in the value of latest hardware and algorithms, consequently boosting its self-development.

3.4 Smart Wireless Communication

The physical layer of communique structures that are wireless, suffers from an extensive form of deterioration, which includes hardware damage, which include local oscillator leakage, amplifier distortion, and channel impairments inclusive of fading, interference, and so on. Because of the high complexity, it can be noticed that end-to-end expansion has in no way been practical in wireless systems [24]. By partitioning the complete chain into the subsequent blocks with simplified model enable with artificial intelligence that have the capability of capturing the same functionality of system. To have a reliable and efficacious communication with the prevention of deterioration of hardware as well as channel, a massive variety of parameters are required to be managed and enhanced together [2] (Table 3).

4 Research Challenges in 6G

6G consists of both ground as well as airborne networks due to which 6G will support communications in 3D area, serving users in 3D, and developing 3D base-stations. This stimulates interest in studying more and more about it, and there are three main areas of study. First, the size and model of 3D-propogation surrounding is required.

Second, newer procedures for 3D frequency and community plans are required for more development. Third, network optimization for controlling mobility, managing resources in 3D, and routing are required [26].

Fundamental 3D performance of 6G systems in terms of rate-reliability-latency trade-offs is needed. Such analysis will quantify the spectrum, energy, and communication requirements that 6G needs in order to support the applications [27].

Several issues arise on utilizing THz and millimeter (mm) wave in 6G. For mm-wave, assisting high mobility will be an open trouble. For THz, new models of propagation and the architectures of the trans-receivers will be required. Higher strength, excessive sensitivity, and low noise figure are the main trans-receiver capabilities required to overcome high THz direction loss. Once those physical layer elements are properly understood, there may be a need to expand new network and hyperlink-layer protocols to optimize the use of cross-frequency sources, and at the same time, taking into consideration, the differentiating and uncertain nature of the mm-wave and THz hertz environments [28, 29].

AI itself is a very huge term which brings ahead many research directions for 6G. AI is called as parent branch which consists of Deep Learning (DL) and Machine Learning (ML) which may be called as its child branch. To have knowledge of AI, there is an utmost need to know about ML and DL with algorithms that would be quite reliable in 6G to deliver the packages to the server. To carry out these critical operations, there should be very little or no latency, should be very reliable, and scalable along with dependable infrastructure [30–32]. The tasks of AI are computationally concentrated and for the most part prepared, created, and sent at server farms with servers which are custom designed. Given the quick development of intelligent portable devices, it is normal that countless applications will be sent at the edge of remote systems. As such, the 6G remote system will be intended to influence propelled remote interchanges and versatile processing innovations to help AI-empowered applications at different edge cell phones [33]. Remarkably, the limit and idleness of remote connections are the key bottlenecks of versatile AI applications because of three reasons. First, to secure protection, some AI applications expect information to be kept at the cell phones as opposed to being transferred to the cloud for model preparing. This has fascinated the ongoing research enthusiasm for on-gadget dispersion preparation. Second, to defeat the asset constraint of edge gadgets, on-gadget appropriate registering gives new open doors by pooling the calculation and capacity assets of various cell phones. For this situation, information rearranging is a key part for trading the registered middle qualities among cell phones. To wrap things up, the heterogeneous blend of the cloud, edge, and end gadgets, gives a dispersed or rather, say, scattered registering condition for both inference and training of DNNs. To empower omnipresent and differentiated versatile AI administrations, 6G is relied upon to give adaptable stages for creating further correspondence along with calculation innovations.

5 Conclusion

The foresight of data rates and also volume is increasing on daily basis by numerous calls for further inventions in processing of the respective signal. The volume of this statistics will be such that superior control information and record analytics will be required to preserve the networks and also to keep them running with effectiveness and high reliability. This article is an unassuming endeavor to give a forward-looking pioneer guide for 6th generation (6G) networks. Newer highlights of 6G development were recognized and also there were some points discussed about empowering advances. This paper gives a description about latest standard body activities which focused at encouraging destiny cordless devices, a good way to use multiple giga hertz bandwidth channels may encourage data rates more than 100 Giga bits per seconds. As the reckoning power steadily increases to meet the processing strength of human-brain. This paper shows an ample amount of applications that require high frequencies upto Tera hertz (THz). 6G visions include three major aspects: mobile ultra-broadband, super Internet of Things (IoT), and Artificial Intelligence. In particular, THz communications is quite promising to support mobile ultra-broadband, symbiotic radio, and satellite-assisted IoT communications which can be used to achieve super IoT, and deep learning and reinforcement learning, which are in turn, two representative AI technologies.

References

1. David K, Berndt H (2018) 6G vision and requirements: is there any need for beyond 5G? IEEE Veh Technol Mag 13(3):72–80
2. Letaief KB, Chen W, Shi Y, Zhang J, Zhang YJA (2019) The roadmap to 6G: AI empowered wireless networks. IEEE Commun Mag 57(8):84–90
3. Zhang Z, Xiao Y, Ma Z, Xiao M, Ding Z, Lei X, Fan P (2019) 6G wireless networks: vision, requirements, architecture, and key technologies. IEEE Veh Technol Mag 14(3):28–41
4. Chen M, Challita U, Saad W, Yin C, Debbah M (2019) Artificial neural networks-based machine learning for wireless networks: a tutorial. IEEE Commun Surv Tutor 21(4):3039–3071
5. Zhang L, Liang YC, Niyato D (2019) 6G Visions: Mobile ultra-broadband, super internet-of-things, and artificial intelligence. China Commun 16(8):1–14
6. Saad W, Bennis M, Chen M (2019) A vision of 6G wireless systems: applications, trends, technologies, and open research problems. IEEE Netw
7. Katz M, Matinmikko-Blue M, Latva-Aho M (2018) 6Genesis flagship program: building the bridges towards 6G-enabled wireless smart society and ecosystem. In: 2018 IEEE 10th Latin-American conference on communications (LATINCOM). IEEE, pp 1–9
8. Raj V, Dias I, Tholeti T, Kalyani S (2018) Spectrum access in cognitive radio using a two-stage reinforcement learning approach. IEEE J Sel Top Signal Process 12(1):20–34
9. Bennis M, Debbah M, Poor HV (2018) Ultra reliable and low-latency wireless communication: tail, risk, and scale. Proc IEEE 106(10):1834–1853
10. Kasgari ATZ, Saad W, Debbah M (2019) Human-in-the-loop wireless communications: machine learning and brain-aware resource management. IEEE Trans Commun 67(11):7727–7743
11. Li Y, Liu J, Cao B, Wang C (2018) Joint optimization of radio and virtual machine resources with uncertain user demands in mobile cloud computing. IEEE Trans Multimed 20(9):2427–2438

12. Park J, Samarakoon S, Bennis M, Debbah M (2019) Wireless network intelligence at the edge. Proc IEEE 107(11):2204–2239
13. AYSHATH SILMIYA MERIL, Basthikodi M, Faizabadi AR (2019) Review: comprehensive study of 5G and 6G communication network. JETIR 6(5). ISSN-2349-5162
14. Arockia Panimalar S, Monica J, Amala S, Chinmaya V (2017) "6G Technology". IRJET 4(9)
15. Ramkumar S (2018) A recent review on growth of mobile generations-case study. ICICSS
16. Mao Y, You C, Zhang J, Huang K, Letaief KB (2017) A survey on mobile edge computing: the communication perspective. IEEE Commun Surv Tutor 19(4):2322–2358
17. Elayan H, Amin O, Shubair RM, Alouini MS (2018) Terahertz communication: the opportunities of wireless technology beyond 5G. In: 2018 international conference on advanced communication technologies and networking (CommNet). IEEE, pp 1–5
18. Tonouchi M (2007) Cutting-edge terahertz technology. Nat Photonics 1(2):97
19. Kalbande D, Haji S, Haji R (2019) 6G-next gen mobile wireless communication approach. In: 2019 3rd international conference on electronics, communication and aerospace technology (ICECA). IEEE, pp 1–6
20. Ma J, Shrestha R, Adelberg J, Yeh CY, Hossain Z, Knightly E, Mittleman DM (2018) Security and eavesdropping in terahertz wireless links. Nature 563(7729):89–93
21. Lovén L, Leppänen T, Peltonen E, Partala J, Harjula E, Porambage P, Riekki J (2019) Edge AI: a vision for distributed, edge-native artificial intelligence in future 6G networks. In: The 1st 6G wireless summit, pp 1–2
22. Chowdhury MZ, Shahjalal M, Ahmed S, Jang YM (2019) 6G wireless communication systems: applications, requirements, technologies, challenges, and research directions. arXiv:1909.11315
23. Giordani M, Polese M, Mezzavilla M, Rangan S, Zorzi M (2020) Toward 6g networks: use cases and technologies. IEEE Commun Mag 58(3):55–61
24. Raghavan V, Li J (2019) Evolution of physical-layer communications research in the post-5G era. IEEE Access 7:10392–10401
25. B5GS (2018) The Brooklyn 5G Summit. https://2018.brooklyn5gsummit.com/
26. Mozaffari M, Kasgari ATZ, Saad W, Bennis M, Debbah M (2018) Beyond 5G with UAVs: foundations of a 3D wireless cellular network. IEEE Trans Wireless Commun 18(1):357–372
27. Kato N, Fadlullah ZM, Tang F, Mao B, Tani S, Okamura A, Liu J (2019) Optimizing space-air-ground integrated networks by artificial intelligence. IEEE Wirel Commun 26(4):140–147
28. Harvey JF, Steer MB, Rappaport TS (2019) Exploiting high millimeter wave bands for military communications, applications, and design. IEEE Access 7:52350–52359
29. Rappaport TS, Xing Y, Kanhere O, Ju S, Madanayake A, Mandal S, Trichopoulos GC (2019) Wireless communications and applications above 100 GHz: opportunities and challenges for 6G and beyond. IEEE Access 7:78729–78757
30. Strinati EC, Barbarossa S, Gonzalez-Jimenez JL, Ktenas D, Cassiau N, Maret L, Dehos C (2019) 6G: the next frontier: from holographic messaging to artificial intelligence using subterahertz and visible light communication. IEEE Veh Technol Mag 14(3):42–50
31. Kasgari ATZ, Saad W (2019) Model-free ultra-reliable low latency communication (URLLC): a deep reinforcement learning framework. In: ICC 2019–2019 IEEE international conference on communications (ICC). IEEE, pp 1–6
32. Kim M, Kim NI, Lee W, Cho DH (2018) Deep learning-aided SCMA. IEEE Commun Lett 22(4):720–723
33. Luong NC, Hoang DT, Gong S, Niyato D, Wang P, Liang YC, Kim DI (2019) Applications of deep reinforcement learning in communications and networking: a survey. IEEE Commun Surv Tutor 21(4):3133–3174
34. Aladsani M, Alkhateeb A, Trichopoulos GC (2019) Leveraging mmWave imaging and communications for simultaneous localization and mapping. In: ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 4539–4543
35. Koenig S, Lopez-Diaz D, Antes J, Boes F, Henneberger R, Leuther A (2013) Nat Photonics 7(12):977

36. Prasad R (2013) Global ICT standardisation forum for India (GISFI) and 5G standardization. J ICT Stand 1(2):123–136
37. Kanhere O, Rappaport TS (2018) Position locationing for millimeter wave systems. In: 2018 IEEE global communications conference (GLOBECOM). IEEE, pp 206–212
38. Yu X, Zhang J, Letaief KB (2018) A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems. IEEE J Sel Top Signal Process 12(2):282–297
39. Chen M, Saad W, Yin C (2018) Virtual reality over wireless networks: quality-of-service model and learning-based resource management. IEEE Trans Commun 66(11):5621–5635

Background Object Removal and Image Inpainting to Fill Irregular Holes

Ayan J. Malhotra, Ashish Chopra, Rajan Dahiya, Pratik Yadav, and Aryan Singhal

Abstract In the current scenario, nearly all deep learning methods to perform image inpainting use a standard convolution network in order to regenerate the holes created by removing the object. However, the result of this procedure is not up to the mark as the resultant image often appears to be distorted and blurry. Post-processing techniques do address this issue but prove to be computationally expensive. We aim to remove the background object from the image while ensuring that the regenerated image has minimal discrepancies. In our approach, we make use of partial convolution in which at every stage of the convolution network, we replace the convolution block with a partial convolution block. We have also focused on producing a high-quality mask dataset which ensures that our model works well on real life images where aberrantly shaped and sized objects are needed to be removed. Qualitative comparisons help analyze the predictions of our model in relation to those made by the standard image inpainting model of OpenCV.

Keywords Image inpainting · Object removal · Partial convolution · CNN

1 Introduction

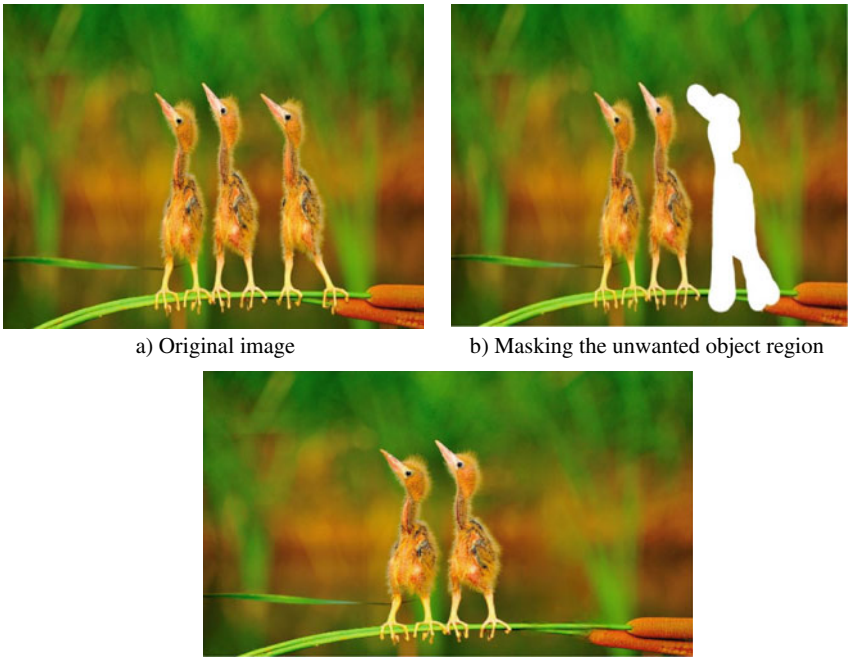
This paper describes a novel approach to remove objects from an image seamlessly. Figure 1 illustrates an example in which an object present in a certain region of the image is removed from the image, and how that region smoothly merges with the neighboring background.

Image inpainting is a widely growing field in computer vision, having several applications including image restoration, editing, and composition. Historically, adept artists have had to perform image inpainting manually. But in recent times, several standard ways to perform image inpainting have arisen through applications such as photoshop.

A. J. Malhotra (✉) · P. Yadav · A. Singhal
Delhi Technological University, 110042 Delhi, India

A. Chopra · R. Dahiya
Samsung Research Institute Noida, 201301 Noida, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_25



a) Original image b) Masking the unwanted object region

c) Resultant image generated by our model after removal of undesired region

Fig. 1 Illustrates an example

Using deep networks (CNNs) for image inpainting is a field that has seen considerable advancement [12]. Due to their adept ability to represent and extensive training, methods relying on CNNs are efficacious while predicting semantically plausible results. However, a large portion of the pre-existing CNN-based image inpainting methods are trained to remove rectangular shapes and do not work appropriately while trying to replace aberrant shaped holes. Numerous post-processing techniques, such as [19], attempted to address this issue but proved to be computationally expensive.

Certain algorithms make use of features belonging to the leftover image to recreate the void rather than depending upon CNNs. PatchMatch [1], a premier method that uses this technique, sequentially looks for the ideal patches to refill the void. Although this algorithm often generates seamless outputs, it is highly dependent on the presence of suitable image statistics and has very little understanding of semantics that can be visualized. Exemplar-based methods which are dependent on image statistics tend to perform poorly on real world data and struggle to generate convoluted structures.

In this paper, we attempt to provide an inpainting approach which works smoothly even while filling irregular holes due to its use of a partial convolutional layer while being trained on a sizable dataset. Partial convolution (PConv) is known to be better at handling irregular holes without producing discrepancies [10]. The trainable masked

images are produced in a random unbiased manner which lead to the approach working well on tangible data.

Our contributions are consolidated below:

(1) We propose a robust deep learning-based inpainting model producing impeccable results due to its usage of a carefully trained partial convolution.
(2) Our approach to create random unbiased shaped masks to train the images produces high quality results even on a challenging dataset such as [9].

2 Related Work

There are many existing non-learning image inpainting techniques but they all depend on propagating information of neighboring pixels to the holes. They use a mechanism of distance field [2]. These methods only give satisfactory results in cases where the hole size is small and the hole shape is narrow. Certain algorithms based on patch-match like [3] function by analyzing the segment of the image not containing the hole, but the computation cost is not feasible as presented in [15].

Some CNN approaches initialize the masked section with the average value of the pixels [14], that is, further processed on by a CNN. Context Encoders [12] embed the 128×128 image with a 64×64 center hole towards a lower dimensional feature space and proceed to decode in the form of a 64×64 image.

Amongst the deep learning techniques, certain approaches ignore the placeholder values. Ulyanov [18] went on to discover that the network does not require the dataset, it can rather just depend on the architecture of the network.

We proceeded to use a convolution function that has been reweighted and allows us to formulate the result only on feasible inputs, [6] has previously used this method in conjugation with a soft attention mask for classifying semantically. Uhrig et al. [17] applied sparsity constant CNNs with a similar convolution technique and a max pooling-dependent mask changing method to complete the process. Whereas, [13] used a Shepard convolution layer in which both the filter as well as mask convolutions make use of an identical convolution layer. This inhibits the hole from regenerating appropriately due to the possible negative entries in the kernel and struggles to cope with larger sized holes. Discussions of other CNN variants is not in the scope of our work.

Our model is vigorous in nature as the masks generated were not from a dataset. We used the open cv library in python to produce random circles, lines, rectangle, etc., of various thickness. As our goal is to remove the marked object from the image where the object can be of random shape and size, the random masks generator which we used is the ideal choice. This also ensures that the model does not overfit to the masks used in training. Most previous works on image inpainting mainly focused on producing the image back to the original image, but our work focuses on removing the object altogether. Hence, our unique approach allows us to remove objects from real world images with ease.

3 Methodology

3.1 Partial Convolution Layer (PConv)

The partial convolution layer is critical to the step of regenerating the masked area. In each iteration, the partial convolution layer makes changes to the mask. The unmasked region is solely responsible for determining the values of the newly created feature map [11]. Assume W to be the weights for the convolution filter and b to be the corresponding filter bias. X is defined as the value of the pixels for the ongoing iteration and M is the mask represented in binary form. Each step of the partial convolution is described with the help of the following equation:

$$x'' = \begin{cases} W^T (X \odot M) \frac{sum[1]}{sum[M]} + b, & if\ sum(M) > 0 \\ 0, & other\ cases \end{cases} \tag{1}$$

Here, $\odot$ represents multiplication (performed element-wise), and sum [1] is identical to sum [M] in shape, but is completely filled with 1's.

Following every partial convolution operation, we proceed to change the mask with the following logic:

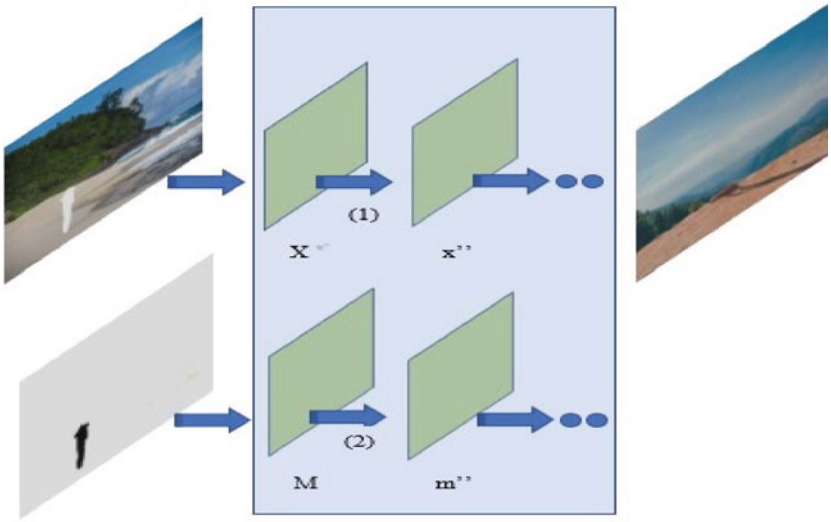
$$m'' = \begin{cases} 1, & if\ sum(M) > 0 \\ 0, & other\ cases \end{cases} \tag{2}$$

Hence, this approach can be easily implemented during the forward pass of our deep learning algorithm. Figure 2 shows the architecture of PConv.

3.2 Model Architecture

Our architecture is in the form of a U-net analogous to the architecture presented in [7]. The simple approach just defines the binary masks. Inputs for every convolution layer are created by combining the previous two feature maps and masks images. To ensure that the masks can be of any size and shape, the training of the model will be done on a vast variety of randomly produced unbiased masks.

The PConv layer is implemented with the help of its existing standard in PyTorch. The mask is created in a straightforward manner by producing an image in binary form that has an identical size to the image being used as the input. Masks are updated with the help of a fixed convolutional layer. To prevent unnecessary values outside the image from affecting the image border, we make use of partial convolution layer padding.



Input image and masked image

Single Update of input image and mask using (1) & (2)

Final output after several updates

Fig. 2 Partial convolution architecture

3.3 Loss Function

To generate a smooth image, we combine multiple losses to train our model. These losses include pixel reconstruction loss, perceptual loss, style loss [5], and total variation loss. These losses are defined in [10] and we calculated the total loss in a similar manner using the equation:

$$L_{total} = \lambda_1 L_{valid} + \lambda_2 L_{hole} + \lambda_3 L_{percep} + \lambda_4 (L_{styke(out)} + L_{styke(comp)}) + \lambda_5 L_{TV} \tag{3}$$

The values of the hyperparameters were decided empirically as follows:

$$\lambda_1 = 1, \lambda_1 = 8, \lambda_1 = 0.06, \lambda_1 = 150, \lambda_1 = 0.12 \tag{4}$$

4 Experiments and Results

4.1 Generating Masked Image Dataset

Our carefully generated masked image dataset plays a significant role in ensuring that our final model is capable of removing objects of all shapes and sizes. Most of the previous work on image inpainting does not perform satisfactorily while trying to remove non-rectangular objects. To counter this issue, we develop masks containing a wide variety of shapes, such as straight lines, circles, and ellipses. These shapes are produced so that they are present in random quantity and in random locations.

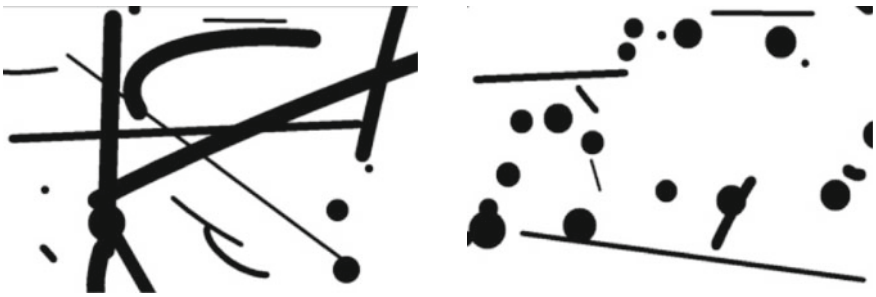
25 K masks are produced which were selected during the training and testing the model while trying to prevent any bias. All the images and generated masks used in the model are of size 512 × 512. Illustrations of the masks generated are shown in Fig. 3.

4.2 Training Procedure

To determine the accuracy of our model, we have used Microsoft coco-val 2014 [9] dataset which has previously been used widely in image inpainting literature [4, 16].

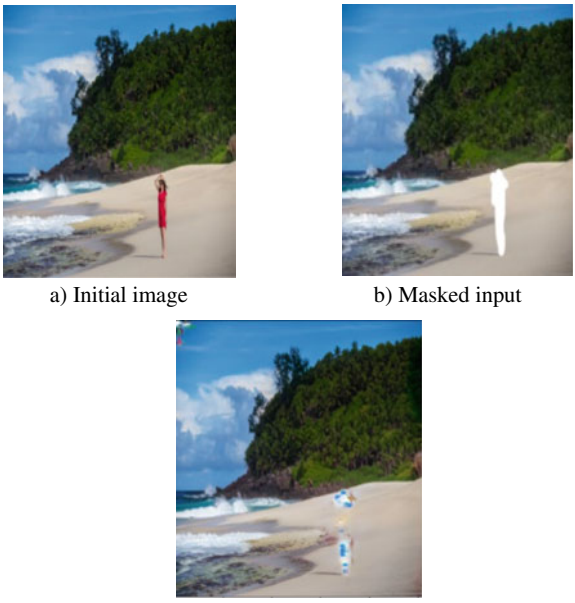
We work on a grand total of 20 K images in which 16 K are randomly selected and partitioned as the training set and the remaining 4 K images are used as the test set.

Adam algorithm has been used to optimize our model [8], where the initial learning weights were set as 3*(1e – 4) and $\beta = 0.5$.



a) b)

Fig. 3 Depiction of masked image dataset



a) Initial image

b) Masked input

c) Resultant image after 10 epochs

Fig. 4 Progression in training after just 10 epochs

The process of training the model is completed in 50 iterations with size of the batch set to 5. We train on a single Nvidia GTX 1080Ti G.P.U. Progression in training after just 10 epochs is depicted in Fig. 4.

4.3 Qualitative Results

In this part, we visually depict the regenerated images produced by our model while also displaying the initial images and their respective masked inputs. To demonstrate the performance of our model in relation to a pre-existing conventional inpainting model, we also showcase the final results achieved by passing the same inputs to the inpainting model present in OpenCV. These comparisons are illustrated in Fig. 5.

Our model clearly delivers satisfactory results commensurate to an existing state-of-the-art inpainting model. Our model is successful at preventing the final image from being distorted and blurry. Figure 5 also shows how our model is adept at removing objects of aberrant shapes and how the neighboring pixels merge seamlessly.

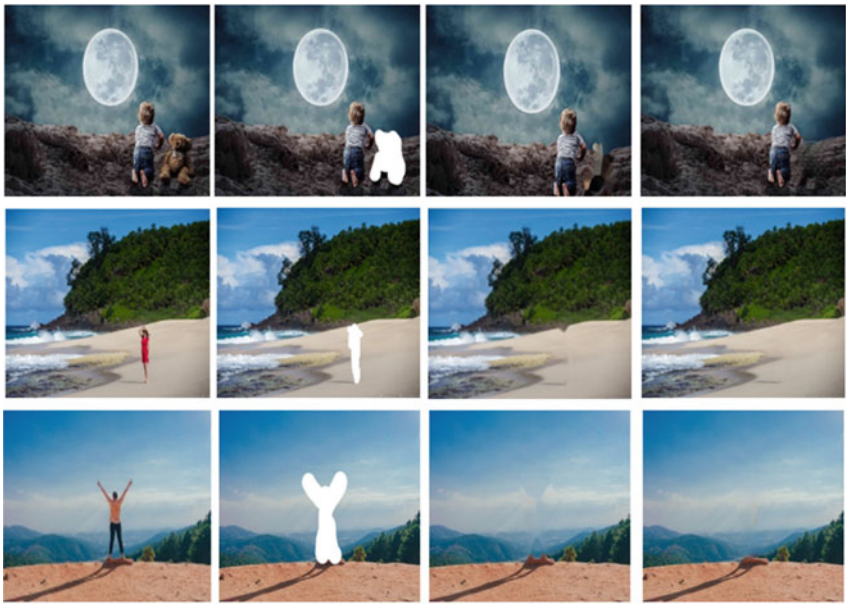


Fig. 5 Comparisons

5 Conclusion

The paper proposed a partial convolution-based image inpainting model. The combination of our unique approach to generate masks and the use of a robust deep learning architecture allowed us to produce satisfactory results. The values set for the various hyperparameters as well the image dataset used for training seem to be working well.

Although the model is able to deal with holes of irregular shapes and sizes, the results produced for extremely large sized holes can be improved upon. However, this is expected as almost any CNN-based method will struggle to perform well with limited input pixels from the original image.

In our upcoming work, we plan on exploring the possibility of training our model on even more challenging datasets and on images with super resolution. We also look forward to reducing the computational expense of our algorithm without compensating on the quality of the generated results.

References

1. Barnes C, Shechtman E, Finkelstein A, Goldman DB (2009) PatchMatch: a randomized correspondence algorithm for structural image editing. In: SIGGRAH, pp 24:1–24:11

2. Bertalmio M, Sapiro G, Caselles V, Ballester C (2000) Image inpainting. In: Proceedings of the 27th annual conference on computer graphics and interactive techniques, pp. 417–424
3. Efros AA, Freeman WT (2001) Image quilting for texture synthesis and transfer. In: Proceedings of the 28th annual conference on computer graphics and interactive techniques. ACM, pp 341–346
4. Elharrouss O, Almaadeed N, Al-Maadeed S, Akbari Y (2019) Image inpainting: a review. Neural Process Lett 51(2):2007–2028
5. Gatys L, Ecker A, Bethge M (2016) Image style transfer using convolutional neural networks. In: CVPR, pp 2414–2423
6. Harley AW, Derpanis KG, Kokkinos I (2017) Segmentation-aware convolutional networks using local attention masks. In: IEEE international conference on computer vision (ICCV), vol 2, p 7
7. Isola P, Zhu JY, Zhou T, Efros AA (2016) Image-to-image translation with conditional adversarial networks. arXiv:1611.07004
8. Kingma D, Ba J (2015) Adam: a method for stochastic optimization. In: ICLR
9. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft COCO: common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) ECCV 2014, Part V. LNCS, vol 8693. Springer, Heidelberg, pp 740–755
10. Liu G, Reda FA, Shih KJ, Wang T-C, Tao A, Catanzaro B (2018) Image inpainting for irregular holes using partial convolutions. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (eds) ECCV 2018, vol 11215. LNCS. Springer, Cham, pp 89–105
11. Li J, He F, Zhang L, Du B, Tao D (2019) Progressive reconstruction of visual structure for image inpainting.. In: Proceedings of the IEEE/CVF international conference on computer vision (ICCV) pp 5962–5971
12. Pathak D, Krähenbühl P, Donahue J, Darrell T, Efros AA (2016) Context encoders: feature learning by inpainting. CoRR, abs/1604.07379
13. Ren JS, Xu L, Yan Q, Sun W (2015) Shepard convolutional neural networks. In: Advances in neural information processing systems, pp. 901–909 (2015)
14. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. In: IJCV, pp 115–211
15. Simakov D, Caspi Y, Shechtman E, Irani M (2008) Summarizing visual data using bidirectional similarity. In: CVPR, pp 1–8
16. Sun T, Fang W, Chen W, Yao Y, Bi F, Wu B (2019) High-resolution image inpainting based on multi-scale neural network. Electronics 8(11):1370
17. Uhrig J, Schneider N, Schneider L, Franke U, Brox T, Geiger A (2017) Sparsity invariant CNNs. In: International conference on 3D vision (3DV)
18. Ulyanov D, Lempitsky V, Vedaldi A (2018) Deep image prior. In: IEEE/CVF conference on computer vision and pattern recognition, pp 9446–9454
19. Yang C, Lu X, Lin Z, Shechtman E, Wang O, Li H (2017) High-resolution image inpainting using multiscale neural patch synthesis. In: The IEEE conference on computer vision and pattern recognition (CVPR), pp 6721–6729

Exploration of Crime Detection Using Deep Learning

Shruti Bhalla and Rajesh Kumar Singh

Abstract There is the need for a device to identify and forecast crimes at a complex period because of the escalation in the occurrence of crimes. The goal of this survey is to learn different deep learning techniques to help identify and forecast crimes using Deep Learning. Noticeable findings from this survey are that pre-processing becomes an essential activity as dataset instances have a significant amount of missing values, and violence does not arise consistently through metropolitan environments but concentrates in particular locations. Predicting crime hotspots is also a very important activity, and adding post-processing would help lower crime rates. The Deep Learning algorithm was commonly used in many areas, including image recognition and natural language processing. With fine tuning, the Deep Learning algorithm produces improved predictive performance. The overall detection depends on the crime data sets which could be found from the UCI library or any standard depository. Through Deep learning algorithm on the data set, it could be analyzed that how the crime patterns are in the particular demography. So, applying deep learning in crime prevention is highly advantageous and could be further reinforced by offering more resources to intelligent systems. This will certainly help to get better findings about crime through deep learnings. This paper explores mainly genetic algorithms to find the crime detection.

Keywords Genetic algorithm · Deep learning · Crime detection · Crime data sets

1 Introduction

As deep learning is the most popular technology at present, it is used in multiple applications [1]. Crime Identification and Prevention is one such technology that is being utilized. The premise behind all of these programmes is that criminals are reasonably predictable; it just involves being able to dig through a large array of

S. Bhalla
Punjab Technical University, Kapurthala, India

R. Kumar Singh (✉)
SUS institute of computer, Tangori, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_26

knowledge and identify trends that are beneficial and law enforcing. A couple of decades back, this sort of data mining was practically unimaginable, but the assumption is that recent advances in machine learning are up to the challenge. In India, 99% of incidents of sexual harassment on women remain unreported and just 13% of instances of shooting have been registered to the police. Just 20 percent of gunshot incidents have been registered on 911 in a developing world like the United States. An automated device that can track the crime [2] itself and warn the police and include intelligence about the crime scene and the potential perpetrators will be very useful in these situations. Such a device will collect and monitor the photographs of the crime scene and determine and advise the police accordingly.

Pre-trained Deep Learning models are models built to help individuals think about algorithms or test out established systems for improved outcomes without complex architecture. And, the deep learning neural network has five layers, namely, Convolution, Max-Pooling and completely linked layer input and output layers, and more parts address the roles of these layers. Due to certain scarce time, memory, and tools such as CPU, too often people use processors to use pre-trained Deep Learning concepts. And opposed to machine learning that lets one code directly, these pre-trained models would have the strongest and reliable outcomes. In surveillance footage, there is a lot of human interaction required to spot arms that appear to be human-prone mistakes.

2 Types of Crime

2.1 Detection of Fraud

A fraud is misdirecting or unfairly profiting from another. A fraud involves any act, exclusion, or concealment, including a breach of, or confidence in, a legal or equitable duty, causing damage to others. Cheque fraud, internet sales, insurance fraud, and credit card fraud are different types of fraud.

2.2 Violent Offence

A violent offence is a crime in which a guilty party threatens to use force on a victim. This includes, for instance, killing or rape, the two crimes of rough act called target. The following are various forms of this crime.

2.3 Traffic Ferocity

High traffic levels are induced by the rising amount of vehicles in towns, meaning that traffic offences are becoming more important, which can cause major road damage and further, injuries that can threaten people's lives. Traffic violation warning systems are required to solve this issue and deter certain effects.

2.4 Sexual Abuse

The chance or threat to violently strike a man is a criminal crime, giving no regard to what touch is ultimately made, insofar as the person understands the danger involved. Sexual harassment levels include:

- Simple Sexual Assault: It entails restricting an individual without unambiguous permission to engage in some form of sexual activity.
- Sexual Harassment with a Firearm: Integrates the usage or threat to an outsider with the use of a firearm or injury.
- Aggravated sexual assault: It occurs whenever, regardless of an attack, the survivor is genuinely hurt, mangled, fiercely battered, or in danger of moving on.

Verbal assault: It is a kind of non-physical, oral ambush that results in the target, instead of physical, significant damage in a passionate, social, and additionally mental pain.

2.5 Cyber-Abuse Abuse

Crimes that are performed by modern media communication networks, such as net and cell, against criminal prosecutions to damage the victims. Online extorting, ATM misrepresentation, wire misrepresentation, paper exchange and theft, intrusion, and so on are various forms. Cyber-crime research is a very serious duty to a country's law enforcement agency. This involves failure in security, or damage to the resources of the PC system, such as records, website pages, or programming.

3 Review of Literature

In recent years, [3] have performed numerous research on the estimation of crime events. This predictive power is built to help deter crime by encouraging the successful operation of police patrols. Data from various fields such as population, finance, and schooling were included in previous studies. Their projection models

handle data similarly from multiple domains. Such approaches have challenges in the estimation of crime events, such as trouble detecting extremely nonlinear associations, redundancies, and similarities between different datasets. This allows environmental background knowledge, such as broken windows theory and crime reduction through urban design, to develop crime prediction models. This paper suggests a feature-level data fusion approach based on a deep neural network (DNN) with environmental background in this article. The dataset is comprised of data gathered in Chicago, Illinois, from numerous online crime information sources, demographic and meteorological data, and photos. Authors pick crime-related data before producing training data by performing statistical analyses. Finally, they train their DNN, consisting of four styles of layers: geographical, temporal, environmental meaning, and layers of representation of joint features. The fusion DNN, combined with critical data derived from multiple contexts, is a result of an effective decision-making method that analyzes data consistency statistically. Experimental output findings suggest that the considered DNN model is more reliable than other prediction models in forecasting the frequency of violence.

In most visual-based monitoring software and protection devices, Target Detection is the main module. Photos and videos play a vital part in offering photographic images of an incident in crime scene investigation. It helps police officers to reconstruct a scene by identifying items relevant to a particular incident for further study. However, for law enforcement officers, the process of identifying items of concern is rather repetitive because of the availability of a vast amount of data. In this study, they present a real-time framework focused on Faster R-CNN (Region-based Convolutionary Neural Network), which automatically identifies artefacts that might be located in an indoor environment. They applied this to a subset of ImageNet comprising 12 entity groups and a Karina dataset to assess the feasibility of the proposed framework. They obtained an average precision of 74.33%, and the mean time taken to detect artefacts per image was 0.12 s [4].

It is easier to deter a crime from taking place than to discover whether or how the crime took place. Much like a vaccine is provided to an infant to avoid illness, with such an elevated crime incidence and brutal crime incidents in today's country, it has been important to provide a prevention scheme that stops crimes from arising. It relates to different strategies such as informing people, building awareness, raising efficacy, and constructive police methods and other preventive tactics through vaccinating population against crime. Driven by two separate current methods to crime prediction, the first offers a visual analytics framework that creates a proactive and analytical context for decision-makers to help them make better choices on resource distribution and implementation. The analysis of crime events relies primarily on the past background of crime and different geospatial and demographic data [1]. They may not take into consideration the rich and quickly expanding social and web networking background that affects events of interest, while it is exciting. Semantic analysis and natural language processing of Twitter posts via latent Dirichlet allocation is the next method, Subject Detection Sentiment Analysis [4, 5]. Yet, all approaches face intrinsic restrictions. Crimes that occur these days have main features

such as periodic replication of crimes, crimes that occur as a consequence of some other operation, and the incidence of crimes pre-indicated by any other results.

The fast and precise detection of criminal behavior is key to securing every home. The incorporation of crime prevention technologies aims to strengthen this protection with the exponential development of smart cities. In order to attain this purpose, a heavy emphasis has been imposed on traditional video monitoring in the past. Which also generates a video data backlog that a supervising official may control. This generates an increasingly heavy burden for supervising officials in broad metropolitan centers, contributing to a rise in the rate of error. To further reduce the workload, solutions have been introduced. At present, auto-regressive models have been used to accurately anticipate violent behavior, but they still have a number of limitations. This paper suggests a solution for processing video stream data by utilizing neural networks in conjunction with a Hybrid Deep Learning algorithm. The proposed technology would be able to detect and analyze illegal behavior efficiently, which would in turn reduce the supervising officials' workloads. It will make for an accurate and adaptable crime prevention framework as introduced via smart city technology [6].

This paper intends to expand the connexion between two images and identify the query picture in the source image by matching the features in the videos by providing a tool for identifying a specific individual or entity. Chowdhary and Rudra [7] the production would be frames matching the function object (not featuring its query) in a specified video. They define a way to use SIFT to locate special feature points in a picture or frame, i.e., the transform method of scale-invariant objects. SIFT is used to obtain distinctive feature points that are invariant to picture scaling or rotation, noise existence, picture lighting shifts, and so on. The picture is recorded for contrast with the feature points contained in the frames after the feature points are identified in a picture. For finding the appropriate query image in the picture, the feature points are compared using homography estimation scan. If the object is not present in the frame, then no output is present.

Crime acts impact our society in many detrimental respects, predicting the sites where potential incidents are most likely to arise will aid security departments across the globe greatly in avoiding such offences. This research introduces an innovative method focused on historical crime data to forecast potential crime hotspots. The sites where any crime has existed was compiled and translated into heat maps in grayscale and used to forecast potential hotspots. A deep learning methodology is used as it offers higher efficiency and quicker outcomes relative to conventional approaches, supplying police departments with real-time data to work more effectively. If adopted, the recommended proposal could be predicted to decrease crime rates in the future [8].

There is a need for the device to identify and forecast crimes at a volatile period owing to the escalation in the incidence of crimes. Noticeable outcomes from this analysis is that pre-processing becomes an essential activity as the dataset instances have a significant amount of missed values, and violence does not appear equally through metropolitan environments but concentrates in particular locations.

Predicting crime hotspots is also a very important activity, and the implementation of post-processing would therefore help reduce the incidence of crime [9].

Crime is one of our society's challenges, violence reduction needs the identification of aggression first, which involves a number of human capital, and can thus not be achieved immediately and omnipresently. It has become possible to diagnose aggression with limited human contact through the advancement of technology. In this article, the concepts and strategies for developing an automatic detection framework for aggression are discussed. The proposed device is capable of rendering the identification of aggression quick, immediate, and omnipresent. It was planned not only to detect aggression at the moment it happens, but also to forecast it for some period before, based on an individual's understanding of verbal and nonverbal signals. In addition, the machine can use numerous external data sources to get information about individuals and the environment and use it in the process of reasoning, thereby improving the accuracy of prediction [10].

Drug-related drug behavior in Taiwan is growing steadily and has a major and harmful social effect. This paper suggests a computer-driven methodology focused on the hypothesis and spatial analysis of "broken windows" to evaluate crime data utilizing software mining algorithms, and thereby forecast new crime hotspots for extra police action. In many areas, including image recognition and natural language processing, the Deep Learning algorithm has been widely implemented. They notice that the Deep Learning algorithm offers stronger predictive outcomes with fine tuning than other approaches for possible crime hotspots, including Random Forest and Naïve Bayes. In addition, by accumulating data with distinct time scales, try to increase model efficiency. This imagines possible crime hotspots on a map to test experimental findings, and examines if the models can classify real hotspots. Finally, they address the applicability of this strategy, and present potential avenues for science [11].

4 Study of Genetic Techniques for Crime Detection

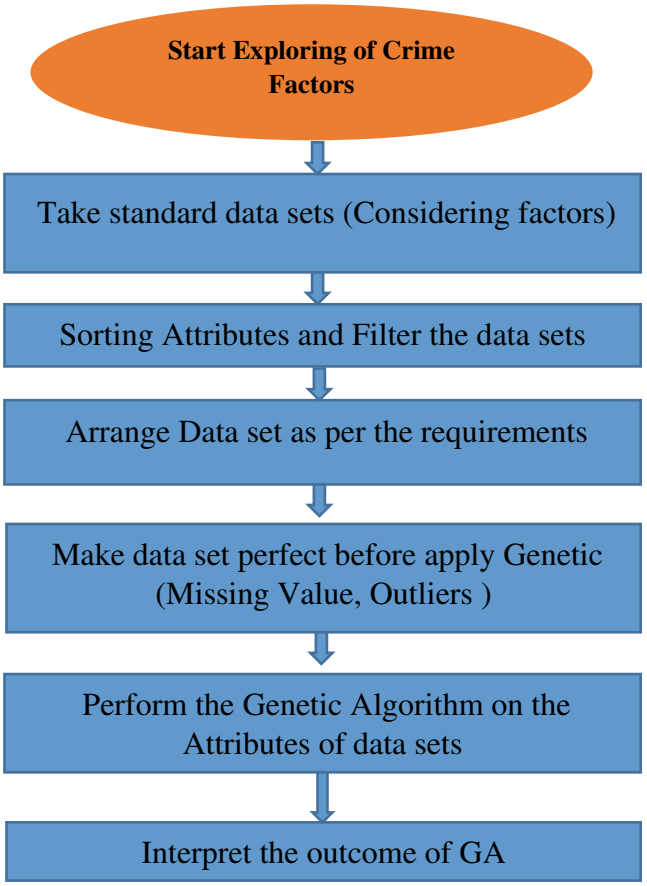
Different methods are used in fraud detection. Here, this paper explores the genetic algorithm in general, and how it works to detect the crime. This paper focuses on genetic algorithm with deep learning algorithm.

The genetic algorithm is a tool for interpreting alteration problems that are both obligatory and unconstrained. It is based on a mechanism of biological preference impersonating natural growth. At any iteration, it creates marks. An effective strategy addresses this. It is connected to the enormous class of revolutionary algorithms that build strategies that strengthen problems such as inheritance.

The following are few steps of this algorithm:

- Initial population implies the beginning of the human phase.
- Fitness feature defines the capacity of the organism to interact with others.
- The screening feature chooses the right user.

As the study of Genetic algorithm apply, the detection of crime draws the following steps.



As in the above flow chart, the procedure shows how the genetic algorithm could be applied over the datasets of crime. This is the way how the crime detection is made possible through genetic algorithms under the study of Deep learning mechanism.

5 Conclusion and Future Scope

There is a need for a system that can detect and anticipate crimes at an evolving moment, contributing to a rise in the rate of crime. The purpose of this study is to learn numerous deep learning techniques such as deep neural network and artificial neural network to identify and forecast crimes. The notable findings of this study show that pre-processing becomes an important task when the survey instances have

a large number of crimes that regularly occur in metropolitan areas. It is also a very valuable practice to forecast crime hotspots, and the introduction of post-processing will also help to reduce the crime incidence. This paper explores the studies on the diverse methods and strategies utilized to detect the crime. Therefore, for certain areas, crime hotspots ought to be predicted in order to assess which places are more vulnerable to crime and the type of crime. This study reveals how the techniques like deep learning algorithm play an important role to understand and detect crime.

References

1. Wang B, Zhang D, Zhang D, Brantingham PJ, Bertozzi AL (2017) Deep learning for real time crime forecasting. arXiv:1707.03340
2. Navalgund UV, Priyadharshini K (2018) Crime intention detection system using deep learning. In: 2018 international conference on circuits and systems in digital enterprise technology (ICCSDET). IEEE, pp 1–6
3. Kang H-W, Kang H-B (2017) Prediction of crime occurrence from multi-modal data using deep learning. PloS One 12(4):e0176244
4. Saikia S, Fidalgo E, Alegre E, Fernández-Robles L (2017) Object detection for crime scene evidence analysis using deep learning. In: International conference on image analysis and processing. Springer, Cham, pp 14–24
5. Azeez J, Aravindhar DJ (2015) Hybrid approach to crime prediction using deep learning. In: 2015 international conference on advances in computing, communications and informatics (ICACCI). IEEE, pp 1701–1710
6. Chackravarthy S, Schmitt S, Yang L (2018) Intelligent crime anomaly detection in smart cities using deep learning. In: 2018 IEEE 4th international conference on collaboration and internet computing (CIC). IEEE, pp 399–404
7. Chowdhary A, Rudra B (2020) Video surveillance for the crime detection using features. In International conference on advanced machine learning technologies and applications. Springer, Singapore, pp 61–71
8. Nair SN, Gopi ES (2020) Deep learning techniques for crime hotspot detection. In: Optimization in machine learning and applications. Springer, Singapore, pp 13–29
9. Bhardwaj AS, Divakar KM, Ashini KA, Devishree DS, Younis SM (2019) Deep learning architectures for crime occurrence detection and prediction
10. Dorogyy Y, Kolisnichenko V, Levchenko K (2018) Violent crime detection system. In: 2018 IEEE 13th international scientific and technical conference on computer sciences and information technologies (CSIT), vol. 1. IEEE, pp 352–355
11. Lin Y-L, Chen T-Y, Yu L-C (2017) Using machine learning to assist crime prevention. In: 2017 6th IIAI international congress on advanced applied informatics (IIAI-AAI). IEEE, pp 1029–1030

Smart Web Investigation Framework

Gundavarapu Mallikarjuna Rao, B. Ramakrishna Reddy, and Peesapati Vishnu

Abstract In this paper, we have developed a framework for web scraping and text analysis. Most of the real-world applications require text analysis and web scraping; however, due to their disjoint behavior, there is no unified frame work that will take care of both needs. In the proposed scalable, portable framework, a centralized toolkit is designed for scraping and text analysis. Further, it allows the analysis of large data in one go and provides inferences about changes in web-based text over time. The framework will give inferences about web-based topics as well as summarize the data in a crisp format which will make it easy for the user to analyze the text-based data of their choosing from the internet. The proposed framework is very much useful when analyzing web-based reviews of various products, entertainment, and the news about a specific topic in www.

Keywords Lemmatization and stemming · TF-IDF score · Latent semantic analysis

1 Introduction

Huge amounts of data are posted on the World Wide Web related to socio, technical, or entertainment. However, the trustworthiness of this data is always questionable [1, 2]. The customer is in a better position to judge the correctness of an interesting topic if they get more elaborative facts on a specific topic. The problem here is how to get them as the information shown by most sites must be seen utilizing an internet browser. They do not give the usefulness of sparing a duplicate of this information for individual use. At that point, the main decision is to physically reorder the information which is an exhausting activity that can take a few hours to finish or even days. Web scratching will be handy here; it is a robotic-based technique, hence avoid physically duplicating the information from the respective sites.

G. Mallikarjuna Rao (✉) · B. Ramakrishna Reddy · P. Vishnu
Gokaraju Rangaraju Institute of Engineering and Technology, Bachupally, Hyderabad, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_27

1.1 Web Scraping

Web Scraping is a method used to extract a large amount of data from websites. This extracted data from the websites is usually stored in a local file in your personal computer (or) in a table format in a database [3, 4]. This process is performed with the help of a bot (or) a web crawler. Some of the techniques used are ae follows:

- **Copy and Paste Manually**: Most of this activity can be automated except the cases where webpages set up blocks to prevent the web crawlers from scrapping the page. Here the only alternative is human involvement.
- **Comparing the Text Pattern**: The UNIX grep command, matching regular expressions in programming languages like Perl or Python, can be a robust idea to scrap the information from the web pages.
- **Socket Programming**: Using HTTP programming, we can retrieve data from static and dynamic webpages by posting HTTP requests to the remote web server.
- **HTML parsing**: Wrapper allows the collection of the bulk data from dynamic/static websites and translates them into relational form. Wrapper program assumes that the input from web pages can be associated with URL schemes which can be used to parse HTML pages and then extract and transform the page content.
- **Document Object Model parsing**: Web browser, like Safari (or) Google Chrome, we can extract the information which is dynamically generated by the client-side scripts. These browsers also parse the web pages into a DOM tree which helps us to extract information from parts of the pages too.
- **Vertical percentage aggregation**: Many companies have devised vertical harvesting platforms. This platform is used to automate the creation of bots with the available knowledge base. The quality of the platform is calculated with the help of the information it will get and the number of the sites it extracted.
- **Recognizing semantic annotation**: The scraped pages can also contain metadata or semantic mark-ups and annotations, which can be used to discover unique snippets of statistics.
- **Computer vision web-page analysis**: There are efforts using machine gaining knowledge of and laptop imaginative and prescient that try to become aware of and extract facts from net pages by deciphering pages visually as an individual might.

1.2 Text Analysis

Data recovery, lexical appraisal to check word recurrence conveyances, test acknowledgment, labeling/comment, measurement extraction, information mining techniques which incorporate connection and affiliation assessment, representation, and perceptive examination are associated with content examination [5, 6]. Content examination ordinarily incorporates the way toward organizing the info's literary substance

(typically sifting, including some inferred phonetic highlights, evacuating others, and resulting consolidation into a database), determining designs inside existing information, lastly separating and deciphering the subtleties. In-Text audit, 'High caliber' typically alludes to a couple of blends of significance, oddity, and intrigue.

1.3 Natural Language Processing

Content pre-handling is customarily an indispensable advance for common language preparing (NLP) undertakings [7]. It changes content into a more noteworthy edible structure with the goal that the device considering estimations can complete better. In NLP, textual content pre-processing is the primary step within the process of building a model.

The various text pre-processing steps are as follows:

- Regular expressions.
- Stop words removal.
- Stemming.
- Lemmatization.
- Tokenization of text.

1.4 Text Classification

This can be performed either manually (accomplished with a human agent's effort reading and categorizing texts) or automatically (including machine learning techniques and algorithms to identify the texts more easily and cost-effectively). Companies receive text facts all the time. Be it emails, chats, social media comments, support tickets, or NPS responses, all these texts are very wealthy sources of information. However, they are not structured, so you need to tag and examine those texts before you can make sense of them and acquire insights.

2 Proposed Frame Work

2.1 Pre-Processing

The retrieved HTML document contains many HTML tags. Using regular language expressions, the HTML document is cleaned of all tags. The content that is meant to be analyzed from the website is generally present in tags or a variation of or tags. The necessary text is then extracted from those tags. This text is then pre-processed

through the removal of stop words, converting all words to lowercase, lemmatization, and stemming. The stops words are generally considered to be conjunctions, prepositions, and determiners. These words are removed. Then lemmatization is done in order to trace the word to its root by comparing the word with a dictionary and then removing its inflectional endings. The stemming is done in order to remove all the word's suffixes and prefixes. This results in only leaving the important parts of the text in the document (Fig. 1).

2.2 Working of the Web-Scraper

The web-scraper is written completely in C. We have written a web-scraping library in C completely from scratch. The library uses Linux socket libraries and the open-source Open SSL library. The web-scraper takes the website name as an input It then sends a DNS request to retrieve the IP address of the server of the website. The scraper initiates a TCP connection with the server of the website. The SSL library is used to create a secure SSL connection on top of the TCP connection. The request is sent to the server, therefore, downloading the entire HTML file of the entire website. The SSL connection is then closed, and the TCP connection is closed next. Figure 2 denotes the web-scraping process.

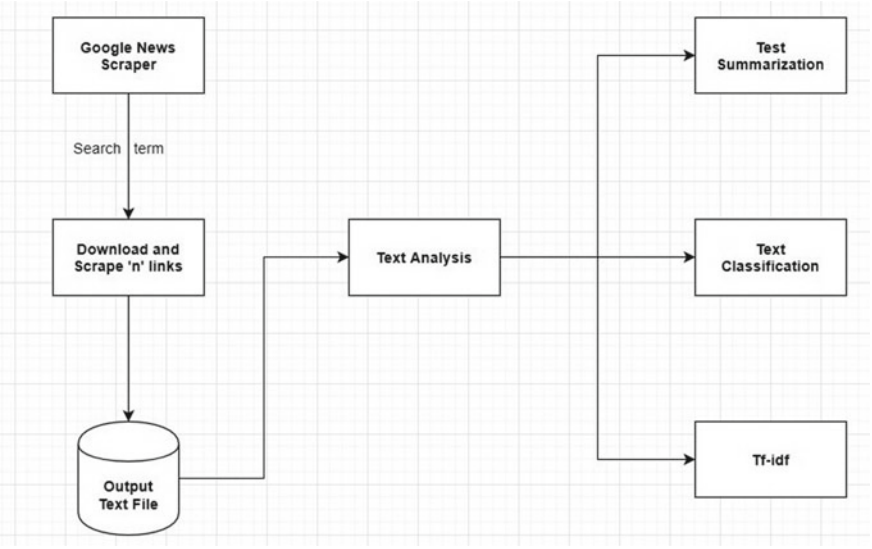
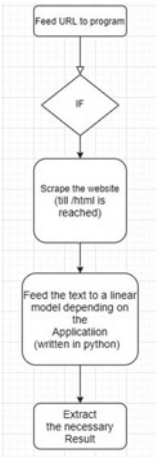


Fig. 1 Proposed framework

Fig. 2 Web scraping

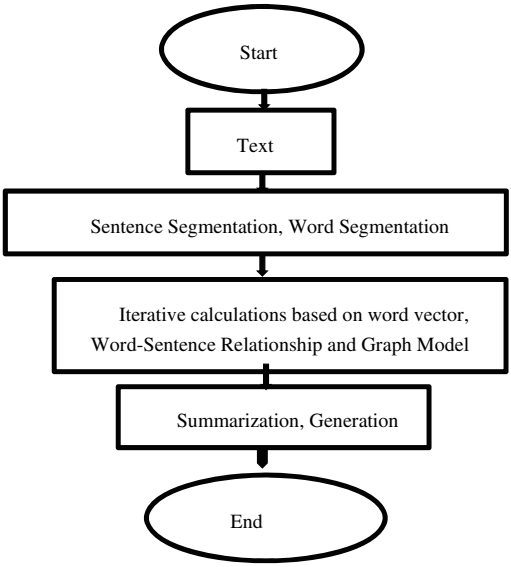


2.3 Text Summarization

The project uses extractive text summary techniques to summarize the document. The flow of the summarization is shown in Fig. 3.

The document is cleaned and pre-processed using regular language expressions which denote the HTML tags as well as punctuation marks. The document is divided into multiple sentences and is then stored in a list of lists format in the program. The sentences are divided into word-based tokens. The term frequencies of each

Fig. 3 Process flow



word are stored in a dictionary with the word being the key and its term frequency being the value. This dictionary of words and their term-frequency is compared to the sentences within the scraped text. Each sentence is then scored by calculating the sum of the term-frequency of the words it holds. Then the respective sentences and their words are stored in a dictionary with the sentence being the keys and their scores being values of the dictionary. An issue with this approach is that longer sentences tend to have bigger scores. To deal with this problem, the lengths of the sentences are fitted to a normal distribution. Then over the normal distribution, only a range of few numbers is considered for summarizing. Depending on the number of words and sentences you want the summarized text to hold, the sentences in the dictionary are selected and outputted together. This makes up the text summary module.

2.4 Sentiment Analysis

The document is cleaned and pre-processed using regular language expressions which denote the HTML tags as well as punctuation marks. The stop words are then removed from the document. The words in the document are put through lemmatization. Using the one-hot encoding, the document is converted into a vector. We have previously created a dataset using news articles and trained on supporting and criticizing news articles and reviews. We have trained a linear regression model on the dataset after one-hot encoding the dataset. The Fig. 3, Flowchart of the process of summarization cleaned and pre-processed document which was retrieved through the scraper is then fed to the model which will then classify the document as being supportive or criticizing.

2.5 TF-IDF Scoring

The document is cleaned and pre-processed using regular language expressions which denote the HTML tags as well as punctuation marks. The stop words are generally considered to be determiners, conjunctions, and prepositions. These occur highly in any kind of document and hold a high term frequency score, but they are not necessary to the document. These are then removed from the document. The words in the document are put through lemmatization. The lemmatization function then scans each term and traces it to the root term of the word effectively removing prefixes and suffixes. The term-frequency of each term is calculated. Depending on the term-frequency of each word in the document and the number of documents retrieved by the scraper, the TF-IDF score is calculated and displayed using

$$W_{i,j} = tf_{i,j} * \log\left({}^{N}/_{df_i}\right), \tag{1}$$

where $tf_{i,j}$ denote the number of occurrences of i in j, df_i is the number of documents containing i, and N is the total number of documents.

2.6 Topic Modeling

Topic Modeling tries to push the prominent words present in the corpus into corresponding topics. We are achieving this using Latent Semantic Analysis or LSA for short. It fits all the words in the document after pre-processing into a Bag of Words Model. After converting the text into a bag of words model, it then applies matrix decomposition on the bag of words model. We used Singular Value Decomposition to break down the bag of words matrix and extracted the corresponding topic—words matrix. We then chose the word with the highest term frequency in the topic as the name of the topic.

3 Experimentation

Python technology: A translator is a type of programming program that executes elective projects. At the point when we compose the Python programs, it changes over source code composed with the guide of the engineer into the middle language that is fairly converted into the local language/contraption language with the expectation to be executed [4] (Fig. 4). Figure 5 denotes the compilation and interpreting process.

Experimentation is performed using corei3 processor of 3 Ghz using Anaconda (Python 3.8), Linux Shell/Windows Command-Prompt, and python-requests, pip version 20.0.3.

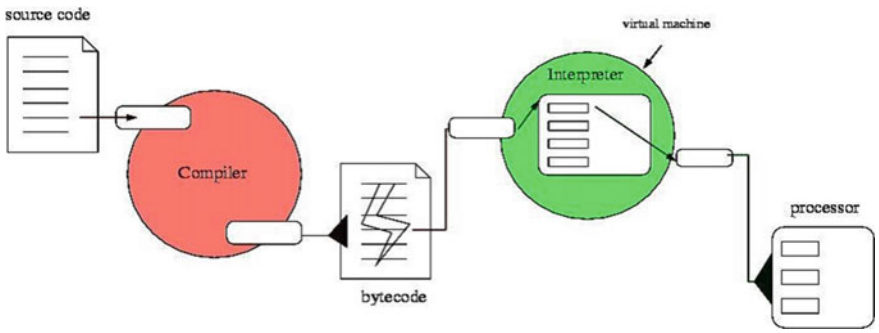


Fig. 4 Object code generation process

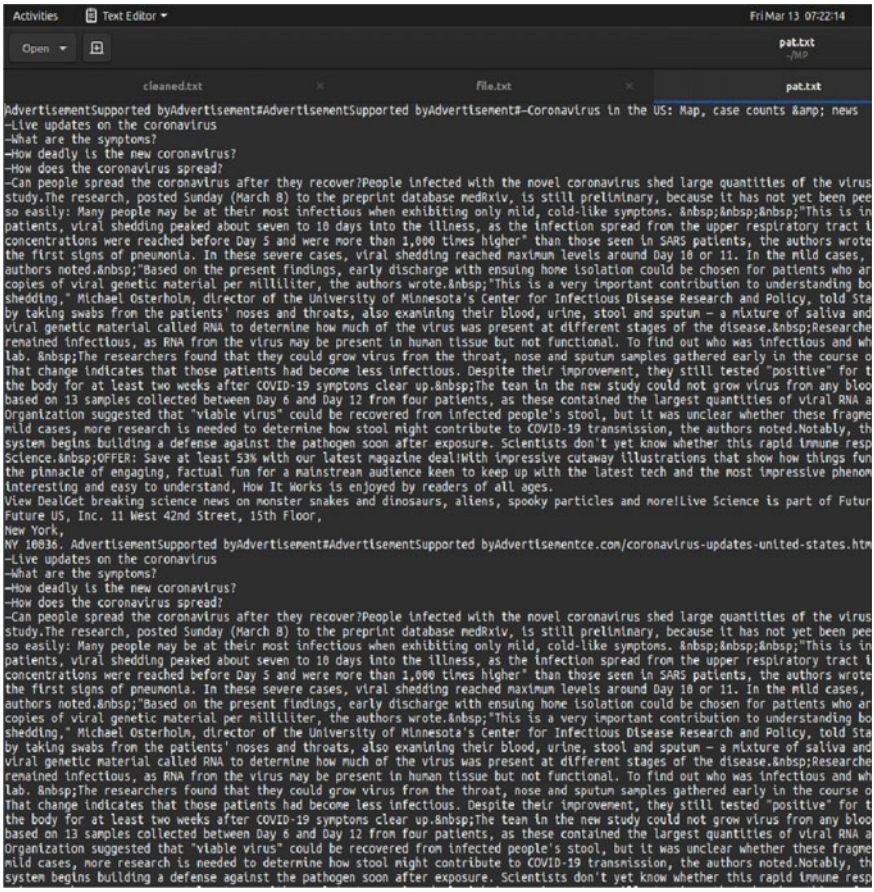


Fig. 5 Web-scraping on keyword corona or Covid-19 (Zoom appropriately)

4 Results and Conclusions

Web Scraping is achieved by connecting to the server of the website, establishing SSL connection, and downloading the web page. Sub-string search algorithms are used to isolate intended elements. Preliminary analysis is required before scraping the website. Depending on the application, the model must be trained on the required datasets to perform specific analysis such as sentiment analysis and cosine similarity. For sentiment analysis, we are using linear classifiers such as linear regression. Using the tools, we have written, and we have constructed a Google news scraper to analyze news reports [8–10]. The architecture designed for the Google news scraper using tools from our framework is represented in Fig. 5.

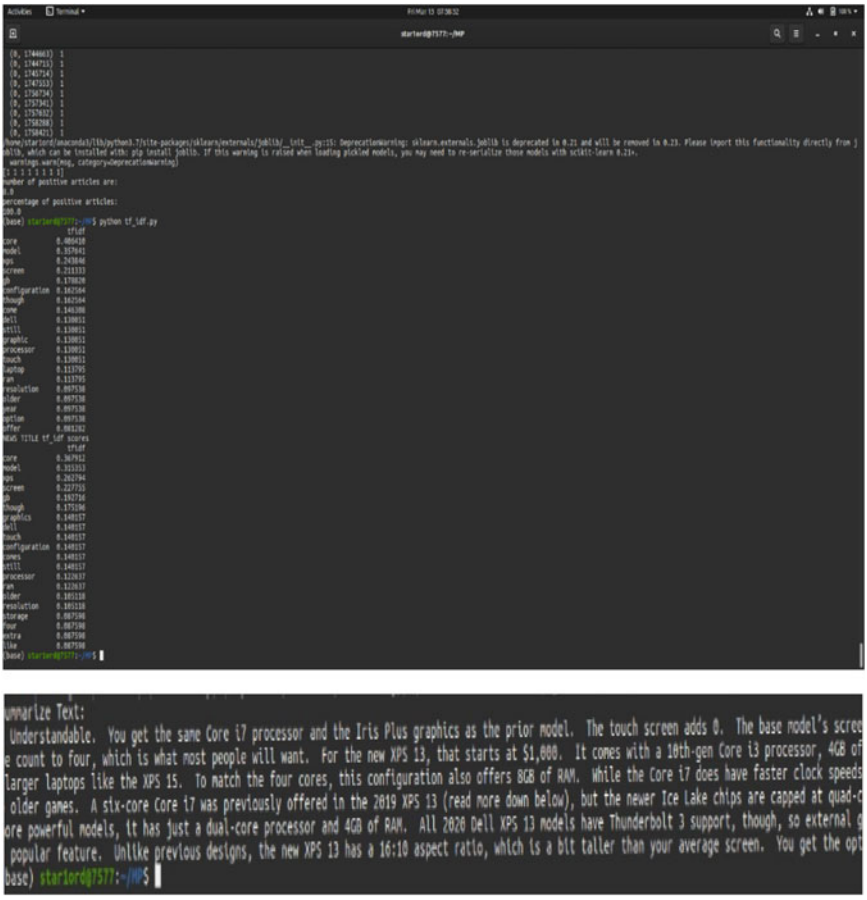


Fig. 6 **a** TF-Idf scores for google news corona, **b** Summarization

We have run the tool on the search 'Dell XPS i3 Laptop Reviews'. Inferring from the sentiments from the results, we got a positive articles percentage of 80. The TF-IDF scores for the results from Google News on Corona are shown in Fig. 6a, and a summarization of the same is shown in Fig. 6b.

References

1. Mäntylä MV, Graziotin D, Kuutila M (2018) The evolution of sentiment analysis—a review of research topics, venues, and top cited papers. Comput. Sci. Rev. 27:16–32. ISSN 1574-0137. https://doi.org/10.1016/j.cosrev.2017.10.002
2. Sree PD, Rao GM (2019) A research on passive forgery detection. Int J Recent Technol Eng (IJRTE) -8(2S11). ISSN: 2277-3878

3. Herrmann M, Hoyden L (2016) Applied webscraping in market research. In: First international conference on advanced research methods and analytics, CARMA2016
4. Hassanien HED (2019) Web scraping scientific repositories for augmented relevant literature search using CRISP-DM. Appl Syst Innov 2(4):37. https://doi.org/10.3390/asi2040037
5. Nassirtoussi AK, Aghabozorgi S, Wah TY, Ngo DCL (2014) Text mining for market prediction: a systematic review. Expert Syst Appl 41(16):7653–7670
6. Adriaanse LS, Rensleigh C (2013) Web of science, scopus and google scholar: a content comprehensiveness comparison. Electron Libr 31(6):727–744
7. Bird S, Klein E, Loper E (2009) Natural language processing with python, 1st edn. O'Reilly Media, Inc.
8. Richard Stevens W (1990) UNIX network programming. PrenticeHall Inc., USA
9. The OpenSSL Project (2003) OpenSSL: the open source toolkit for SSL/TLS. www.openssl.org
10. OpenSSL Foundation, I (2020) /Docs/Index.Html. Openssl.org. https://www.openssl.org/docs/. Accessed 9 May 2020

Detection of Driver Distraction Using Convolutional Neural Network

R. Sumalatha, Ch. Sravani, and M. Supriya

Abstract Driver distraction is a major factor in vehicle accidents. Day by day the percentage of vehicle crashes is rising due to several activities performed by the driver. Many researchers have developed various algorithms to detect driver distraction. In this paper, an algorithm is proposed to detect driver distraction in real time. To evaluate the proposed algorithm, the dataset is used which was provided by State Farm through a Kaggle competition. The predefined convolutional neural network (CNN) namely resnet50 is used to extract the features of testing images and training images. The extracted features are fed to a Support vector machine (SVM) to classify the normal images and distracted images. The proposed approach classification accuracy achieved 84% for normal images and 87% for distracted images. A real time hardware model is developed by using Arduino Uno to detect the driver distraction. The output hardware modules such as LED, LCD display, and buzzer are used to alert the driver.

Keywords Driver distraction · CNN · SVM · Arduino uno

1 Introduction

Nowadays major accidents are occurred due to driver lack of concentration and distraction. In general, drivers are engaged with operating sound system usage of mobile, drinking water, and chatting with passengers. Long hours of driving cause the driver tiredness and, consequently, reduce his/her concentration on driving. Driver distractions not only do the accidents and also reason to put other life's in danger. Hence driver drowsiness and distraction are very important factors in accidents. Most of the researchers proposed algorithms to detect the condition of an eye, mouth, face, and yawning for alerting the driver safety.

R. Sumalatha (✉) · Ch. Sravani · M. Supriya
Electronics and Communication Engineering, Vardhaman College of Engineering, Hyderabad, Telengana, India
e-mail: r.sumalatha@vardhaman.org

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_28

2 Related Work

Ahlström et al. discussed various algorithms for driver distraction in real time [1]. Eraqi et al. have proposed convolutional neural networks to detect face, hands, and skin, and also weights are calculated via a genetic algorithm [2]. Fan et al. have used Gabor wavelet transform to extract text features from the mouth in order to detect the yawning for monitoring driver fatigue [3]. Fletcher et al. developed a proto-type system for estimating driver observations and detecting drivers' distraction based on eye gaze [4]. Ji et al. implemented a proto-type model for monitoring driver vigilance with various computer vision algorithms and under different illumination conditions, genders, and ages also [5]. Kutila et al. have proposed a system to detect the drivers by combining stereo image and lane tracking images with rule-based and support vector machine (SVM) classification methods [6]. D'Orazio et al. have proposed a neural classifier algorithm to recognize the eyes from an image with various eye colors, glasses, and different light environments [7]. Tango et al. presented various classification methods for driver distraction based on machine learning techniques [8]. Zeng et al. have discussed the strategy of attention budget for effective driver distraction detection [9].

3 Proposed System

The proposed model is as shown in Fig. 1. It has four phases as dataset preparation and splitting, preprocessing, feature extraction through CNN, and classification.

- The database consists of two classes; one is normal images, and another one is distracted images.
- The database images are split into training and testing images. The training and testing images features are extracted using resnt50 pretrained CNN features are network.

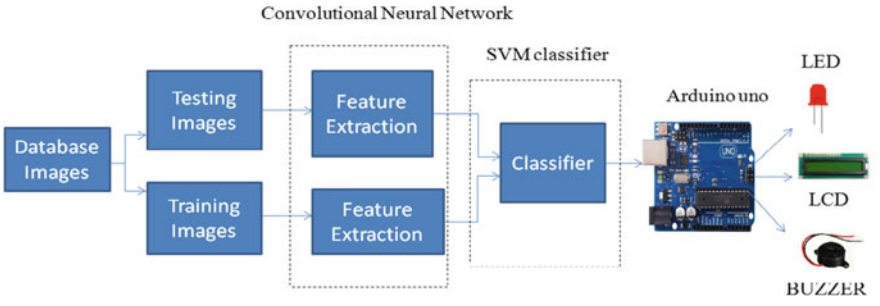


Fig. 1 Proposed block diagram

- By using the SVM classifier, the driver condition is classified as normal or distracted.
- The classified output is sent to the hardware module, and the output will be displayed in the display unit.

3.1 Dataset

The dataset used in this paper was provided by State Farm through a Kaggle competition [10]. This dataset consists of 10 classes; one class belongs to safe driving images, and the other nine classes are related to distracted images with different activities like texting, talking on phone, operating a sound system, drinking, eating, and chatting with passengers and so on. In this paper, we consider only two classes; one is normal, and another nine classes together are considered as distracted images.

3.2 Feature Extraction

In this paper, the Resnet50 CNN model is used to classify driver status. Figure 2 shows different distracted driving actions. The architecture of Resnet50 CNN consists of convolutional, pooling, Rectified Linear Unit (ReLU), and fully connected layers. The Resnet50 architecture is shown in Fig. 3.

Fig. 2 Ten normal and distracted drivers images

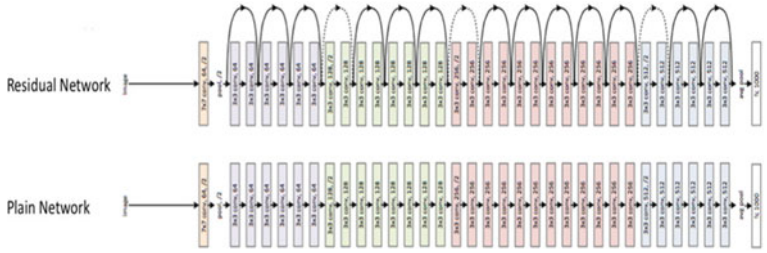


Fig. 3 Resnet50 architecture

Fig. 4 The convolution layer

3.2.1 Convolutional Layer

The input color images with the size 640 × 480 × 3 are fed to the convolutional layer. This layer extracts the features from input images. The first convolutional layer extracts the low-level features and the next layers extract the middle and high-level features from the images. The first layer performs a convolution operation between the input image and filter to produce a feature map. This feature map is given as input for the next layer (Fig. 4).

3.2.2 Pooling Layer

The pooling layer reduces the dimensionality of the feature map. In this category, there are two layer options:

1. Max pooling.
2. Average pooling.

Max pooling finds the maximum value from the part of an image enclosed by the filter, and an average pooling computes the average of all the values from the part of an image enclosed by the filter.

3.2.3 ReLU Layer

ReLU is a non-linear operation. This layer removes every negative value from the filtered images and replaces it with zero's.

$$\begin{aligned} f(x) &= x \quad if \quad x > 0 \\ &= 0 \quad if \quad x < 0 \end{aligned} \tag{1}$$

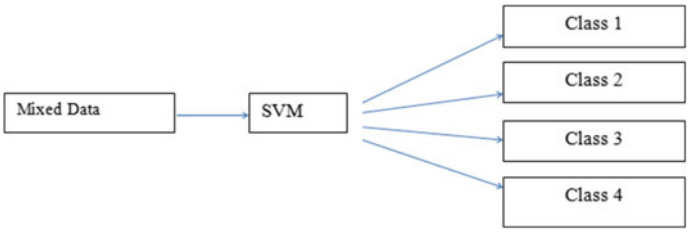


Fig. 5 SVM classifier

3.2.4 Fully Connected Layer

A Fully Connected layer (FC) establishes the connection between each filter in the earlier layer to every filter in the subsequent layer. The FC layer provides the feature map to the softmax activation function for classification.

3.2.5 Softmax Activation Function

The softmax activation function is used to get probabilities of the input being. Finally, the obtained probabilities of the object in the image belonging to the different classes.

$$f_i(\mathrm{z}) = \frac{e^{z_i}}{\sum_k e^{z_k}}, \tag{2}$$

where $z_i = i$th input.
$k =$ no of inputs.

3.3 SVM Classifier

SVM is used to classify the driver distraction condition. It uses the hyper plane to classify images. The multi-class classification using SVM is as shown in Fig. 5.

3.4 Arduino Uno

Arduino/Genuino Uno is a microcontroller board based on the ATmega328P as shown in Fig. 6. The board consists of digital and analog input-output pins. In this paper, those pins are used to interface with LED, LCD, and buzzer.

Fig. 6 Arduino Uno board

Fig. 7 Light-emitting diodes in various colors

3.5 Light Emitting Diode

In this paper, LED (Light Emitting Diode) is used for indication purposes to drivers. It is a small light-emitting semiconductor component. It consists of two terminals; both are cathode and anode. Figure 7 shows different colors of LED.

3.6 Liquid Crystal Display (LCD)

An LCD is used for the display of driver status. In this paper, 16 × 2 LCD is used for display purposes. The 16 × 2 LCD is shown in Fig. 8.

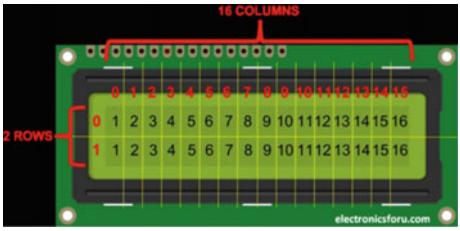


Fig. 8 16 × 2 LCD display

Fig. 9 Buzzer

3.7 Buzzer

A **buzzer** is used to alert the driver if he/she is distracted from driving. A buzzer will give continuous beeep. sound if the driver is distracted. The buzzer is connected to Arduino Uno board. The basic buzzer used in this paper is as shown in Fig. 9.

4 Results and Discussions

Figure 10 shows when the input image is a non-distracted image (Normal), and then the LED does not glow and LCD displays "NORMAL". Figure 11 shows when the input image is a distracted image, and then the LED glows and LCD displays "DISTRACTED". Table 1 gives the information about the accuracy in two different conditions. Tables 2 and 3 are the confusion matrix in two cases.

Fig. 10 Driver is in normal condition

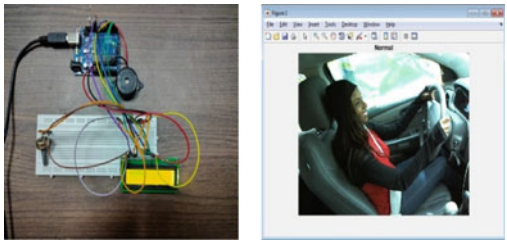


Fig. 11 Driver is in distracted condition

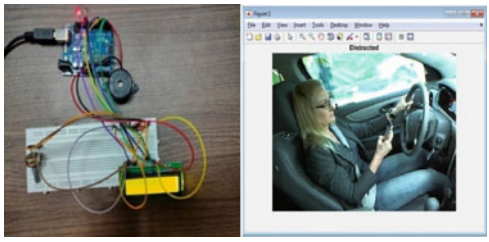


Table 1 Image classification accuracy

Image	Accuracy (%)
Normal Image (C0)	84
Distracted Image (C1)	87

Table 2 Confusion Matrix of the normal input image

	C0	C1
C0	0.7397	0.2603
C1	0.0476	0.9524

Table 3 Confusion Matrix of the distracted input image

	C0	C1
C0	0.9111	0.0889
C1	0.1524	0.8176

5 Conclusion

In worldwide, a major number of accidents are occurred due to driver distraction. To decrease the accidents, it is essential to recognize the main causes of it. In this paper, the resnet50 pretrained network is used to extract features from images. The proposed system efficiently detects the driver distraction condition using the resnet50 CNN model. The proposed system detects the driver condition in the normal case with 84% accuracy and the distracted condition with 87% accuracy.

References

1. Ahlström C, Kircher K (2010) Review of real-time visual driver distraction detection algorithms. In: Proceedings of the 7th international conference on methods and techniques in behavioral research, vol 2 pp 1–4
2. Eraqi HM, Abouelnaga Y, Saad MH, Moustafa MN (2019) Driver distraction identification with an ensemble of convolutional neural networks. J Adv Transp 1–12
3. Xiao F, Yin B-C, Sun Y-F (2007) Yawning detection for monitoring driver fatigue. In: Proceedings of the sixth international conference on machine learning and cybernetics
4. Fletcher L, Zelinsky A (2009) Driver inattention detection based on eye gaze–road event correlation. Int J Robot Res 28(6):774–801
5. Ji Q, Yang X (2002) Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. Real-Time Imaging 8:357–377
6. Kutila M, Jokela M, Markkula G, Rue MR (2007) Driver distraction detection with a camera vision, pp VI-201–VI-204
7. D'Orazioa T, Leoa M, Guaragnellab C, Distante A (2007) A visual approach for driver inattention detection. Pattern Recognit Soc 40:2341–2355
8. Tango F, Botta M (2013) Real-time detection system of driver distraction using machine learning. IEEE Trans Intell Transp Syst 14(2):894–905

9. Zeng J, Sun Y, Jiang L (2010) Driver distraction detection and identity recognition in real-time. In: 2010 second WRI global congress on intelligent systems. Wuhan, pp 43–46
10. Sultan I (2016) Academic purposes. https://www.kaggle.com/c/state-farm-distracted-driver-detection/discussion/20043{#} 114916

NDN Content Poisoning Attack Mitigation Using Fuzzy-Reputation Based Trust

Vishwa Pratap Singh and R. L. Ujjwal

Abstract Content poisoning attack is a primary cause of denial of service attack in current TCP/IP Internet architecture. It is affecting one of the proposed content-oriented future Internet architecture known as Named Data Network (NDN). Unlike TCP/IP networks, where an attacker attempts to poison DNS Content, the attacker targets NDN routers, where the Content is cached to fulfil future requests for the same Content. NDN Content poisoning attacks can be mitigated by means of signature authentication. NDN suggests signature verification at consumer nodes instead of at routers as signature verification is a resource-consuming process. Trust-based schemes are recognized to detect various types of attacks as they are crisp, strong, and computationally inexpensive. In this article, we offer an outline of reputation-based trust approaches, address their architecture tradeoffs, and conclude that by operating in addition to current credential-based systems. We have also proposed a mechanism to mitigate NDN content poisoning attacks using the Fuzzy-Reputation based trust model.

Keywords Named Data Network · Content poisoning attack · NDN trust

1 Introduction

Content has become the most important on the Internet. Nowadays, consumers are just concerned about Content regardless of where it is stored or originated. The disparity between existing TCP/IP Internet architecture that operates as a point-to-point network and consumer specifications motivated scientists to come up with a new Internet architecture in which Content is of the highest importance [1]. In 2010, NDN was proposed as a potential future content-oriented Internet architecture [2]. NDN assigns unique hierarchal names to the Content, and users can access data using names

V. P. Singh (✉) · R. L. Ujjwal
University School of Information Communication and Technology, Guru Gobind Singh Indraprastha University, Delhi, India

R. L. Ujjwal
e-mail: ujjwal@ipu.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_29

instead of location, unlike IP networks. There are two types of packets in NDN, i.e., Interest packet (*I_Packets)* and Data packet (*D_Packet*). Users request content using *I_Packets*, and the user gets data in the form *D_Packet*. Caching is one of the main features of NDN, and NDN routers have the ability to cache Content. Data passing through routers can be stored in a way that goes from user to customer, and the cached content is used to satisfy subsequent demands for the same Content. NDN routers have three types of Data structures, namely, Pending Interest Table (PIT), Content store (CS), and Forwarding Information Base (FIB). Whenever the consumer sends interest packets to the NDN router (Fig. 1), it first checks the availability of the Content in its content store. If Data is not available in CS, Routers search in their PIT for previous entries for the same Content. If the previous entry is available, then Router inserts an incoming Interest interface, and if the entry is not available, then creates a new entry in PIT and forwards the packet to the next routers using FIB. Various routing protocols fill the FIB. It is one of the essential features of NDN that separates routing from forwarding. Separation of routing and forwarding allows developers to design different protocols for routing and forwarding.

The next forwarding node is identified using the longest prefix match performed at FIB. Data packets are returned by the producer to the customer by tracing back incoming paths while travelling from the producer to the user, data is cached by in-path routers. Pervasive caching enables efficient content distribution and decreases latency. NDN has also introduced a new layer "Security" to its architecture, and all security functions are managed in this layer. NDN do not secure communication

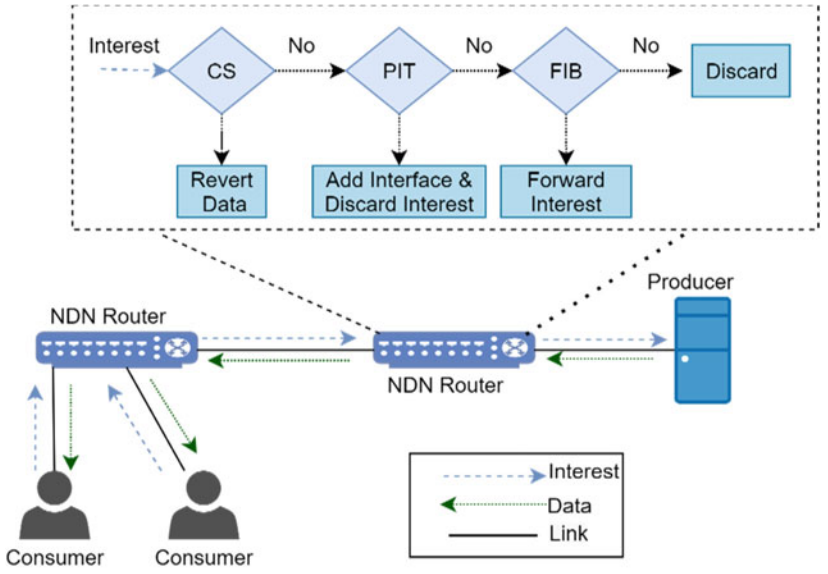


Fig. 1 NDN interest processing

channel instead, the philosophy is to secure the Content. NDN consumers concatenate digital signature with the Content that can be verified by in-path routers or consumers. The producer generates a public–private key pair using a specified key generation algorithm, and the private key is used to digitally sign the producer's Content. NDNSim [3] supports Signature HMAC With Sha256, Signature Sha256 With ECDSA, Signature Sha256 With RSA, and DigestSha256. SNDN digital signatures include the entire Content (Content, Name, MetaInfo, SignatureInfo) specified in Signature Value. Signature verification is the same as to the systems in which the router or consumer can request a public key from a network, which he gets as a special Data packet. In the entire process, it is necessary to authenticate the public key itself, which can be done by key certificates of trust models. NDN uses PKI (Public key infrastructure) [4] to build Trust. The verification process prevents bad or corrupted content from being cached in the router, but signature verification is tidies and resource-consuming process. Each piece of data is encrypted in NDN, which provides a higher degree of protection. The NDN architecture can avoid most of the security attacks that are common in the TCP/IP Internet, but its design has also introduced some new types of security attacks related to caches like cache pollution attacks, content poisoning attacks, denial of service attacks, and various new attacks on privacy like timing attack.

Pervasive caching has also introduced vulnerabilities that lead to denial of service attack and privacy attack [5] in NDN. There are primarily three types of DoS in NDN, Interest flooding, Cache pollution attack, and Cache poisoning attack. In a cache pollution attack, the attacker attempts to insert non-popular Content in the CS, which reduces the cache hit ratio. CPA is of two types Location disruption attack (LDA) and false locality attack (FLA) [6]. In LDA, the attacker repeatedly requests different non-popular Content that squeezes out popular Content from the CS and cache non-popular content. In FLA, the attacker repeatedly requests a set of non-popular Content and hence creates false popularity for the unpopular Content. In Interest flooding attack, the attacker floods the network with random Interest which consumes routers resources. In Content poisoning attack (Fig. 2), the motive of the attacker is to inject corrupt/poisoned Content with a valid signature into the router's cache. Corrupted data can be inserted by (a) compromised router, (b) Malicious provider, (c) collaboration of Bad provider, and consumer (Fig. 3). Content poisoning can inject two types of Content into the router (a) Corrupted Content (b) fake Content. Corrupted Content with a valid signature can be easily identified by signature verification, but in fake Content, a lousy router with fake Content and valid signature cannot be identified easily. To avoid the dissemination of poisoned Content across the network, a digital signature is inserted in every piece of Content and validated either by routers or by end-hosts. However, because the authentication process requires a high overhead, it is not practical for routers to validate all Content at the line rate.

In this paper, we focused on mitigating content poisoning attacks. In the next section, we have described various reputation-based trust approaches, address their architecture tradeoffs, and conclude by operating in addition to current credential-based systems, and these approaches will improve trust and security in NDN. In

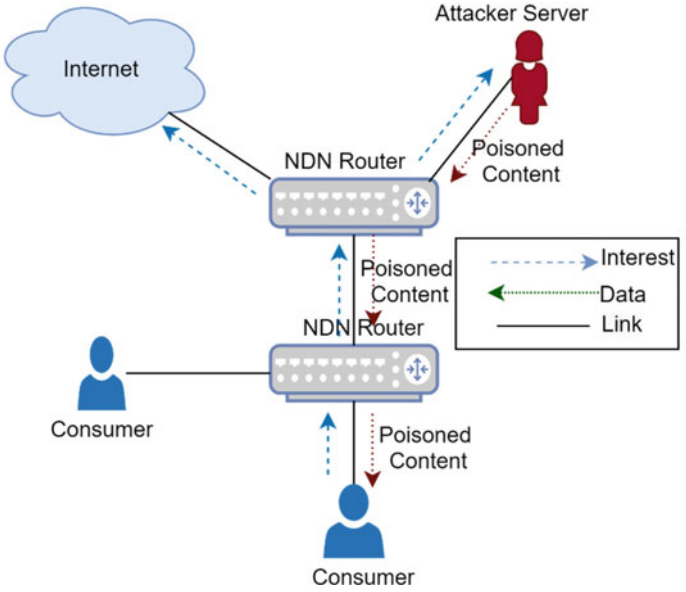


Fig. 2 Content poisoning attack

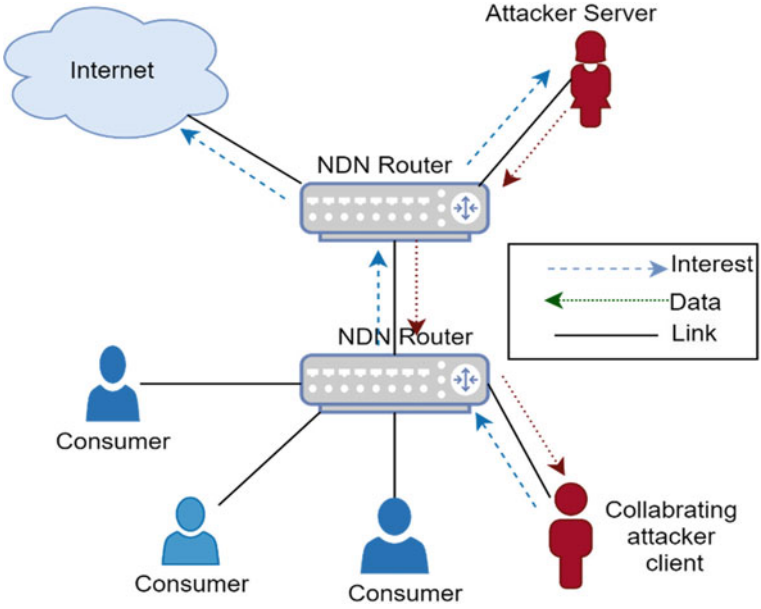


Fig. 3 Collaborative content poisoning attack

Sect. 2, we have discussed previously proposed reputations-based systems and explained why they apply to NDN. Efforts made by researchers for detection and mitigating a Content poisoning attack. In Sect. 3, we have proposed a Fuzzy-Reputation Based Trust Model and discussed the effectiveness of the proposed model using various logical scenarios. We concluded our paper in Sect. 4 and provided a discussion on results and remarks.

2 Reputation Based Trust Models

Reputation-based trust models use "trust values" given to the service or products by users in a dispersed way and used to provide ratings or rankings. Trust values from various interacting partners are aggregated and further used to provide feedback to the system. In this section, we have various Trust-based systems proposed to use in NDN and their implications. The attacker takes advantage of the absence of in-network signature verification to launch a cache poisoning attack [7]. NDN routers are not compulsory to perform a signature verification, whereas consumers are compelled to do so because of the following reasons [8]:

- The computation cost of signature verification (cryptographic calculations) is infeasible and burdensome at line speed.
- NDN does not provide any unified trust model, and applications are free to design their own trust models.
- Fetching certificate chain introduces additional cost.

Because of the above-specified reasons, reputation-based trust mechanisms have been investigated and researched as an alternative to cryptographic digital signature verification schemes. The absence of certificates in trust management systems carries with it a range of advantages including no certificate management, no certificate lookup, administration, and no certificate revocations. The trust management schemes free of the certificate can be commonly categorized as follows:

- Self-certified trust management
- Certificate less trust management
- Identity-based trust management
- Certificate-based trust management.

Ghali et al. [9] proposed a cache mitigation cache poisoning attack using Exclusion Filters. The authors indicated that customers should verify the authenticity of the Content using digital signatures, and if Content is found to be invalid or fraudulent, then create an Interest Packet and mention excluding field. On receiving interest packet with exclusion field, router ranks the Content according to the exclusion filters. More number on exclusion means that lesser ranking and low ranking Content do not get cached and forward the Interest to next routers. The proposed exclusion mechanism is able to detect and mitigate cache poisoning attacks efficiently, but it also has vulnerabilities, as if an attack collides with an attacker consumer then the

consumer will not put exclude field even in a corrupted packet. Even this field can be used to launch another DOS attack as a bad consumer can put exclusion filter even invalid Content. In reference [10], authors have proposed to use credibility values of the content provider with trust value to the Content. When a router receives a packet with exclusion filed, it reduces the credibility value of the Content provided as well as trusts the value of the Content. This mechanism improves the mitigation of CPA with selective caching. Wu et al. [11] proposed a scheme to improve previously proposed trust-based schemes using distributed Trust. They have argued that if a router itself is corrupting the Content, the above-said schemes cannot mitigate the CPA. For fighting router-oriented cache poisoning attack detection and mitigation, they proposed that the credibility value of routers should also be calculated using the number of exclusion filters for a packet received from a particular router. All the routers maintain the credibility value of adjacent routers and forward packets on the basis of credibility values.

Interest Key Binding (IKB) [12] rule is another method for checking the validity of Content. IKB uses the NDN Interest packet filed "PublisherPublicKeyDigest" and use to it sent hashed public key of the producer with the Interest packet. A router calculates the Hash of the public key available in the Data packet and matches that hash with the corresponding PIT entry. If hash matches, then it is cached for further requests fulfilment. Ntuli et al. [13] proposed a trust model to detect adversaries attempting cache snooping. In a cache snooping attack, an attacker used an exclusion field and send interest packets by excluding certain contain and on the basis of reply from the router attacker guesses the data caches in the router. For detecting this attack, the authors considered the fact that there is a high number of Interests with exclusion enabled in a short period from a particular client. Further authors used this fact for calculating the reputation of the consumer (trust value). The attacker can be identified if reputation values go below the threshold. Reputation-based trust model also used for mitigating Interest flooding attacks (IFA). ICRP [14] detects on and off IFA. The scheme calculates the trust value of the consumer based on the number of Interest sent in unit time. The scheme also keeps records of the previous behavior of the attacker and using previous and current trust values detects IFA. Reputation-based trust management schemes are used in various domains, and there are several design options available for designing and implementing these schemes. In reference [8], the author has analyzed various reputation-based trust schemes on the basis of dimensions, dissemination, calculation, and namely, formulation. Reputation-based trust formulation consists of two parts: reputation measure and model used for aggregation of trust values. Reputation-based trust value can be binary, string, continuous, and vector. The mathematical model aggregates values in the forms of ratings, fuzzy logic, probabilistic, and flow-based. There is a tradeoff between various design options that can be used according to the domain of the application. The following are various tradeoffs discussed in [8].

- Approaches based on Centralization and decentralization
- Operation or transaction-based Trust
- Global or local trust calculations

- Rank or threshold-based systems
- Rule-based and incentive-based system
- Ranking of Network Entities or Content.

In an NDN trust management system, a fundamental question is whether the ratings or reputation value should be connected to the disseminated Content itself or to the entity forwarding the information. When applying credibility values to providers, they can serve as a condition for (proactively) preventing transactions with untrustworthy providers, or caching whereas, determining the consistency of the service by itself may be helpful to customers or to detect unwanted or harmful activities. In the centralized system, the trustworthy central authority is responsible for gathering ratings, measuring credibility scores, saving, and communicating them to network institutions. This approach reduces measurement complexity since there is no needs to measure credibility by each node. In decentralizing, methods are developed by measuring ratings separately for each node and then by distribution protocol (optionally) distributed between the nodes. This method avoids the single failure point issue but still calls for trustworthy estimates and credibility values of the individual nodes. The transaction-based method depends on reliable knowledge obtained from transactions. Authentication of data integrity and provenance authentication can be considered as trust indicators, as given by the default trust management scheme of NDN. Opinion-based Trust is formed on an entity's subjective opinion about another on the grounds of its previous experiences. Mechanisms based on opinion operate with lower overhead and less resistance to attacks. Localized methodology preserves privacy and is best adapted for decentralized contexts, whereas the global method produces more generic, more precise outcomes as calculations represent the quantity of evaluations. Rank-based methods provide a quantitative rating of the credibility of all individuals. Rank-based systems are used in NDN to prioritize valid Content over fake, but the drawback is that high priority data objects are not always best. The threshold-based approach eliminates individuals that do not meet the necessary criterion. Due to its rigidity, it can be more potent if it is helpful in making caching and forwarding decisions to detect an attack.

3 Proposed Fuzzy-Reputation Based Trust Model

The aim of the proposed trust model is to prevent NDN from on-and off-cache poisoning attacks. Fuzzy logic is fundamentally simple to understand, versatile, and receptive to imprecise data, and it can be beneficial for NDN trust management for the following reasons:

- In order to quantify the Trust in NDN, fuzzy systems can be used to assess the value of the information. Deciding and choosing which Data is more important than other data, or which Data is required more quickly, is a highly humanistic phenomenon that fuzzy logic is capable of modelling [15].

- There are circumstances in which Trust is not evident because of uncertainty. The credibility value in this region lies between the limits of absolute confidence and complete misconfidence; thus, in these situations, the fuzzy technique is very effective for trust decisions.

3.1 Trust Matrices

In proposed mechanisms, the final trust value relies on two factors, i.e., the Trust calculated using direct interaction of a node with another node. It also defines how much a particular host is able to perform the expected task. For example, a router expects from the data provider that the data provider will provide uncorrupted and corrected data and that can be identified using a number of negative feedback, popularity, etc. The second factor is the correctness and usability of trust values gathered from other nodes (other than the node on which the router have to make a decision about content caching or correctness). This factor depends on the trustworthiness of cooperating nodes, similarity, direct or indirect connection, etc.

3.2 Trust Calculations

Whenever NDN nodes get any content, it needs to decide whether to cache that Content or not. The caching decision can be made using the reputation value of the incoming content node. The reputation value is the sum of reputation value it gets from its own calculation, value got from neighbors, and values got from the stranger nodes (nodes that are not directly connected). The node can also assign different weight-age to reputation value it got from other nodes like A, B, and C. For the trust value, two more matrices should be applied (α,β), which are the similarity of nodes and activity value. The reputation value can be calculated as the following equation:

$$\begin{aligned} Repu(Y \Rightarrow X)_{Previous} = {} & ARepu(Y \Rightarrow X) \\ & + B\frac{\sum_i \alpha_i Repu(Y \Rightarrow X)_i}{\sum_i \alpha_i} \\ & + C\frac{\sum_i \beta_i Repu(Y \Rightarrow X)_j}{\sum_i \beta_i} \end{aligned} \tag{1}$$

We have restricted trust values between 0 and k, so A + B + C should be equal to 1. The reputation value also reduce with time, so we have defined a factor τ, such that

$$Repu(Y \Rightarrow X)(t) = InitialValue - (FinalValue)e^{((t-t_0)/\tau)} \tag{2}$$

We have also defined two thresholds, upper, and lower trust. The lower trust value can be defined as miss trust, and the node will not cache Content from a host whose value lies below *miss_trust,* and the same upper value is *high_trust*, Content from *high_trust* value host will be cached by a node. In a condition where trust value lies between *miss_trust* and *high_trust*, the caching decision will be taken by using a fuzzy approach defined in the next section. $Repu(Y \Rightarrow X)$ can be trusted if:

- $Repu(Y \Rightarrow X) \geq high\_trust \Rightarrow Y$ can be trusted
- $Repu(Y \Rightarrow X) \leq miss\_trust \Rightarrow Y$ cannot be trusted
- $miss\_trust \leq Repu(Y \Rightarrow X) \leq high\_trust \Rightarrow$ Y Decision will be taken using fuzzy approach.

If any host is newly added into the system then its value fluctuates rapidly for the initial period so the system has to wait till its reputation stabilizes or provides an initial value for initialization. The weight factors α, β are multiplied by a host to the trust value from other hosts that defined the capability of nodes in providing recommendations. α, β defines similarity and activity values. Similarity values *Sim_value* defines similarity of hosts in calculating reputations; the value is calculated as

$$\alpha_i = a \times Sim\_value + b \times ACT\_Value. \tag{3}$$

'a' and 'b' are the weight factor given to similarity and activity values

$$\text{Sim\_value} = 1 - \sqrt{\frac{\sum (u - v)^2}{25n}} \tag{4}$$

u: Reputation value of cooperating host in initiator host reputation vector.
v: the reputation value of cooperating host in the target host reputation vector.
n: total number of cooperating hosts.

The second factor is activity value (*ACT_Value*) estimate the activity level of the cooperating hosting past interval. It is calculated using the total number of interactions of a host from the other hosts' reputation vectors.

$$\text{Act\_Value}(X) = \frac{\sum Int\_from_X}{\sum \text{Int} from_{hosts}} \tag{5}$$

$\sum Int\_from_X$: Sum of all past interactions reported by other hosts with a particular host.

$\sum \text{Int} from_{hosts}$: Sum of all interactions.

In order to calculate the values of time the calculation of decay factor is importation this proposal, we have used dynamic decay factor which depends on the reputation of the host. Initially, a nominal value should be applied to the decay factor, and then the value of τ increases or decreases with the reputation (consistency) of the host.

Table 1 Reputation vector

HostId	Status	RepValue	Weightfactor	Decay	Host interaction with (Int With)	Interaction from (Int From)	Attempts
12E4	Neighbour	4.5	0.6, 0.8	3425	56	58	78
–	–	–	–	–	–	–	–
1568 V	Neigh	5.6	0.5, 0.7	5434	54	67	89

For the dynamic factor, a minimum and maximum decay factor have to be defined. Decay factor can be calculated as [16].

1. $Difference(Diff) = NewValue - OldValue$
2. $If|Diff| \leq 1, then\tau_y = \tau_y \times round(5 - 4 \times |Diff|)$
3. $If Diff > 1, then\tau_y = \frac{\tau_y}{2 \times round(Diff)}$
4. $If - Diff > 1, then\tau_y = \tau_y \times 2 \times round(|Diff|)$
5. $NewValue$ is$Limited\tau_y inrange[\tau_{min}, \tau_{max}]$

All the interaction values are stored in the reputation vector defined in Table 1.

3.3 Fuzzy Approach

We have used the fuzzy system for two reasons, first to incorporate the 'importance factor', and the second one is to take decisions when aggregated reputation value is between *miss_trust* and *high_trust.*

(a) **The Importance factor**
The importance logic introduces the humanitarian concept in the system. The importance factor defines the criticality of the information. The more critical the information, the more danger it might have. In NDN, we consider Content necessary when it is requested by many interfaces. If the Data is important, it is appropriate to verify the integrity of the host where the information is coming from. The second factor is time.

(b) **Vague area decision**
Vague areas or grey areas are where the trust values lie between *miss_trust* and *high_trust*. For a taking grey area reason decision, the algorithm used a variable θ. The value of θ is calculated on the basis of the previous two trust value

$$\theta = \begin{cases} 1, & \text{if } truatvalueincreasing \\ 0, & \text{if } trustvalueisnotinanytrend \\ -1, & \text{if } trustvaluedecreasing \end{cases}$$

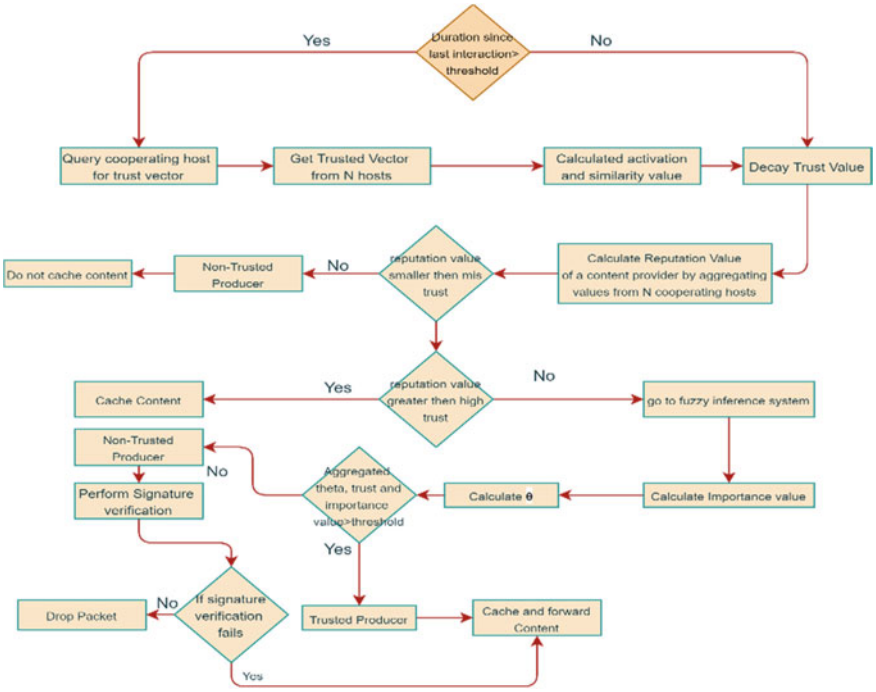


Fig. 4 Flowchart of the proposed approach

For taking a final grey area decision, we have considered total of three factors, the first is the important factor defined above, the second is the θ value, and the third is the trust value. If the trust value is approaching towards high Trust, then we will allot it a positive point, and if it is close to miss Trust then we provide it with a negative value.

(c) **Caching and forwarding Decisions**

In the proposed approach, if the calculated trust value is more than the predefined high Trust then Content if cached and forward to other requesting nodes, and if it is lower than the missing trust then the packet is simply discarded. In conditions where trust value lies between high Trust and miss trust, if the fuzzy system considers the Content as trusted, then Content is cached and forwarded to other nodes otherwise router goes content signature verification process for verifying its integrity (Fig. 4).

4 Conclusion

The paper proposed a trust model which is comprehensive and general. The proposed model incorporated a fuzzy system in the reputation-based trust model, which can be properly applied and used for mitigating content poisoning attacks. The model is

unique in the incorporation of different principles that are essential for the estimation of credibility values and subsequent confidence decisions. The model also applies principles of resemblance, reputation, behavior, and collaboration between hosts. Furthermore, the proposed approach encourages the transmission of trustworthy Content among network nodes. In this paper, we have not presented the simulation result. Future studies will be simulated for the proposed framework with different topologies, such as DFN, XC, and AS3967.

References

1. Singh VP, Ujjwal RL (2020) A walkthrough of name data networking: architecture, functionalities, operations and open issues. Sustain Comput: Inform Syst 100419
2. Zhang L, Estrin D, Burke J, Jacobson V, Thornton JD, Smetters DK, Papadopoulos C (2010) Named data networking (ndn) project. In: RelatórioTécnico NDN-0001, Xerox Palo alto research center-PARC, vol 157, p 158
3. Afanasyev A Moiseenko I, Zhang L (2012) ndnSIM: NDN simulator for NS-3. University of California, Los Angeles, technical report, 4
4. Hamdane B, Serhrouchni A, Fadlallah A, El Fatmi SG (2012) Named-data security scheme for named data networking. In 2012 third international conference on the network of the future (NOF). IEEE, pp 1–6
5. Singh VP, Ujjwal RL (2019) Privacy attack modeling and risk assessment method for name data networking. In Advances in computer communication and computational sciences. Springer, Singapore, pp 109–119
6. Singh VP, Ujjwal RL (2020) Gini impurity based NDN cache pollution attack defence mechanism. J Inf Optim Sci 1–11
7. Kim D et al (2017) Security of cached content in NDN. IEEE Trans Inf Forensics Secur 12: 2933–2944.
8. Kapetanidou IA, Sarros CA, and Vassilis T (2019) Reputation-based trust approaches in named data networking. Future Int 11(11): 241
9. Ghali C, Tsudik G, Uzun E (2014) Needle in a haystack: Mitigating content poisoning in named-data networking. In: Proceedings of NDSS workshop on security of emerging networking technologies (SENT)
10. Rezaeifar Z, Wang J, Oh H (2018) A trust-based method for mitigating cache poisoning in name data networking. J Netw Comput Appl 104:117–132
11. Wu D, Xu Z, Chen B, Zhang Y (2016) What if routers are malicious? Mitigating content poisoning attack in NDN. In Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA, Tianjin, China, 23–26 August 2016
12. Ghali C, Tsudik G, Uzun E (2014) Network-layer trust in named-data networking.. ACM SIGCOMM Comput Commun Rev 44(5):12–19
13. Ntuli N, Han S (2012) Detecting router cache snooping in named data networking. In Proceedings of the 2012international conference on ICT convergence (ICTC), Jeju Island, Korea, 15–17 October 2012
14. Umeda S, Kamimoto T, Ohata Y, Shigeno H (2015) Interest flow control method based on user reputation and content name prefixes in named data networking. In Proceedings of the 2015 IEEE Trustcom/BigDataSE/ISPA, Helsinki, Finland, 20–22 August 2015
15. Tajeddine A et al (2011) Fuzzy reputation-based trust model. Appl Soft Comput 11(1):345–355
16. Tajeddine A et al (2007) PATROL: a comprehensive reputation-based trust model. Int J Internet Technol Secur Trans 1(1–2):108–131

PID Controller Design for Continuous Stirred Tank Reactor (CSTR) Temperature Control

Bhumica Dutta and Vijay Kumar Tayal

Abstract Continuous Stirred Tank Reactor (CSTR) is an extensively used reactor in process control industrial applications and is the most researched topic for control engineers. The objective of this paper is to analyze the temperature control of CSTR with and without designing a PID controller so that the output tracks the changes in flow rate in the jacket and maintains the temperature of the tank reactor. All the designing, simulation, and calculations are done with the aid of MATLAB and SIMULINK.

Keywords CSTR (Continuous Stirred Tank Reactor) · PID (Proportional Integral Derivative) control · Tuning methods · Transient response · Eigenvalues

1 Introduction

In the chemical process, manufacturing units and various other industries distillation columns, heat exchangers, pumps, and CSTRs are commonly used. To control [1] their performance, PID controllers are considered as one of the options for 50 years. PID controllers basically consist of three individual units viz, a proportional gain (K_p), an integrator with the suitable value of gain of the integral controller (K_I), and a derivative controller with gain (K_D). The basic structure of the PID controller is given in Fig. 1.

The PID [2, 3] controller like any other controller works on the error between the input and the desired output and takes suitable action. This has many advantages over other controllers such as improvement in time performance specifications, makes the system faster, and is easy to tune. For best results, the PID controller should be perfectly tuned. The tuning of PID controller parameters is a tedious task.

The CSTRs are basically batch reactors in which feed is supplied continuously, and the output is also obtained continuously. It consists of a motor, cooling jacket,

B. Dutta (✉) · V. Kumar Tayal
Electrical and Electronics Engineering Department, Amity University Uttar Pradesh, Noida, India

V. Kumar Tayal
e-mail: vktayal@amity.edu

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_30

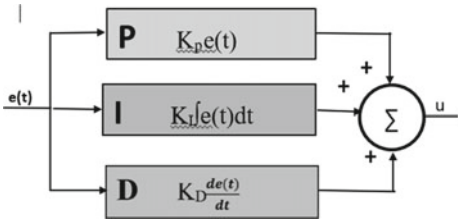


Fig. 1 Representation of PID controller

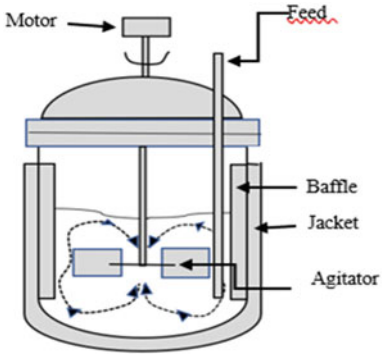


Fig. 2 Basic construction of CSTR

baffle, and agitator used for mixing the reactants in the tank. The basic construction is shown in Fig. 2.

CSTRs [1, 4] have highly nonlinear behavior or dynamics due to which it is the most researched topic. The process taking place in the reactor is considered to be exothermic that heat is liberated when reactants are mixed in the reactor. Thus, the heat reaction rate becomes faster, and even more heat is liberated which raises the temperature; hence, a cooling jacket is provided. Therefore, to maintain the temperature of CSTR is very crucial, and the same is tried to achieve by using a PID controller in this paper.

2 Mathematical Modelling

To [5, 6] obtain the mathematical model of any system first of all we need to define the input and output variables. In this system, the following input variables have been considered:

- Rate of feed flow in the reactor tank.
- Rate at which fluid flows in the jacket.
- Temperature at which feed is supplied.
- Temperature of fluid in the jacket.

The output variables considered are as follows:

Table 1 Description of the parameters used in the process

Terms	Description
V	Volume of tank
Q	Rate of heat transmission
Cp	Specific heat capacity
A	Heat transfer area
Ref	Value indicating reference
J	Jacket
I	Input/inlet
F	Rate at which volume changes
F	Final value
J_i	Input to the jacket
P	Density of liquid
T	Time
T	Temperature in degree
T_t	Temperature of tank
T_j	Temperature of jacket
U	Coefficient of heat transmission rate
Q	UA (T_j–T_t)

- Temperature at which liquid is in the tank
- Height or volume of liquid present in the tank
- Rate at which fluid flows from the tank that is the output.

The various parameters used in the mathematical modelling have been described in Table 1.

The principles used while developing the mathematical equations are as follows:

- Balance of material about the jacket and the reactor tank.
- Balance of energy about the jacket and reactor tank.

First, considering the balance of material about the reactor tank:
Accumulation rate = rate at which input flows – the rate at which output flows.

$$\frac{d(VP)}{dt} = F_i P - FP. \tag{1}$$

Second, considering the material balance around the jacket.

$$\frac{d(Vj P)}{dt} = F_{ji} P - F_j P_j. \tag{2}$$

Third, considering the energy balance around the tank.

$$V_t P_t Cp_t \frac{dTt}{dt} = P_t F_t Cp_t T_i - P_t F_t Cp_t T_t + Q \tag{3}$$

Fourth, considering energy balance around the jacket.

$$V_j P_j Cp_j \frac{dTj}{dt} = P_j F_j Cp_j T_{ji} - P_j F_j Cp_j T + Q \tag{4}$$

With the help of the above four equations, the following mathematical equations are developed

$$\frac{dTt}{dt} = \frac{Ft(Tj - Tt)}{Vt} + \frac{UA(Tj - Tt)}{VtCptPt} \tag{5}$$

$$\frac{dTj}{dt} = \frac{Fj(Tji - Tj)}{Vj} - \frac{UA(Tj - T)}{VjPjCpj} \tag{6}$$

3 Steady State Conditions

These conditions are determined by equating Eqs. (1) and (2) to 0, which gives us the following values:

$F_f = 1$ ft^3/min.
$P_j\ Cp_j = 61.3$ Btu/deg F – ft^3.
PCp = 61.3Btu/deg F – ft^3.
$T_f = 125$ deg F.
$T_{jf} = 150$ deg F.
$T_{is} = 50$ deg F.
$v = 10$ ft^3.
$v_j = 1$ ft^3.

4 Modelling in State Space Representation

For [7, 8] obtaining the state-space model, Eqs. (1) and (2) are converted into linear form with the use of the Taylor series. Here, only linear terms are considered rest of all terms are neglected.

By substituting these steady-state values in Eqs. (1) and (2), we get

$$\frac{dTt}{dt} = 0.4T_t + 0.3T_j - 7.5F_t + 0.1T_{ti} \tag{7}$$

$$\frac{dTj}{dt} = 3\ T_t - 4.5\ T_j + 50F_j + 1.5T_{ji} \tag{8}$$

The state-space representation is given by

$$\left[\frac{dX}{dt}\right] = [\mathrm{A}][\mathrm{X}] + [\mathrm{B}][\mathrm{U}] \tag{9}$$

$$[\mathrm{Y}] = [\mathrm{C}][\mathrm{X}] + [\mathrm{D}][\mathrm{U}] \tag{10}$$

where

- X represents the matrix of state variables.
- U is the input matrix
- Y is the matrix referring to the output.
- The state matrix is given by A.
- The input matrix is given by B.
- The output matrix is given by C.
- D is known as the translational matrix.

$\mathrm{X} = [\mathrm{T} - \mathrm{T_s}; \mathrm{T_j} - \mathrm{T_{js}}]$.
$\mathrm{U} = [\mathrm{F_j} - \mathrm{F_{js}}; \mathrm{F} - \mathrm{F_s}; \mathrm{T_i} - \mathrm{T_{is}}; \mathrm{T_{jin}} - \mathrm{T_{jins}}]$.
$\mathrm{Y} = [\mathrm{T} - \mathrm{T_s}; \mathrm{T_j} - \mathrm{T_{js}}]$.
$\mathrm{A} = [-0.4\ 0.3; 3\ -4.5]$.
$\mathrm{B} = [0\ -7.5\ 0.1\ 0; 50\ 0\ 0\ 1.5]$.
$\mathrm{C} = [1\ 0; 0\ 1]$.
$\mathrm{D} = [0\ 0\ 0\ 0; 0\ 0\ 0\ 0]$.

Now, converting the above state-space model into transfer function, we get the following results

$$\frac{Tt(s)}{Ft(s)} = \frac{-7.5s - 33.75}{s2 + 4.9s + 0.9} \tag{11}$$

$$\frac{Tt(s)}{Fj(s)} = \frac{15}{(s2 + 4.9s + 0.9)} \tag{12}$$

$$\frac{Tt(s)}{Tti(s)} = \frac{(0.1s + 0.45)}{s2 + 4.9s + 0.9} \tag{13}$$

$$\frac{Tt(s)}{Tji(s)} = \frac{0.45}{s2 + 4.9s + 0.9} \tag{14}$$

$$\frac{Tj(s)}{Ft(s)} = \frac{-22.5}{s2 + 4.9s + 0.9} \tag{15}$$

$$\frac{Tj(s)}{Tti(s)} = \frac{0.3}{(s2 + 4.9s + 0.9)} \tag{16}$$

$$\frac{Tj(s)}{Fj(s)} = \frac{(50s + 20)}{s2 + 4.9s + 0.9} \tag{17}$$

$$\frac{Tj(s)}{Tji(s)} = \frac{(1.5s + 0.6)}{(s2 + 4.9s + 0.9)} \tag{18}$$

Since we are focussing to maintain the temperature of the tank by controlling the fluid flow rate in the jacket, the transfer function used for the main analysis is given by

$$\frac{Tt(s)}{Fj(s)} = \frac{15}{s2 + 4.9s + 0.9} \tag{19}$$

5 Design of PID Controller

The general mathematical representation of the PID controller is as shown in Eq. (20).

$$u(t) = K_p e(t) + K_I \int e(t)\, dt + K_D \frac{de(t)}{dt}, \tag{20}$$

where u(t): controlled variable.

e(t): error signal.

The values of K_p, $K_{i,}$ and K_d are determined with the help of MATLAB.

There [9, 10] are various ways that can be used for determining the values of these constants also known as tuning methods like Zieglar Nichols Tuning Method, Cohen Coon Method, ITAE method, ultimate cycling method, etc.

Zieglar Nichols Tuning Method: In [11, 12] this method, first the process reaction curve is made by giving step input to the system and observing the output with time. Then from the graph of output V/s time, we calculate the values of L (lag ratio) and T (time constant). The values of K_p, K_i, and K_d are obtained as follows:

$K_p = \frac{1.2T}{L}$; $T_i = 2L$; $K_d = 0.5L$.

Cohen–Coon Tuning Method: This [13] method uses process gain, dead time, and time constant derived from step analysis of the system, as shown in Table 2.

t: time constant, td: dead-time, and gp: process gain.

Table 2 Cohen coon tuning parameters

K_p	T_I	K_D
$\frac{1.35}{gp}(\frac{t}{td} + 0.185)$	$2.5td\frac{t+0.185td}{t+0.611td}$	$0.37td\frac{t}{t+0.185td}$

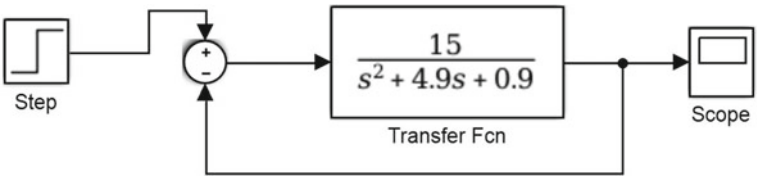


Fig. 3 Simulink Model of CSTR without PID controller

Ultimate Cycling Method: Start [14, 15] by setting the values of I and D gains to zero. After that increase the value of K_p from zero to maximum (K_u) value so as to achieve stable and consistent oscillations.

$K_p = 0.6\ K_u$, $T_i = 0.5T_u$, $K_d = \frac{Tu}{8}$ [T_u: oscillation period].

ITAE Method: This [16, 17] method involves minimizing the ITAE performance index, which is also known as optimization.

To calculate the values of these gains manually is a difficult task; hence, we have gone with auto-tuning.

The values of K_p, K_i, and K_d that are obtained for this process are given as follows:

$K_p = 1.081$, $K_{d\,=}\ 0.2247$, $K_i = 0.742$.

6 Block Diagram

- **Without PID controller** (Fig. 3).
- **With PID controller** (Fig. 4)

7 Stability Analysis

This has been observed that [18] the proposed control design yields in a stable system as shown by eigenvalue analysis. The eigenvalues of the system matrix are negative. The following MATLAB code gives the eigenvalues associated with the proposed system:
A = [−0.4 0.3; 3 −4.5];
x = eig(A);
−0.1911, −4.7089.
As both the eigenvalues are negative, hence, the system is stable.

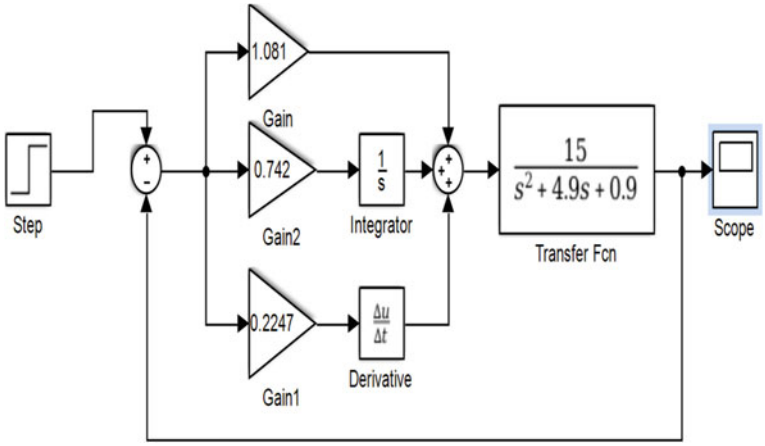


Fig. 4 Simulink Model of PID controller for CSTR

8 Simulation Results

- **Without PID controller** (Fig. 5)
- **With PID controller** (Fig. 6)

9 Results and Discussion

There are various factors that can be used as disturbance for affecting the working of CSTR such as rate of feed flow in the reactor tank, the rate at which fluid flows in the jacket, the temperature at which feed is supplied, and the temperature of the fluid in the jacket. In this work, the rate at which fluid flows in the jacket is used as the disturbance. The proposed technique is aimed at maintaining the temperature of the reactor tank, first without using a PID controller and then by incorporating it into the system.

The comparative analysis of different time response parameters with and without PID controller are shown and are also compared with the results of the reference paper [5] (with PID controller) in Table 3.

10 Conclusion

CSTR is a highly non-linear process, and maintaining its temperature is a difficult task. The reaction taking place in it might get affected due to a change in temperature

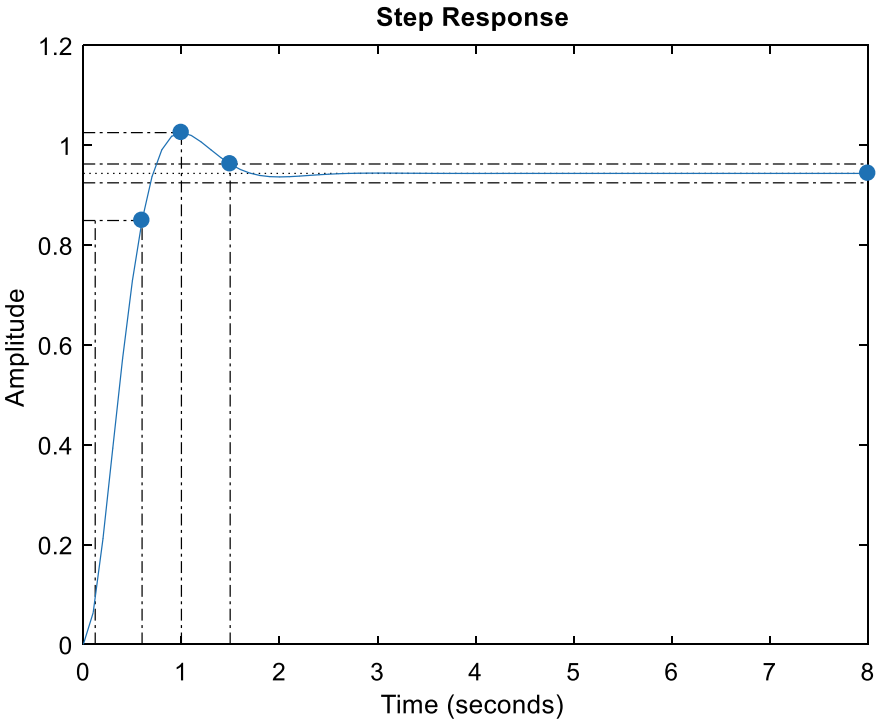


Fig. 5 Response of CSTR process without PID controller

in the reactor and hence resulting in hampering of output. From the analysis, this has been concluded that the delay time is more when the PID controller is not connected. Moreover the steady-state error is 5.7% indicating the output has not followed the changes in the input. On the other hand, using a PID controller with CSTR results in excellent improvement of the output of CSTR under step changes in input flow rate. The delay time is reduced. Further, it was successfully able to track the input changes that is change in fluid flow rate in the jacket and maintaining the temperature in the tank giving the value of steady-state error as 0%. Hence, using the PID controller gave much improved results. On the other hand when the results were compared with the reference paper [5], it showed that we got improved results including less delay time, less peak overshoot, and much less settling time which proved that our system was much more efficient.

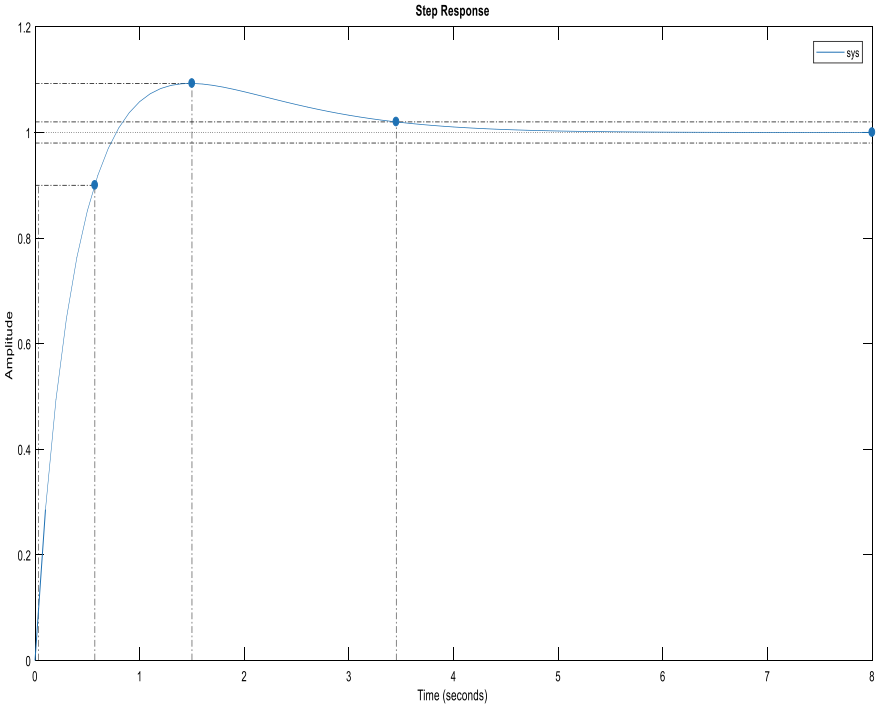


Fig. 6 Response of CSTR process with PID controller

Table 3 Comparison of time-domain parameters of the CSTR system

Parameter	Without PID controller	With PID controller	Author results with PID controller
Delay time (T_d)	0.344 s	0.2 s	0.25 s
Percentage peak overshoot	8.66%	9.29%	12%
Rise time	0.478 s	0.537 s	0.46 s
Settling time	1.5 s	3.45 s	6.5 s
Steady-state value	0.943	1	Not given
Steady-state error	5.7%	0%	Not given

References

1. Rubi VA, Deo A, Kumar N (2015) Temperature control of CSTR using PID controller. IJECS 4:11902–11905
2. Cohen GH, Coon GA (1953) Theoretical consideration of regard control‖. Trans ASME 75:827–834
3. Ruiyao Gao, Coon GA (1953) Theoretical consideration of regard control. Trans ASME 75:827–834

4. Rajavathsavai D,Khapre A, Munshi B (2014) Study of mixing behaviour of CSTR using CFD. Braz J Chem Eng 31:119–129; Red Tank Reactor Athema
5. Kabila V, Glan Devadhas G (2015) Comparative analysis of PID and fuzzy PID controller performance for CSTR. IJST 8(23
6. Luyben WL (2004) Temperature control of CSTR by manipulation of fresh feed. Ind Chem Eng 20:6430–644
7. Tayal VK, Lather JS (2015) Reduced order H∞ TCSC controller & PSO optimized fuzzy PSS design in mitigating small signal oscillations in a wide range. Int J Electr Power Energy Syst (Elsevier) 68:123–131
8. Dwivedi Y, Tayal VK (2017) Dynamic stability improvement of alkali fuel cell integrated system using PSO optimized PID control design. In: IEEE international conference on recent developments, control, automation and power engineering (RDCAPE 2017), pp 499–505
9. Ang KH, Chong G, Li Y (2005) PID control system analysis, design and technology. IEEE Trans Control Syst Technol 13(4):559–576
10. Xiong Q, Cai WJ, He MJ (2007) Equivalent transfer function method for PI/PID controller design for MIMO processes. Int J Process Control 17:665–673
11. Jyothi SN, Arvind S, Chidambaram M (2001) Design on PI/PID controllers for systems. Indian Chem Eng 43:288–293
12. Nikravesh M, Farell AE, Stanford TG (2000) Control of non-isothermal CSTR with time varying parameters with dynamic neural networks. Chem Eng J 76:1–16
13. Rahmat MF, Yazdani AM, Movahed MA, Mahmoudzadeh S (2011) Temperature control of a CSTR by means of two different intelligent strategies. Int J Smart Sens Intell Syst 4(2):244–267
14. Zieglar JG, Nichols NB (1942) Optimum settings for automatic controllers. Trans ASME 64:759–765
15. Chen F, Yang Z (1993) PM algorithm of self-tuning and its formula derivation. Acta Autom Sin 19:736–739
16. Shyamalagowri M, Rajeswari R (2013) Model predictive control design for non-linear process control reactor case study: CSTR, vol 7, pp 88–94
17. Yadav V, Tayal VK (2018) Optimal controller design for a DC motor using PID tuner. IEEE international conference on power energy, environment intelligent control (PEEIC), 2018 13–14 April
18. Sandeep gkp (2012) Temperature control of CSTR using PID & PID controller. Trans IJARCSSE 2(5)

SPSS Analysis and Performance Evaluation of Health Clubs

Pushpendra Neniwal and Vikas Kapoor

Abstract The society is suffering from various health-related issues due to many reasons. Due to advancements in automation, people are becoming lazy day by day, and the participation in physical activities is less. Due to this, people are joining gym for maintaining their health. The number of fitness industries is increasing worldwide day by day. So, there is a competition going on in various fitness industries across the world. In this paper, assessment of the quality of various fitness centers of Jodhpur city has been done. Both descriptive and inferential statistical analysis has been performed. SERVQUAL is utilized to measure the quality of service measures.

Keywords SERVQUAL · Indian fitness center · Customer service · Customer perception · Customer expectation · SPSS software

1 Introduction

In recent decades, people are seeing the negative aspects of the fast pace of life, fast food, air pollution, and many other consequences of the prevailing changes of life in the twentieth century; a trend of healthy lifestyle support has emerged [1]. In the present context, the customers are very much concerned about their living style. Standardization of various fitness service centers have taken place nowadays [2]. Since IT industries demand long sitting hours, extensive desktop working, and sedentary life style, the large population of ITES professionals are drawn towards gyms, health, and fitness centers [3].

P. Neniwal (✉) · V. Kapoor
Department of Production and Industrial Engineering, MBM Engineering College, Jodhpur, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_31

2 Research Methodology

Both descriptive and inferential statistical analysis [4–9] of data can be utilized to perform various operations over data [10]. In this study, quality of service of various fitness centers at Jodhpur city has been evaluated [11]. The convenience sampling method was applied in Talwalkars Health and Fitness center [12]. Respondents were chosen by the convenience method. Data has been collected by means of questionnaires during February 2019 [13]. The data was analyzed by means of Paired t-tests using the SPSS [14]. In our case, the customer expresses two opinions which are collected by pre-test (Expectation) or post-test (Perception). The following are the advantages of paired t-test:

- Greater control over confounding variables
- To detect a (statistically significant) difference.

3 Material and Methods

In this study, 50 subjects were randomly selected from the fitness center. The subjects were asked to complete the questionnaire during February and April 2019 [15]. Ages of the subjects were between 20 and above 40 years old.

4 Paired T-Test

This test compares two mean values that may be from the same individual, object, or related units [16]. In our case, the customer expresses two opinions which are collected by pre-test or post-test. This test can also be termed as Repeated Measures t Test, Dependent t Test [17].

Hypothesis: The following Hypothesis was formulated in order to observe the association between the Expectations and Perception among the customers:

- Null hypotheses, Ho: $\mu x = \mu y$,
- Alternative hypothesis, H1: $\mu x \neq \mu y$ (2-tailed).

Table 1 shows the respective perception means, expectation means, gap means, and t-values regarding sustainable health and fitness center service quality.

Gap scores have been calculated by using Eq. (1) which is again reproduced below:

$$\text{GAP SCORE} = \text{PERCEPTION} - \text{EXPECTATION} \tag{1}$$

- If, P-E ≥ 0, satisfactory.
- P-E ≤ 0, not satisfactory.

Table 1 The paired samples t-test for the data

Dimension	Item no	Expected service			Perceived service			Gap	Paired	
		Mean	Mode	SD	Mean	Mode	SD	P-E	t-value	Sig
Tangibility	1	6.32	6	0.620	4.92	5	1.306	−1.40	6.79	0.000
	2	5.74	6	0.803	5.00	5	1.030	−0.74	4.32	0.000
	3	6.38	6	0.602	5.10	5	1.249	−1.28	6.46	0.000
	4	4.98	5	0.958	4.90	5	1.035	−0.08	0.468	0.642
	Mean	**5.85**			**4.98**			**−0.87**		
Reliability	5	6.16	6	0.738	4.82	5	1.172	−1.34	7.64	0.000
	6	6.42	7	0.702	5.12	6	1.287	−1.3	7.00	0.000
	7	6.26	6	0.750	5.14	6	1.178	−1.12	5.74	0.000
	8	6.08	6	0.695	4.98	5	1.269	−1.1	6.06	0.000
	9	5.34	5	0.798	4.94	6	1.331	−0.4	2.02	0.049
	Mean	**6.05**			**5.00**			**−1.05**		
Responsiveness	10	6.02	6	0.958	4.86	5	1.229	−1.16	5.40	0.000
	11	6.30	6	0.839	4.84	5	1.149	−1.46	8.73	0.000
	12	6.54	7	0.705	4.76	4	1.254	−1.78	7.77	0.000
	13	6.30	7	0.952	4.96	5	1.211	−1.34	6.17	0.000
	Mean	**6.29**			**4.85**			**−1.43**		
Assurance	14	6.04	7	0.946	4.88	5	1.287	−1.16	7.25	0.000
	15	6.42	7	0.810	5.28	5	1.143	−1.14	6.84	0.000
	16	6.22	6	0.789	5.04	5	1.159	−1.18	6.39	0.000
	17	6.30	7	0.762	5.14	5	0.947	−1.16	6.39	0.000
	Mean	**6.24**			**5.08**			**−1.16**		

(continued)

Table 1 (continued)

Dimension	Item no	Expected service			Perceived service			Gap	Paired	
		Mean	Mode	SD	Mean	Mode	SD	P-E	t-value	Sig
Empathy	18	5.82	7	1.172	4.84	6	1.218	−0.98	4.52	0.000
	19	6.34	7	0.847	5.50	5	0.994	−0.84	6.08	0.000
	20	6.04	6	0.832	4.78	5	1.282	−1.26	7.81	0.000
	21	5.96	6	0.781	4.98	6	1.169	−0.98	6.66	0.000
	22	6.42	7	0.609	5.08	5	1.259	−1.34	6.79	0.000
	Mean	**6.11**			**5.03**			**−1.08**		
Overall	**Mean**	**6.10**			**4.99**			**−1.11**		

Table 2 Means of SERVQUAL dimensions

TAexp	TAper	RLexp	RLper	REexp	REper	ASexp	ASper	EMexp	EMper
6.32	4.92	6.16	4.82	6.02	4.86	6.04	4.88	5.82	4.84
5.74	5.00	6.42	5.12	6.30	4.84	6.42	5.28	6.34	5.50
6.38	5.10	6.26	5.14	6.54	4.76	6.22	5.04	6.04	4.78
4.98	4.90	6.08	4.98	6.30	4.96	6.30	5.14	5.96	4.98
		5.34	4.94					6.42	5.08

Table 2 represents the Means of SERVQUAL Dimensions of Tangibles, Reliability, Responsiveness, Assurance, and Empathy.

Table 3 shows the Paired Sample t-test data which has come out from SPSS software. Here, TAexp indicates expectation of TANGIBLES; TAper indicates perception of TANGIBLES, and so on.

There is no need of improvement. It requires maintaining the same service to the customers which fulfill the customers' needs and satisfy the customers' expectations.

Table 3 Paired samples test

(a)

		Mean	Std. deviation	Std. error mean	95% confidence interval of the difference	
					Lower	Upper
Pair 1	TAexp–TAper	0.87500	0.60274	0.30137	– 0.08410	1.83410
Pair 2	RLexp–RLper	1.05200	0.37963	0.16978	0.58063	1.52337
Pair 3	REexp–REper	1.43500	0.26096	0.13048	1.01975	1.85025
Pair 4	ASexp–ASper	1.16000	0.01633	0.00816	1.13402	1.18598
Pair 5	EMexp–EMper	1.08000	0.21071	0.09423	0.81837	1.34163

(b)

		t	D_f	Sig. (2-tailed)
Pair 1	TAexp–TAper	2.903	3	0.062
Pair 2	RLexp–RLper	6.196	4	0.003
Pair 3	REexp–REper	10.998	3	0.002
Pair 4	ASexp–ASper	142.070	3	0.000
Pair 5	EMexp–EMper	11.461	4	0.000

5 Recommendations and Conclusion

Using SPSS software, it is noted that largest significance occurs at "Material Associated with the Service," and second largest significance occurs at "Error Free Records" [18]. It means that there is no further need to change or to further investigate those two questionnaires. Rest of the factors need improving or are in need of further investigations.

5.1 Recommendations for Health and Fitness Center

- The staff needs to respond to the customer's major issue regarding the complaints. Many problems are solved in gym and very few further efforts are required. However, in Cardio Section, the gym needs to provide at least three trainers and to follow up after the completion of workout of the customer by simply asking a small question as to how be the workout? Or go randomly in different shifts and to ask the customers how the workout is going, how they are coping?
- Make sure that trainer gives proper attention to every client for that one trainer should provide guidance to a similar group of members and make competitiveness strategy to achieve the target?
- At least one helper is always available at every gym floor area to clean the sweat on the machine or floor.
- Take monthly follow up of the clients. Inform the members through message or in person that how much weight they lost or gained in a month, and maintain a feedback card where the client is free to mention the remark regarding service in that particular month.
- Monthly follow up of every client regarding measurement, weight loss, or gain and any complaints, whether they are comfortable or not. What they like most and what they like least about the gym. Take the data and follow up with the action as soon as possible.
- Solve the problem of clients as soon as possible, do not take too much time, and after solving the problems, take feedback of the client that he/she is satisfied or not. Make sure that the client feels that he/she gets personal attention because personal attention is a big reason why people chose your gym over other gyms.
- Make sure trainers do not provide personal attention to the female clients because this is the big issue coming from the male clients, that is, trainers give more personal attention to the female clients as compared to male clients. To overcome that, the CCTV camera recording be scrutinized randomly at different timing to ensure that this is not happening. Another issue is similar that some trainers are more engaged with particular clients on a regular basis due to some personal interaction between client and trainer outside the gym. This may also be improved by proper inspection from time to time and to ensure that same trainer does not engage with the same client for long on each and every day.

- A small issue is raised from some clients that when they workout, they hear the very same song day after day on particular timing which becomes monotonous and boring. So, it would not hurt to make sure that new songs are played or the entire playlist of songs is shuffled after particular time duration.

References

1. Afthinos Y, Theodorakis ND, Nassis P (2005) Customers' expectations of service in Greek fitness centers. Manag Serv Q 15(3):245258
2. Blaikie N (2000) Designing social research. Blackwell Publishers, New York
3. Farrell D, Laboissiere MA, Rosenfied J (2005) Sizing the emerging global labour market. McKinsey Q 3:93–103
4. Joshi A, Sabitha AS, Choudhury T (2018) Crime analysis using K-means clustering. In: Proceedings—2017 international conference on computational intelligence and networks, CINE 2017. https://doi.org/10.1109/CINE.2017.23
5. Sharma D, Sai Sabitha A, Choudhury T (2019) Discovering the unknown patterns of crop production using clustering analysis. In: Advances in intelligent systems and computing, vol 841. https://doi.org/10.1007/978-981-13-2285-3_70
6. Raj G, Sharma S, Choudhury T (2018) Load analysis in SDN for distributed topologies. In: Proceedings of the 2nd international conference on green computing and internet of things, ICGCIoT 2018. https://doi.org/10.1109/ICGCIoT.2018.8753099
7. Mehra S, Choudhury T (2018) Sentiment analysis of user entered text. proceedings of the international conference on computational techniques. Electron Mech Syst CTEMS 2018. https://doi.org/10.1109/CTEMS.2018.8769136
8. Mittal V, Gupta S, Choudhury T (2018) Comparative analysis of authentication and access control protocols against malicious attacks in wireless sensor networks. In: Smart computing and informatics. Springer, Singapore, pp 255–262
9. Chauhan N, Agarwal R, Garg K, Choudhury T (2020) Redundant IAAS cloud selection with consideration of multi criteria decision analysis. Procedia Comput Sci 167:1325–1333
10. Frew M, McGillivray D (2005) Health clubs and body politics: aesthetics and the quest for physical capital. Leis Stud 24(2):161–175
11. Gillison FB, Skevington SM, Sato A, Standage M, Evangelidou S (2009) The effects of exercise interventions on quality of life in clinical and healthy populations: a meta-analysis. Soc Sci Med 68(9):1700–1710
12. Lalic H, Kalebota N, Kabalin M (2006) Measures for achieving recruits' enhanced fitness: a transversal study. Collegium Antropol 30(3):585–592
13. Lloyd C (2008) Recruiting for fitness: qualifications and the challenges of an employer-led system. J Educ Work 21(3):175–195
14. Lloyd HC, Reynolds KL (2003) The consequences of dysfunctional customer behaviour. J Serv Res 6(2):144–161
15. Macintosh E, Doherty A (2007) Reframing the service environment in the fitness industry. Manag Leis 12(4):273–289
16. Malhotra T (2001) Statistical methods. 7th edn. lowa State University Press, Ames, lowa
17. Serbes S, Cengiz C, Sivri M, Filiz T (2017) Health-related fitness knowledge of middle school students in public and private schools. Monten J Sports Sci Med 6(1):29–35
18. Tansik DA (1990) Managing human resources issues for high-contact service personnel. In: Bowen DE, Chase RB, Cummings TG (eds) Service management effectiveness. Jossey-Bass, Scan Francisco

A Comprehensive Effectual Load-Balancing Method in Cloud Computing

Sandeep Srivastava, Vivek Kumar Sharma, Rajeev Sharma, Bhupesh Kumar Singh, Jitendra Nath Singh, and Tanupriya Choudhury

Abstract In the era of continuous economic growth, application development strategies demand strong computing infrastructure that may inculcate a high cost in return to the firms. The solution of this cost problem is Cloud Computing which assures the on-demand delivery of computing resources. But every solution has hidden challenges, and this paper focuses on identifying the significant challenges in a cloud computing environment. One among them is distributing workloads across multiple computing resources. This paper addresses the fundamental difficulties of load-balancing in the cloud environment. Meanwhile, the paper further discusses the probable improvised techniques used to overcome cloud system problems. The gist of paper lies in solving the aforementioned problems by including a chain of logical analysis and generating the algorithm that leads to the appropriate solution. The prime focus is on job queue making strategy that appropriately allocates various jobs to CPUs on the basis of assigned priority or without priority. It also deals with some of the major problems of load-balancing in cloud environment like a timeout. To conclude, this paper exhibits how this approach partially fits into the notable AWS and GAE cloud architecture. This paper presents the various load-balancing problems in cloud computing environment and holds a clear intention to provide readers an overview of issues faced and also simulating further interest to pursue more advanced research in it.

S. Srivastava · V. K. Sharma
G L Bajaj Institute of Technology & Management, Greater Noida, India

R. Sharma · J. N. Singh
Galgotias University, Greater Noida, India
e-mail: sharma.rajeev@galgotiasuniversity.edu.in

J. N. Singh
e-mail: drjnsingh@galgotiasuniversity.edu.in

B. K. Singh
Arba Minch University, Arba Minch, Ethiopia
e-mail: dr.bhupeshkumarsingh@amu.edu.et

T. Choudhury (✉)
Department of CSE, University of Petroleum and Energy Studies (UPES), Dehradun, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_32

Keywords Cloud · Load-balancing · Migration · Job-iteration · Timeout · Super-node

1 Introduction

Cloud is defined as a large distributed computing system which shares resources, software, and information on-demand, like public utility to a large number of users. Cloud computing [1–3] is an evolution of Virtualization, Utility computing, Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS), and Platform-as-a-Service (PaaS). To enhance the response time and ensure the utilization of resources, the process of Load-balancing is performed which reassigns the total load stored in the individual nodes of the collective system. Cloud system is the collection of many heterogeneous systems [4, 5]. The systems actually consist of servers and clients. Clients' request for the resources must be provided as soon as possible. The CPU of the server must process the client's request without too much delay. CPU located in different places have different task loads. Again, every CPU has different speed. So, the task completion rate of every CPU changes. To make the system efficient, the system must have prepared a process where the CPUs work in such a way that the collective speed of the system is increased. So, majorly, the focus relies on balancing these loads in cloud computing. And, it is too much difficult to manage cloud computing without balancing loads. There are several other resources which will be load-balanced, like.

i. Network edges and provides s-DNS, FTP, and HTTP
ii. Making Connections through smart switches
iii. Execution through computer system task
iv. Storage resources right of entry to application object Instances.

The remaining part of the paper is further categorized into various sections as mentioned: Sect. 1.1 discusses about overall architectural components and various architectures of cloud system. In the Sects. 1.2 and 2, various difficulties of cloud systems including load-balancing are discussed, and a solution-based approach is provided in Sect. 3. Section 4 discusses about the QTS service in cloud. Finally, Sect. 5 highlights some research issues and Sect. 6 gives the conclusion of this paper.

1.1 Cloud Architecture

Cloud services [6–8] are categorized as per application's requirement into the following mentioned models: SaaS, PaaS, IaaS, and Data as a Service (DaaS), and these models are demonstrated using a specific cloud infrastructure [9–11].

A. Software-as-a-Service (SaaS)
Cloud consumers select a hosting environment to release their applications and perform execution through the networks from various clients (e.g., Web browser, PDA, etc.) by application users and other end users. Cloud consumers do not surpass the cloud infrastructure that is dedicated to provide multi-tenancy system architecture. As in SaaS, to enhance efficiency in terms of speed, security, availability, disaster recovery, maintenance availability and to achieve economies of scale, different cloud consumers' applications fits in a single logical environment.
In some cases, it is the Contact Volume Editor that checks all the pdfs. In such cases, the authors are not involved in the checking phase.
Applications that show the strength of SaaS include SalesForce.com, Google Mail, Google Docs, and many more to go.

B. Platform-as-a-Service (PaaS)
Cloud consumers require a development platform to develop cloud services and applications. To strengthen PaaS, full support mechanism of "Software lifecycle" is incorporated (e.g., SaaS) directly on the PaaS cloud. SaaS and PaaS can be differentiated easily on the basis of application developed. SaaS only hosts completed cloud applications, whereas PaaS hosts both completed and in-progress cloud applications providing the platform as well. PaaS therefore requires a supporting application hosting environment and development infrastructure along with programming environment, various tools, configuration management, and other valuable resources to accomplish the tasks allotted. Google AppEngine exemplifies PaaS (Fig. 1).

C. Infrastructure-as-a-Service (IaaS)
Cloud consumers require an infrastructure and resources to deploy the applications. This demands the direct usage of IT infrastructures which further includes processing, storage, networks, and other basic computing resources required for completing a task. To integrate/decompose physical resources and to meet the growing or shrinking resource demand from cloud consumers, virtualization is deployed as a concept in IaaS. Virtualization implements a strategy which focuses on setting up independent virtual machines (VM) and ensure that VM are isolated from both the underlying hardware and other VMs. It must be noted that strategy differs a lot from the multi-tenancy model, which performs the transformation of the application software architecture in order to

Fig. 1 Basic cloud architecture

allow execution of multiple instances (from multiple cloud consumers) a single application (i.e., the same logic machine). An example of IaaS is Amazon's EC2.

Database DaaS' working style is different from traditional storage interfaces such as RDBMS and file systems. It offers table-style abstractions that are designed to support the storage and retrieval of huge amount of data within a very compressed timeframe, often too large, too expensive, or too slow for most commercial RDBMS to cope with. Amazon S3, Google Big Table, and Apache H Base, demonstrate the working strategies of DaaS.

D. Data as a Service (DaaS)

Cloud consumers require a separate service to deliver the virtualized storage on demand. This is incorporated using a separate Cloud service—data storage service. Though it is observed that DaaS is a special type IaaS. The reason to avail this service is associated with costs involved. Usually, the on-premise enterprise database systems are bound with a prohibitive upfront cost in dedicated server, software with license, post-delivery services, and in-house IT maintenance. DaaS allows a consumer to question about the cost that is paid for the service. Consumers pay for what they are actually using rather than the site license for the entire [12, 13].

1.2 Issues in Cloud Systems

Cloud computing [14–16] is highly scalable which allows it to fulfill the need of infinite computing resources on demand. This in return ensures the work satisfaction of a large number of customers. There are several difficulties associated with the services of cloud computing. They are listed below.

Security and Privacy Issues

There are several common security threats in cloud computing paradigm. The assurance of information security and data privacy is of utmost importance. This requires the identification of organizational and technical issues related to keeping cloud services so that the security and privacy are not compromised at any level. Data used by Banks, medical, and research facilities is highly sensitive in nature and demands the secure policies for security and privacy [10].

Infrastructure

This deals with the issues related to hardware layer of cloud services along with the software that are responsible to operate the embedded hardware. The proposal presented here elaborates the mentioned issue.

Data Handling

This category deals with the data storage type of problems like data segmentation and recovery, data resiliency, data fragmentation and duplication, data retrieval, data provenance, data anonymization and placements, etc.

In this paper, major focus is on the symmetric distribution of workload among processing cloud systems. It is commonly found that the CPUs are not working with proper distribution of workloads. Sometimes, a set of CPUs stay idle or less loaded while others are too much loaded. Though the system has the high functioning CPUs, but this type of behavior of the system makes the performance too much degraded from the actual or theoretical value. So, several cloud systems face the challenges of maintaining its speed by sharing loads or tasks to the idle or less busy CPUs.

2 Problems of Load-Balancing

For a cloud system, the tasks are distributed over the system, and every system has two parts (a) Task Reception and (b) Task Migration. These two are also known as "Node States Information" [17]. Here, node is actually the system CPU or server. For a general task given to the node, the expected waiting time is (at node i)

$$M_i(n) = q_i(n) \cdot n_s \tag{1}$$

where, $M_i(n)$ is the wait time for job, q_i is the queue length of the task, and t_s is average time or completion of a task.

Now, the load-balancing concept says that there are communicating nodes in between two local domains to communicate the data with the global domain. These nodes are actually the nodes which are overlapped in more than one domains. In this type of load sharing, N is Number of nodes, $V = \{1, 2\ldots N\}$ a set of nodes in a system.

Here, we calculate a maximum load to be executed in the node and then proceed to share the load to the other nodes.

For the algorithm, the node's speed and task completion rate are also calculated. But, in this instance, some possible drawbacks are not calculated.

Here, those drawbacks are,

(1) If the communicating node fails, then the whole load-balancing will be stopped.
(2) For two domains, there is a possibility of communication delay.
(3) Communication overhead problem.

This paper focuses on this particular problem of waiting time of the tasks or jobs. The given approach is to solve waiting problem as well as simplistic approach to the communication delay [18–21].

3 Proposed Approach

The proposed approach for effective load-balancing is divided into three sub-sections. They are as follows.

3.1 Queue Making and Job Making Processes

For every CPU, there is a speed limitation. This speed is a function depending on variables like instruction set, cache, clock speed, bandwidth, generated heat, and heat dissipation. Depending on all these variables, the clock speed is calculated and then the average task completion time is also calculated.

According to the task completion average time, there is a queue for every CPU. This queue is actually a list of tasks which also contains a number called "threshold." This threshold is a general indicator which tells the task that beyond this point, the tasks have to wait longer than the expectation. Thus, the task is ready to be migrated to the node which is idler. Using these three processes will increase the speed of the system. It also converts the system into a backup planned system which is adaptive to the speed of the nodes and also to the queue and server status.

If we try to think that there are n numbers of CPUs and k number of tasks with their own different characteristics. So, we can get a working algorithm for queue making in the following way:

CPU rearrangement

This task allows you to arrange all the available CPUs from high to low in according to their functional speed.

Rearrange Job-requests

Priority basis

i. If there is any priority assigned to the jobs, then the jobs should be rearranged according to the priority. In this rearranged list, the highest priority job will be in the first position and the least one at the last position.
ii. If any job is not assigned with priority, then the job is set to execute with the lowest priority and the priority index is updated.

Without priority basis

When there is no priority assigned for the listed jobs, then the jobs are arranged by their estimated completion time. The estimated time depends on code of lens, loops, inner loops, and resources required. Jobs are rearranged from highest to lowest order.

CPU-allocation

The rearranged CPU list gets the rearranged job list, and then the CPUs are given the jobs, as like first CPU in the list gets the first job and the last CPU gets the last

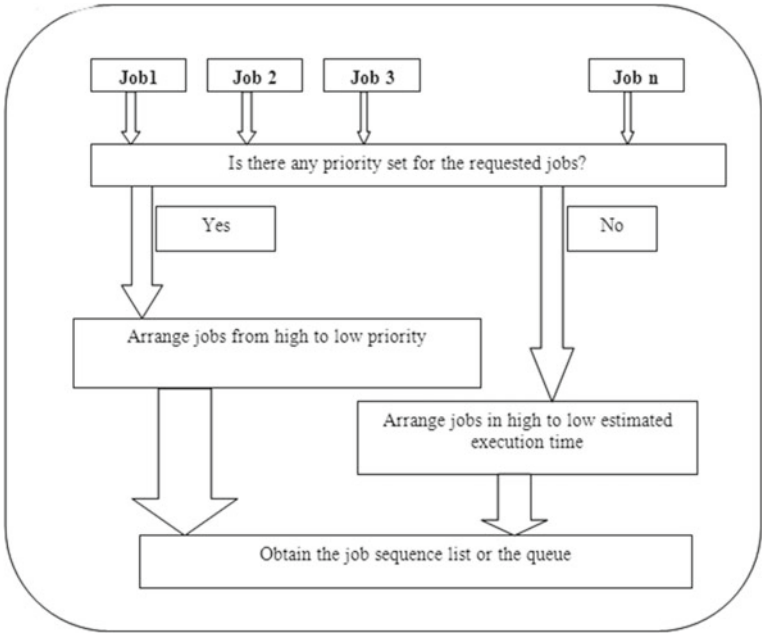


Fig. 2 Job-making process

job until all the CPUs are occupied, then the list of jobs start to allocate themselves from the start of the CPU list.

Completion time estimation

The speed of the CPU and the required CPU cycle of the job give the general completion time of the job.

Runtime new request allocation

When a new job is arrived, then the job is included to the initial list of jobs that are not being executed by the CPUs. Then, the list goes through the step (ii) to (iv). This way a new job is inserted in the existing list of jobs and allocated to the available CPUs.

The above job making and queue making processes are represented in the Figs. 2 and 3, respectively.

3.2 Timeout Chart

For a particular task queue, one task is migrated to a node but then the task is delayed because of longer waiting time. Again, this may happen that searching or migration

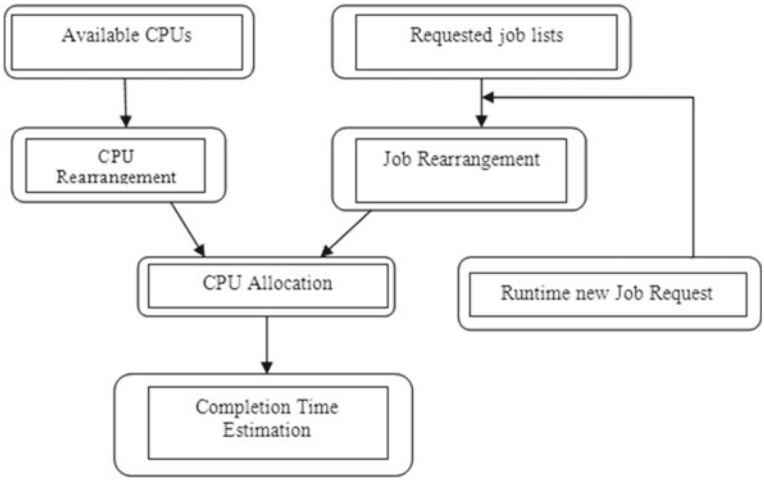


Fig. 3 Queue-making process

takes so long time that it is actually exceeding the first node completion time. The third scenario such that one task is migrated and processed very quickly but the transfer time for the task is too high to wait for the other task. To overcome all these problems, there must a timeout chart available which will work as a look ahead table and also to take a decision whether the task should be migrated or not. Now, if it is seen that the total system is not working and it is becoming too slow, then also this timeout chart work as a very good reference to eliminate or terminate jobs. That is, if the job completion is not executing properly, then the chart provides us the time of termination, and then, we can decide which jobs are to be completed within this time and which are to be terminated. This process also checks the priority and job types. As if there is no priority set initially, then we can have many different jobs. Among those jobs, some are reversible or have roll back option, others may not have that option. In that case, if a job without rollback option is being executed, then that job will not be terminated before completion.

3.3 Cloud Node as Proxy-Node

All the communicating nodes of the global domain of cloud are connected to one super-node with data base of required characteristic of these nodes. In such a case, if a Failure happens in one of these nodes, then the super-node will provide all the information to make the system work properly. Here, this super-node is actually a proxy node which keeps all the copy of the information and updates it from time to time. But it is not searched in every case as it is part of the large database. But it is

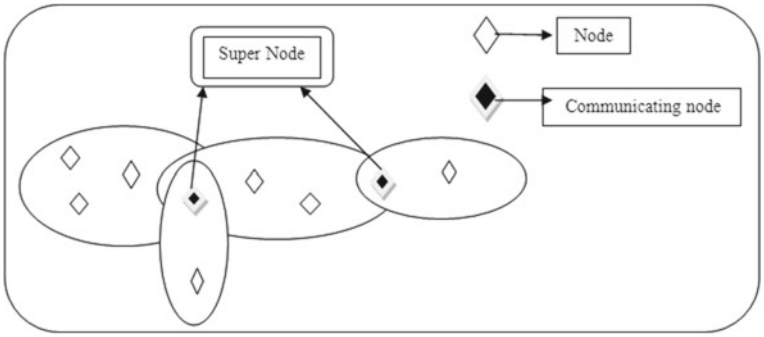


Fig. 4 Functionality of super-node

updated without delaying the system that the system may have a backup plan at the time of failure. The entire process is shown in Fig. 4.

4 QTS in Cloud Computing

The term QTS is actually a combination of those three processes which have been discussed previously. In cloud environment, the two mostly available and used cloud architectures are Amazon Web Services (AWS) [22] and Google App Engine (GAE) [23]. In the proposed model, the QTS comes in between web server and VaR Analytic Server (GAE) and EC2 Instance VaR Server (AWS).This process is shown in Fig. 5. This is the proposed approach of the paper to improvise the migration process so that the performance gets better or the system.

According to Figs. 6 and 7, the process of QTS is applied in GAE google app engine [23]. These Figs. 6 and 7 describe what QTS are, and when and how to use

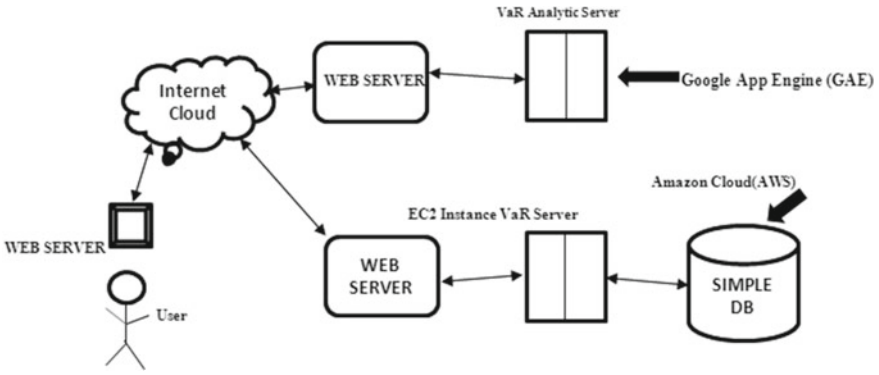


Fig. 5 Working process of AWS/GAE without QTS

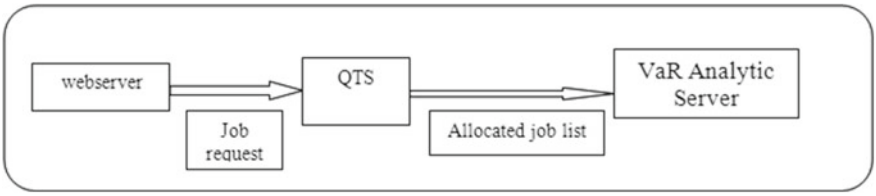


Fig. 6 QTS applied in GAE

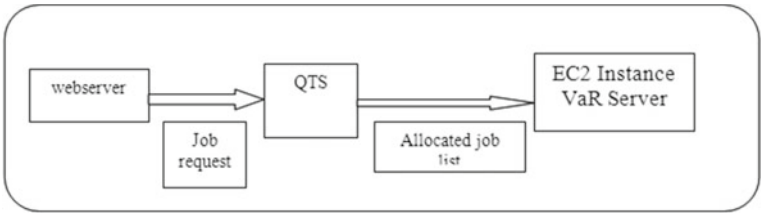


Fig. 7 QTS applied in AWS

them. Job queues let applications perform work, called *tasks*, asynchronously outside of a user request. If an app needs to execute work in the background, it adds tasks to job *queues*. The tasks are executed later, by worker services. The QTS applied in AWS environment with job queue implementation and work flow is represented by Fig. 7.

5 Result

In this paper, the essentials of cloud computing to facilitate the research challenges faced in load-balancing are provided. It has identified the pros and cons of cloud computing. It further elaborates the process of allocating the jobs to suitable CPUs following a dynamic approach. Other features addressed so far are: measure the delay and terminate a process, if required, and migrate the process to other CPUs for the efficient usability of the server. Finally, this paper adapts an approach that is evaluated theoretically and mathematically.

The paper involves an algorithm to increase the acceptance of the prescribed approach. For the best optimization of job allocation and migration to appropriate CPUs, manual intervention through the various tables is also implemented along with available dynamic scenarios. Undoubtedly, cloud computing is a cutting edge innovation in the field of computer science with some drawbacks associated with it. This idea of paper is to strictly follow the cloud architecture without expecting any high level changes. It suggests a contribution to the existing process for an efficient and improvised approach.

6 Future Research Issues

This paper also tries to highlight the changes happening in the field of load-balancing algorithms across the world.

i. Maximum resource utilization is one of the main objectives in any distributed load-balancing system. In future, this load-balancing algorithm can be associated with some optimization techniques (PSO, ACO, GA, etc.) to give more flexibility to distributing the loads and improving the performance of the system.
ii. Besides load-balancing, fault-tolerance in cloud system needs to be incorporated in order to remove security and usage issues and to design the approach more cost-effective and successful.
iii. The ongoing cloud load adjusting arrangements are concentrating on Green cloud point in load adjusting instruments by considering the difficulties like: Reducing the vitality and force utilization, reducing carbon emanation, and furthermore, lessening cost for the clients because of vitality effectiveness subject. There must be devised some dynamic load-balancing algorithm which can set the load degree high and low dynamically. In future research issues, some sort of fuzzy logic may be associated with load-balancing techniques to make it more flexible.

7 Conclusion

This paper provides the essentials of cloud computing to facilitate the research challenges faced in load-balancing. It has identified the pros and cons of cloud computing. It further elaborates the process of allocating the jobs to suitable CPUs following a dynamic approach. Other features addressed so far are: measure the delay and terminate a process, if required, and migrate the process to other CPUs for the efficient usability of the server. Finally, this paper adapts an approach that is evaluated theoretically and mathematically. Undoubtedly, cloud computing is a cutting edge innovation in the field of computer science with some drawbacks associated with it. The idea of this paper is to strictly follow the cloud architecture without expecting any high level changes.

References

1. Chauhan N, Agarwal R, Garg K, Choudhury T (2020) Redundant IAAS cloud selection with consideration of multi criteria decision analysis. Procedia Comput Sci 167:1325–1333
2. Dewangan BK, Agarwal A, Choudhury T, Pasricha A, Chandra Satapathy S (2020) Extensive review of cloud resource management techniques in industry 4.0: issue and challenges. Softw: Pract Exp

3. Wadhwa M, Goel A, Choudhury T, Mishra VP (2019) Green cloud computing-a greener approach to IT. In: 2019 international conference on computational intelligence and knowledge economy (ICCIKE), 760–764
4. Patil U, Shedge R (2016) Improved hybrid dynamic load balancing algorithm for distributed environment. Int J Sci Res Publ 3(3). ISSN 2250-3153
5. Hsu CH, Liu JW (2010)Dynamic load balancing algorithms in homogeneous distributed system. In: Proceedings of the 6th international conference on distributed computing systems, pp 216–223
6. Kaur A, Raj G, Yadav S, Choudhury T (2018). Performance evaluation of AWS and IBM cloud platforms for security mechanism. In: 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS), pp 516–520
7. Kumra S, Choudhury T, Nhu NG, Nalwa T (2018) Challenges faced by cloud computing. In: Proceedings of the 2017 3rd international conference on applied and theoretical computing and communication technology, ICATccT 2017. https://doi.org/10.1109/ICATCCT.2017.8389105
8. Gupta A, Choudhury T, Lal R (2017) An efficient scheme to secure cloud with diversified fortified mechanisms. In: Proceedings of the 2017 international conference on big data analytics and computational intelligence, ICBDACI 2017. https://doi.org/10.1109/ICBDACI.2017.8070829
9. Charkraborty S, Chaudhary N et al (2016) An study of new dynamic load balancing approach in cloud environment. World J Technol Manag 3(4). wjcat-2016-040302
10. Ghanam Y, Ferreira J, Maurer F (2016) Emerging issues & challenges in cloud computing—a hybrid approach. J Softw Eng Appl 5:923–937
11. Turab NM et al (2018) Cloud computing challenges and solutions. Int J Comput Netw Commun (IJCNC) 5(5)
12. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of Weiler-Atherton algorithm to address loose polygons. J Sci Ind Res 78:771–774
13. Kumar V, Choudhury T (2019) Real-time recognition of Malignant skin lesions using ensemble modeling. J Sci Ind Res (JSIR) 78:148–153
14. Khanna A, Kero A, Kumar D (2016) Mobile cloud computing architecture for computation offloading. In: 2016 2nd international conference on next generation computing technologies (NGCT). IEEE, pp 639–643
15. Khanna A, Goyal R, Verma M, Joshi D (2018) Intelligent traffic management system for smart cities. In: International conference on futuristic trends in network and communication technologies. Springer, Singapore, pp 152–164
16. Khanna A, Anand R (2016) IoT based smart parking system. In: 2016 International conference on internet of things and applications (IOTA). IEEE, pp 266–270
17. Neelakantan P (2017) An adaptive load sharing algorithm for heterogeneous distributed system. Int J Res Comput Sci 3(3):9–15 (A Unit of White Globe Publications). . ISSN 2249-8265. www.ijorcs.org, https://doi.org/10.7815/ijorcs.33.2017.063
18. Kotkondawar RR, Khaire PA, Akewar MC, Patil YN (2014) A study of effective load balancing approaches in cloud computing. Int J Comput Appl 87(8)
19. Joshi G, Verma SK (2015) A review on load balancing approach in cloud computing. Int J Comput Appl 119(20)
20. Yadav AV, Mohammad F (2018) Different strategies for load balancing in cloud computing environment: a critical study. Int J Sci Res Eng Technol 3(1)
21. Google Cloud Platform. https://cloud.google.com/appengine/ docs
22. Devi DC, Uthariaraj VR (2016) Load balancing in cloud computing environment using improved weighted round robin algorithm for non-preemptive dependent tasks. Sci World J (Hindawi)
23. Marinescu DC (2018) Cloud computing: theory and practice. Morgan Kaufmann, ISBN-13: 978-0124046276
24. Katyal M, Mishra A (2018) A comparative study of load balancing algorithms in cloud computing environment. Int J Distrib Cloud Comput 1(2)

25. DoddiniProbhuling L (2013) Load balancing algorithms in cloud computing. Int J Adv Comput Math Sci 4(3):229–233. ISSN 2230-9624. http://bipublication.com.
26. Mohamed Shameem P, Shaji RS (2015) A methodological survey on load balancing techniques in cloud computing. Int J Eng Technol (IJET) 5(5). ISSN: 0975-4024
27. Mellon C, Lewis C (2010) Basics about cloud computing. Softw Eng Instit
28. Buyya R, Broberg J, Goscinski AM, Cloud computing principles and paradigms
29. Amazon Web Services Whitepapers. http://aws.amazon.com/ de/whitepapers/

Code Smell Refactoring for Energy Optimization of Android Apps

Reeshti, Rajni Sehgal, Deepti Mehrotra, Renuka Nagpal, and Tanupriya Choudhury

Abstract Reducing the amount of Mobile device energy usage to conserve environmentally friendly resources and maintain a reasonable level of energy consumption is an important issue for information and communication Industry. There are many opportunities to reduce the energy consumption at various levels from hardware, OS, machine code to application level. Innumerable research is going on the optimization of low-level software; e. g., Upgrades machine code. In software engineering, best approach to optimize applications of energy usage is to detect and remove the errors from the code, which when executed with the code increase the energy consumption. In this study, different mobile applications are considered from two different public repositories, namely, GitHub and F-droid app store, and code smells are detected and corrected using different refactoring techniques to evaluate the effect on energy consumption.

Keywords Code smells · Refactoring · Energy · Jdeodorant

1 Introduction

The rising energy use of computing devices such as smartphones generates growing demand for more energy efficiency. Reducing the amount of mobile [1, 2] device

Reeshti · R. Sehgal (✉) · D. Mehrotra · R. Nagpal
Amity University, Uttar Pradesh, Noida, India
e-mail: rsehgal@amity.edu

Reeshti
e-mail: reeshtithukral@gmail.com

R. Nagpal
e-mail: rnagpal1@amity.edu

T. Choudhury
Department of Informatics, School of CS, University of Petroleum and Energy Studies (UPES), Dehradun, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_33

energy usage to conserve environmentally friendly resources and maintain a reasonable level of energy consumption is an important issue for information and communication Industry. There are many opportunities to reduce the energy consumption at various levels from hardware, OS, machine code to application level. Innumerable research is going on the optimization of low-level software; e. g., Upgrades machine code. Another solution is the optimization [3, 4] of hardware components to reduce mobile device energy consumption. In software engineering, best approach to optimize applications of energy usage is to detect and remove the errors from the code, which when executed with the code increases the energy consumption. Errors also called as code smells are introduced in the code at the time of evolution of code in the maintenance phase. Changing the source code to enhance the quality of the software is considered as perfective maintenance. When code evolves in the maintenance phase without following the blue printing of the software architectures, it can lead to many code smells in the software, so it can increase any device's energy consumption. Changing an application's power consumption profile can have a huge effect on the capacity of end-users to use their device. This paper is working towards the Green Mining idea, an effort to quantify the energy consumption when code is executing with code smells and without code smells. Green Mining aims to help software engineers to minimize the usage of power consumption of their own software by measuring the power consumption effect after the software update.

In this study, java-based mobile application is taken from two repositories, Git hub repository and F-droid app store. The applications from these repositories are of various types, such as some applications from the entertainment field, others from the news field, etc. Different applications impact energy consumption differently, as in this paper. In order to find out the effect of code smells on energy consumption, code smells are detected with an open eclipse-based software named as Jdeodorant. Code smells are eliminated by applying appropriate automated refactoring techniques. In this study, two different studied tools are used to refactor the code (i) To remove the code smells from the java-based application, a tool named Jdeodorant is used (ii) To remove the code smells from XML file, a tool named oxygen is used. After eliminating the code smells by applying the automated refactoring technique, energy consumption is again measured and impact of refactoring on energy consumption is evaluated.

This paper is divided into five sections; section 2 presents the work done by various authors; in Sect. 3, experimental work is given; Sect. 4 presents the results of the experimental work; and Sect. 5 presents the conclusion.

2 Literature Survey

The ongoing expansion of users of embedded systems and mobile apps, the dependency on efficient battery life has caught the attention of developers and the researcher to investigate about the diverse factors that can alter and enhance the battery life of an embedded system.

This section contains the literature review of several research papers on the impact of code smells on the embedded systems and how refactoring can help in improving the legacy code. Reducing redundancy of test cases generation using code smell detection and refactoring [5] detects and refactors three major code smells using DART technique to reduce the number of test cases. An Eclipse Plug-in Tool for Generating Test Cases from Source Codes[6] deals to ease the complexity faced by the developers by coming up with manual creation of test cases and checking the functionality of systems, if it caters the user requirements or not. Leveraging Model-Driven Engineering for Energy Optimization [7] uses MoDisco to dynamically perform reverse engineering framework to generate a model out of a Java program using the standard Structured Metrics Metamodel (SMM). On the Impact of Code Smells on the Energy Consumption of Mobile Applications, [8] determines (i) the extent to which code smells affect the source code methods of mobile apps and influence energy efficiency, and (ii) if refactoring operations are applied to remove them directly improves the energy efficiency of refactored methods. Empirical Evaluation of the Energy Impact of Refactoring Code Smells [9] explored Open Source Java ORM-based applications to conclude that refactoring code smells is a vital process to largely improve energy efficiency of the software. The magnitude of the impact depends upon the properties of applications like size, age, etc. Using Automatic Refactoring to Improve Energy Efficiency of Android Apps, [10] analyzed code smells in 140 free and open-source apps using automatic refactoring tool Leafactor. As an outcome, it fixed 222 energy-related anti-patterns and improved the energy footprint of 45 apps. Multiclass classification of mobile applications as per energy consumption [11] Comprehensive review of power consumption in mobile apps using ML tools using Bayesian classification, K-nearest neighbor, decision tree, j458 Random Forest etc. A systematic literature review: Refactoring for disclosing code smells in object oriented software [12] systematically reviews already existing relating to refactoring and code smells and anti-patterns. An Empirical Investigation into Code Smells Elimination Sequences for Energy Efficient Software [13] studies the Object Oriented software systems to analyze the sequence of refactoring and eliminating the diagnosed code smells by refactoring three major code smells using automated tool JDeodorant. Investigation for Software Power Consumption of Code Refactoring Techniques [14] analyzed the power consumption of the 63 refactoring techniques given by Fowler's [15], and after analysis, recommended 33 techniques among them as energy-efficient refactoring ones.

Based on the work done by different authors in the filed on energy optimization, an experiment is conducted which is given in Sect. 3 to find out the effect on code smells on energy consumption.

3 Experimental Study

During this study, two different tools, namely, JDeodorant and Oxygen are explored to detect and correct the code smells through refactoring (Table 1).The tools obtain

Table 1 Tools used during the study

Tools	Language supported	Type	Version used	Automated refactoring	Link to code directly
JDeodorant	Java	Eclipse plugin	v5.0.76	Yes	Yes
Oxygen	XML	Stand-alone	Editor 22	Yes	Yes
Gradle build tool	Java	Eclipse Plug-in	6.4.1	No	No

a file alone, a kit, or a full project based on Android [16, 17] as input and search for appropriate files, i.e., source files Java and XML. To review these files and create a new and improved version, it uses automated refactoring software.

Two separate engines are used: compiling Java-based files first, and exploring XML files second. The Java file refactoring tool is implemented in the JDeodorant open-source project which is an Eclipse plugin to automatically refactor Java application code. They are useful in extensive code cleanup to help get "shorter, easier to maintain and easier to base code expressing. JDeodorant does not allow the refactoring of XML-based files, so it has deployed and implemented a separate refactoring tool called Oxygen XML editor. The architecture of tools used to refactor an android project is shown in Fig. 1.

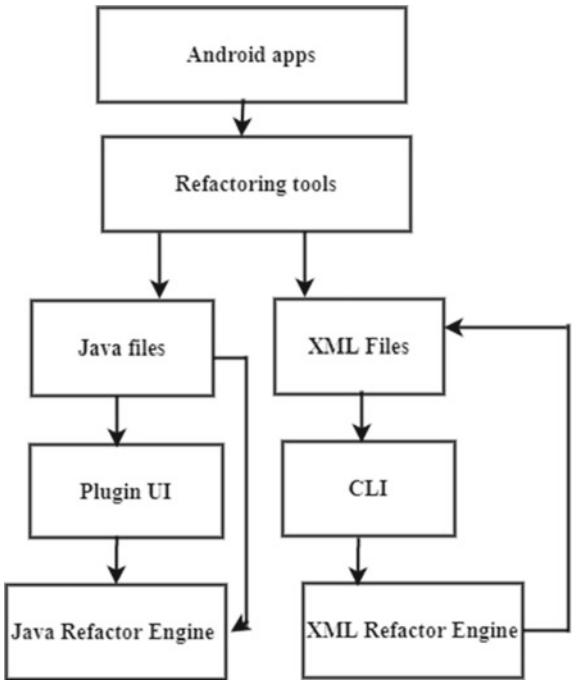


Fig. 1 Architecture diagram of the automatic refactoring tools

Table 2 Applications from different categories are taken from public repositories

S.no	App category	No. of apps studied
1	Utilities	3
2	Entertainment	2
3	Games	2
4	News	2
5	Productivity	2
6	Lifestyle	1
7	Social networking	2
8	Education	2
9	Travel	1
10	Health and fitness	1
11	Business	2
	Total apps	20

To study the effect of code smells on embedded systems, various Free and Open Source Software (FOSS) are studied for android apps. The following methodology is opted to find out the energy consumption for the android-based application.

Step1: The first step towards the experiment is collection of data which is taken from two public repositories, namely (i) F-droid app store which is a repository for free and open-source software (FOSS) applications for the Android platform, and (ii) Secondly, the applications are taken from the GitHub which is public repository. Applications collected from these repositories are from different categories like Utilities, entertainment, games, news, productivity, social networking, games, health and fitness, Business as given in Table 2.
Step2: To understand the effect on energy consumption, code smells are evaluated using an eclipse-based tool, Jdeodorant, from these applications.
Step3: Next step towards the energy evaluation is to correct the detected code smells is refactoring. In this study, to correct the code smells, an automated tool Jdeodorant is used as given in Table 3.

4 Results

In this paper, three tools have been employed that are used to detect and refactor five major code smells, namely, god class, feature envy, long method, and type checking, and code smells from the source code of various android apps on a smart phone with the features mentioned in Table 4.

To detect the code smells in java-based files, an automatic tool IntelliJDeodorant tool is used, while on the other hand, to refactor XML-based files, a tool Oxygen is used. The results so obtained are tabulated in Table 5. Analysis of source code

Table 3 Code smells detection and refactoring techniques

S.no	Code smells detected	Refactoring method
1	Long method	**Extract method** refactoring technique allows to move the fragment of code that can be grouped together, place it in a separate method block, and replace the previous code with method calling
2	Duplicate code	**Extract method** for both groups, followed by Pull Up Field for the fields you are pulling up in the process
3	Feature Envy	**Move method** refactoring technique allows placing the method from one class to the class where the method is put to use more often
4	God class	**Extract class** refactoring technique is applied when a class begins to overweigh with too many methods and its application becomes vague. It involves creating a fresh class and moving methods and the data to new class
5	Type checking	**Replace type code** with State/Strategy refactorings

Table 4 Specification of smart phone

Battery voltage	3300 mAh battery with dash charging
Battery type	Li-Polymer
Processor	Snapdragon 845
CPU	4 × 2.8 GHz Kryo 385
Architecture	64 bit
Storage	8 GB RAM + 128 GB ROM
Camera	16 + 20 MP dual camera
Screen	Optic AMOLED 6.28″ display

before refactoring and after refactoring is performed. From the data collected, we see the reduction in computation time of several apps which is significantly dependent on the hardware of the device, graphical user interface (GUI) of the android apps, and power consumption. For the purpose of accuracy of the data collected for power consumption before and after refactoring, we perform the study for 10 min for five times each to take the average reading, thus calculating the impact on the power consumption.

The pattern so observed during the study is largely dependent on the hardware of the device on which the study is performed and not the software, but we can conclude that the pattern remains the same with similar set of android apps because the refactoring technique deployed optimizes the computation time which reduces competitively upon refactoring code smells like god class and long method.

As from the Table 5, it is observed that energy consumption is significantly improved in the application, namely, entertainment application, gaming application, educational, and travel applications.

Table 5 Effect on energy consumption before and after refactoring

S.no	App category	No. of apps studied	Power consumed (average)		Impact on power consumption
			Before refactoring (in mJ)	After refactoring (in mJ)	
1	Utilities	3	16,876.67	16,227.67	Low
2	Entertainment	2	871,001	850,792	High
3	Games	2	1,058,692	1,063,031	Very high
4	News	2	164,437	162,223	Low
5	Productivity	2	693,193	668,722	Medium
6	Lifestyle	1	29,001	28,100	High
7	Social networking	2	610,420	622,208	Medium
8	Education	2	822,346	808,214	High
9	Travel	1	29,901	27,901	High
10	Health and fitness	1	67,222	66,021	Medium
11	Business	2	221,793	208,302	Medium
	Total apps	20			

5 Conclusion

Smart phones are very essential to people these days and the battery life is a vital concern for the proper functioning like communication and operability anywhere anytime. If the battery life is poor, then a big constraint comes up in this communication cycle. This study helps to understand the impact of four code smells on the power consumption patterns in various categories of android apps. We target the classes, methods, and sections of the applications that mostly impact the battery life, and we categorize the impact as low, medium, and high, and conclude that the behavior of the app varies with the removal of code smells by using automated tools, JDeodorant and Oxygen, with recursive iteration. The improvement in the power consumption depends on the correctness of the tools used, hardware of the device, GUI, number, and the type of code smells present. The android apps that have a lot of God classes in them, have very high impact on the power consumption, do not show significant change in the power consumption even after multiple refactoring applied. This is because breaking down of god class into smaller classes leads to recursive call overhead. The energy behavior in apps with medium impact on power consumption can be improved by refactoring god class first, then feature envy, and at last, long method. The maintainability of apps with low impact on power consumption show considerable drift after few iterations of refactoring using automated tools. The results so obtained may vary with the significant change in the hardware, but show same patterns with the type of apps classified on the basis of usability largely.

References

1. Khanna A, Sah A, Choudhury T (2020) Intelligent mobile edge computing: a deep learning based approach. In: International conference on advances in computing and data sciences, pp 107–116
2. Gianey H, Choudhury T, Elngar A (2019) Node credit based efficient flooding (NCBEF) method for mobile Ad-hoc networks. EAI Endors Trans Ind Netw Intell Syst 6(20)
3. Dewangan BK, Agarwal A, Choudhury T, Pasricha A (2020) Cloud resource optimization system based on time and cost. Int J Math Eng Manag Sci 5(4). https://doi.org/10.33889/IJMEMS.2020.5.4.060
4. Chakraborty S, Kumar V, Choudhury T (2018) Analysis of external content plagiarism using character swarm optimization. In: Smart innovation, systems and technologies, vol 78. https://doi.org/10.1007/978-981-10-5547-8_39
5. Ibrahim R, Ahmed M, Nayak R, Jamel S (2018) Reducing redundancy of test cases generation using code smell detection and refactoring. J King Saud Univ Comput Inf Sci. https://doi.org/10.1016/j.jksuci.2018.06.005
6. Ibrahim R, Ahmed M, Jamel S (2019) An eclipse plug-in tool for generating test cases from source codes. In: Proceedings of the 2019 Asia Pacific information technology conference (APIT 2019). Association for Computing Machinery, New York, NY, USA, pp 25–30. https://doi.org/10.1145/3314527.3314535
7. La Fosse TB (2019) Leveraging model-driven engineering for energy optimization. JDOC19
8. Palomba F et al (2019) On the impact of code smells on the energy consumption of mobile applications. Inf Softw Technol 105:43–55
9. Verdecchia R et al (2018) Empirical evaluation of the energy impact of refactoring code smells. ICT4S
10. Cruz L, Abreu R (2018) Using automatic refactoring to improve energy efficiency of android apps. arXiv:1803.05889
11. Mehrotra D et al (2018) Multiclass classification of mobile applications as per energy consumption. J King Saud Univ-Comput Inf Sci
12. Singh S, Kaur S (2017) A systematic literature review: refactoring for disclosing code smells in object oriented software. Ain Shams Eng J. https://doi.org/10.1016/j.asej.2017.03.002
13. Dhaka G, Singh P (2016)An empirical investigation into code smell elimination sequences for energy efficient software. In: 2016 23rd Asia-Pacific software engineering conference (APSEC). IEEE
14. Park JJ, Hong JE, Lee SH (2014) Investigation for software power consumption of code refactoring techniques. SEKE
15. Opdyke WF (1992) Refactoring object-oriented frameworks 361–396
16. Sarita, Mukherjee S, Choudhury T (2020) An android-based mobile application to help alzheimer's patients. Adv Intell Syst Comput 1090. https://doi.org/10.1007/978-981-15-1480-7_83
17. Chauhan A, Verma S, Sharma S, Choudhury T (2018) Healthcare information management system using android OS. In: Proceedings—2017 international conference on computational intelligence and networks, CINE 2017. https://doi.org/10.1109/CINE.2017.29
18. Marinescu R, Ganea G, Verebi I (2010)InCode: continuous quality assessment and improvement. In: 2010 14th European conference on software maintenance and reengineering, Madrid, pp 274–275. https://doi.org/10.1109/CSMR.2010.44.
19. Fowler M (1999) Refactoring: improving the design of existing code. Addison-Wesley Professional
20. Lee S-J et al (2016) Co-changing code volume prediction through association rule mining and linear regression model. Expert Syst Appl 45:185–194
21. Sehgal R, Nagpal R, Mehrotra D (2019) Measuring code smells and anti-patterns. In: 2019 4th international conference on information systems and computer networks (ISCON), Mathura, India, pp 311–314. https://doi.org/10.1109/ISCON47742.2019.9036304

22. Palomba F et al (2017) Lightweight detection of android-specific code smells: the a doctor project. In: 2017 IEEE 24th international conference on software analysis, evolution and reengineering (SANER). IEEE
23. Azeem MI et al (2019) Machine learning techniques for code smell detection: a systematic literature review and meta-analysis. Inf Softw Technol 108:115–138
24. Cruz L, Abreu R (2019) Catalog of energy patterns for mobile applications. Empir Softw Eng 24(4):2209–2235
25. Pecorelli F et al (2019) Comparing heuristic and machine learning approaches for metric-based code smell detection. In: 2019 IEEE/ACM 27th international conference on program comprehension (ICPC). IEEE
26. Habchi S, Rouvoy R, Moha N. On the survival of android code smells in the wild. In: 2019 IEEE/ACM 6th international conference on mobile software engineering and systems (MOBILESoft). IEEE
27. Chan-Jong-Chu K et al (2020) Investigating the correlation between performance scores and energy consumption of mobile web apps. In: Proceedings of the evaluation and assessment in software engineering, pp 190–199
28. Cruz L et al (2019) Do energy-oriented changes hinder maintainability?. In: 2019 IEEE international conference on software maintenance and evolution (ICSME). IEEE
29. Salza P et al (2019) Third-party libraries in mobile apps. Empir Softw Eng 1–37
30. Ayala I, Amor M, Fuentes L (2019)An energy efficiency study of web-based communication in android phones. Sci Programm 2019
31. Rasool G, Azhar A (2020) Recovering android bad smells from android applications. Arab J Sci Eng 1–27
32. Ergasheva S et al Metrics of energy consumption in software systems: a systematic literature review. IOP Conf Ser Earth Environ Sci 431:202

Human Detecting Sensors and End-To-End Security Model for Design and Manufacturing of IoT-Based Disinfectant Sanitizer Tunnel: An Innovation Against Covid-19

Ramesh Chandra Panda, Rajdeep Chakraborty, Tanupriya Choudhury, K. E. Mathivanan, and Sudeshna Chakraborty

Abstract COVID-19 pandemic affected many people world-wide. It has become a global issue. The major priority of the scientific community is to control this pandemic. The infection can spread through various ways, mainly by touching a contaminated or infected surface or person-to-person contact (droplets from the oral and nasal passage from an infected person). Increase in the rate of infection can by caused by inadequate hygiene practices and cleanliness. The virus is active up to few hours on various surfaces and time may vary according to the nature of the surface. The virus can continue to spread rapidly by contact surfaces and air. Prevention of further viral spread can be attained by effective disinfection of surface and air. It is suggested that sanitizers (0.1% sodium hypochlorite) can effectively arrest COVID-19 virus within a minute. This paper describes a strong IoT-based disinfectant tunnel which is used to disinfect external surface of objects, clothes, or even human skin to provide protection against COVID-19 in social or public places such as malls, schools, hospitals, airports, and offices. Primarily focusing on the significance, structural design, and functioning of the tunnel. To make the tunnel cost-effective and sustainable, solar panels and steel rods have been used. The disinfectant tunnel is highly efficient as it offers no contact disinfection or sanitization. This sensitization tunnel must detect human to sanitize, this paper illustrates a novel approach of using sensors for human detection, and since this project is based on IoT

R. C. Panda
Synergy Institute of Technology, Bhubaneswar, Odisha, India

R. Chakraborty (✉)
Netaji Subhash Engineering College, Kolkata, West Bengal, India

T. Choudhury
Department of Informatics, School of CS, University of Petroleum and Energy Studies (UPES), Dehradun, India

K. E. Mathivanan
Madras Institute of Technology, Chennai, Tamilnadu, India

S. Chakraborty
Department of Computer Science Engineering, Sharda University, Greater Noida, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_34

as backbone, so security model must be there to guard against malicious user. This paper also illustrates end-to-end security model.

Keywords Disinfecting tunnel · Sanitization · COVID-19 disease · Sustainable and cost-effective · Sensors · IoT · End-to-end security model

1 Introduction

Corona virus refers to a large group of viruses that infects humans and animals. A new strain of corona virus causes COVID-19 [1, 2]. To reduce the spread of infection, wearing PPE kits, cleaning hands, and disinfecting people as well as contaminated surfaces are the only possible ways. A useful tool which has emerged to decrease the spread of the virus is "social distancing." In public places, mainly, it is quite difficult to maintain the social discipline. Precautions should be taken and it is only an effective way to deal with the virus. One of the significant and effective methods to protect ourselves on individual basis is disinfection. Disinfection [3] is defined as the process to eliminate pathogenic microorganisms. Quaternary ammonium compounds, alcohols, phenolics, chlorine and chlorine compounds, peracetic acid, formaldehyde, hydrogen peroxide, formaldehyde, and glutaraldehyde act as disinfectants. A powerful and effective disinfecting system is required for reducing the spread of the COVID-19 virus and for bringing back the economy to normalcy. Considering the large population, public places which are heavily crowded, and busy lifestyle, there is a need for sustainable approach for disinfection to combat the spread of infection. One of the cost-effective and sustainable methods [4] is DT (Disinfecting Tunnel) [5] or sanitizing tunnel [6]. A disinfecting or sanitizing tunnel is a walk-through pathway with continuous release of sanitizer or disinfecting agent. It can be used as an efficient way to stop the spread of the virus in public places. An IoT-based disinfecting tunnel can be used to effectively manage contactless disinfection of the surfaces to reduce the risk of infection. After the lockdown is lifted fully, the tunnel can effectively help in the movement of public. Hence, the tunnel is IoT-based, that is, there is no intervention of human in arresting the virus.

Section 2 describes structural design of this project, Sect. 3 gives functioning/pseudo-code of the design, Sect. 4 illustrates sensor-based approach for human detection and also illustrates end-to-end encryption model, Sect. 5 gives the discussions, and Sect. 6 draws the conclusion.

2 Design of Disinfecting Tunnel

Section 2.1 describes the significance of the sanitizers and Sect. 2.2 describes the structural design.

2.1 Significance of Sanitizers

A fluid substance which is designed to kill germs on skin, surface, or object is known as sanitizer. It can also be categorized as household tunnel (hand sanitizer) which kills the germs on skin and surfaces to prevent the spread of diseases. Since there is a tremendous rise in the number of COVID-19 cases with every passing day, all the public places being open and crowded, sanitizer has become the most important part of human life. Being aware of maintaining not only hand hygiene but the entire body surface has become necessary. Sanitizing the entire body surface before entering as well as leaving a workplace, shopping mall, hospitals, etc., can become easier by the implementation of disinfecting or sanitizing tunnel. Awareness among individuals and intelligent use of the resources gifted to us by science can lead to a reduction in infection.

2.2 Structural Design

There have been good measures undertaken during the current corona virus disease (COVID-19) pandemic [7]. The most common measure is called "disinfection tunnel" (DT) or "sanitization tunnel" (ST). They are being stationed outside of the crowded places like vegetable markets, offices, bus and railway stations, and hospitals. The people may walk-through and into this tunnel then and also even ride through and into them on two-wheelers.

Now Sect. 2.2.1 will give component of the structure and Sect. 2.2.2 will illustrate procedure for construction.

2.2.1 Component of the Structure

These are movable structures, made of steel, with the distance varying from 12 to 25 ft. It has a pump (working pressure of 90PSI, maximum pressure of 125 PSI), an adaptor having an input of 110–240 v ac as well as an output 14 v DC, 2.5 amp, and a battery which requires:

- Charging voltage-27c
- Charging mode-12–13.8 v
- Stand by float vol-1404-15.0 v
- Cycle (boost Vol)-1404-15.0 v
- Charging current-405 amps
- The solution is at a height of 2 m (Open flow—1.8 Ipm, Max Amp—102 A, Voltage = 24VDC).

Fig. 1 Design of the sanitization tunnel

2.2.2 Procedure for Construction

The skeleton of the structure is prepared by assembling steel rods in a cuboidal frame and is covered by solar panels on the top. A section holding the bottle of the disinfectant is placed at the adjacent end. A battery and an inverter is connected to direct the flow of current. The cuboidal structure has nozzles attached on the corners of the frame connected to a pump, which directs the flow of disinfectant from the nozzles. The center of the tunnel has a rotor which helps in the rotation of the person standing on it. It is an IoT-based system, which has sensors, assuring the no contact entry and exit from the tunnel. Figure 1 shows the structural design of this sanitization tunnel.

3 Functioning of This Sanitization Tunnel

Sodium Hypochlorite solution will be sprayed in these tunnels as mist. It can be of two types—dynamic and static. The person will be rotated inside the station for 10–15 min which is a static model, the disinfectant will be sprayed through nozzles which are arranged in a whole of the circumference. In the dynamic model, the person will move for 10–12 ft. where the disinfectant will be sprayed throughout their path. The IOT-based tunnel, where the sensors are being attached, will be activated when the person enters. For power supply, it has a battery and solar panels—a sustainable approach (utilization of solar energy—a renewable resource).

The key features of the functioning of this sanitization tunnel are:

Fully Automatic

This sanitization tunnel is fully automatic, which means when a human enters this tunnel, he gets sanitized automatically.

Operates with sensors

This tunnel automatically detects human with sensors [8] installed inside it.

It is not closed at the back to avoid contact after sanitization

This is another feature to avoid contact after sanitization.

Sanitizer tunnel for Office, Home, Public Places, Shops

Since this sanitization tunnel is very cost-effective, so, it can be used in many places.

Color-Multicolor

This sanitization tunnel comes in multicolor.

Material-Steel

Steel is used as the construction material.

Solar Assisted

Solar panel is fitted above this tunnel to charge the battery installed.

The key features of installed Battery are:

- Charging voltage-27c
- Charging mode-12–13.8 v
- Stand by float vol-1404-15.0 v
- Cycle (boost Vol)-1404-15.0 v
- Charging current-405 amps.

Key features of the adaptor used in this sanitization tunnel are:

- Input-110–240 v ac
- Output 14 v DC, 2.5 amp.

The key features of the sanitization pump are:

- Working pressure: 90 PSI
- Max Pressure: 125 PSI
- Solution height-2 m
- Open flow-1.8l pm
- Max Amp-102 A
- Voltage = 24 VDC.

Figure 2 gives the pseudo-code of main module. So, first is setup module which makes the sanitization tunnel in setup phase after each loop. The loop module repeatedly sanitizes person by person. It basically works based on the distance of that person from the sensor. If the distance is less than 10 cm, then the motor starts sanitization and waits for delay, else motor stops and waits for a delay.

Figure 3 gives the circuit design of this IoT-based [9–12] sanitization tunnel.

```
int triggered_pinslot = 2;
int eecho_pinslot = 3;
int mootaar = 10;
int timenow;
int actualdistance;

void setup ( )
{
Seriall.begin (9550);
Pin2Mode (triggered_pinslot, OUTTPUTT);
Pin2Mode (eecho_pinslot, INPPUTT);
Pin2Mode mootaar, OUTTPUTT);
}

void loop ( )
{
digitallWritee (triggered_pinslot, VERYHIGH);
delayMicroseconds (15);
digitallWritee (triggered_pinslot, TOOLOW);
timenow  = pulseIn (eecho_pinslot, VERYHIGH);
actualdistance = (timenow * 0.034) / 2;

if (actualdistance <= 15)
{
Seriall.println (" Motor on");
Seriall.print ("actual distance = ");
Seriall.println (actualdistance);
digitallWritee (mootaar, VERYHIGH);
delay (650);
}
else {
Seriall.println ("Motar off ");
Seriall.print ("actual distance = ");
Seriall.println (actualdistance);
digitallWritee (mootaar, TOOLOW);
delay (555);
}
}
```

Fig. 2 Main-module pseudo-code

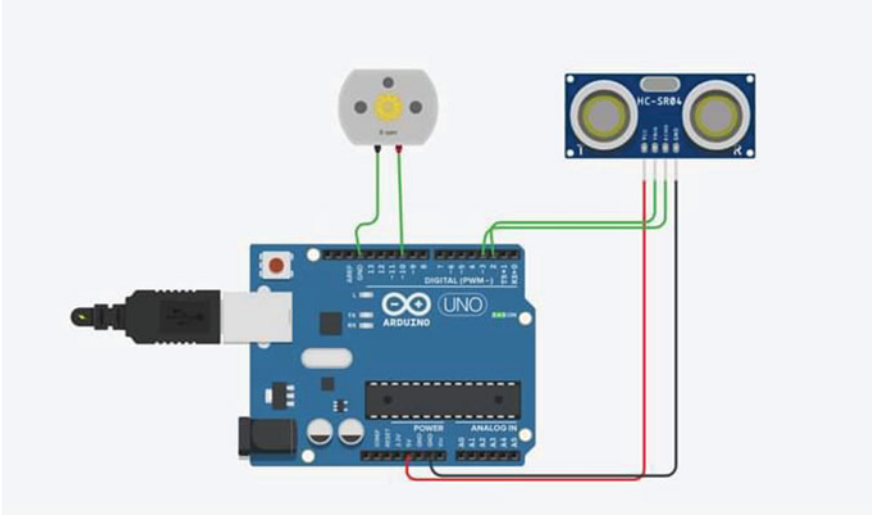


Fig. 3 IoT-based circuit

4 Sensors and Security

Sensors are the main component of this design. 12vdc or 240ac Walk-Through Sanitizer Tunnel Sensor, Sensing Distance: 3 Mtr is used in this design. Table 1 gives the pre-processing time, processing time, and storage requirements for human recognition methods [13] using sensors.

Table 2 illustrates the building blocks of End-to-End Encryption model in IoT security. So, the Security Service Layer of E2EE [14] model consists of authenticated encryption with forward secrecy and backward compatibility. It also uses MAC and

Table 1 Pre-processing time, processing time, training time, and storage requirements

Method	Pre-processing training time (ns)			Storage requirements	Processing time (ns)		
	RRSS	p-Fold	Lio		RRSS	p-Fold	LIO
BDM	27.95	26.65	22.55	Mean to be found, Covariance to be found, CCPDF to be found	5.55	5.75	5.66
RBA	2812.22	3775.77	3455.14	All the rules	0.68	0.99	0.87
LSM	7.11	9.93	5.55	Average of training vector to be found for each class used	0.35	0.33	0.33
K-NN	3.33	11.55	10.25	To find all training vectors	111.35	365.25	188.35
DTW_1	7.77	9.99	6.55	To find average of training vectors of each class used	87.22	85.23	86.65
DTW_2	3.33	12.65	11.66	All training vectors to be found	117.55	158.95	163.22
SVM	7558.17	13,317.88	11,198.61	Use of SVM models	18.55	7.33	8.22
ANN	299,815	222,278	224,267	Determination of network structure and to find connection weights	0.05	0.55	0.55

Table 2 Building blocks of E2EE model in IoT security

Used security service provider layer	Used key management provider layer	Use cryptographic primitive provider layer	Use math provider layer
Authenticated encryption	Key generation	Symmetric: AES	Multiple precision arithmetic
Forward secrecy	Key distribution	Asymmetric: Rabin, RSA, ECC, and ElGamal	Finite field arithmetic
Backward compatibility	Key agreement	Hash: SHA-1, SHA-3	To find Modular number Exponentiation
MAC	KDF	NA	problem on Elliptic curve cryptography and arithmetic
HMAC	NA	NA	NA
Multiple encryption	NA	NA	NA
Modes of operation	NA	NA	NA

HMAC algorithms and multiple encryptions of blocks with modes of operations such as CBC.

The key management layer of E2EE model gives services like key generation, distribution, and agreement between two ends.

Cryptographic primitive layer uses the actual encryption using AES for symmetric encryption, RSA, ElGamal, and ECC for asymmetric encryption, and SHA algorithm for authentication.

Finally, mathematics layer of E2EE model uses finite field arithmetic, modular exponentiation, and elliptic curve arithmetic for implementation of higher layers of security just discussed.

5 Discussion

The tunnel is unique in its own way and can be considered as the best option at the time of pandemic. The key features are:

A. IoT-based [15] system—The future of technology lies in artificial intelligence and IoT. Since it has intelligent sensors, it can assure no contact entry and exit, which reduces surface contamination.
B. Sustainable—Solar energy, which is a renewable resource, is being utilized making it a sustainable approach.
C. Cost-effective—The cost of setting up these contraptions depends on basic models (15,000 Indian rupees) to high-end steel and chrome tunnels with

various sensors (20,000 Indian rupees) which is approximately 5,000–10,000 rupees cheaper than the already launched models.

D. Readily available—Having a simple model and easily available components, it is easy to manufacture, and implant, making it one of the most readily available sanitizing tunnels.
E. Sensors—This design uses effective sensors as given in Sect. 4 to automatically operate in this sanitization tunnel.
F. End-to-End Encryption Model—This sanitization tunnel is IoT controlled, so, security mechanism must be there. E2EE Model is considered the best security mechanism which is used in this sanitization tunnel.

6 Summary and Conclusion

The futuristic approach towards this tunnel can be conversion of the cuboidal frame into simple brackets, one solar panel being fitted on the top. Fitting the pump, battery, and inverter on the side wall of the tunnel. An additional box fitted on the side wall for storing the extra sanitizers. Capillary action can be used for the discharge of sanitizer from the nozzle and one rotating base being fitted at the bottom which makes the person rotate without rotating itself. Looking at the current scenario, there is a significant need for disinfecting our body surfaces as well as the objects around us. It acts as a reliable option to maintain overall hygiene while continuing with our day-to-day necessities. The tunnel being fully automatic, sensor-operating, opened from one end (to avoid contact after sanitization), solar-assisted, and steel-body makes it a sustainable, cost-effective, and intelligent option.

References

1. Chang L, Yan Y, Wang L (2020) Coronavirus disease 2019: coronaviruses and blood safety. Transfus Med Rev
2. Yan Y, Chang L, Wang L (2020) Laboratory testing of SARS-CoV, MERS-CoV, and SARS-CoV-2 (2019-nCoV): current status, challenges, and countermeasures. Rev Med Virol 30(3):e2106
3. Lin Q, Lim JY, Xue K, Yew PYM, Owh C, Chee PL, Loh XJ (2020) Sanitizing agents for virus inactivation and disinfection. View e16
4. Verhoef PC, Bijmolt TH (2019) Marketing perspectives on digital business models: a framework and overview of the special issue
5. Maurya D, Gohil MK, Sonawane U, Kumar D, Awasthi A, Prajapati AK et al (2020) Development of autonomous advanced disinfection tunnel to tackle external surface disinfection of COVID-19 virus in public places. Trans Indian Natl Acad Eng 5(2):281–287
6. Mallhi TH, Khan YH, Alotaibi NH, Alzarea AI (2020) Walkthrough sanitization gates for COVID-19: a preventive measure or public health concern?. Am J Trop Med Hyg tpmd200533
7. He H, Harris L (2020) The impact of Covid-19 pandemic on corporate social responsibility and marketing Philosophy. J Bus Res
8. Yun J, Lee S-S (2014) Human movement detection and identification using pyroelectric infrared sensors. Sensors 14:8057–8081. https://doi.org/10.3390/s140508057

9. Chhabra AS, Choudhury T, Srivastava AV, Aggarwal A (2018) Prediction for big data and IoT in 2017. In: 2017 International conference on infocom technologies and unmanned systems: trends and future directions, ICTUS 2017. https://doi.org/10.1109/ICTUS.2017.8286001
10. Krishnan DSR, Gupta SC, Choudhury T (2018) An IoT based patient health monitoring system. In: Proceedings on 2018 international conference on advances in computing and communication engineering, ICACCE 2018. https://doi.org/10.1109/ICACCE.2018.8441708
11. Khurana M, Choudhury T, Malik P (2019) A review on network security challenges and the internet of things (IoT). In: Proceedings of the 4th international conference on contemporary computing and informatics, IC3I 2019. https://doi.org/10.1109/IC3I46837.2019.9055675
12. Narang S, Nalwa T, Choudhury T, Kashyap N (2018) An efficient method for security measurement in internet of things. In: 2018 international conference on communication, computing and internet of things (IC3IoT), pp 319–323
13. Choudhury T, Dangi G, Singh TP, Chauhan A, Aggarwal A (2018) An efficient way to detect credit card fraud using machine learning methodologies. In: Proceedings of the 2nd international conference on green computing and internet of things, ICGCIoT 2018. https://doi.org/10.1109/ICGCIoT.2018.8753077
14. Lee J, Sungsook RC, Kim KK (2017) Security analysis of end-to-end encryption in Telegram. In: Proceedings security AO, Corpus ID: 26205922
15. AjazMoharkan Z, Choudhury T, Gupta SC, Raj G (2017) Internet of things and its applications in E-learning. In: 2017 3rd international conference on computational intelligence & communication technology (CICT), pp 1–5

Four-Factor Authentication with Emerging Cybersecurity for Mobile Transactions

Sanyam Jain, Raju Gautam, Shivani Sharma, Ravi Tomar, and Tanupriya Choudhury

Abstract Cybersecurity is very much essential for Mobile Transactions to complete seamlessly. Mobile Commerce (Mcom.) is the very basic transaction type, which is very commonly used (two in five people use mobile as transaction medium). To secure this, there are various technologies used by this research. The four "factors" formally known as Multi-Factor-Authentication are: two of them are Traditional methods (User Login password and One Time Password (aka OTP)) with addition of Geolocation and Facial Recognition. All the data is converted to a text file, which is hidden in an image (using Babushka algorithm). The end-point then decrypts the image using the same algorithm.

Keywords Babushka · 4FA · Mobile transactions · Facial recognition · OTP

1 Introduction

This research paper intends to make mobile transactions more secure using multi-factor authentication utilizing mobile resources (sensors) optimally at the time of transaction.

S. Jain
Research Scholar, Quantum Information and Computation, Dept. of Physics, IIT Jodhpur, Karwar, Rajasthan, India
e-mail: sanyamjaincs@gmail.com

R. Gautam · S. Sharma · R. Tomar (✉) · T. Choudhury
School of Computer Science, University of Petroleum and Energy Studies (UPES), Dehradun, India
e-mail: ravitomar7@gmail.com

R. Gautam
e-mail: rajugautam45@gmail.com

S. Sharma
e-mail: 1998shivisharma@gmail.com

T. Choudhury
e-mail: tanupriya1986@gmail.com

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_35

The whole module (Program) will run on the Bank's gateway server. The current technology we use in 2018 (2-Factor Authentication) can be extended with these two more factors, namely, Geolocation and Facial recognition. The old methods are very easy to break. Once a hacker (Thief) gets your sim, everything is leaked. According to an experiment (Research by Wired), once a sim is compromised (so as sms), one can get your Google. There are various methods then to retrieve your text message, e.g., Push Bullet is used to monitor your mobile activity to computer to increase mobility. There are various proved Methods to crack sim Text Messages using intermediary networks like "Man in the Middle attack." Thus, two more methods need to be increased to make transactions more secure; there is geolocation matcher and facial recognition matcher, which altogether matches the user current location of point of purchase as well as the face, which is stored in bank database.

2 Till Now

There are various techniques used by banking institutions to secure mobile transactions. Banks usually come with two basic solutions called 2FA which includes user details (generally user ID password and another is one time password) for end customer/user. These details are generally breakable at very low-level computation. End user on the bank website sets user name and password and the 4–6 digit code is received on phone registered with bank. Till now, there is no additional security for this. Every mobile transaction directly or indirectly depends on only mobile phone. This strategy worked well when there were people with no smartphones. The user detail gets encrypted and transferred to bank in secured environment. There are high rates of compromising and cracking bank transactions. Day by day, banks update their security algorithms, even then also the transactions are being compromised either by networks or by phishing user authenticity credentials [1–3].

3 Proposed Work

To overcome these low-level security measures, in this paper, two advanced methods are proposed so as to wipe hacking of mobile transactions. User after performing basic two factors, i.e., user credentials and OTP, it will get redirected to the gateway consisting of geolocation (Performed background) and Facial Recognition in main stream. All will be done on the smartphone/smart hand-held device consisting of GPS (To match location) and Front Camera (For facial recognition). The Computation part will run on Bank server, the customer will only get the instance of program. Starting from second factor, the geolocation part, user will have to update the current location of performing transaction to the bank database which is personalized for each user of bank—a dashboard. The same location is matched with the bank server at the time of purchasing by the customer. This will run in background and user will not even get

notified about the location factor. (This is done to make transaction faster and less informative so as user thinks of three factors). All the data gets converted to a text file. Text file consists of four things: Username + Password + Image ASCII code (Fetched from facial Recognition) + GeoLocation (Of the mobile phone). The text file is then embedded/encrypted with the image file and sent to the Merchant. Merchant or Receiving bank then Decrypts the Image file, gets the user information, and authenticate that. This way the transaction is completed and successfully transfers money. The actual program should deploy on the server, which executes several instances for each invoking user, and same methodology occurs every time. The proposed work is different from among all the existing work by "Babushka/Matryoschka" algorithm. The algorithm has been developed under many trainings to get text file encoded to image file in the meta data of the image. The encryption uses HMAC-SHA256 to authenticate the hidden data. We cannot ignore traditional approaches completely, the underlying architecture of login, and mobile phone OTP is important when user does not want to add more security measures [4–6].

4 Methodology

This section of the paper consists of the working of all four factors describing each one by one as follows:

1. User Login and Password ➔ this is a very basic authentication factor used to verify the user against the database. The local database or internet database anything can work to store and verify the user credentials. (Here we used sqlite3 for Ubuntu). The username and password are matched with database and are verified and proceeded to the next step.
2. Mobile Authentication or One Time Password ➔ OTP helps the bank to authenticate user accessibility to the point of sale as to determine the ownership or privacy for the original user to continue with the transaction. OTP may be 4 or 6 digit code as per standard. OTP is now submitted to the portal. OTP is hosted by the Twilio API [7]. There are numerous services for the same. OTP is just requested to the Twilio server and user buffers a hash key for same. Now, since hash can only be broken by Twilio service, thus hash and the OTP received are matched on the device.
3. Facial Recognition ➔ The Third Milestone is the Facial recognition system as the authentication factor.
   After authenticating OTP, the user will be asked for his/her image capture, the camera automatically takes it, and stores it in the file system for the future authentication purpose. Now, the webcam analyzes the registered user's face in the organization database through the webcam, and identifies his features, thereby unlocking the third milestone as secured user. Facial recognition is used along with OpenCV library as the source tool for identifying the Human's

face (Link: https://opencv.org). This technique allows very less chance to the attacker to bypass a particular system.

4. Geolocation (Of the mobile phone) ➔ the geolocation part, user will have to update the current location of performing transaction to the bank database which is personalized for each user of bank—a dashboard. The same location is matched with the bank server at the time of purchasing by the customer. This will run in background and user will not even get notified about the location factor. This creates the integrity of the location from which a user operates. The smartphone app (here: Location Tracker for android) will be in sync to the organization or bank database server which will be operating on hosted website. The User, manually after the OTP validation, will update the Location Tracker app with the Update button on it, this will set the latitude, the longitude of the user's current location based on the GPS, and this will be reflected in the database server. This is a crucial step for a user whenever attacker tries to use VPN as a layer to infiltrate into the user account. The security will be hard to crack as mobile device GPS is ON and this matches the transaction done from the application used to perform transaction.

4.1 Existing Security

The traditional approach towards doing transaction or logging in is entering username and strong alphanumeric passwords, but they are vulnerable to different attacks like social engineering, rainbow table, brute force attack, etc. This approach also has OTP as a second factor sometimes, but they are liable to Man in the Middle attack which causes leaking of the transferred OTP. There is a need to replace this technique with some additional liabilities. Devoted programmers have little issue bypassing through the weaker executions, either by capturing codes or by misusing account-recuperation frameworks. The guarantee of two-factor started to disentangle right off the bat. By 2014, lawbreakers focusing on Bitcoin administrations were discovering courses around the additional security, either by catching programming tokens or more detailed record recuperation plans. Now and again, assailants followed telephone bearer accounts specifically, setting up a minute ago call-sending game plans to capture codes in travel [8–12].

4.2 Proposed Security

4FA aims at higher assurance in comparison to single-factor schemes. Our proposed system uses multiple factor which acts as an upper model in the shortcomings of alphanumeric and OTP factors. The 4FA will be using facial recognition and the geolocation feature as extra milestone in addition to the above-mentioned technique. Geolocation creates the integrity of the location from which the user operates (Fig. 1).

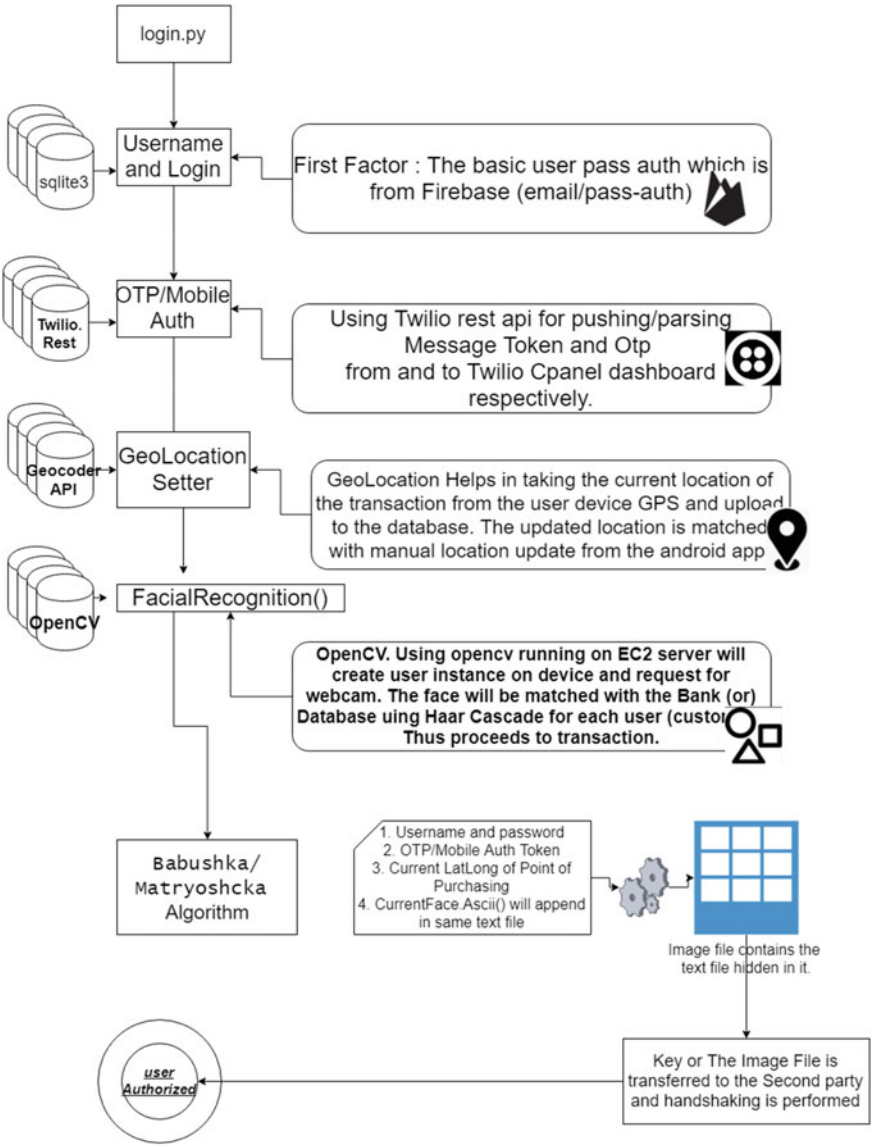


Fig. 1 The working model

1. If OTP is verified, system asks for face recognition and run geolocation check in background automatically.
2. If OTP is not verified, system stops transaction.
3. If the face recognition fails, system stops transaction.

4. The user verifies location through app, if geolocation verifies, then proceed the user, else system stops transaction.

Essentially, all the components of the 4FA module work together to identify a person. This implies that even if one of the components gets an error, the complete process of authentication will be terminated. For example, if in the middle of a transaction the program is unable to match the geolocation, the program will terminate the transaction, and hence, the transaction stays blocked at that point in time [13–16].

4.3 Matryoschka Algorithm

At last, the application will collect all data in a text file, and the text file is sent to the merchant for successful transaction. The intermediary is "Babushka/Matryoschka" algorithm [17]. Matryoschka are basically sets of traditional Russian wooden dolls of decreasing size placed one inside the other. A Matryoschka doll can be opened to reveal a smaller figure of the same sort inside, which has, in turn, another figure inside, and so on. The Russian Matryoschka Museum recently exhibited a collection of similarly designed Matryoschka sets, differing only in the number of nested dolls in each set. Unfortunately, some over-zealous (and obviously unsupervised) children separated these sets, placing all the individual dolls in a row. There are n dolls in the row, each with an integer size. You need to reassemble the Matryoschka sets, knowing neither the number of sets nor the number of dolls in each set. You know only that every complete set consists of dolls with consecutive sizes from one to some number m, which may vary between the different sets. Matryoschka algorithm invokes Encryption HMAC-SHA256 bit [18] to authenticate the hidden data which further implements the XTEA algorithm. To hide the sensitive data. All the previous authentication factors will be passed through this encryption tunnel in order to hide information in the image form. The text will not be visible throughout the transfer of image, due to the hard coded encryption by HMAC-SHA256 and XTEA algorithm. The non-significant values of the pixels on image are used to store all sensitive information being encrypted by XTEA [19] algorithm which is being sent to the server for completing the transaction. The SHA in turn is being implemented to check the integrity of the data sent.

1. The Username and password from first factor are captured.
2. The data from second factor will be, i.e., OTP will be selected.
3. The data from third factor, i.e., facial recognition, the image captures will be used, but first it is converted to ASCII ART of their image.
4. The data from fourth factor, i.e., latitude and longitude will be captured.

All the above-mentioned three points will be first stored in a text .txt file as the primary source, and then this algorithm automatically encrypts that text data inside the selected image, so as to preserve its identity, and then it is being sent to the merchant. All the above-mentioned four points will be first stored in a text .txt file

as the primary source, and then this algorithm automatically encrypts that text data inside the selected image, to preserve its identity.

Syntax: python babushka.py –hide –m <mac-password> -k <password> <secret> <change>

But first, the organization has to install python library called pillow.py using "sudo pip install pillow," this will install all the dependencies required for Matryoschka. The file then gets encrypted and stored, into a mentioned image, as a factor for high security. Now, let us say if a merchant wants to decode that image, there is a module called DecodeNow.py, which will only be used by merchant for the pure decoding purpose.

Syntax: python babushka.py –open –m <mac-password> -k <password> <image>

4.4 Software Engine Specification

1. The software described in this application is based on the python client integrated with third party messaging service API, i.e., twilio.
2. OTP is just requested to the Twilio server and user buffers a hash key for same and is secured for sending OTP messages.
3. Python being more popular language can be integrated in any system making four-factor authentication more scalable.
4. The API use makes it easy for system to communicate with other services. One of the API being used is Geolocation API of google.
5. The geolocation API uses different techniques to identify user's current location like public IP address of the device, cell tower IDs, GPS information, a list of Wi-Fi access points, signal strengths, and MAC IDs (Wi-Fi and/or Bluetooth).
6. The open source android technology is used for the geolocation factor. The app location tracker of our institution will be in sync to the organization/bank database server which will be operating on hosted website.
7. The machine leaning technique is being implemented on the system for the face recognition factor through an image dataset, which in turn are trained through a variety of positive and negative images, thereby generating a common standard format, i.e., XML which contains facial feature (Cascade). This format can be used with any technology thereby making the system scalable and flexible.
8. The XML recognizes a real-time face. And Frame of the dataset can be stored as for further use.
9. PyQt4 is used to give the graphical Dialog Box Layout to the system performing transactions. Though it can be modified as per Application's need because of Open Source development.

5 Result

The system is developed with integration of different modules. The complete transaction costs at end user is 30 s. And complexity at bank server costs *O(n)*. The result was observed under various cloud servers from AWS. A small bank type environment was created to perform the transactions and handle mobile level processes. Successful transactions were recorded by the Amazon Dynamo DB. Tests were performed specifically for mobile transactions, and hence, the bank systems have a prior information about the device of the transaction using ip address.

6 Conclusion

The use of multifactor authentication helped to achieve the purpose of more secure electronic transactions. The four-factor validation arrangement prepares client by giving solid validation to expansive scale. The 4FA could unquestionably decrease the recurrence of online fraud and data leakage furthermore, other online blackmails in the light of the way that the causality secret key could never again be adequate to give a criminal access to user information.

References

1. Ravi T, Sastry HG, Prateek M (2020) A novel protocol for information dissemination in vehicular networks. In: Lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 11894. LNCS, Springer, Cham, pp 1–14
2. Tomar R, Prateek M, Sastry HG (2017) A Novel Approach to multicast in VANET using MQTT. Ada User J 38(4):231–235
3. Tomar R, Sastry H, Prateek M (2020) Establishing parameters for comparative analysis of V2V communication in VANET. J Sci Ind Res (JSIR) 79(01):26–29
4. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of Weiler-Atherton algorithm to address loose polygons. J Sci Ind Res 78:771–774
5. Choudhury T, Aggarwal A, Tomar R (2020) A deep learning approach to helmet detection for road safety. J Sci Ind Res (JSIR) 79(06):509–512
6. Kumar V, Choudhury T (2019) Real-time recognition of malignant skin lesions using ensemble modeling. J Sci Ind Res (JSIR) 78:148–153
7. Pyhton Documentation for Twilio, The Twilio Python Helper Library. https://twilio.com
8. Sehrawat A, Choudhury TA, Raj G (2017) Surveillance drone for disaster management and military security. In: Proceeding—IEEE international conference on computing, communication and automation, ICCCA 2017. https://doi.org/10.1109/CCAA.2017.8229846
9. Kumar S, Ravi T (2018) The role of artificial intelligence in space exploration. In: 2018 international conference on communication, computing and internet of things (IC3IoT). IEEE, pp 499–503
10. Bansal P, Aggarwal B, Tomar R (2019) Low-voltage multi-input high trans-conductance amplifier using flipped voltage follower and its application in high pass filter. In: 2019 international

conference on automation, computational and technology management (ICACTM). IEEE, pp 525–529
11. Gupta SS, Thakral A, Choudhury T (2018) Social media security: analysis of threats and security measures. In: Proceedings on 2018 international conference on advances in computing and communication engineering, ICACCE 2018. https://doi.org/10.1109/ICACCE.2018.8441710
12. Khurana M, Choudhury T, Malik P (2019) A review on network security challenges and the internet of things (IoT). In: Proceedings of the 4th international conference on contemporary computing and informatics, IC3I 2019. https://doi.org/10.1109/IC3I46837.2019.9055675
13. Singh RK, Panchal VK, Singh BK (2019) Analysis of GA performance on its various parameters for solving travelling salesman NP-hard problem. Int J Comput Sci Eng 7(1):509–512. ISSN:2347–2693(UGC-63193)
14. Singh RK, Panchal VK, Singh BK (2019) Analysis of the various population sizes in genetic algorithm for solving N-queen NP-hard problem. Int J Res Electron Comput Eng IJRECE 7(1):564–566. ISSN:2393–9028(Print)| ISSN:2348–2281. (UGC(44816)
15. Singh BK, Luthra M (2019) MAHI as a sketching language. Int J Comput Sci Eng IJCSE 7(3):1102–1109. ISSN:2347–2693. (UGC-63193)
16. Mbunge E (2018) A robust and scalable four factor authentication architecture to enhance security for mobile online transaction
17. Matroschka algorithm or Babushka algorithm. https://github.com/samy280497/Matroschka
18. SHA-256, Secured Hashmap algorithm. Wikipedia.org
19. XTEA Algorithm, eXtended TEA. Wikipedia.org

FPGA Implementation of Image Processing Filters using IoT

Ajay Rupani, Sayamuddin Ahmed Jilani, F. H. A. Shibly, and Tanupriya Choudhury

Abstract The greater part of the articles or substances that are available around us will specifically or by implication be accessible on the system in the present scenario of IoT. To associate with Internet of Things, different sensor network advancements and Radio Frequency Identification (RFID) will be produced. By these imaginative innovations, data and correspondence frameworks are impalpably encompassed in nature around clients. In this paper, appropriate filter selection work is done using web server application by means of Raspberry-Pi. Verilog code is written in Xilinx Vivado software to design various image processing filters.

Keywords FPGA · IoT · Image processing

1 Introduction

The term IoT was first presented by Kevin Ashton in 1999. IoT can be utilized for various applications like transport, medicinal services and utilities and so on. Be that as it may, the importance of 'Things' has been enhanced as innovation advanced the goal of seeming well and good data without the help of human association continues as before [1]. Internet of things is a framework of interconnected objects, advanced and mechanical machines, substances, figuring gadgets and individuals with uncommon aptitudes. System architecture has been shown in the below Fig. 1.

A. Rupani (✉)
JIET Jodhpur, Jodhpur, India

S. A. Jilani
Makaut West Bengal, Kolkata, India

F. H. A. Shibly
South Eastern University of Sri Lanka, Oluvil, Sri Lanka
e-mail: shiblyfh@seu.ac.lk

T. Choudhury
Department of Informatics, School of CS, University of Petroleum and Energy Studies (UPES), Dehradun, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_36

Fig. 1 System architecture of IoT [2]

The architecture of IoT [3–5] contains various things associated in an immediate or roundabout way, regardless of whether it might be any transportation gadget or any object or any computing gadget or any Wi-Fi gadget, in a system with or without human intercession [6].

1.1 Image Processing Operations

The filtering of image data is a well-known procedure utilized as a part of all image processing systems [7]. The decision of the filter is decided by:

- Behaviour of the filter.
- The way how the filter perform the task.
- Type of the data [8–10].

1.1.1 Mean Filter

This filter is generally utilized as a part of graphics programming, commonly to reduce the noise present in an image and diminish detail [11]. The average filter is also known as mean filter. The mathematical function of mean filter is given by:

Fig. 2 Original image

$$f(i,j) = \frac{1}{3*3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

1.1.2 Gaussian Filter

The mathematical function for Gaussian filter is given by:

$$f(x,y) = \frac{1}{16} *$$

0	1	0
1	9	1
0	1	0

The Figure shows the example of a Gaussian filter (Figs. 2 and 3):

1.1.3 Edge Detection

A process of computing [12–14] the gradient of image intensity at every point of the image is known as edge detection [15]. The edge identification depends on traversing the whole picture with the help of a filter in horizontal and vertical direction and is in this way generally cheap regarding computations [16].

The Fig. 4 shown below demonstrates a case of this procedure.

Fig. 3 Output of Gaussian filter

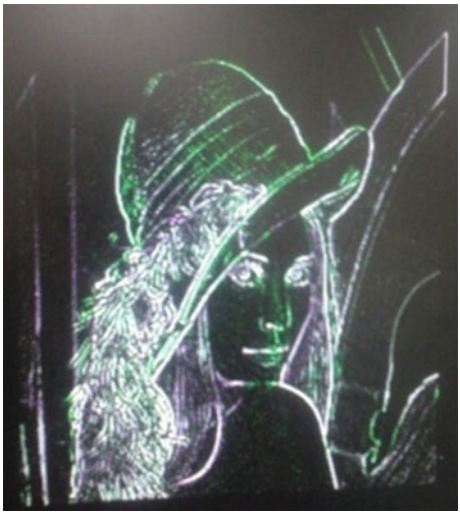


Fig. 4 Edge detection operation result

The mathematical function for the edge detection operation is given by:

$$f(x,y) = \begin{array}{|c|c|c|}\hline +1 & +1 & +1 \\ \hline +1 & -8 & +1 \\ \hline +1 & +1 & +1 \\ \hline \end{array}$$

2 Related Work

IoTs [17–19] enabled VLSI Design can be designed using Verilog, and we can use raspberry-pi as an IoT device to make the web server-based application [20] (Fig. 5).

Interfacing of FPGA development board with Raspberry-Pi model 3B has been done to propose the following task in Fig. 6.

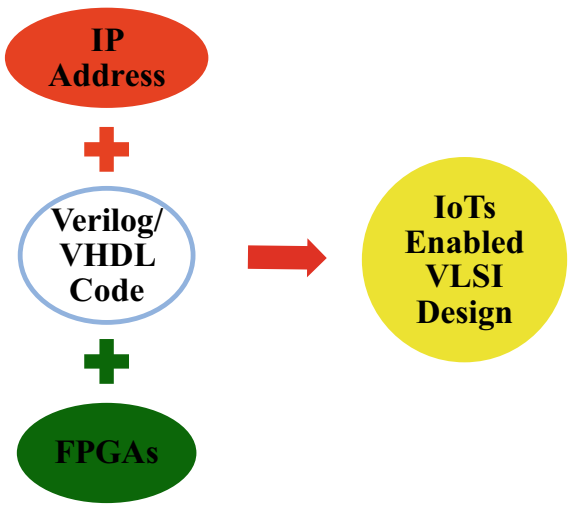


Fig. 5 IoTs enable VLSI design [21]

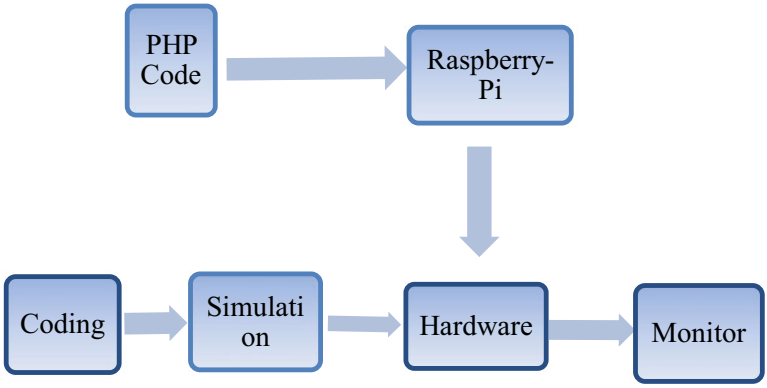


Fig. 6 Related work [16]

3 Methodology

In the proposed system architecture, a decoder logic circuit selects the appropriate filter operation which a user can want to use. After selecting appropriate filter operation, user can click on the appropriate choice using web server application as shown in Fig. 7.

The web server application is designed using PHP code, i.e. shown in Fig. 8. After decoding the correct filter option, user can select the proper filter operation using this web server application. There are five types of filters designed in this research paper.

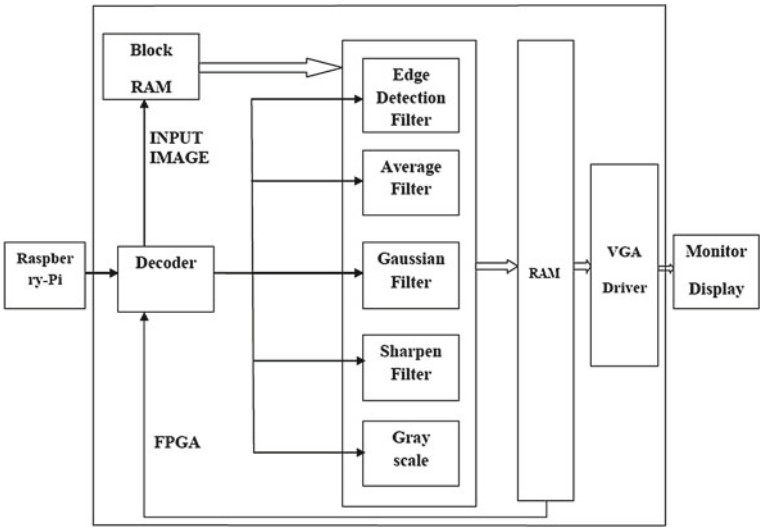


Fig. 7 Proposed system architecture

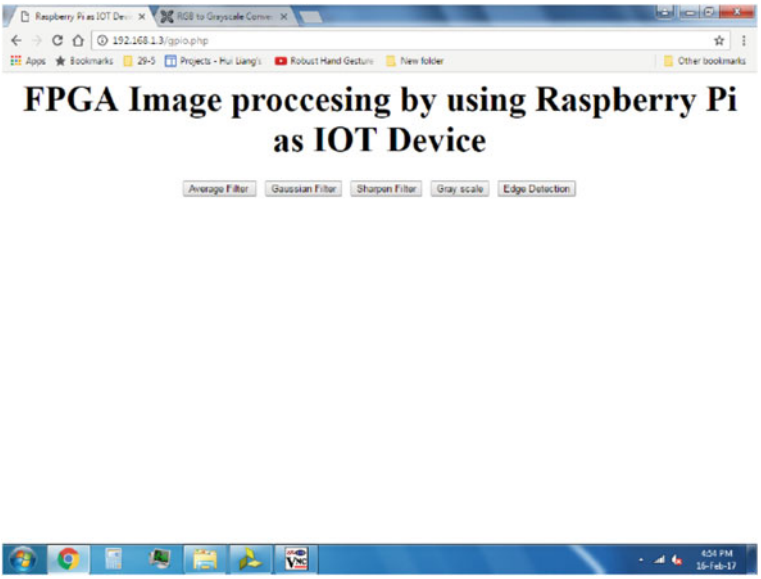


Fig. 8 Web server application snapshot

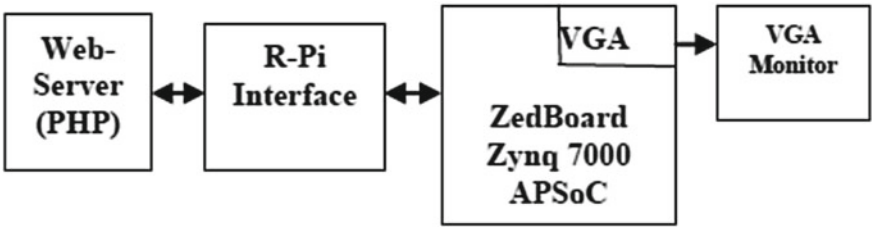


Fig. 9 Block diagram

The input image is loaded through Block RAM. If a user clicks on Average Filter Option, the average filtered output of that image will appear on the screen (Fig. 8).

The block diagram of overall operation is presented in Fig. 9.

The FPGA implementation for proposed architecture is shown in Fig. 10.

4 Analysis and Result

We have designed and implemented various image processing filters in Xilinx Vivado software using Verilog HDL. Also, hardware implementation has been done to verify our Verilog code. The Chip Utilization report has been obtained that is shown in Fig. 11.

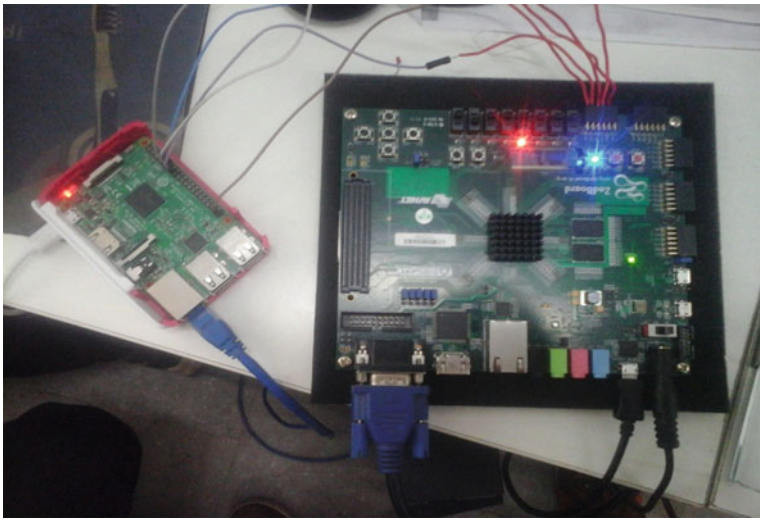
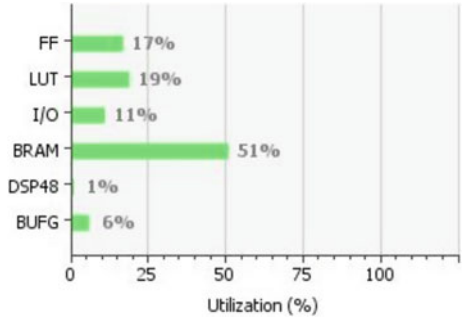


Fig. 10 FPGA implementation of proposed system

Fig. 11 Chip utilization report



The input and output images are displayed as follows in Figs. 12, 13, 14, 15, 16 and 17.

Figure 12 shows the original image (Lena) and other figures represent the output of various image processing filters.

5 Summary and Conclusion

In the Performance analysis [22–24], we have gone through all the parameters which we have discussed in theory and which we have observed from the software during the designing. We have obtained appropriate results using web server-based application. FPGA implementation and its interfacing with Raspberry-Pi have been justified.

Fig. 12 Original image (Lena)

Fig. 13 Gray scale filtered image

We have obtained the chip utilization report along with the output of various filters using web server-based application. More filters can be designed and implemented in future to enhance this method. IoT will be effective in the coming years because of the following reasons:

- Broadband web is widely utilized in the light of the fact that the media transmission framework is getting to be plainly productive.
- By improving remote and sensor innovation, the generation of sensors which worked in Wi-Fi abilities made associating gadgets considerably less expensive to prior. Along these lines, it would be an astute choice to utilize IoT.
- Use of advanced cell phones and other registering gadgets are widely high and, as a result of that, various media transmission organizations are endeavouring to fulfil their clients by giving the best administrations by enhancing framework.

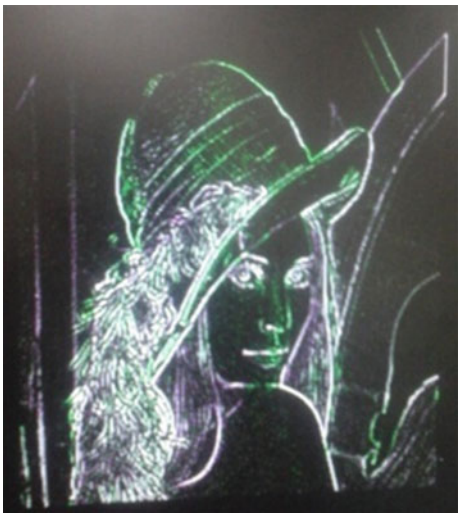


Fig. 14 Edge detection

Fig. 15 Average filter

- Because of goliath system of heterogeneous gadgets, IoT some of the time likewise needs to isolate some correspondence from the current channel. What is more, it is likewise less expensive and effortlessly achievable on account of IoT contrasted with conventional system.

Fig. 16 Output of sharpening

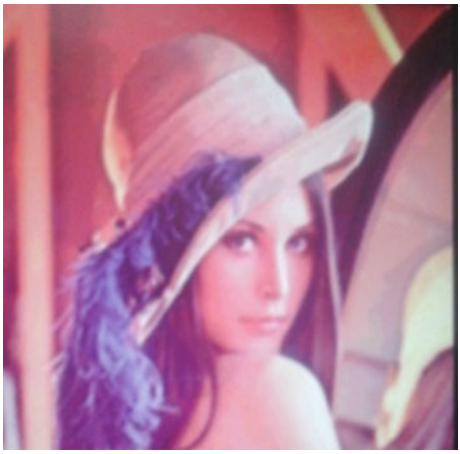


Fig. 17 Output of Gaussian filter

References

1. Girau R, Martis S, Atzori L (2017) Lysis: a platform for IoT distributed applications over socially connected objects. IEEE Internet Things J 4(1):40–51. https://doi.org/10.1109/JIOT.2016.2616022
2. Mouradian C et al (2020) An IoT platform-as-a-service for NFV-based hybrid cloud/fog systems. IEEE Internet Things J 7(7):6102–6115. https://doi.org/10.1109/JIOT.2020.2968235
3. Chhabra AS, Choudhury T, Srivastava AV, Aggarwal A (2018) Prediction for big data and IoT in 2017. In: 2017 international conference on Infocom technologies and unmanned systems: trends and future directions, ICTUS 2017, 2018-January. https://doi.org/10.1109/ICTUS.2017.8286001
4. Krishnan DSR, Gupta SC, Choudhury T (2018) An IoT based patient health monitoring system. In: Proceedings on 2018 international conference on advances in computing and communication engineering, ICACCE 2018. https://doi.org/10.1109/ICACCE.2018.8441708

5. Khurana M, Choudhury T, Malik P (2019) A review on network security challenges and the Internet of Things (IoT). In: Proceedings of the 4th international conference on contemporary computing and informatics, IC3I 2019. https://doi.org/10.1109/IC3I46837.2019.9055675
6. Ding J, Nemati M, Ranaweera C, Choi J (2020) IoT connectivity technologies and applications: a survey. IEEE Access 8:67646–67673. https://doi.org/10.1109/ACCESS.2020.2985932
7. Coady J, O'Riordan A, Dooly G, Newe T, Toal D (2019) An overview of popular digital image processing filtering operations. In: 2019 13th international conference on sensing technology (ICST), Sydney, Australia, pp 1–5. https://doi.org/10.1109/ICST46873.2019.9047683
8. Singh RK, Panchal VK, Singh BK (2019) Analysis of GA performance on its various parameters for solving travelling salesman NP-Hard Problem. Int J Comput Sci Eng 7(1):509–512, ISSN:2347-2693 (UGC-63193)
9. Singh RK, Panchal VK, Singh BK (2019) Analysis of the various population sizes in genetic algorithm for solving N-Queen NP-Hard problem. Int J Res Electr Comput Eng IJRECE 7(1):564–566, ISSN:2393-9028(Print)| ISSN:2348-2281(Online). (UGC-(44816)
10. Singh BK, Luthra M (2019) MAHI as a sketching language. Int J Comput Sci Eng IJCSE 7(3):1102–1109, ISSN:2347-2693. (UGC-63193)
11. Thanh DNH, Hieu LM, Engínoğlu S (2019) An iterative mean filter for image denoising. IEEE Access 7:167847–167859. https://doi.org/10.1109/ACCESS.2019.2953924
12. Tomar R, Sastry HG, Prateek M (2020) A novel protocol for information dissemination in vehicular networks. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 11894, LNCS. Springer, Cham, pp 1–14
13. Tomar R, Prateek M, Sastry HG (2017) A novel approach to multicast in VANET using MQTT. Ada User J 38(4):231–35
14. Tomar R, Sastry HG, Prateek M (2020) Establishing parameters for comparative analysis of V2V communication in VANET. J Sci Ind Res (JSIR) 79(01):26–29
15. Xiong C, Chen L, Pang Y (2010) An adaptive bilateral filtering algorithm and its application in edge detection. In: 2010 international conference on measuring technology and mechatronics automation, Changsha City, pp 440–443. https://doi.org/10.1109/ICMTMA.2010.41
16. Kwon S-Y, Kim Y-H, Lee Y-H, Kim M-J (2018) Architecture of object recognition to implement in hardware. In: 2018 international conference on computational science and computational intelligence (CSCI), Las Vegas, NV, USA, pp 421–425. https://doi.org/10.1109/CSCI46756.2018.00085
17. Narang S, Nalwa T, Choudhury T, Kashyap N (2018) An efficient method for security measurement in internet of things. In: 2018 international conference on communication, computing and Internet of Things (IC3IoT), pp 319–323
18. AjazMoharkan Z, Choudhury T, Gupta SC, Raj G (2017) Internet of things and its applications in E-learning. In: 2017 3rd international conference on computational intelligence & communication technology (CICT), pp 1–5
19. Choudhury T, Dangi G, Singh TP, Chauhan A, Aggarwal A (2018) An efficient way to detect credit card fraud using machine learning methodologies. In: Proceedings of the 2nd international conference on green computing and Internet of Things, ICGCIoT 2018. https://doi.org/10.1109/ICGCIoT.2018.8753077
20. Asoba S, Supekar S, Tonde S, Siddiqui JA (2020) Advanced traffic violation control and penalty system using IoT and image processing techniques. In: 2020 2nd international conference on innovative mechanisms for industry applications (ICIMIA), Bangalore, India, pp 554–558. https://doi.org/10.1109/ICIMIA48430.2020.9074949
21. Li S, Wang W, Pan W, Hsu CJ, Lu C (2018) FPGA-based hardware design for scale-invariant feature transform. IEEE Access 6:43850–43864. https://doi.org/10.1109/ACCESS.2018.2863019
22. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of Weiler-Atherton algorithm to address loose polygons, J Sci Ind Res, 78:771–774

23. Choudhury T, Aggarwal A, Tomar R (2020) A deep learning approach to Helmet detection for road safety. J Sci Ind Res (JSIR) 79(06):509–512
24. Kumar V, Choudhury T (2019) Real-time recognition of malignant skin lesions using ensemble modeling. J Sci Ind Res (JSIR) 78:148–153

An Overview on State-of-Art in Mid-Air Writing and Recognition Systems

Aradhana Kumari Singh, Lalit Kane, Abhirup Khanna, and Tanupriya Choudhury

Abstract Moving hand in the air with an intention to convey a character, symbol, or word is a dynamic gesture. These gestures are characterized by projected trajectories of the hand movement. Automated recognition of these air trajectories is open research due to inherent complexities. Research works in this area range from isolated character trajectory recognition to continuous word trajectory recognition. Attempts have been made to develop in-air writing systems in order to mimic natural pen-paper writing. At the other end, algorithms have been reported which presume the existence of unwanted, ligature strokes within the legitimate strokes. Right from the acquisition of the trajectories to the spotting of the legitimate gesture boundaries is principal challenges. This paper gives a comprehensive overview of various schemes employed for developing Mid-Air Writing (MAW) systems. An intensive walk through the literature suggests that most of the concerned schemes are variations to Hidden Markov Model (HMM), Dynamic Time Warping (DTW), and Longest Common Subsequence (LCS) techniques. However, some novel schemes targeting MAW systems have also been discussed in this paper.

Keywords HCI · Hand trajectory · Continuous stream data · HMM · DTW · LCS

1 Introduction

In the modern scenario of human–computer interaction, touchless gesture recognition is becoming an area of interest. Hand Gesture Recognition (HGR) can take a variety of forms depending upon the nature of the applications [1–5]. Further, the trajectory input can be touch-based such as a glove and electro sensor pen [6–9] or it could be touchless [10–13]. The predominant factor deciding a gesture are hand posture, posture sequences, and movement of the hand in the air, which eventually forms trajectories. For instance, fingerspelling recognition exemplifies posture recognition,

A. K. Singh · L. Kane · A. Khanna · T. Choudhury (✉)
Department of CSE, University of Petroleum and Energy Studies (UPES), Dehradun, India

A. K. Singh
e-mail: aradhana.singh@ddn.upes.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_37

sign language gestures denote posture sequences, and articulations in the air (mid-air writing [14]) in order to convey a character or word are characterized by motion trajectories. Since, in the mid-air writing [15] system, the user is allowed to write in the air with fingers or hand that leads to creating a trajectory forming a single character stroke and this is considered as a one-stroke finger gesture. Moreover, the system does not have any visual aid for the writing area in mid-air. Hence, character/word articulated by subject is highly superimposed over other with the inclusion of ligature stroke. The ultimate aim of such type of problem is to recognize the arbitrary connection and superimposed character in single-stroke finger gestures.

The acquired mid-air gesture trajectory contains 3D point information, which needs to be translated in 2D if needed, for the recognition process. In vision-based mid-air trajectory recognition, the subject articulates a character, symbol, or word in the air in front of a 2D/2.5D camera/sensor. The intercepted trajectory of coordinates is recognized by an automated scheme working in the background. The recognized results can be deployed as inputs to the public interaction terminals, e.g., railway inquiry and bank tellers. Other possibilities to utilize the recognized input can be in secured individual authentication and deaf and dumb (specially-abled) interaction.

However, the real-time solutions have just begun to be developed and the pertinent literature is still full of offline, experimental work. Clearly, the emphasis is still on the development of real-time air writing systems that can be deployed in a practical environment.

Several schemes attempt to recognize these different traits with a common strategy, few are feature-based [16–20] and other approaches are exemplary based [21–27]. However, real-time systems require specifically customized methods to render solutions.

This review paper intends to give a survey on the state-of-art and legendary schemes in the area of mid-air character/word recognition. To the best of our knowledge, the presented review is the first of its kind on the recognition of mid-air writing. The observed general workflow for such a system is depicted in Fig. 1.

Trajectory Acquisition

The development of multimedia technology catalyses the need for touchless systems as input in place of touch-based input devices. Such type of scheme demand 2D/2.5D sensor camera calibration for trajectory acquisition. As some of the literature used a 2D camera for video stream acquisition. Hence, they additionally require color-based segmentation as well as a banalization step for fingertip tracking which is sometimes not much accurate, as it does not provide depth information. While in 2.5D camera depth information is available, this additional information helps to track the fingertip more precisely. With the calibration of a 2.5D camera gesture, the trajectory can be acquired which is further used for pre-processing. Pre-processing steps require segmentation, feature extraction, and centroid calculation for further trajectory creation. In practical systems, low-cost depth sensors, e.g., leap-motion controller, Kinect Xbox, Action Pro, etc., have predominantly been used for trajectory acquisition.

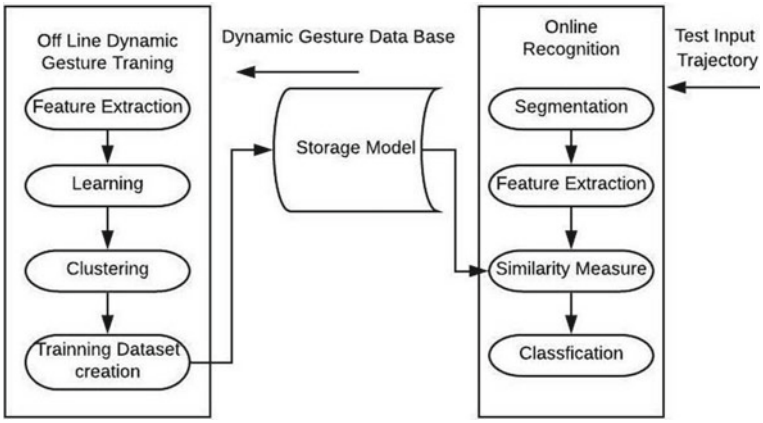


Fig. 1 General workflow for the trajectory recognition system

Taxonomy of Mid-Air Writing Systems

Gesture-based trajectory recognition system can be implemented for two broad types of gestures. Firstly, for isolated gestures, gesture starting points are pre-cut. Such schemes are fully off-line. The second category of recognition systems is based on continuous trajectory recognition on the fly, which can be implemented practically. Moreover, the trajectory written on the fly could be a character or a word-based recognition algorithm. In the literature, character/digit recognition [25] and word recognition with a limited corpus of the word [28] are emphasized separately.

As of now, based on literature, the classification approaches are feature-based and template-based. Feature-based classification approach deals with statistical analysis or state-based probability computation. In this context, modified Hidden Markov Model (HMM), Condition Random Field, and Modified Quadratic Discriminative function are considered for the classification scheme. Further, template-based approach uses Dynamic Time Warping, Longest Common Sub-sequence, edit distance, etc.

Hand Trajectory features

A hand trajectory is a collection of flow vectors, represented by $F = \{f_1,..f_t,...f_T\}$. Flow vector is nothing but the location of trajectory points and dynamics, $f_t = [x, y, V_{x,y}, \theta_{x,y}]$. The distance between the two points can be calculated by Euclidean distance as $d(f_t, f_T)$. Resampling of the trajectory is required to have equally spaced vectors.

Commonly Used Features

Hand position is the fundamental feature that is simple to extract and highly informative. Some notations are in a frame k; the hand position $P_k = (x, y, z)$ and the frame rate t_k are obtained. Subsequently, velocity v_k and acceleration a_k are calculated. Chain code also provides the direction of hand in 3D space. Twelve directional code

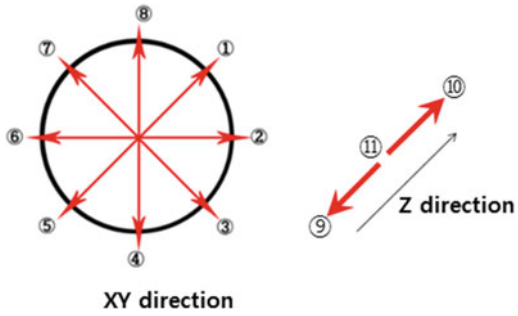


Fig. 2 Chain code [29]

is considered for Chain code. XY directional has eight directions and two codes for z direction. The 3D position achieved from a depth camera consists of the value of three axes in projective co-ordinates. The co-ordinates x and y represent the pixel position of the hand in the XY plane, and z represents the real distance [29] as shown in Fig. 2.

In most of the works, researchers have converted 3D feature vectors into 2D projective coordinates. Normalization is applied on hand positions to make them invariant to translation while velocity, acceleration, and chain code are left intact due to their inherent invariance to translation. The organization of the rest of the manuscript is as follows: Section II discusses conventional approaches for mid-air trajectory recognition. While Section III discusses the recent work done in the area of the trajectory recognition system. Section 4 lists some available datasets. Further, Sect. 5 lists comparative efficiency results of different research done. Section 6 concludes the literature for this manuscript. Section 7 discusses some open problems/Future challenges for Gesture Recognition System (GRS).

2 Conventional Approaches for Trajectory Recognition

In dynamic gesturing, Mid-Air Writing (MAW) has a vision-based tracking algorithm that calculates the centroid position of the hand. The sequence of centroid forms a gesture trajectory. Descriptive features are extracted from the trajectory obtained. Based on the features obtained, the classification algorithm should be able to find a meaningful interpretation either as a gesture or as a meaningless stroke [30]. Based on the physical perception, MAW systems can be classified as touch-based and touchless systems. The former requires users to make contact with a device, while the latter uses 2D or 2.5D cameras to acquire continuous trajectory.

Modified Dynamic Time Warping (DTW)

DTW (Dynamic Time Warping), one of the traditional approaches for sequence matching (e.g., DNA matching, signal matching), has been demonstrated to match a

predetermined template of exemplary gesture trajectory [29, 31–33]. DTW promises optimal alignment between two-time series.

Sohn et al. [29] presented a framework for mid-air drawn [34] digits recognition in which they used a K-NN classifier in combination with DTW alignment. Using a comparative analysis with various distance measures, it is established that DTW offers better performance on irregular sequences, provided that the training set is large. In a similar work, Vikram [33] used DTW to recognize mid-air characters by using time series of 3D positions as input, which is acquired by a leap motion camera.

Modified Longest Common Subsequence (LCS)

One more approach in this trend is the Longest Common Subsequence (LCS) method of alignment that finds the similarity between a pair of full and partial trajectories. Trajectories may reside in different dimensional spaces. It has a natural advantage over usual similarity measures, e.g., Euclidean, Manhattan, Mahalanobis, etc., that require fix dimensional data strings. In tone with DTW, LCS attempts to line up a test and template sequence temporally, using feature distance cost, and finds all the points with good matching are accepted. Moreover, it ignores outlier matching unlike in the case with DTW [25]. The LCS algorithm was initially developed to find the sub-word sequence in documents and medical fields. In the context of MAW systems, modifications to the basic LCS have been reported to find similarities between sub-gestures. Here, it is assumed that a gesture is composed of sub-segments, which are denoted as most discriminating segments (MDS). Work by Stern et al. [25] is tested on pre-isolated digits 0–9 and achieved 98% efficiency with $O(\log K) + O((K + B)\ mn)$ time complexity. Sample trajectories for each Arabic digit have been observed to learn MDSs. A similar pattern of the work by Nyirarugira and Kim [23] which tunes LCS with rough set theory achieved 96.4% accuracy on English alphabet and digits with $O(nm)$ efficiency of their algorithm. Input trajectory streams are searched for the most suitable MDS sequences to determine their labels using LCS. Three dimensional trajectory patterns have been projected into 2D space as a preliminary step. A probabilistic approach on the same line is introduced by Frolova and Stern [26]. Here, the 3D trajectory is projected into the 2D space using PCA in a pre-processing step.

Hidden Markov Model (HMM)

The statistical scheme HMM can generate models based on some set of input sequences called states. This memoryless model takes the outcome of the previous state as input to the next. The transition from one state to another state depends on the transition probability. Used predominantly in application areas like automated speech recognition, handwriting recognition, etc., HMM has also been applied in air trajectory recognition. Schick et al. [35] proposed a vision-based system and used Hidden Markov Models with multi-camera 3D hand tracking. They mainly focused on recognizing small continuously written words as well as a single character. Word boundaries are established by moving the hand forward or backward. Elmezain et al. [36] proposed a system that performs gesture spotting and recognition simultaneously based on HMM. Mean-shift algorithm and Kalman filtering are

used in preprocessing stage for smoothening the trajectory, and orientation dynamics is taken as a feature. Using trained HMM model gestures/non-gestures are classified and finally, the Viterbi algorithm is used for the recognition process. With this approach, digits from 0 to 9 are classified to claim an accuracy of 95.78%.

The main disadvantage with HMM is the requirement of a large number of samples and significant time in training to calibrate the model.

Conditional Random Field (CRF) Classifier

CRF is a type of discriminative classifier [37–40] that models decision boundaries between different classes. Some of the application areas are NLP, part of speech, etc. CRF is used to predict any sequence in multiple variables that depend on each other. Although CRF and HMM both model the sequential data, they are still different. Where HMM are generative and give output by modeling the joint probability distribution, CRFs are discriminative and model the conditional probability distribution. MacDorman et al. [19] identified that features are more prompting part for gesture spotting. With this, they modified the motion chain code to accurately spot the gesture. Orientation, length of an ellipse least-squares fitted to motion-trajectory points, and the position of the hand are taken as the features. These features are classified using CRFs.

Modified Quadratic Discriminant Function (MQDF) classifier

Quadratic Discriminant analytics (QDA) is an enhancement over the LDA (Linear Discriminant Analysis), where it can classify the pattern that cannot be linearly classified by LDA. In the context of MAW systems, Long and Jin [41] used a Quadratic classifier with Modified Quadratic Discriminant Function (MQDF) for the classification of air written characters. They aim to embed the MQDF classification approach with small hand-held devices such as personal digital assistants and mobile phones. However, due to the large parameter size and error rate, MQDF cannot be embedded into small memory devices. LDA can give a good result with less classification error. Furthermore, vector quantization (VQ) can be used for parameter compression. HCL2000 dataset is used for training and testing.

3 Recent Approaches for Trajectory Recognition

The recent approaches for air writing recognition [42, 43] are classified as the techniques that emphasize building air writing system and the techniques which attempt to recognize continuous air written trajectories with no options for writing frontend. In the latter case, some techniques customize the conventional schemes [25] while others devise newer schemes with their own heuristics [18].

Attempts in Building Air Writing Interfaces

Li and Zhang [44] gave a framework for a mid-air writing system in which they tried to find out the fingertip in two modes; frontal and side using depth stream of

Kinect depth sensor. Oriented gradient feature-based fingertip detection mechanism has been adopted for frontal mode. Additionally, the Kalman filter is used to predict the position of the fingertip. Similarly, for the side mode, the "Choose-To-Trust Algorithm" is proposed for hand segmentation. Also, they used a multi-objective algorithm for successful tracking and detection of the algorithm. Toward building an air-writing interface, Mukharjee et al. [45] proposed a system that identifies a hand in writing mode by efficient fingertip detection. The efficacy of the system is justified by recognizing a set of air-written trajectories such as characters and digits. To develop a writing interface, a faster Region-Convolution Neural Network (R-CNN) framework is used for hand region detection. Distance weighted curvature entropy assist to locate the fingertip. KCF (Kernelized Correlation filter) tracking algorithm to detect hand region in each frame.

Kane and Khanna [18] used skeleton-based endpoints to designate air-writing controls. In the extension, depth clues are utilized in implementing pen-up and pen-down actions in order to determine a hand in writing mode. EPS (Equi-Polar Signature) is derived from circular grid normalization of the trajectory points, which is invariant to translation, scale, rotation, and stroke directions. A template matching-based approach is used for EPS evaluation. The system shows its efficacy on Arabic digits, English lower-case alphabets, and a set of symbols. Agrawal et al. [46] proposed a fingertip tracking and recognition algorithm and used both Kinect and leap motion sensor for data acquisition. The algorithm was tested for characters as well as a lexicon of words and achieved an efficiency of 97.5%. For the preprocessing step they used segmentation and normalization on data points. Which is further being classified by SVM and DTW approaches. Starting and stopping criteria are identified by writing speed.

Table 1 lists some of the work proposed methodology for the MAW system. Digits and Alphabet are taken as elementary for the recognition process. On the other hand, some have targeted Chinese as well string numerals (group of digits). Table 1 also shows the efficiency and complexity obtained. Kumar et al. [14] used the ROVER framework for recognizing Devanagari and Latin words within limited no. of corpus, and Mukherjee et al. [45] have used R-CNN for localizing the fingertip for tracking purposes. Singha et al. [15] proposed a combinational feature approach to classify characters. They used a variation-based gesticulation pattern in a dynamic hand gesture recognition system. They also created an on-line dataset "NITS hand gesture database III" in which they a red tape on the index finger to identify motion of trajectory in a video clip. They also proposed a Right Sector Trajectory feature and Left Sector Trajectory Feature in fusion with the conventional featuring feature, which gives the improved recognition result.

Schemes Targeting Continuous trajectory recognition

Kumar et al. [14] proposed a lexicon-free approach for 3D handwriting recognition using classifier combination. They use a leap motion depth sensor to acquire a video stream. A classifier called ROVER (Recognize Output Voting Error Reduction), which is a combination of multiple classifiers and BLSTMNN (Bidirectional Long

Table 1 A brief description of MAW recognition methods on the research year, methodology adopted, and efficiency and complexity claimed

Author, year	Feature extracted	Target gestures	Key methodology	Efficiency (%)
Stern et al. [25]	1. Direction 2. Coordinates 3. Centroid of hand	Digits	MDSLCS (Most discriminating segment LCS)	92.6
Frolova and stren [34]	1. Direction 2. Coordinates 3. Centroid of hand	Digits	MPLCS (Most probable LCS)	97
Kumar et al. [14]	1. Writing direction 2. Curvature 3. Convex hall	Latin, Devanagari script (more then 200 characters)	BLSTMNN used with ROVER framework	72.25 [for Latin], 71.86 [for Devanagari]
Zeng et al. [20]	1. Co-ordinates 2. Acceleration 3. Direction of finger	Digits, English lower case alphabets	Radial basis function (RBF)	95.1 for digits, 92.9 for alphabets
Li et al. [44]	1. Co-ordinates 2. Gradient 3. Direction of finger	Chinese 6,763 characters, English lower and upper alphabets, digits	Modified quadratic discriminant function (MQDF)	90
Nyirarugira et al. [23]	Orientation	English lower case alphabets and digits	Modified LCS, Normalized LCS, rough set theory	96.4
MacDorman et al. [19]	Orientation, length of the major axis of ellipse (small/medium/Large), Position of hand (top/middle/bottom)	Digits only	Modified motion chain code, and conditional random field for classification	88.9
Kane and Khanna [18]	Angle and radial of trajectory for circular raster, Equi-Polar system (EPS)	0–9, Spacial symbol(e.g., +, ×,), English lower case alphabets	EPS (Equi-Polar System) for trajectory feature calculation, and KNN for classification	94.1
Mukherjee et al. [45]	Fingertip, velocity of hand, distance weighted curvature entropy, to determine termination criteria	0–9, A–Z, a–z	R-CNN	96.1

(continued)

Table 1 (continued)

Author, year	Feature extracted	Target gestures	Key methodology	Efficiency (%)
Jin et al. [47]	Character stroke, orientation	String numeral	Character stroke finding by segmentation	81
Choudhury and Sarma [16]	Linear eccentric, flatness, area of ellipse	Vowel and number for assamese alphabet	Start and end point determination under least square fitting ellipse criteria	Minimum error rate (ER) is 0.04%

Short Term Memory Neural Network), is used for classification. The approach is tested on self-developed Devanagari and English words corpuses.

Choudhury and Kumar Sharma [16] recognize air-written vowels and numbers of the Assamese alphabets. Geometric features used are the linear eccentricity of the ellipse, its flatness, and area. The fundamental idea to spot start, end, and intermediate trajectory is to determine the behavior of these three feature parameters. Their experiments reported a minimal error rate of 0.04%.

Zeng et al. [20] used Deterministic-learning technique by using the leap motion sensor. Motion patterns of different gestures are modeled by the constant Radial Basic Function (RBF) neural network. Yang et al. [48] identified an HMM-based method to recognize single hand gestures. Gesture images are acquired by a common web camera and skin color-based segmentation is applied to detect hands. Further, a state-based spotting algorithm is used to split continuous gestures. Hand position, velocity, size, and shape are the features extracted for each gesture. HMM classifies the spotted gesture.

Jin et al.'s [47] work recognized numeral strings. For this purpose, strings are split into segments or sub-strokes which could be a complete character, its part, or an undesired ligature stroke. The appropriate segment can be identified by the dynamic programming approach. To avoid the ambiguous classification of legitimate strokes in ligature strokes (false negatives), the quadratic discriminating function is used. The final classification is carried out using a Support Vector Machine. CASIA-OLHWDB database is used for experimental purposes. This work, however, cannot identify long numeral strings efficiently.

4 Datasets

For experimental analysis concerning the implementation of a MAW system, open datasets available are quite a few. Table 2 lists several open datasets. The last column

Table 2 List of open datasets and results obtained by them

Dataset, way of acquisition, and content description	Work by and results obtained	Link available
CASIA-OLHWDB [47], Kinect, 10,141 digits available in depth stream	Jin et al. 80% (String Level) [47]	http://www.nlpr.ia.ac.cn/databases/handwriting/Home.html
DAIR, Kinect [46], Depth frames carrying Alphabets	Aggarwal et al. [46], 96% Kane and Khanna [49], 97%	https://cvit.iiit.ac.in/research/projects/cvit-projects/online-handwriting-recognition-using-depth-sensors
NITS dataset [15], RGB Camera, 16,000 gesture for alphabet, digits	Singha et al. [15], 91.07%	https://joyeetasingha26.wixsite.com/nits-database
Kumar et al. [14], leap motion controller, 2000 Latin word, 50 Devanagari	Kumar et al. [14]	https://sites.google.com/site/iitrcsepradeep7/resume

of the table provides the link for the available dataset. NITS dataset [15] used video-based datasets having fixed starting points used red tape on the fingertip. CASIA-OLHWDB [47] have online pre-isolated Chinese datasets for the MAW system. Dataset of air handwritten (DAIR) [46, 49] contain articulations of the English alphabet. IIITDMJ is a dataset of 100 articulations each for the English alphabet, Arabic digits, and symbols that can also be obtained from the authors [49]. Kumar et al. [14] dataset provided an online 3-D points MAW system.

5 Conclusion

This paper intends to give a brief and precise survey on the state-of-art and legendary schemes in the area of mid-air character/word recognition. It is observed that schemes targeting air-writing need to be more precise in detecting the hand in writing mode. An evaluation framework is also necessary to collect the haptic feedback for the interface. While the practical systems are evolving with the availability of low-cost depth sensors, modifications to the schemes to allow co-articulations and articulations for continuous trajectories of words are present requisites. As far as continuous trajectory recognition is concerned, most of the work is focusing features extraction-oriented determination of gesture boundaries. With the previously reported usage of state-based models like HMM for classification, interest is shifting to deep learning-based strategies [45]. Some schemes are adapted for specific gesture sets (e.g., most of the work circumvents around Arabic digits) due to their lexical nature. Adaptability to multiple gesture sets is still an issue with continuous trajectory recognizers.

As per the literature, most of the work on MAW systems is either based on exemplary matching approaches or based on state-based probabilistic models. In most of the works, several assumptions have been made, e.g., fixed started points, uniform

styles of writing. In addition, word size in the corpuses is also fixed. Implementation of the MAW system in real-time is a challenge in itself. The fundamental open problem is to spot the ligature strokes within the continuous stream. Work on single-stroke gestures is quite general; however, the research on articulation and recognition of multi-stroke characters and words is scarce. Popular touchpad gesture recognizers like $1-Recognizer have also been reported to be used in MAW systems [18]. However, they still have the potential to be improvised for multi-stroke characters and word recognition.

References

1. De Rosa M, Fuccella V (2017) Handwriting on smartwatches : an empirical 47(6):1100–1109
2. Caputo FM, Prebianca P, Carcangiu A, Spano LD, Giachetti A (2018) PT US CR. Comput Graph
3. Chen YV, Hassan R, Huynh W, Mcgraw T, Qian ZC (2018) Unobtrusive touch-free interaction on mobile devices in dirty working environments, pp 1–10
4. Zhang X, Huang Y, Zou Q, Pei Y, Zhang R, Wang S (2020) A hybrid convolutional neural network for sketch recognition. Pattern Recognit Lett 130:73–82
5. Cheng H, Dai Z, Liu Z, Zhao Y (2016) An image-to-class dynamic time warping approach for both 3D static and trajectory hand gesture recognition. Pattern Recognit 55:137–147
6. Misra S, Laskar RH (2019) Development of a hierarchical dynamic keyboard character recognition system using trajectory features and scale-invariant holistic modeling of characters. J Ambient Intell Humaniz Comput 10(12):4901–4923
7. Amit M, Tzvi G, Sigal B, Ilana N (2018) The scaling of the gripper affects the action and perception in teleoperated grasping via a robot-assisted minimally invasive surgery system
8. P. O. Box and P. O. Box (2018) Perception and action in remote and virtual environments, pp 67–68
9. Chang HJ, Garcia-hernando G, Tang D, Kim T (2016) Spatio-temporal hough forest for efficient detection—localisation—recognition of fingerwriting in egocentric camera. Comput Vis Image Underst 148:87–96
10. Sykora P, Kamencay P, Hudec R (2014) Comparison of SIFT and SURF methods for use on hand gesture recognition based on depth map. AASRI Procedia 9(Csp):19–24
11. Yun SN, Zhong LX (2011) experimental research on handwritten character written in the air recognition based on computer vision, pp 530–533
12. Bag S, Harit G, Bhowmick P (2014) Recognition of Bangla compound characters using structural decomposition. Pattern Recognit 47(3):1187–1201
13. Zeng W, Zhang Z (2012) Multimedia at work microsoft Kinect sensor and its effect, pp 4–10
14. Kumar P, Saini R, Pratim P, Pal U (2018) A lexicon-free approach for 3D handwriting recognition using classifier combination. Pattern Recognit Lett 103:1–7
15. Singha J, Misra S, Laskar RH (2016) Neurocomputing effect of variation in gesticulation pattern in dynamic hand gesture recognition system. Neurocomputing 1–12
16. Choudhury A (2018) A novel approach for gesture spotting in an assamese gesture-based character recognition system using a unique geometrical feature set. In: 2018 5th international conference on signal processing and integrated networks, pp 98–104
17. Chiu L, Hsieh J, Lai C (2018) Person authentication by air-writing using 3D sensor and time order stroke context, vol 2, no iii. Springer International Publishing
18. Kane L, Khanna P (2016) A framework to plot and recognize hand motion trajectories towards development of non-tactile interfaces. Procedia Comput Sci 84:6–13
19. Macdorman KF, Iwahori Y (2014) A novel set of features for continuous hand gesture recognition

20. Zeng W, Wang C, Wang Q (2018) Hand gesture recognition using Leap motion via deterministic learning
21. Chiang C, Wang R, Chen B (2017) Recognizing arbitrarily connected and superimposed handwritten numerals in intangible writing interfaces. Pattern Recognit 61:15–28
22. Tsai T, Hsieh J, Chen HC, Huang S (2017) Reverse time ordered stroke context for air- writing recognition
23. Nyirarugira C, Kim T (2015) Signal processing: image communication Stratified gesture recognition using the normalized longest common subsequence with rough sets. Signal Process Image Commun 30:178–189
24. Niitsuma Y, Torii S, Yaguchi Y, Oka R (2015) Time-segmentation and position-free recognition of air-drawn gestures and characters in videos
25. Stern H, Shmueli M, Berman S (2013) Most discriminating segment—longest common subsequence (MDSLCS) algorithm for dynamic hand gesture classification. Pattern Recognit Lett 34(15):1980–1989
26. Frolova D, Stern H, Berman S (2013) Most probable longest common subsequence for recognition of gesture character input, 43(3):871–880
27. Islam R, Mahmud H, Hasan K, Rubaiyeat HA (2016) Alphabet recognition in air writing using depth information
28. Kumar P, Saini R, Pratim P, Debi R, Dogra P (2016) 3D text segmentation and recognition using leap motion. Multimed Tools Appl
29. Sohn M, Lee S, Kim D, Kim B, Kim H (2012) A comparison of 3D hand gesture recognition using dynamic time warping, pp 418–422
30. Baig F, Khan MF, Beg S (2013) Text writing in the air, vol 0316
31. Stefan A, Athitsos V, Alon J, Sclaroff S (2008) Translation and scale-invariant gesture recognition in complex scenes
32. Doliotis P, Stefan A, Mcmurrough C, Eckhard D, Athitsos V, Comparing gesture recognition accuracy using color and depth information
33. Vikram S (2013) Handwriting and gestures in the air, recognizing on the fly, pp 1179–1184
34. Frolova D, Stern H, Berman S (Jun. 2013) Most probable longest common subsequence for recognition of gesture character input. IEEE Trans Cybern 43(3):871–880
35. Schick A, Amma C, Vision-based handwriting recognition for unrestricted text input in mid-air, pp 217–220
36. Magdeburg O, Elmezain M, De AA (2009) Hand trajectory-based gesture spotting and recognition using HMM. In: Elmezain M, Al-Hamadi A (eds) Bernd michaelis institute for electronics, signal processing and communications (IESK), pp 3577–3580
37. Verma A, Shukla A, Choudhury T, Chauhan A (2020) Recommendation system for prediction of tumour in cells using machine learning approach. Commun Comput Inf Sci (CCIS) 1229. https://doi.org/10.1007/978-981-15-5827-6_18
38. Sabitha S, Choudhury T (2018) Proposed approach for book recommendation based on user k-NN. Adv Intell Syst Comput 554. https://doi.org/10.1007/978-981-10-3773-3_53
39. Jain KN, Kumar V, Kumar P, Choudhury T (2018) Movie recommendation system: hybrid information filtering system. In Intelligent computing and information and communication. Springer, Singapore, pp 677–686
40. Bhasin S, Choudhury T, Gupta SC, Kumar P (2017) Smart city implementation model based on IoT. In: Proceedings of the 2017 international conference on big data analytics and computational intelligence, ICBDACI 2017. https://doi.org/10.1109/ICBDACI.2017.8070836
41. Long T, Jin L (2008) Building compact MQDF classifier for large character set recognition by subspace distribution sharing. Pattern Recognit 41(9):2916–2925
42. Singh N, Sharma T, Thakral A, Choudhury T (2018) Detection of fake profile in online social networks using machine learning. In: Proceedings on 2018 international conference on advances in computing and communication engineering, ICACCE 2018. https://doi.org/10.1109/ICACCE.2018.8441713
43. Gupta SS, Thakral A, Choudhury T (2018) Social media security: analysis of threats and security measures. In: Proceedings on 2018 international conference on advances in computing

and communication engineering, ICACCE 2018. https://doi.org/10.1109/ICACCE.2018.8441710
44. Li K, Zhang X (2014) A new fingertip detection and tracking algorithm and its application on writing-in-the-air system, pp 457–462
45. Mukherjee S, Ahmed SA, Dogra DP, Kar S, Roy PP (2019) Fingertip detection and tracking for recognition of air-writing in videos. Expert Syst Appl 136:217–229
46. Aggarwal R, Swetha S, Namboodiri AM, Sivaswamy J, Jawahar CV (2015) Online handwriting recognition using depth sensors, pp 1061–1065
47. Jin X, Wang Q, Hou X, Liu C (2014) Visual gesture character string recognition by classification-based segmentation with stroke deletion visual gesture character string recognition by classification-based segmentation with stroke deletion
48. Yang Z, Li Y, Chen W, Zheng Y (2012) Dynamic hand gesture recognition using hidden Markov models. In: IEEE 7th international conference on computer science & education (ICCSE), pp 360–365
49. Kane L, Khanna P (2017) Vision-based mid-air unistroke character input using polar signatures. IEEE Trans Human-Mach Syst 47(6):1077–1088
50. Leem SK, Khan F, Cho SH (2019) Detecting mid-air gestures for digit writing with radio sensors and a CNN. IEEE Trans Instrument Meas 69(4):1066–1081
51. Yanay T, Shmueli E (2020) Air-writing recognition using smart-bands. Pervasive Mob Comput 66:101183
52. Ayachi N, Kejriwal P, Kane L, Khanna P (2015) Analysis of the hand motion trajectories for recognition of air-drawn symbols. IEEE fifth international conference on communication systems and network technologies, pp 505–510

Global Skew Detection and Correction of Document Image Based on Least Square Method and Extensive Connected Component Analysis

Faisal Imran, Md. Ali Hossain, Md. Al Mamun, Bhupesh Kumar Singh, and Tanupriya Choudhury

Abstract Document images can have numerous noises that should be filtered to eliminate those unutilized objects. Skewness is liable not to understand the texts clearly in the document image. A morphological technique is being conducted to preprocess the image. Connected component analysis along with the bounding box approach is introduced to figure out a line. A number of longest connected components are extracted from numerous illustrations. In order to detect a precise skewed angle, the least square method is applied to the logically selected regions to acquire the best fitting line. The skewed angle is discovered by averaging angles of best-fitting lines. The proposed system is already applied on the printed document image and got tremendous accuracy during execution.

Keywords Bounding box · Morphological dilation · Structure elements · Connected components · Longest connected component · Line fitting

1 Introduction

At present all the documents are being stored in digital format. This document can be in the following format: PDF, JPG, PNG, etc. In this paper, we are working on JPG image files which are surely document images. Unconsciously the images are rotated or skewed when taking pictures from digital devices. Most of the time the document is unreadable due to this unfamiliar skew. Though it is visual in the human eye, the machine will be confused for the OCR system [1]. The texts will not be recognized as they are skewed. So the image needs to be preprocessed, and this skew must be

F. Imran (✉) · Md. Ali Hossain · Md. Al Mamun
Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh, India

B. K. Singh
Arba Minch University, Noida, Ethiopia
e-mail: dr.bhupeshkumarsingh@amu.edu.et

T. Choudhury
Department of Informatics, School of CS, University of Petroleum and Energy Studies (UPES), Dehradun, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_38

removed. After the image is restored, it will be readable for the machine as well as human view too.

The proposed framework along with elaborative preprocessing stages are organized in Sect. 3. Experimental results will be analyzed in Sect. 4.

2 Literature Review

Skew detection can be partitioned into local and global skew detection. The application of these techniques depends on the nature of the processing. If we want to apply on local text regions in the image, then local skew detection, as well as correction method, is applicable. The method is based on the characters, lines, and words of the document image which is proposed by Boukharouba and Abdelhak [2]. In the global skew detection technique, the method is performed on the whole document image. Most of the skew detection methods have been suggested based on global skew detection. For skew identification, several methods have been proposed. Some of them are as follows: projection profile [3–7], Hough transform [8], nearest neighbor clustering [9], and interline cross-correlation [10]. Ganai [3] proposed a traditional method which is the projection profile approach. It is being constructed on a horizontal projection profile. Maximum variations of the projection profile are calculated. The value is taken to find out the best position of text lines. The final output is the projection angle which will be used to rotate properly the skewed document image.

There are many existing algorithms that are developed to gain high accuracy in order to find the skew angle correctly. The horizontal projection was used initially to detect skew in the document image. But vertical projection was introduced for Latin characters to get the better result which was proposed at [6]. Bounding box minimization criteria had also been reintroduced. Connected components are used to get the skew angle [4]. Horizontal projection along with vertical projection was used concurrently to calculate the skew angle. The bounding box technique was also used in the meantime. The skew angle can be found through background information [7] of the scanned document image. The background information of scanned documents is non-textual features that are distinguished from text areas. Dilation operations accompanied by some statistical methods were combined to detect global skew which was proposed by Sharfuddin and Soora [11].

Morphological skeleton operation and Probabilistic Hough Transform (PPHT) [8] were being used to detect and correct skew in the historical document image. In order to reduce the redundant pixels, a morphological skeleton is applied. Probabilistic Hough Transform (PPHT) identifies the lines in the document image. The global skew angle is then detected and corrected by their proposed algorithm. As projection profile methods are restricted to detect the skew angle of the document image in between $\pm 10^{\circ}$ and $\pm 15^{\circ}$. Hough transform works better in the binary digital image and for some context, the image must be transferred into the Hough domain [9] to find the maximum peak for the skew detection.

In our method, we have used Connected Components (CCs) analysis with the help of the bounding box approach. Each line was considered as the longest CC, and the five longest CCs were picked to apply the proposed methodology. In order to detect the skew angle of a text line, we need to find out the reference text line and it will be estimated from the left side of LCC. This experiment is being applied to English, Bangla, and Arabic language dataset. The experimental result is showing high accuracy in contrast to other skew detection techniques [12–16].

3 Proposed Framework

Document images are scanned with any scanning devices (e.g., digital Camera, scanner). The images are found in digital format and can be color or grayscale. If the RGB image is said as a color image, then it will have three color components (red, green, and blue) with R rows and C columns. For the grayscale image, the pixel values are ranged between 0 and 255 with R rows and C columns. The image matrix M can be presented as $M(i, j) \in \{0, 1, 2, 3,\ldots, 255\}$, where $i = 1, 2, 3,\ldots, R$ and $j = 1, 2, 3,\ldots, C$. The binary image can be found by converting the grayscale image into it which contains only two values (0 and 1).

The proposed system comprises the following stages:

Step 1: Convert the image into the binarized image. Smaller components such as comma and semicolon are removed. If the smaller component has less than five pixel values, it will be removed.
Step 2: Select structuring elements and apply the morphological operation to the image.
Step 3: From many Connected Component (CC), find out the largest connected component (LCC).
Step 4: Estimate reference text line using line fitting technique.
Step 5: Finding the Centroid of individual connected components.
Step 6: Calculate Global Skew Angle and Rotate Image According to the Angle.

The steps are shown in Fig. 1 visually.
The steps are abbreviated as follows.

3.1 Preprocessing

The input image is read through any digital device (Fig. 2).

The RGB or grayscale image is converted into a binary image. The color components are quantized into two levels which are 0 and 1. Non-text contents (Comma, Semicolon, dots, noises) must be removed for greater accuracy. The connected components which length or dimension is less than the size of smaller objects are removed (Fig. 3).

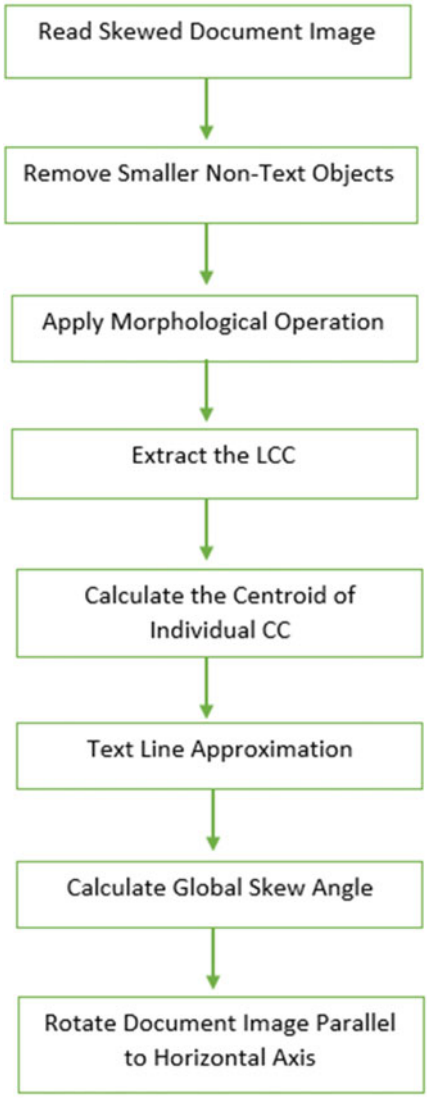


Fig. 1 Steps of skew detection and correction

3.2 Apply Morphological Operator

There are different types of structuring elements such as square, circular, rectangular, etc. For the document image, we are using square-shaped structuring element of size 3 $\times$ 3. Words will be connected together by applying morphological dilation operation.

If the input image is denoted as f_{ij} and square-shaped structuring element is SE, then the resultant image X_c will be

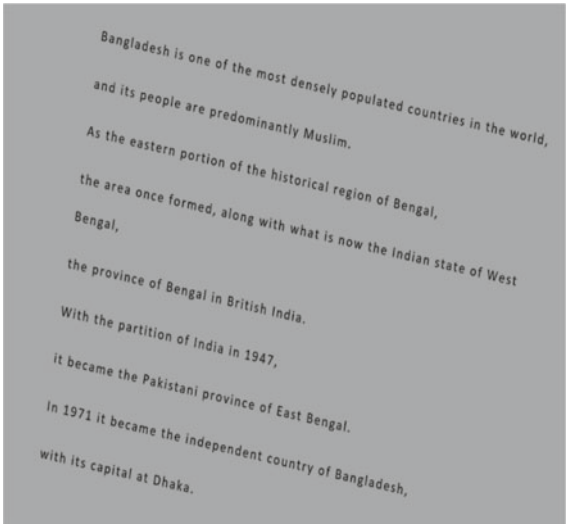


Fig. 2 Input image with noise

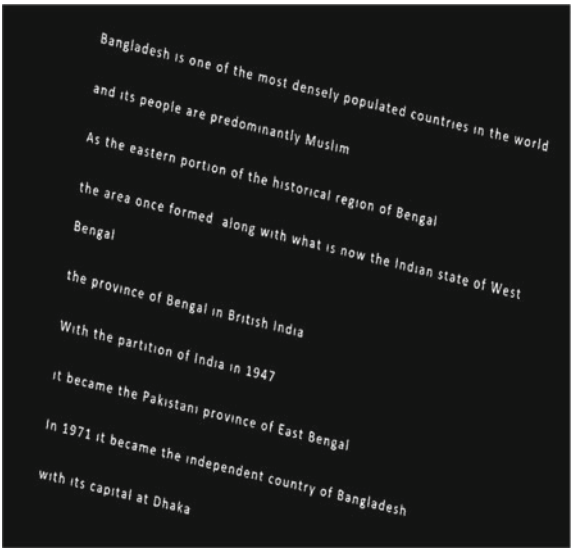


Fig. 3 Binary image with noise reduction

$$X_c = f_{ij} \oplus SE. \tag{1}$$

Morphological dilation is used here for connecting components as shown in Fig. 4.

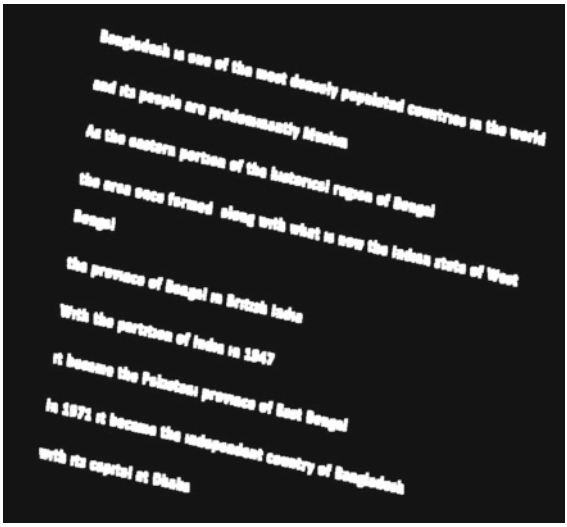


Fig. 4 Morphological dilation on the binary image

3.3 Finding Longest CC

After obtaining the output, X_c we can extract the longest CC. Analyzing the connectivity from the leftmost bounding box to the next one and so on, the length can be found easily. We are considering the five longest CC for reducing complexity and for greater accuracy (Fig. 5).

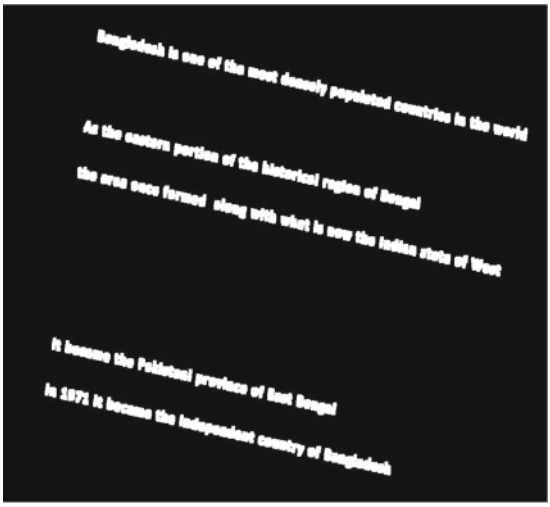


Fig. 5 Longest connected component

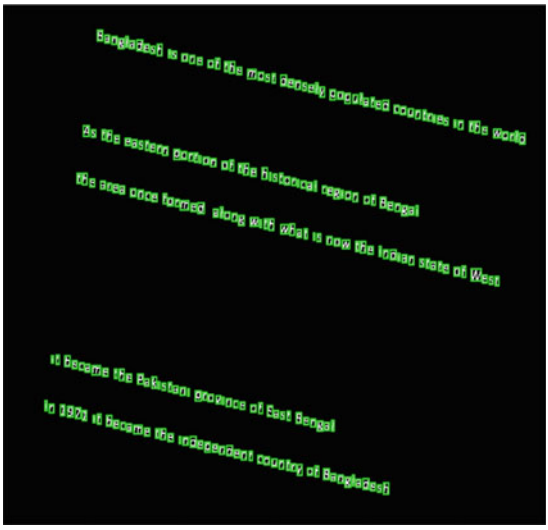


Fig. 6 Bounding box of individual CC

3.4 Finding Centroid of Individual CC

The bounding box of every character has their own centroid [10]. A line has several characters as its feature. For each and every character, there should have a constructive bounding box that should have a centroid. We are considering the centroid of the bounding box of every character.

The green rectangle in Fig. 6 is representing the bounding boxes of every character whereas the red-colored star symbol denotes centroids of individual bounding boxes (Fig. 7).

Individual bounding boxes have four coordinates. Two of them are initially defined which are $(X_{\min}, Y_{\min})$ and $(X_{\max}, Y_{\max})$. From these two points, the rest of them are found easily which are $(X_{\min}, Y_{\max})$ and $(X_{\max}, Y_{\min})$.

The centroid of individual CC is calculated from the following equation:

$$C_{cc} = \left(\frac{X_{\min} + X_{\max}}{2}, \frac{Y_{\min} + Y_{\max}}{2} \right). \tag{2}$$

3.5 Text Line (Reference) Estimation by Least Square Method

Usually, the centroid of connected components forms straight lines. But the centroid points are not exactly linear in fashion for the skewed document image [17]. In order to draw a straight line, a best line fitting method should be introduced.

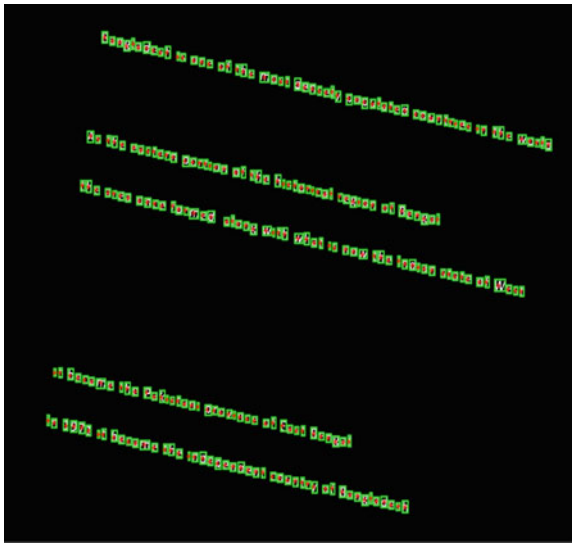


Fig. 7 Centroids of the individual bounding box

The equation of a straight line is as follows:

$$y = Mx + B. \tag{3}$$

From the given set of data, the line of best fit can be found which will definitely be a straight line. The best-fitted line should have a distinct characteristic. The line must pass through nearly all of the scatterplots. Also, the number of plots above and below the line should be equidistant (approx.).

There is so many line fitting algorithms but the least square method is being used for reducing complexity. Suppose the scatterplots are given as follows:

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n).$$

The following formula can be used in order to find the slope of the line:

$$M = \frac{\sum_{i=1}^{i=n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{i=n}(x_i - \bar{x})^2}, \tag{4}$$

where (x_i, y_i) denotes the scatterplots,

$\bar{x}$: average or mean value of x values, and

$\bar{y}$: average or mean value of y values.

The following equation finds the y-intercept of the best-fitted line.

$$B = \bar{y} - M\bar{x}. \tag{5}$$

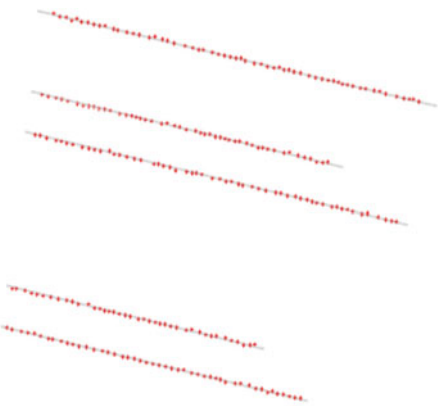


Fig. 8 Line fitting of CCC scatterplots

The slope of the line, M, and y-intercept, B, are used to find coordinate values of the best-fitting line. Best fitting lines of the five longest connected components are visualized as Gray-colored lines over centroids of all bounding boxes in individual lines (Fig. 8).

3.6 Calculate Global Skew Angle and Rotate Image According to the Angle

For every fitted line, a slope, M, has been calculated. Using the following equation, the skew angle can be detected:

$$\theta = \arctan(M). \tag{6}$$

The angle of the five-fitted lines would be A, B, C, D, and E. Required global skew angle $\bar{\theta}$ is found by averaging the angle of five longest CC.

The whole document image is skewed with this global skew angle, $\bar{\theta}$.

4 Experiment Results and Comparison

The proposed technique has been applied to Arabic, English, and Bengali document images. Skewed samples were taken as input and after rotation, the document images were viewed as output. To remove unwanted noise, preprocessing steps were applied. Connected component analysis [18–23], bounding box approach, and least square method were used to detect text lines. The skewed angle may be positive or negative. The slope of a line indicates whether a line lies in a negative direction or positive direction shown in Fig. 9. In our example, the average global skew angle was 15.13°

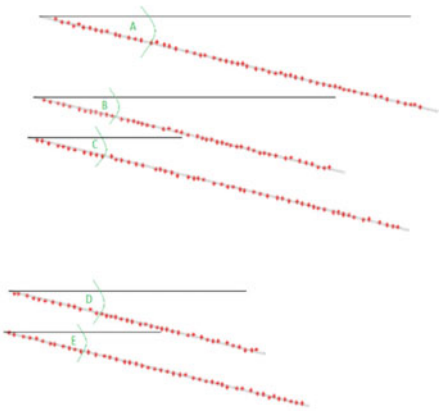
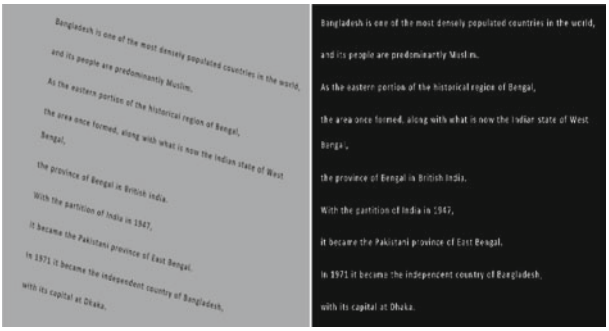


Fig. 9 Computing angles from least-square line fitting

which was used to de-skew the document image though the original skew angle was 15.15°. The calculated accuracy rate is being obtained above 99% (Fig. 10).

In Table 1, a comparison with the proposed method has been given. Due to the iterative process in the projection profile method, better efficiency cannot be provided as said by Sadri et al. [24]. The cross-correlation method is applied to the printed and



(a) **(b)**

Fig. 10 Experimental results

Table 1 Comparison with Existing Methods

Proposed By	Document type	Angle	Method
Sadri et al. [24]	Printed document	±15°	Projection profile
Rehman et al. [25]	Printed, handwritten	±10°	Cross-correlation method
Kumar et al. [26]	Printed document	±45°	Hough transformation
Proposed Method	Printed and handwritten English document image	±45°	Least square method, connected component analysis, and bounding box approach

handwritten document. The cross-correlation method can de-skew the document image in between ±10° [25] and the computational cost is high in this contrast. Kumar et al. [26] proposed Hough transform technique which method can de-skew ±45° angle but it will work on the printed document image. The proposed method can correct the skew angle ±45° for printed and handwritten English document images. By analyzing this comparison given in Table 1, this technique generates high efficiency for printed and handwritten document images in contrast to others.

5 Conclusion

The experimental result is indicating that the accuracy of the proposed technique is high in contrast to others. This is not well suitable for handwritten documents having multiple skewed angles in a single line. As the global skewed angle was found from the best-fitted line by minimizing the orthogonal distance between fitted line and centroid of every CC, the skewed angle was accurately found.

References

1. Susanto R, Putri FP, Wiratama YW (2018) Skew detection based on vertical projection in latin character recognition of text document image. Int J Eng Technol 7(4.44):198–202
2. Boukharouba A (2017) A new algorithm for skew correction and baseline detection based on the randomized Hough Transform. J King Saud Univ-Comput Inf Sci 29(1):29–38
3. Ganai AF, Koul A (2016) Projection profile based ligature segmentation of Nastaleeq Urdu OCR. In: 2016 4th international symposium on computational and business intelligence (ISCBI). IEEE, pp 170–175
4. Rajput GG, Ummapure SB, Patil P (2016) Separation of touching or overlapping lines from handwritten document images using histogram and connected component analysis. Int J Comput Appl, Natl Conf Digit Image Signal Process 0975–8887
5. Kumar H, Shivakumara P (2007) Skew detection technique for binary document images based on hough transform. Int J Comput Inf Eng 1(8):2401–2407
6. Papandreou A, Gatos B (2011) A novel skew detection technique based on vertical projections. In: 2011 international conference on document analysis and recognition. IEEE, pp 384–388
7. Mascaro AA, Cavalcanti GD, Mello CA (2010) Fast and robust skew estimation of scanned documents through background area information. Pattern Recogn Lett 31(11):1403–1411
8. Boudraa O, Hidouci WK, Michelucci D (2017) An improved skew angle detection and correction technique for historical scanned documents using morphological skeleton and progressive probabilistic Hough transform. In: 2017 5th international conference on electrical engineering-boumerdes (ICEE-B). IEEE, pp 1–6
9. Bafjaish SS, Azmi MS, Al-Mhiqani MN, Radzid AR, Mahdin H (2018) Skew detection and correction of Mushaf Al-Quran script using hough transform. Int J Adv Comput Sci Appl 9(8):402–409
10. Zhang Z, Zhang C, Shen W, Yao C, Liu W, Bai X (2016) Multi-oriented text detection with fully convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4159–4167

11. Mohammed SW, Soora NR (2018) Global skew detection and correction using morphological and statistical methods. In: Computational vision and bio inspired computing. Springer, Cham, pp 556–568
12. Singh RK, Panchal VK, Singh BK (2019) Analysis of GA performance on its various parameters for solving travelling salesman NP-hard problem. Int J Comput Sci Eng 7(1):509–512. ISSN:2347-2693(UGC-63193)
13. Singh RK, Panchal VK, Singh BK (2019) Analysis of the various population sizes in genetic algorithm for solving N-queen NP-hard problem. Int J Res Electron Comput Eng IJRECE 7(1):564–566. ISSN:2393-9028(Print)| ISSN:2348-2281(Online). (UGC-44816)
14. Singh BK, Luthra M (2019) MAHI as a Sketching Language. Int J Comput Sci Eng IJCSE 7(3):1102–1109. ISSN:2347-2693. (UGC-63193)
15. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of Weiler-atherton algorithm to address loose polygons. J Sci Ind Res 78:771–774
16. Kumar V, Choudhury T (2019) Real-time recognition of malignant skin lesions using ensemble modeling. J Sci Ind Res (JSIR) 78:148–153
17. Valy D, Verleysen M, Sok K (2016) Line segmentation approach for ancient palm leaf manuscripts using competitive learning algorithm. In: 2016 15th international conference on frontiers in handwriting recognition (ICFHR). IEEE, pp 108–113
18. Joshi A, Sabitha AS, Choudhury T (2018) Crime analysis using K-means clustering. In: Proceedings—2017 international conference on computational intelligence and networks, CINE 2017. https://doi.org/10.1109/CINE.2017.23
19. Sharma D, Sai Sabitha A, Choudhury T (2019) Discovering the unknown patterns of crop production using clustering analysis. In: Advances in intelligent systems and computing, vol 841. https://doi.org/10.1007/978-981-13-2285-3_70
20. Raj G, Sharma S, Choudhury T (2018) Load analysis in SDN for distributed topologies. In: Proceedings of the 2nd international conference on green computing and internet of things, ICGCIoT 2018. https://doi.org/10.1109/ICGCIoT.2018.8753099
21. Mehra S, Choudhury T (2018) Sentiment analysis of user entered text. In: Proceedings of the international conference on computational techniques, electronics and mechanical systems, CTEMS 2018. https://doi.org/10.1109/CTEMS.2018.8769136
22. Mittal V, Gupta S, Choudhury T (2018) Comparative analysis of authentication and access control protocols against malicious attacks in wireless sensor networks. In: Smart computing and informatics. Springer, Singapore, pp 255–262
23. Chauhan N, Agarwal R, Garg K, Choudhury T (2020) Redundant IAAS cloud selection with consideration of multi criteria decision analysis. Procedia Comput Sci 167:1325–1333
24. Sadri J, Cheriet M (2009) A new approach for skew correction of documents based on particle swarm optimization. In: 2009 10th international conference on document analysis and recognition. IEEE, pp 1066–1070
25. Rehman A, Saba T (2011) Document skew estimation and correction: analysis of techniques, common problems and possible solutions. Appl Artif Intell 25(9):769–787
26. Kumar D, Singh D (2012) Modified approach of hough transform for skew detection and correction in documented images. Int J Res Comput Sci 2(3):37

Improved Pattern of Adaptive Rood-Pattern Search Algorithm for Motion Estimation in Video Compression

Rahul Reddy Nadikattu, Rahul Bhandari, and Pawan Whig

Abstract Due to fast growing digitalization, huge amount of data, especially video, is generated all over the world which is very difficult to store, process, and transmit. Many international coding standards are used to compress high quality video content at low bitrates for wide range of video applications. This research is to provide an enhanced efficiency of coding to provide higher compression capabilities with high perceptual quality. Motion Estimation is an important as well as computational step in video coding. This paper presents improved version of pattern in Adaptive Rood-Pattern Search (ARPS) algorithm for estimating a motion to compress a video. This improved version is comprised of Rectangular Shaped Search (RSS) Pattern which gives better results which are based on number of computational points for searching pattern as well as PSNR values.

Keywords Video compression · Motion estimation · PSNR

1 Introduction

Digitalization is the need all over in today's world. Because of this, saving and transmitting a video data will be a biggest challenge. Video compression [1–3] is the essential tool for compressing a video on mobiles, laptops, TV, and many other gadgets. Many coding standards are available for compressing a video such as H.26x, MPEG-x, etc. All these standards are recommended by ITU-T/ISO organization [4, 5]. The main thing is to compress a video without affecting the quality of video with the help of removing redundancy between/inside frames. There are mainly three techniques to reduce redundancy in the data. Temporal redundancy is reducing between frames in

R. R. Nadikattu
University of the Cumberlands, Williamsburg, Kentucky, USA

R. Bhandari (✉)
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
e-mail: er.rahulbhandari1987@gmail.com

P. Whig
V.I.P.S, Delhi, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_39

time domain by motion estimation and motion compensation. Spatial Redundancy is used to remove a redundancy within frames (i.e., called image processing) by transform coding. Statistical Redundancy is used for removing redundancy while compressing a data by entropy coding [5, 6]. This paper comprises to eliminate temporal redundancy between successive frames for compressing a video. To reduce temporal redundancy, Motion Estimation plays a lead role in video compressing [7–9].

2 Related Work

The process of motion estimation (ME) is to generate motion vectors that decide each and every motion of anticipated frame from preceding frame. It also calculates object movements in an image sequence of a video using motion vector in video compression [10–12]. This process can absorb a lot of time of total time of encoding in many standards. This paper is dedicated to study as well as develop efficient motion estimation algorithm; with respect to reducing time for encoding and also improve efficiency of an encoding [13].

Nowadays, populous fast ME algorithms [14–16] have been designed to estimate the motion such as 3-step search (3SS), new 3-step search (N3SS), 4-step search (4SS), diamond search (DS), and many more algorithms [13]. These algorithms are also known as Block-based Motion Estimation (BME). In BME, each frame can be divided into several blocks with equal size, i.e., 8×8. The idea behind BME is to search a block in reference frame with the help of motion vector through current frame [17–20]. This represents the displacement of block among successive frames using minimization of cost measures (Sum of Absolute Difference (S.A.D), Mean Square Difference (M.S.D), and Mean Absolute Difference (M.A.D)) [21].

2.1 Block-Based Motion Estimation Algorithms

The prominent technique in motion estimation method is BME in video coding [17–20]. In this technique, a matching will do between the current and estimated block within successive frames respect to searching window. This gives a minimum distortion value with the use of cost measures [22]. The efficient block-based motion estimation algorithm is Adaptive Rood-Pattern Search (ARPS) algorithm among all other algorithms. The process of ARPS algorithm [23, 24] is:

Step 1. Start with left point in macro-block using predicted motion vector. This step will directly search local region through raster scan order for finding minimum weight.

Step 2. The predicted motion vector is shown in Fig. 1. This point is checked by M.A.D. as locations of rood-pattern. The rood-pattern has step size which

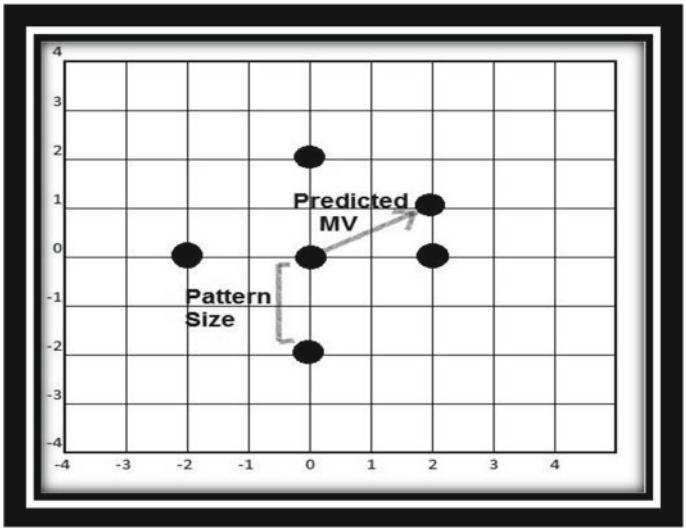


Fig. 1 Procedure of adaptive rood search pattern [23, 25]

is equal to maximum of both coordinate points, i.e., $S = \text{Max}\ (|x|, |y|)$. For all macro-blocks, the rood-pattern step size of 2 pixels is fixed.

Step 3. The point that has least value of M.A.D. becomes an origin in local search. In this area, a compact, fixed, and unrestricted; and small Diamond Searching Pattern (SDSP) is used.

Step 4. This pattern is repeated until least value is found at center of SDSP.

Step 5. If least value of M.A.D. at center of rood-pattern, search completely stops halfway at any moment.

3 Proposed Pattern in Algorithm for Block-Based Motion Estimation Algorithm

This proposed algorithm is an improved pattern in Adaptive Rood-Pattern Search (ARPS) algorithm. In this, small rectangular shaped searching (RSS) pattern is used instead of small diamond search pattern as well as large rectangular shaped searching (RSS) pattern is used instead of large diamond search pattern for better results (Figs. 2, 3, and 4).

The procedure of improved pattern in ARPS algorithm is:

Step 1. Initially start with leftmost column in raster order for calculating minimum weight in each macro-block for finding step size.

Step 2. Check this point using M.A.D. value as locations of rood-pattern. The rood-pattern has step size which are fixed at 2 pixels in each macro-block such as large rectangular search pattern (LRSP).

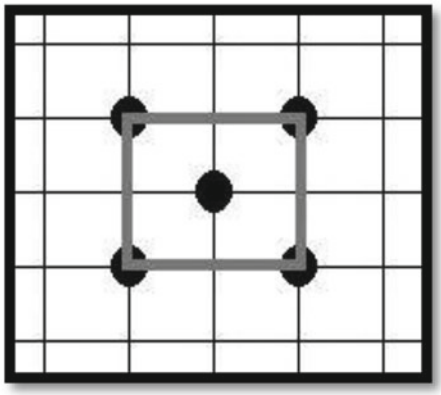


Fig. 2 Small rectangular shaped searching pattern used in proposed algorithm

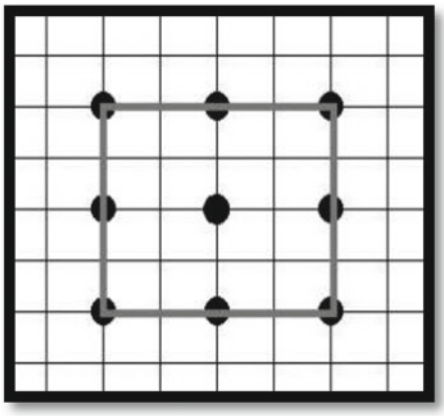


Fig. 3 Large rectangular shaped searching pattern used in proposed algorithm

Step 3. The points with least weight become an origin in searching window. In this area, a compact, fixed, and unrestricted; and small rectangular search pattern (SRSP) is used.

Step 4. Repeat this pattern until least value of M.A.D. is found at center of SRSP.

Note: If least value at center of rood-pattern, search completely stops halfway at any moment.

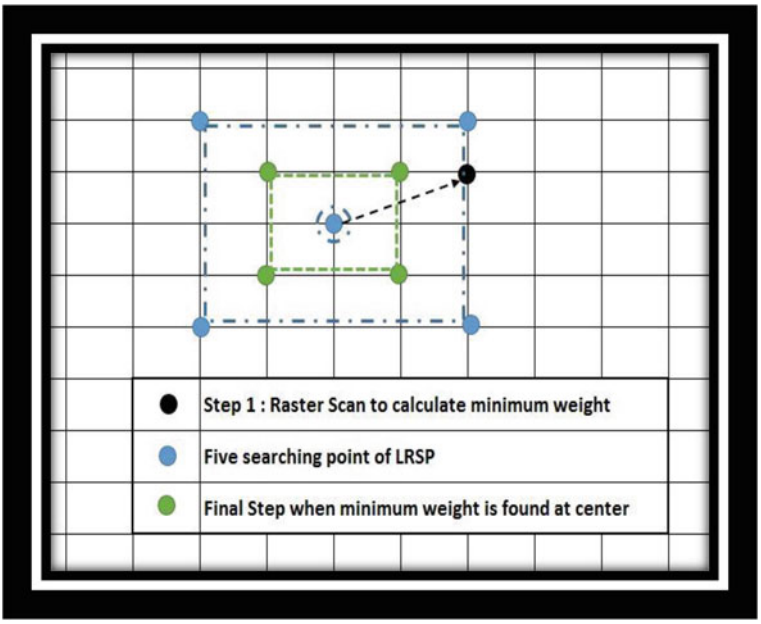


Fig. 4 Proposed pattern in adaptive rood-pattern search algorithm

4 Analysis and Result

With the objective of measuring the performance of proposed algorithms on basis of searching points and PSNR values. This result of proposed algorithms is simulating in MATLAB R2010a. All the performances are carried out on standard video sequences of 50 frames [26].

In Fig. 5, it is clearly seen that the average value of points at the time of searching per macro-block in IARPS is 7.145276, and the average value of points at the time of searching per macro-block in ARPS is 8.120915. From this value, we can say that IARPS is better performing compared to ARPS algorithm in terms of searching points.

In Fig. 6, it is clearly seen that the average PSNR value in IARPS algorithm is 16.286381 dB, and the average points of searching per macro-block in ARPS algorithm is 16.535421 dB. From this value, we can say that IARPS algorithm is slightly better performing compared to ARPS algorithm in terms of value of PSNR.

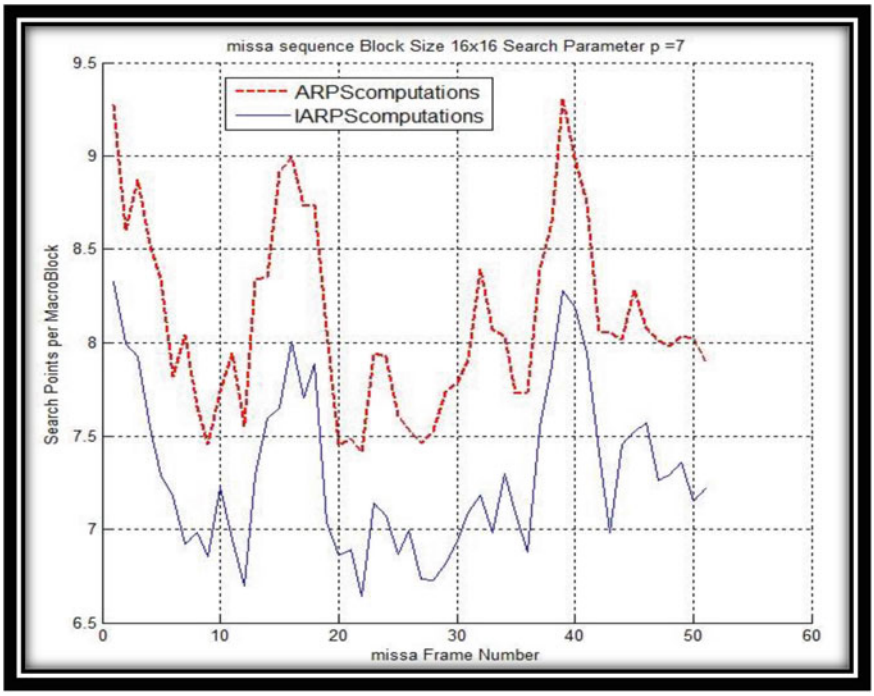


Fig. 5 Performance of a proposed algorithm in terms of number of searching points per macro-block

5 Summary and Conclusion

In this paper, block-based ME algorithms are implemented in MATLAB, and the results are shown in Figs. 3 and 4. From these figures, it is clearly seen that the average searching points per macro-block in ARPS algorithm is 8.120915, and the average points while searching per macro-block in IARPS algorithm is 7.145276. While in terms of PSNR value, average PSNR value of ARPS algorithm is 16.535421 dB and average PSNR value of ARPS algorithm is 16.286381 dB. From this value, it is inferred that this proposed pattern in ARPS performed efficiently based on both number of searching points per macro-block and PSNR value. In future, computational cost can also be reduced to improve the efficiency of the estimation for smoothening the motion in video. If parallelism is done at the frame level, then overall computational cost can be minimized in motion estimation algorithms for compressing a video.

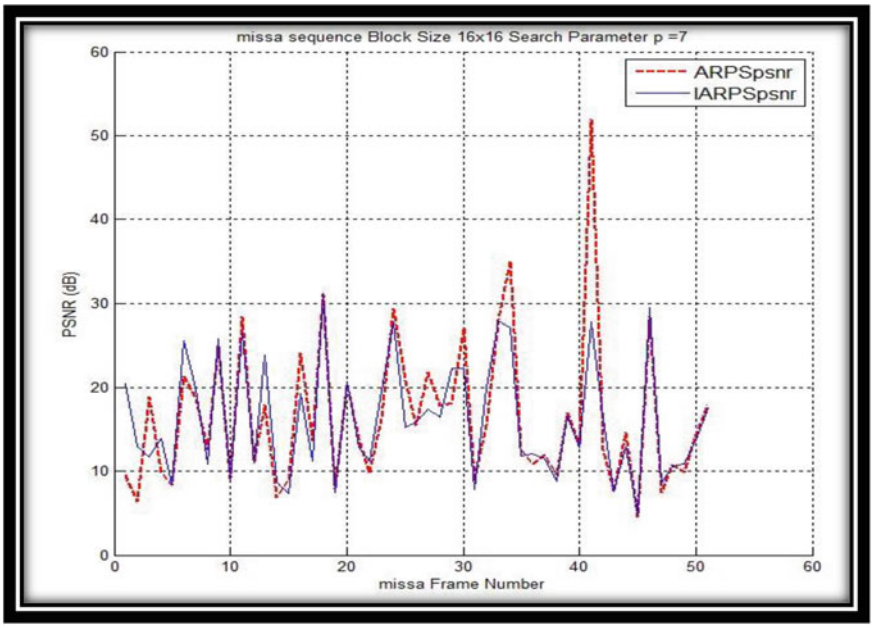


Fig. 6 Performance of proposed algorithm in terms of PSNR value

References

1. Rai A, Chand Gupta S, Choudhury T (2018) Image compression using block truncation coding algorithm. In: Proceedings on 2018 international conference on advances in computing and communication engineering, ICACCE 2018. https://doi.org/10.1109/ICACCE.2018.8441702
2. Khatri A, Choudhury T, Singh TP, Shamoon M (2019) Video analytics based identification and tracking in smart spaces. In: 2019 international conference on contemporary computing and informatics (IC3I), 261–267
3. Mittal A, Khan FS, Kumar P, Choudhury T (2018) Cloud based intelligent attendance system through video streaming. In: Proceedings of the 2017 international conference on smart technology for smart nation, SmartTechCon 2017. https://doi.org/10.1109/SmartTechCon.2017.8358587
4. Mathur A, Bhandari R, Mishra A (2019) An improved block-based motion estimation algorithm for compressing a video. In: Lecture notes on data engineering and communications technologies, vol 26. Springer, Cham. https://doi.org/10.1007/978-3-030-03146-6_21
5. Wien M (2015) High efficiency video coding, signals and communication technology. © Springer-Verlag, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44276-0_2
6. Ratnottar J, Shrivastav M, Joshi R (2012) Comparative study of motion estimation & motion-compensation for video compression. IJETTCS 1(1). ISSN 2278-6856
7. Tomar R, Sastry HG, Prateek M (2020) A novel protocol for information dissemination in vehicular networks. In: Lecture notes in computer science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol 11894. LNCS, Springer, Cham, pp 1–14
8. Tomar R, Prateek M, Sastry HG (2017) A novel approach to multicast in VANET Using MQTT. Ada User J 38(4):231–235

9. Tomar R, Sastry H, Prateek M (2020) Establishing parameters for comparative analysis of V2V communication in VANET. J Sci Ind Res (JSIR) 79(01):26–29
10. Parashar N, Soni R, Manchanda Y, Choudhury T (2018) 3D modelling of human hand with motion constraints. In: Proceedings of the international conference on computational techniques, electronics and mechanical systems, CTEMS 2018. https://doi.org/10.1109/CTEMS.2018.8769229
11. Dhamija J, Choudhury T, Kumar P, Rathore YS (2017) An advancement towards efficient face recognition using live video feed: for the future. In: 2017 3rd international conference on computational intelligence and networks (CINE), pp 53–56
12. Kunwar V, Sabitha AS, Choudhury T, Aggarwal A (2019) Chronic kidney disease using fuzzy C-means clustering analysis. Int J Bus Anal 6(3). https://doi.org/10.4018/IJBAN.2019070104
13. Gao Y, Liu P, Jia K (2014) An adaptive motion estimation scheme for video coding. Hindawi Publishing Corporation. Sci World J, 14 pages Article ID 381056. https://doi.org/10.1155/2014/381056
14. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of Weiler-atherton algorithm to address loose polygons. J Sci Ind Res 78:771–774
15. Choudhury T, Aggarwal A, Tomar R (2020) A deep learning approach to helmet detection for road safety. J Sci Ind Res (JSIR) 79(06):509–512
16. Kumar V, Choudhury T (2019) Real-time recognition of malignant skin lesions using ensemble modeling. J Sci Ind Res (JSIR) 78:148–153
17. Sri MS (2018) Motion estimation using block matching algorithm. Int J Adv Eng Res Sci 5(5):346–348. ISSN: 2349-6495(P) | 2456-1908(O)
18. Bull DR (2014) An introduction to video coding. Academic Press Library in Signal Processing, Elsevier Ltd. https://doi.org/10.1016/B978-0-12-420149-1.00001-6
19. Ahmed Z, Hussain AJ, Al-Jumeily D (2012) Edge detection for fast block-matching motion estimation to enhance mean predictive block matching algorithm. In: Proceedings of the IEEE international symposium on innovations in intelligent systems and applications (INISTA), pp 1–5
20. Hui-Yu H, Shih-Hsu C (2011) Block motion estimation based on search pattern and predictor. In: Proceedings of the IEEE symposium on computational intelligence for multimedia, signal and vision processing (CIMSIVP), pp 47–51
21. Jakubowski M, Pastuszak G (2013)Block-based motion estimation algorithms—a survey. Opto-Electron Rev (Springer)
22. Manikandan LC, Dr. Selvakumar RK (2014) A study on block matching algorithms for motion estimation in video coding. Int J Sci Eng Res 5(7). ISSN 2229-5518
23. Nie Y, Ma KK (2002)Adaptive rood pattern search for fast block-matching motion estimation. IEEE Trans Image Process 11(12):1442–1448
24. Zhang Y, Zhang C, Fan R (2018) Fast motion estimation in HEVC inter coding: an overview of recent advances. In: Proceeding APSIPA annual summit and conference 2018. ISBN: 978-988-14768-5-2 ©2018 APSIPA
25. Richardson IEG (2010) The H.264 advanced video compression standard, 2nd edn. Wiley, UK
26. Gonzalez R, Woods R, Eddins S (2009) Digital image processing using MATLAB, 2nd edn. Gatesmark Publishing

Modelling and Simulation of Distance Relays with Auto-Recloser Function Using PSCAD: An Active Learning Tool

Anupama Prakash, Pallavi Choudekar, and Nasir ul Islam Wani

Abstract Protection is an utmost important part of a power system as it helps in its proper and safe operation. In order to comprehend the complexity of it, modern simulation tools can be used as an active learning tool along with classroom teaching. The simulation tools considerably improve the level of understanding of the theoretical concepts without expensive lab-setups. This paper presents a methodical approach using simulation tool PSCAD to enhance the teaching–learning experience of transmission line protection using distance relay with the auto-recloser feature. The response of the protection scheme is studied and analyzed when the temporary and permanent fault occurs in the system.

Keywords Auto-recloser · Distance relay · Permanent faults · PSCAD · Simulation · Temporary fault

1 Introduction

The objective of a protection scheme is to maintain the stability of the power system by isolating only the components that are under fault, whilst leaving the maximum part of the network as possible still in operation. On the occurrence of the fault in the system, the protective relay detects the fault and generates a trip signal to operate the breaker. There are various types of protection relays. Distance relay is mostly used as a protection device for transmission lines. They respond to the impedance between the relay location and the fault location. As the impedance per km of a transmission line is fairly constant, these relays respond to the distance to a fault on the transmission line.

A. Prakash · P. Choudekar · N. I. Wani (✉)
Electrical and Electronics Engineering Department, AMITY University, Noida, Uttar Pradesh, India

A. Prakash
e-mail: aprakash1@amity.edu

P. Choudekar
e-mail: pachoudekar@amity.edu

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_40

About 80–90% of the faults occurring are transient in nature. The timing of opening and closing of the breaker depends upon the nature of the fault, i.e., temporary or permanent. A circuit breaker (CB) is equipped with an auto-recloser mechanism that automatically closes the breaker which is opened due to a fault. The auto-recloser allows a selected number of attempts to restore service after adjustable time delays. For example, a recloser may have two or three "fast" reclose operations with a few seconds delay, then a longer delay and one reclose; if the last attempt is not successful, the recloser will lockout and require human intervention to reset. If the fault is a permanent fault then the auto-recloser will exhaust its pre-programmed attempts to re-energize the line and remain tripped off until manually commanded to try again. In distribution systems, a maximum of three consecutive reclosures is allowed. The multiple reclosures help in burning out the objects that have caused the fault, say, a tree branch, and thus clearing out the fault. In contrast, in transmission systems, because reclosures impose an arduous duty to the CB and other elements of the power system due to high levels of fault MVA, only one reclosure is permitted [1, 2].

There is another feature of reclaim time is added to the relay operation to prevent the failure of the circuit breaker during multiple faults. It is important for the CB to regain its breaking strength before handling another fault otherwise CB will not be able to provide protection against the fault. If the reclosing shot has been carried out and the line is energized and a new fault occurs before the reclaim time has elapsed, the auto-reclosing equipment is blocked and a signal for definite tripping of the breaker is obtained. After the reclaim time has elapsed, the auto-reclosing equipment returns to the starting position, and a new reclosing sequence can occur [3–5].

In order to understand the concept of protection, it is important for the students to have a practical experience of the same by implementing the theoretical concept on a practical system. Since it is very difficult to mimic the actual power system in a scaled-down manner in the laboratory, the student is often left with misunderstood concepts and misinterpretations. Simulation is an effective alternative to scaled-down hardware to illustrate the various relaying concepts of the power system. Simulation tools such as MATLAB, PSCAD, and PSIM, can be used as teaching aids to augment the understanding of the power system relaying concepts. The advantage of simulation tools is that the learner can actually observe the instantaneous voltages and currents, the overvoltage and under-voltage conditions, the transient spikes produced by the fault conditions, unbalance created, operation, and effect of protective devices [6–8].

The main contribution of this paper is that without any hardware model we can do the fault analysis of the system. The simulation can help in the overall fault analysis of the system. Since it is very difficult to mimic the actual power system in a scaled-down manner in the laboratory, the student is often left with misunderstood concepts and misinterpretations. Simulation is an effective alternative to scaled-down hardware to illustrate the various relaying concepts of the power system. Simulation tools such as MATLAB, PSCAD, and PSIM, can be used as teaching aids to augment

the understanding of the power system relaying concepts. The advantage of simulation tools is that the learner can actually observe the instantaneous voltages and currents, the overvoltage and under-voltage conditions, the transient spikes produced by the fault conditions, unbalance created, operation, and effect of protective devices. This paper presents a systematic approach for the teaching of the transmission line protection, using distance relays with auto-recloser, through simulation to undergraduate students. PSCAD has been used as the simulation tool, though the method is similar if any other tool such as PSIM and NETPLAN are used. The distance relay is modelled in PSCAD as described by mathematical equations. The auto-recloser function is modelled to help the students to understand its operation for permanent and temporary faults.

Section 1 describes the fault and benefits of simulation tools in overall fault analysis and the overall effectiveness of the simulation tools in the relaying concepts of the power system. Section 2 explains the different types of faults and the model of distance relays and their installation in the power system. The distance relay is modelled in PSCAD as described by mathematical equations. Section 3 talks about various case studies of the functioning of the distance relays, a typical 230 kV, 100 km transmission line modelled in PSCAD.

2 Model of Distance Relay for Transmission Line Protection

Distance relays are typically installed for the protection of the transmission lines. A distance relay, as its name implies, has the ability to detect a fault within a pre-set distance along an overhead transmission line or the underground power cable from its location. Every power line has a resistance and reactance per kilometre related to its design and construction; hence, its total impedance is a function of its length.

Table 1 lists the presence of sequence components, of voltage and current, during various faults. It is clear that the positive(+ive) sequence component is the only component that is present during all faults. Thus, it would be prudent to measure +ive sequence impedance between the relay location and the fault so as to cater to every fault [9, 10].

Table 1 Presence of sequence components in various faults

Fault	Positive sequence	Negative sequence	Zero sequence
L-G	Yes	Yes	Yes
L-L	Yes	Yes	No
L-L-G	Yes	Yes	Yes
L-L-L	Yes	No	No

Line-Line fault: The ratio of difference of +ive and –ive sequence voltages and currents gives the desired value of +ive sequence impedance between the relay location and the fault point as shown in Eq. 1.

$$Z = \frac{(V_{a1} - V_{a2})}{(I_{a1} - I_{a2})}. \tag{1}$$

The line voltages V_a, V_b, and V_c can be expressed in terms of zero (V_{a0}), positive (V_{a1}), and negative (V_{a2}) sequence components as shown in Eqs. 2–4.

$$V_a = V_{a0} + V_{a1} + V_{a2} \tag{2}$$

$$V_b = V_{b0} + V_{b1} + V_{b2} \tag{3}$$

$$V_c = V_{c0} + V_{c1} + V_{c2} \tag{4}$$

Since $V_{a0} = V_{b0} = V_{c0}$; $V_{b1} = a^2 V_{a1}$, $V_{c1} = aV_{a1}$, $V_{b2} = aV_{a1}$, $V_{c2} = a^2 V_{a1}$
Therefore,

$$V_{a1} - V_{a2} = \frac{V_b - V_c}{a^2 - a'}; \text{ where } a = 1\angle 120^{\theta} \tag{5}$$

Similarly,

$$I_{a2} = \frac{I_b - I_c}{a^2 - a} \tag{6}$$

Hence,

$$\frac{V_{a1} - V_{a2}}{I_{a1} - I_{a2}} = \frac{V_b - V_c}{I_b - I_c} = Z_1 \tag{7}$$

Thus, a distance measuring unit with voltages of ($V_b - V_c = V_{\text{bc}}$) and current of ($I_b - I_c$) will measure positive sequence impedance up to fault point Z_1, in case of phase *b*–*c* faults as shown in Eq. 7. Similarly, V_{ab}, $I_a - I_b$ and V_{ca}, $I_c - I_a$ for phase *a*-to-*b* and phase *c*-to-*a* faults [9].

Line-Ground Faults: In case of ground faults,

$$V_{a0} + V_{a1} + V_{a2} = I_{a1}z_1 + I_{a2}z_2 + I_{a0}z_0 \tag{8}$$

However,

$$V_{a0} + V_{a1} + V_{a2} = V_a \tag{9}$$

From Eqs. 8 and 9,

$$V_a = I_{a1}z_1 + I_{a2}z_2 + I_{a0}z_0 - I_{a0}Z_1 + I_{a0}Z_1 \tag{10}$$

Since,

$$\mathrm{Z2} = \mathrm{Z1}V_a = (I_{a1} + I_{a2} + I_{a0})_{Z1} + I_{a0}z_0 - I_{a0}Z_1 \tag{11}$$

$$V_a = I_a Z_1 + (Z_0 - Z_1)I_{a0} \tag{12}$$

$$V_a = (I_a + \frac{Z_0 - Z_1}{Z_1} I_{a0})_{Z_1} \tag{13}$$

$$I_{a0} = \frac{I_a + I_b + I_c}{3} \tag{14}$$

Since, $I_{res} = I_a + I_b + I_C$

$$V_a = (I_a + \frac{Z_0 - Z_1}{3Z_1} I_{res})_{Z_1} \tag{15}$$

$$Z_1 = V_a/(I_a + \frac{Z_0 - Z_1}{3Z_1} I_{\mathrm{res}}), \tag{16}$$

where $k = \frac{Z_0 - Z_1}{3Z_1}$

$$Z_1 = V_a/(I_a + kI_{\mathrm{res}}) \tag{17}$$

Thus, a distance measuring unit with voltages of (V_a) and current of $(I_a + k\, I_{\mathrm{res}})$ will measure positive sequence impedance up to fault point Z_1, in case of phase a-to-g faults as per Eq. 17. Similarly, V_b, $(I_b + k\, I_{\mathrm{res}})$ and V_c, $(I_c + k\, I_{\mathrm{res}})$ for phase *b*-to-*g* and phase *c*-to-*g* faults.

Figure 1 shows the distance relay logic for the protection of the transmission line from all types of faults. The OR gate will issue the trip command to the circuit breaker in case of fault [9]. The positive sequence impedance is calculated in real time and when it is below the pre-set value, the distance relay gives a command to operate the circuit breaker. This model is used in developing the distance relay logic in PSCAD.

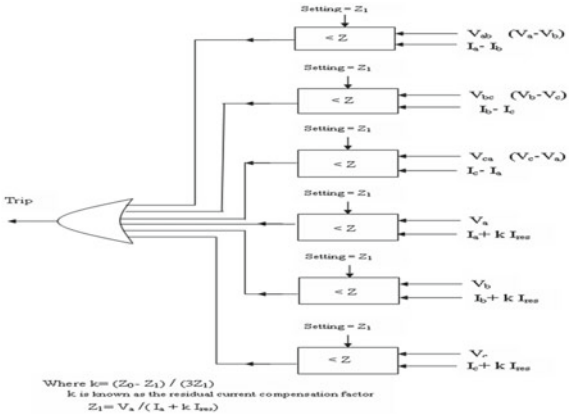


Fig. 1 Complete protection of a three-phase line using distance relay

3 Case Studies

In order to study the functioning of the distance relays, a typical 230 kV, 100 km transmission line has been modelled in PSCAD [3, 4]. The transmission line is connected between two buses Bus1 and Bus2. The generator G1is rated 100 MVA, 230 kV, 60 Hz, and generator G2 is rated 100 MVA, 230 ∠20°kV, 60 Hz which are connected to bus1 and bus2, respectively. The single line diagram of the 100 km long transmission line provided with the breakers at both ends is shown in Fig. 2. The fault occurs at 60 km from bus1. The fault type can be controlled through the rotary switch, and the duration of fault can be controlled through timed fault logic.

The distance relay is developed in PSCAD as per the logic discussed in section II. The logic for one phase to ground (A–G) and one line to line (A–B) is shown in Fig. 3. Similar logics are developed for other phases to ground and line-to-line faults.

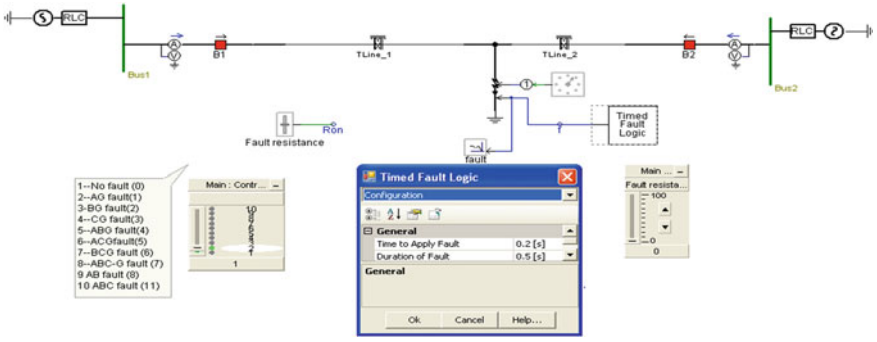


Fig. 2 Single line diagram of transmission line with fault logic block

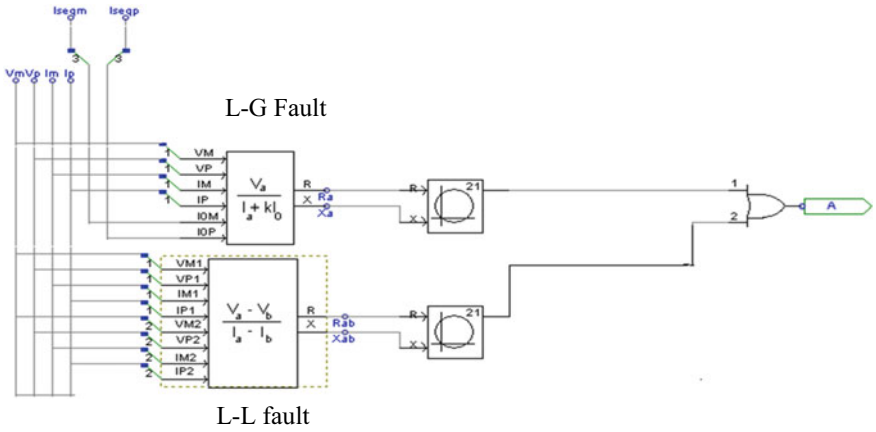


Fig. 3 Distance relay logic in PSCAD

The logic of auto-recloser is also needed to be developed in PSCAD, as it is not available as a built-in function. The function of auto-recloser is to close the breaker, after it is opened by the relay on the occurrence of a fault, with a delay of 1 s. If the fault is temporary and gets cleared in 1 s., the breakers remain closed and the supply to the system is restored. If it is a permanent fault, on closing the breaker, the relay again senses the fault and opens the breaker permanently. The logic building for the auto-recloser function, for different cases, is presented in the subsequent sections, along with detailed results.

3.1 Differrent Cases

CASE 1: Distance protection with auto-recloser function for temporary fault.

On the occurrence of fault, the distance relay detects the fault and issues trip command to the circuit breakers BRK1 and BRK2. Since the breaker operating time should be 60 ms, the two breakers open to clear A–G fault at 0.26 ms. After a delay of dead time (1 s) due to auto-recloser function, the circuit breaker closes its contacts and the contacts remain closed as the fault is the transient fault and is cleared at 0.8 s.

Fault specifications:

- *Fault (A-G) is at transmission line T1*
- *Fault duration- 0.2–0.5 s.*

The voltage and current waveforms captured at bus1, for this fault, are as shown in Fig. 4. Since it is an A-G fault, the voltage of phase-A has dropped and the current of phase-A has increased from 0.2–0.5 s as seen in Fig. 4a, b, respectively.

The distance relay issues the trip command to the breaker as shown in Fig. 5. The circle in Fig. 5a shows the region of operation of the distance relay and the fault

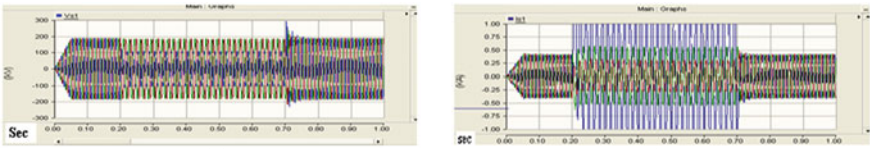


Fig. 4 **a** Voltage waveform at bus 1 for temporary A–G fault. **b** Current waveform at bus 1 for temporary A–G fault

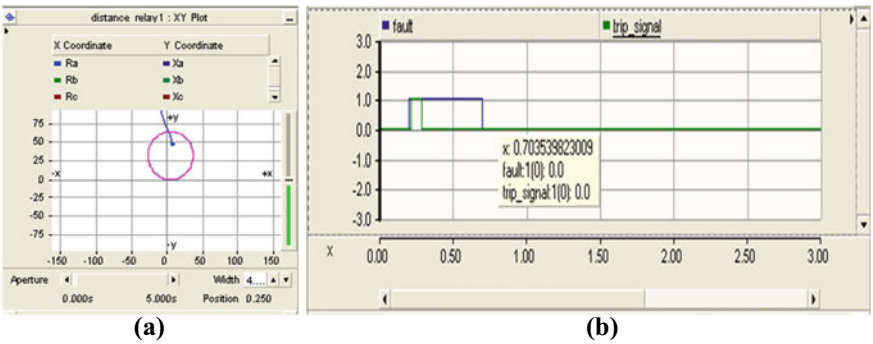


Fig. 5 Fault and trip signal waveform of distance relay

location. Since the fault is within the relay reach, the relay operates. The trip signal issued by the relay is shown in Fig. 5b.

The auto-recloser logic is developed in PSCAD as shown in Fig. 6 and discussed below.

Block-1 is the module designed to give the magnitudes and phases of the voltages and current measured by multi-meter connected to bus1 of Fig. 2 along with the

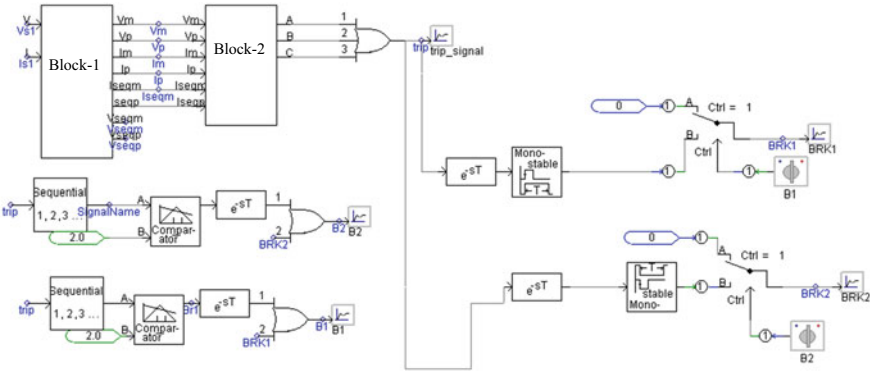


Fig. 6 Auto-recloser logic

magnitudes and phase of the sequence component of the same currents and voltages. Block-2 is the distance relay that is designed as per the logic discussed in section II to produce the trip signal on the occurrence of the fault. The trip signal is given to BRK1 and BRK2 supported with auto recloser function logic. However, the operation of the breakers (shown in Fig. 6) is controlled by signals B1 and B2 which are obtained from the OR gate used in the sequential circuit.

For temporary fault, the trip signal is a pulse as shown in the graph. On receiving the trip signal time delay blocks (T.D.-1 and T.D.-2) give a time delay of 60 ms so that breakers open after a delay of 60 ms. The mono-stable blocks (M.S-1 and M.S-2) make the duration of trip pulse to 1 s. This ensures that the breaker B1 and B2 remain open for one second. The trip signal is also given to the sequential block which increments by one on the change in the input, i.e., trip signal. Thus at 0.2 s., the output of the sequential block is 1 and again at 0.28 s when the trip signal becomes zero from 1, the output is 2. In case of a temporary fault, a trip signal is only one pulse and hence the output of the sequential block will not be more than 2. The output of the sequential block is given to the comparator whose one input is set at 2. It is configured to give an output of 1 when sequential output is more than 2, and output 0 if it is less than 2. In case of temporary fault, the outputs of COM-1 and COM-2 are zero. These outputs are given to time delay blocks (T.D.-3 and T.D.-4) to get a time delay of 1 s. The delayed output from the time delay block and the signal from Breaker is given to the OR gate. The outputs of two OR gates (2 and 3) will control the operation of breakers B1 and B2. Since in this case, the outputs of T.D-3 and 4 are zero, the breaker operation is controlled through the signals BRK1 and BRK2. The sequence of operations for the transient fault and their corresponding output waveforms are as summarized below:

- Time of Faultoccurrence −0.2 s (Fig. 7a)
- Distance Relay initiates trip signal at 0.2 s (Fig. 7a)
- Mono-stable vibrator gives a signal at 0.26 s (after a delay of 60 ms) for a period of 1 s. So that beaker does not close before 1 s of opening (Fig. 7a).
- The output of sequential is 2 due to two operations; first opening and then closing of breakers (Fig. 7b).
- The output of the comparator circuit is 0 since the output of sequential is 2 (Fig. 7c).
- Breakers B1 and B2 operations are due to signal BRK1 and BRK2 received from relay logic. Thus Breakers open at 0.26 s and remains open till 1.26 s. After 1.26 s. breakers will close (Fig. 7d).
- Since the fault duration is 0.5 s, it gets cleared before 1.26 s, and thus relay will not issue any trip signal to the breakers and they continue to remain close.

CASE 2: Distance protection with auto-recloser function for permanent fault**.**
Fault specifications:

- ***Fault (A-G) is at transmission line T1***
- ***Fault starts at 0.2 s.***

A permanent a-g fault is created on the transmission line at a distance of 60 km from the bus1. The fault is initiated at 0.2 s. The voltage and current waveforms

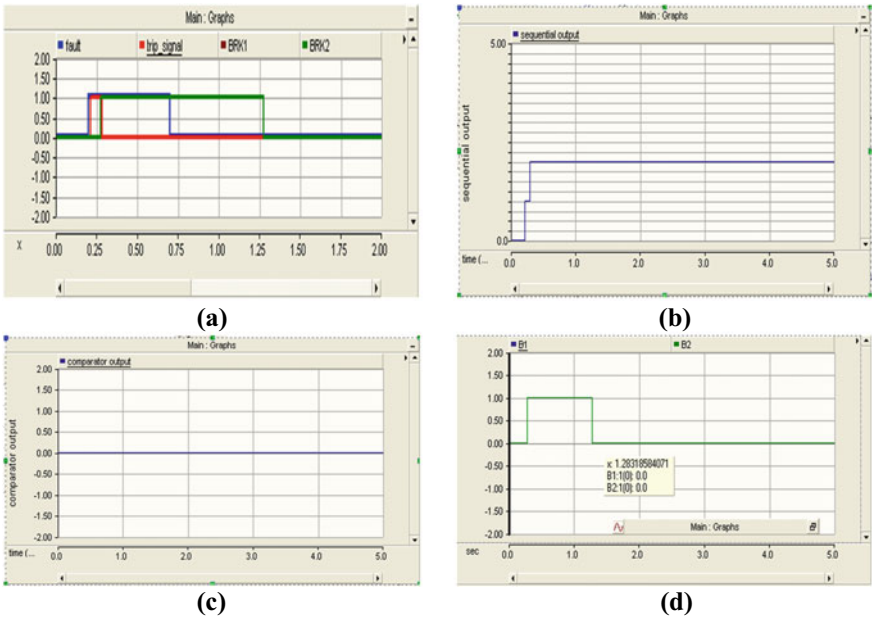


Fig. 7 **a** Waveform of fault, trip signal, and mono-stable vibrator at 0.2 s. **b** Waveform of sequential output. **c** Waveform of the comparator output. **d** Waveform of breakers B1 and B2 operation

captured at bus1, for this case, are as shown in Fig. 8. Since it is an A-G fault, the voltage of phase-A has dropped and the current of phase-A has increased at 0.2 s as seen in Fig. 8a, b respectively.

The distance relay issues the trip command to the breaker as shown in Fig. 9. The circle in Fig. 9a shows the region of operation of the distance relay and the fault location. Since the fault is within the relay reach, the relay operates. The trip signal issued by the relay is shown in Fig. 5b.

For permanent fault also, breakers close after 1 s of opening, as in the case of temporary fault. On closing of breakers, transmission system gets energized and the fault is again seen by the relay. The relay again issues a trip signal to the breakers

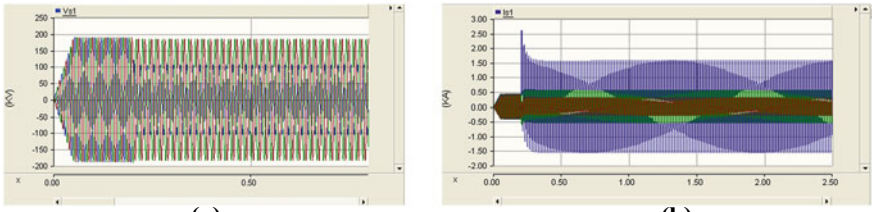


Fig. 8 Voltage and current waveforms at bus 1 for permanent A–G fault

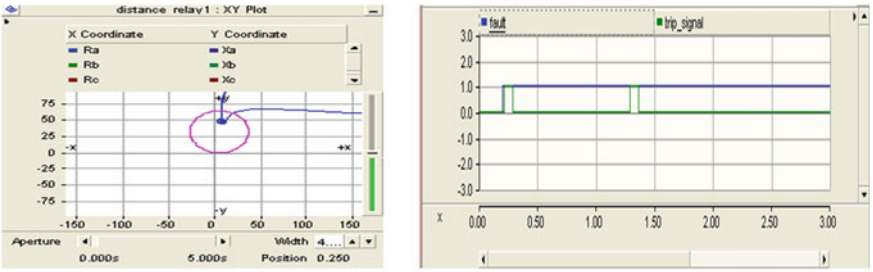


Fig. 9 Fault and trip signal waveform of distance relay

as shown in Fig. 9b. Thus in case of permanent fault, the trip signal consists of two pulses. The output of the sequential block will be four as seen from the waveform as shown in Fig. 10. Thus the output of the comparator will be high. This is given to

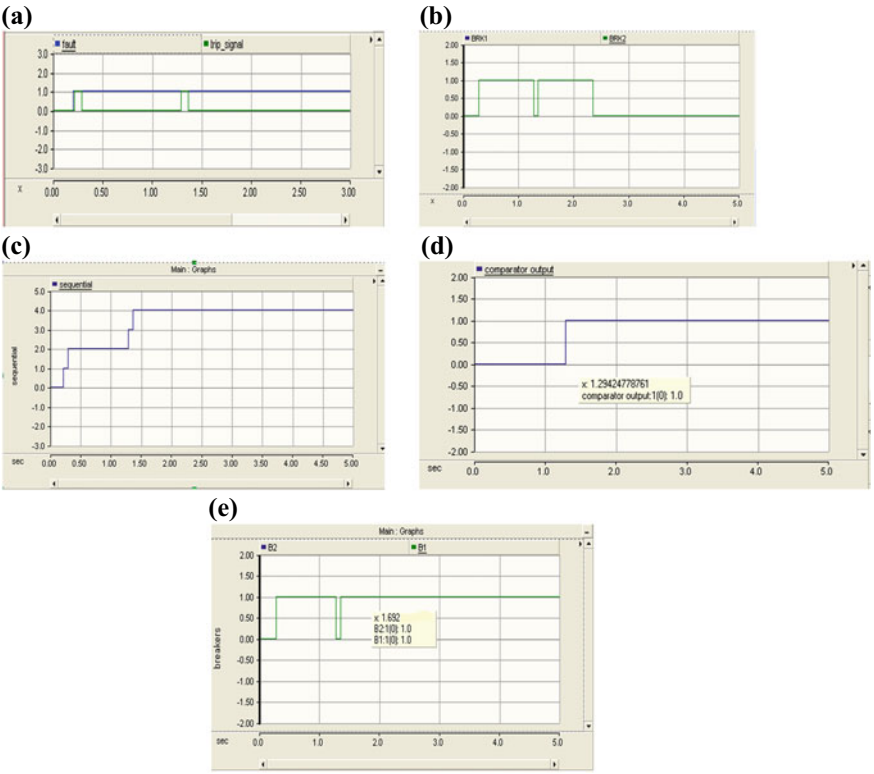


Fig. 10 **a** Waveform of the permanent fault and two pulses of the trip signal. **b** Waveform of the mono-stable vibrator. **c** Waveform of sequential output. **d** Waveform of the comparator output. **e** Waveform of the breakers B1 and B2 operation

the OR gate after a delay of 60 ms. This logic captures the second pulse of the trip and opens the breaker after 60 ms of the second trip signal.

Initially, the breaker operation is controlled from the first trip signal and after the first recloser; it is controlled by the second trip signal which is processed properly by the developed logic. The processed signal is the output of the OR gate. It thus opens the breaker again permanently. The sequence of operations for the permanent fault and their corresponding output waveforms are as summarized below:

- Fault occurrence −0.2 s (Fig. 10a).
- Distance Relay initiates trip signal at 0.2 s (Fig. 10a).
- Mono-stable vibrator gives a signal at 0.26 s (after a delay of 60 ms) for a period of 1 s. So that beaker does not close before 1 s of opening (Fig. 10b).
- Breaker opens at 0.26 s and remains open till 1.26 s. After 1.26 s. the breaker will close (Fig. 10b).
- Since the fault is permanent, the relay will again issue a trip signal to the breakers (Fig. 10a).
- Mono-stable vibrator again gives a signal at 0.26 s (after a delay of 60 ms) for a period of 1 s (Fig. 10b).
- The output of sequential logic is 4 as it sees four operations; first opening, closing then again opening and closing of breakers (Fig. 10c).
- The output of the comparator circuit is 1 since the output of sequential is 2 (Fig. 10d).
- Breakers B1 and B2 operations are due to signal from comparator logic. Thus Breakers open permanently at 1.32 s (Fig. 10e).

Thus the second trip signal is detected by a sequential-comparator logic circuit. A delay of 60 ms is provided to the trip signal, in order to open the breakers after a delay of 60 ms of the second trip and the breakers are open permanently.

CASE 3: Distance protection with reclaim function.

The auto-recloser function with the reclaim time feature is also simulated to make the students understand the operation of the protection system in case of multiple faults. Reclaim time is the time between two consecutive auto reclosers. If the second fault occurs before the elapse of reclaim time auto recloser function will be disabled and CBs will remain open. If the second fault is initiated after the lapse of reclaim time, the auto-recloser function will be carried out as usual.

In order to implement reclaim time logic, it is important to measure the time between two trip signals. If the time between two trip signals is less than 3.0 s, then auto recloser function is not initiated and breakers which are remain opened by the relay remain opened. If time between two trip signals is more than 3.0 s, then the auto recloser function is initiated and the breakers will be closed to check whether the fault is temporary or permanent.

Figure 11a shows two sequential blocks with comparator, sequential block gives 1 on the increment of the input and comparator gives high when the magnitude of the input signal crosses the set integer value. Sequential block-1 and comparator-1 give high when the first trip signal is issued by the relay on the occurrence of the

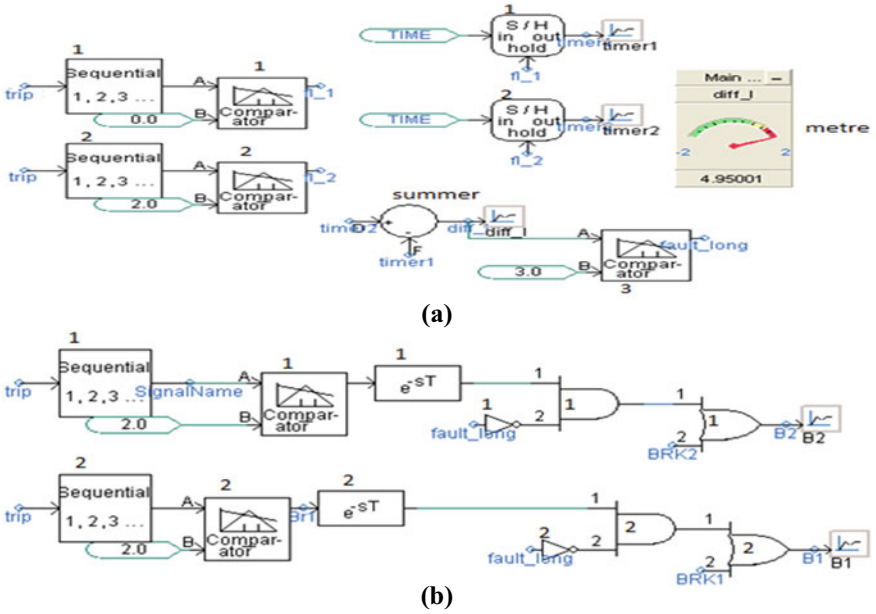


Fig. 11 **a** Reclaim time logic. **b** Auto-recloser logic with reclaim time function

first fault, this time is captured by sample hold circuit. Similarly, the combination of sequential block-2 and comparator-2 gives high when the second fault occurs and that time is captured by sample hold circuit-2, the difference between these time is calculated by summer, and its output is given to comparator-3 if the difference in the time of these two times is greater or equal to 3 faults is considered an as long fault, and the auto-recloser function is restored.

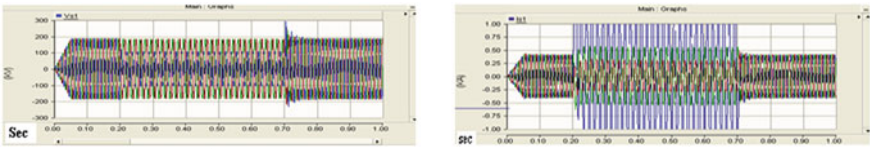
Figure 11b shows the auto-recloser logic with the reclaim time feature. One AND gate is introduced to take care of reclaim logic. One input of AND block is to enable the auto-recloser function, and the other input is NOT of long fault, i.e., when the long fault occurs, auto-recloser signal is allowed to pass through AND gate, and that signal is given to OR gate of with other signals of breaker and if it is not a long fault then only breaker signal operates without auto-recloser function.

Fault specifications of the first fault:

- First Fault (A–G) is at the transmission line T1.
- First Fault duration-0.2–0.5 s.

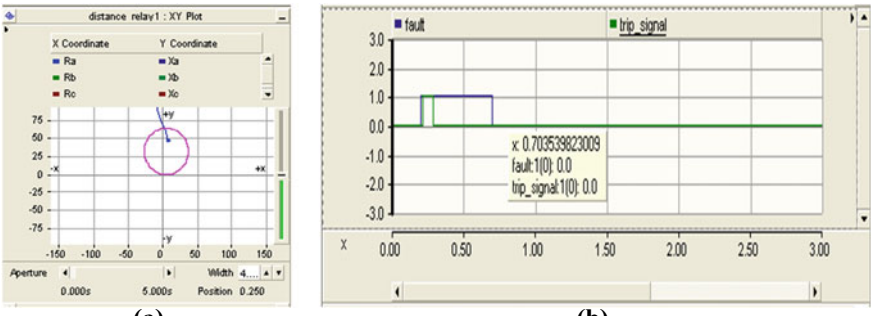
The voltage and current waveforms captured at bus1, for this fault, are shown in Fig. 12. Since it is an A-G fault, the voltage of phase-A has dropped and the current of phase-A has increased from 0.2 to 0.5 s as seen in Fig. 12a, b respectively.

The distance relay issues the trip command to the breaker as shown in Fig. 13. The circle in Fig. 13a shows the region of operation of the distance relay and the



(a) (b)

Fig. 12 A voltage waveform at bus 1 for temporary A–G fault. **b** Current waveform at bus 1 for temporary A–G fault



(a) (b)

Fig. 13 **a** Fault and trip signal waveform of distance relay. **b** Second fault (A–G) before 3.0 s of the occurrence of the first fault

fault location. Since the fault is within the relay reach, the relay operates. The trip signal issued by the relay is shown in Fig. 13b.

The first fault is temporary, so the breaker is closed after 1 s and the second fault occurs before 3 s (as shown in Fig. 14) so the auto-reclosure function is not restored and breakers remain opened as seen in Fig. 15.

Here the first fault is temporary and the second temporary fault occurs after three seconds of the occurrence of the first fault breakers are closed after one second since the auto reclosure function is restored (Fig. 16).

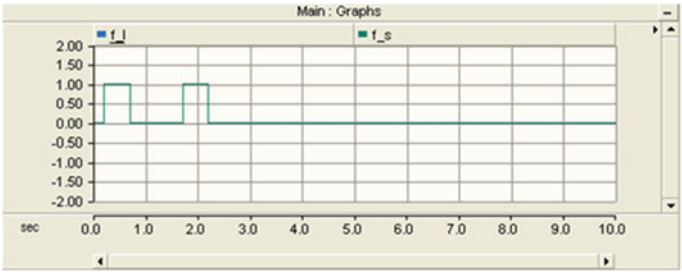


Fig. 14 Graph shows short fault with the difference of less than 3 s between two faults

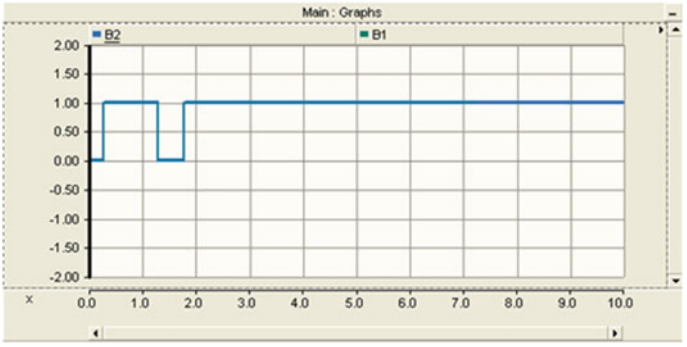


Fig. 15 Breaker operation for two consecutive faults within 3 s

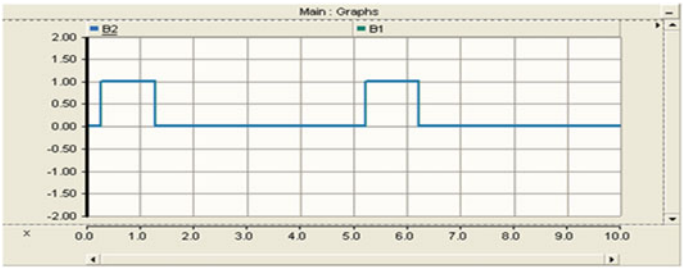


Fig. 16 Breaker operation for the second fault occurring after 3 s of the first fault

The case studies discussed above clearly illustrate the operation of distance relays with an auto-recloser facility under transient and permanent faults. The case studies simulate a real power system, thus enhancing the theoretical concept of auto-recloser, to the student. Some extended exercises which the facilitator can assign to the students are as follows:

- Simulation of the above system for different types of faults and study the behaviour of the distance relay.
- Change the location of the fault and simulate.
- Change the relay setting and observe.
- Compare the system performance, with and without auto-recloser facility.

4 Conclusion

This paper presents a systematic approach for the teaching of the transmission line protection, using distance relays with auto-recloser, through simulation to undergraduate students. PSCAD has been used as the simulation tool though the method is similar if any other tool such as PSIM and NETPLAN are used. The distance relay

is modelled in PSCAD as described by mathematical equations. The auto-recloser function is modelled to help the students to understand its operation for permanent and temporary faults. PSCAD is an easy tool, due to the vast library capabilities present and the drag and drop approach used to build networks. Similar simulation exercises can be developed for other power system applications.

References

1. Dorf RC (ed) (1993) The electrical engineering handbook. CRC Press, Boca Ratonp, p 1319. ISBN 0-8493-0185-8
2. Phadke A, Thorp JG (1990) Computer relaying for power systems. Wiley. ISBN 0471920630
3. Thakre MP, Kale VS (2014) Distance protection for long transmission line using PSCAD. VNIT Nagpur, Int J Adv Eng Technol 6(6):2579–2586
4. Lui JF, Wilson PL, A PSCAD/EMTDC simulation study for protective relay. Manitoba HVDC Rsearch Centre
5. Rambabu M, Venkatesh M, SivaKumar JSV, AyyaRao TSLV (2015) Three zone protection by using distance relays in SIMULINK/MATLAB. Int Res J Eng Technol (IRJET) 02(05)
6. Muller C, Eng P (2005) Power system computer aided design user's guide. Manitoba HVDC Rsearch Centre
7. Dommel (1986) EMTP Reference manual. Bonneville Power Administration
8. Makwana S, Makwana V (2019) Simulation and hardware implementation of over-current relay used for transmission lines. In: 2019 3rd international conference on trends in electronics and informatics (ICOEI), Tirunelveli, India, pp 1189–1193. https://doi.org/10.1109/ICOEI.2019.8862569
9. Paithanakar YG (1998) Transmission network protection theory and practice. Published by MARCEL DEKKAR INC,
10. Anderson PM (1998) Power system protection. IEEE Press Powering engineering series, USA

Resume Data Extraction Using NLP

Umang Goyal, Anirudh Negi, Aman Adhikari, Subhash Chand Gupta, and Tanupriya Choudhury

Abstract Extracting valuable and relevant information from a potential employee's CV to ease the hiring process for employers, automating data extraction, parsing documents of multiple formats, and storing data in a standardized relational database model. The user uploads a single résumé or multiple résumés into the program, the program accepts multiple formats (.pdf, .doc, .rtf, etc.) and converts it into a standard text format which is later parsed for the required information, and the extracted data is organized in a standard defined format. The user can then download the extracted information in the .CSV format.

Keywords Data extraction · Résumé parsing · Natural language processing · Résumé CSV · Data mining

1 Introduction

Due to the nonexistence of a standard résumé format a plethora of formats exists, which hinders the automation of the data extraction process. Since every new résumé may have a different definition and location for various attributes, it is a seemingly difficult task to model a program to fetch the required data values. Every day companies, firms, recruiting, organizations face the challenge of processing several résumés and shortlisting the most desirable and deserving candidates, often the process requires eyeballing and needless to mention hundreds if not thousands of man-hours. The techniques of Natural Language Processing (NLP) offer a rather complex but effective solution to this complicated problem. NLP methods such as chunking, parsing, and regex facilitate the definition of language grammar and extraction of data by the application of custom grammatical rules, and it also enables us to

U. Goyal · A. Negi · A. Adhikari · S. C. Gupta (✉)
Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India
e-mail: scgupta@amity.edu

T. Choudhury
Department of Informatics, School of CS, University of Petroleum and Energy Studies (UPES), Dehradun, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_41

bring structure to the immensely unstructured data. Data points can later be allotted to different fields in the newly formed structured database creating room for machine learning-based smart recruitment programs and exhaustive data analytics.

The objective is to develop a program capable of extracting data from multiple unstructured résumés reconstructing it into structured CSV format and also present the extracted information to potential employers.

2 Pre Processing

Pre-processing of data is a noncritical yet crucial step in the process of Data Mining. It essentially is a technique used to reconstruct crude data into a convenient and efficient format. Customary crude data is often largely erroneous, malicious, conflicting, and even inadequate. Data pre-processing is predominantly used to eradicate such issues. Following are the major steps involved in the implementation of data pre-processing:

Cleaning: It involves conjugating the consistent parts of data by filling gaps, eliminating noise and inconsistency.
Integration: Clustering [3] of similar data and elimination of conflicts takes place in this step.
Transformation: It involves theorization, assembly, and distribution of data.
Reduction: Compression and contraction of data and presentation of the data model in the data warehouse.

3 Literature Review

There has been a lot of job searching nowadays. The online market has flourished by providing various platforms for our youth in order to find the right designation for their career. Websites like LinkedIn, Recruvia.com, Monster.com, and many more have provided a much easier platform to search for online jobs as per one's need. The journey from applying for a job to land in one is a hectic and competitive task in itself. The process is also difficult for the recruiters since they have to select the right talent for their organization. Curriculum Vitae plays an important role in this recruiting process. It is a quick image of your experience for them, and hence, many companies have a number of members to manually go through the resumes of the applicants and extract information from them. This extraction of information could lead to a serious amount of resources and time. This manual process needs an automated process in order to get the right information and talent with less time complexity. Therefore, we have used Machine Learning techniques [1, 7, 8] and Natural Language Processing (NLP) [2] in order to extract vital information such as name, contact, qualifications, and years of experience from the resumes [4–6].

In machine learning [9, 10], data is divided into three types: Structured Data, which is in form of tables, generally used in databases and CSV's. Unstructured Data comprises all the random data in or day to day lives, such as WhatsApp and

Twitter. Semi-Structured Data: The data which is presented in the form of structured data, but does not fully follow the rules of structured data. Resumes are the type of semi-structured data and here in our project, we will use machine learning techniques [11] to work on semi-structured data. Natural Language Processing is the field of computer science and Artificial Intelligence [13] where we deal with human language and computers. The basic aim of this project is to make our machine [12] understand the grammar of resume and be able to extract desired information from the resumes.

The project works in the Extract Transform Load (ETL) tool where extraction of data from the resumes, transform it into the desired format and load it in either a new database or existing database. The project is divided into three phases. The first phase is dividing our text into tokens and using them to make the grammar for our machine. The second phase is deriving grammar using regular expressions. Regular expressions or regex are the search patterns that the machine understands. They are usually in string format. The third phase is loading the data into the desired format.

Previous works have been done before this project that included deriving grammar for qualification and personal details. By this project, we have tried to enhance the automation and included the extraction of the applicant's experience. Future use of this project could bring an all automation procedure in the field of job hiring and talent hunt. This process could save resources for a particular organization which is the goal of every company.

3.1 Methods and Materials

Python

Python is the fastest growing and worlds most popular coding language. People from different disciplines use python for data analysis and visualization, artificial intelligence and machine learning, and automation. Any repetitive task like copying files and uploading to servers, working with excel spreadsheets, pdfs, and parsing websites can be automated all by a Python script.

Python is fast and easy to code. The big task can be performed with comparatively fewer lines of code. It is a high-level language, so programmers need not worry about memory management. It is a cross-platform. It has a huge community. We used python for scrapping websites and resumes in formats like PDF and DOC. Python provides really amazing libraries and modules for this task.

NLP

NLP or Natural Language Processing gives the computer the ability to understand natural human language. It is a combination of computer science and linguistics. When a person writes or speaks, he does not follow one syntax like the programming languages do, a person can use a wide range of vocabulary for the same word and different formation of sentences that mean the same. Only 15% of the data that is being generated is in the structured form. So, it is very challenging to come up with one particular method to understand and parse human language.

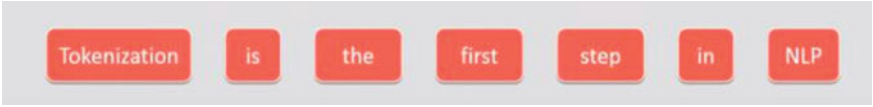


Fig. 1 Sentence divided as a token

So in NLP we try to break sentences into parts of speech and then into noun and verb phrases. So a parse tree is made making relations between different entities and the computer tries to understand that. NLP is used for text mining to derives meaning full insights from the huge chunk of unstructured text data to perform an actionable task, making voice assistants like Siri and Ok Google, for building chatbots, sentimental analysis, and language translation.

Python provides a module called NLTK for NLP. We used the NLTK module in python for scrapping resumes to extract information as resume format varies from person to person.

TOKENIZATION

In tokenization, strings are broken down into tokens, every token is a small structural unit that can be used for the process of tokenization. Taking an example (Fig. 1).

"Tokenization is the first step in NLP" as seen is divided into seven tokens and is very useful in Natural Language Path. It is a process of re-figuring out the data into a smaller and compactable format that has all the attributes of regular/original data without compromising its security. Tokenization in Natural Language Processing is chopping sentences into the most possible meaningful chunks after removing all the non-needed characters such as punctuations. Tokenization is important in Natural Language Processing as it tends to give the basic knowledge of the information of the unstructured, semi-structured data [14, 15].

STEMMING

Generally, stemming refers to normalizing the words in their base form or to their root form. There are various forms of words with the common base word. Stemming tends to extract the base word from various forms of the word and helps us identify the base word from the unstructured, semi-structured data (Fig. 2).

All of the words in the figure above have a single base word "Affect". The algorithm (Stemming) cuts the beginning or the end of the word and takes into account the list of common suffixes and prefixes found in an infected word. This can be successful in some cases but not always. There was some form of words that the

Fig. 2 Extraction of the root word

Stemming algorithm was unable to catch. Those exceptions are fixed by the upcoming processes.

LEMMATIZATION

It takes the architectural analysis of the word into consideration. In order to do so, it is essential to have a well-articulated dictionary that the algorithm could look through to link the form back to its initial word or the root word, also known as the lemma.

Essentially what lemmatization does is it groups together differently inflected forms of a word called lemma which is also similar to stemming since it maps multiple words into a single common root. The main contrast between Lemmatization and Stemming is that the output of lemmatization is the proper word. For instance, a Lemmatizing algorithm would map 'going', 'went', and 'gone' into 'go', while that is not what the output would be for Stemming. Once the tokens are available, and the tokens are divided into their root form; the POS tags are the next step.

POS TAGS

The grammatical type of a word is generally referred to as a part of speech (POS) tag; it may be a verb, adjective, noun, article, adverb, and many more, and essentially it is an indication of the functionality of meaning and grammar in a sentence. A word may have a single or even multiple POS based on the context in which it is used (Fig. 3).

Taking a sentence for instance: "Google something on the internet". In this sentence, "Google" is used as a verb even though it is a proper noun. This is an example of the certain limitations while processing the Natural Language. There are various forms of words that are used differently in day-to-day uses. This creates a major confusion to the algorithm while putting tokens into consideration for grammar.

CHUNKING

Chunking is the process of picking up the individual pieces of information and grouping them into a bigger piece. In the context of Natural Language Processing, Chunking means the grouping of words and tokens into chunks, so as we can see here, we have "pink" as an adjective, "panther" as a noun, and "the" as a determiner

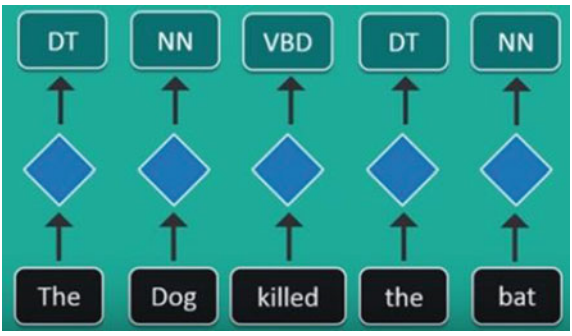


Fig. 3 Part of speech tag division

Fig. 4 Chunking of a sentence

and all of these are together chunked as noun phrase. Now, this helps in getting the insights and meaningful information from the given text (Fig. 4).

Chunking is grouping data into more understandable chunks that could have various possibilities in meaning. It is the same as sets and subsets. Sets have various subsets with subsets having subsets of themselves. It refers to a more detailed explanation as we go deeper. For example, if we say, "I have a car", it is a bigger chunk with underlying details as to what is the brand of the car? Details about the engine, what is the mileage, and many more details.

3.2 Proposed Methodology

Tokenization

It the first stage for any NLP task. It is a part of pre-processing the data. Here we break down the whole text document into tokens. This process can be used to split the document into a list of words, lines, or sentences. To make the search process more efficient, we split the entire document into words, lines, and sentences. Then we tagged them according to the parts of speech tagging (Figs. 5 and 6).

Parsing

```
lines = [el.strip() for el in document.split("\n") if len(el) > 0]
lines = [nltk.word_tokenize(el) for el in lines]
lines = [nltk.pos_tag(el) for el in lines]
```

Fig. 5 Use of .strip() function

```
sentences = nltk.sent_tokenize(document)
sentences = [nltk.word_tokenize(sent) for sent in sentences]
tokens = sentences
sentences = [nltk.pos_tag(sent) for sent in sentences]
dummy = []
for el in tokens:
     dummy += el
tokens = dummy
```

Fig. 6 NLTK tokenizer function

Here we do the syntactic analysis of the data. This defines the structure of the data. We make a parse tree or the whole data and form phrases like noun phrases verb phrases and then form patterns for required data that has to be extracted. We have labeled the data with pos tags and now we form a pattern using regular expressions to target the specific details that are needed. So we parse the whole data to search for the set of data that matches the pattern. For example, for extracting a person's phone number we look for ten digits coming consecutively, and for the name, we look for two or three proper nouns coming together. So like this, we analyze what is the common pattern that particular data follows, recognize that pattern and then match that with data. Some examples of regular expression patterns are used.

Pattern for name

See (Fig. 7).

Pattern for number

See (Fig. 8).

Pattern for educational institute

See (Fig. 9).

Cleaning.

Removing the outliers in the extracted data. It is possible that some unwanted data can also be extracted while parsing the data. So it is important to remove the data that is not intentionally targeted. So, the extracted are further analyzed. More refining patterns are made to eliminate the unrequired data. Finally, we have the most accurate match of the pattern of data that is targeted.

```
grammar = r'NAME: {<NN.*><NN.*><NN.*>*}'
```

Fig. 7 Pattern for name

```
Pattern used for phone numbers:
pattern = re.compile(r'([+(]?\d+[)\-]?[ \t\r\f\v]*[(]?\d{2,}[()\-]?
[ \t\r\f\v]*\d{2,}[()\-]?[ \t\r\f\v]*\d*[ \t\r\f\v]*\d*[ \t\r\f
\v]*)')
# Any phone number can start with +91- or (91) -> [+(]? \d+ [)\-]?
# \s* is any whitespace character - careful, use [ \t\r\f\v]*
instead since newlines are trouble
# General number formats after extension are 123 456 7890 or 12345
67890 or 1234567890 or 123-456-7890, hence 3 or more digits
```

Fig. 8 Pattern for educational institute

```
instiregex=r'INSTI: {<DT.>?<NNP.*>+<IN.*>?<NNP.*>?}'
```

Fig. 9 Pattern and value for instiregex

4 Outcome and Results

{"basics": {

"name": "Manas Kashyap", "label": "Junior Chef", "picture": "",
"email": "kashyapcook@gmail.com",
"phone": "(+91) 9156249492",
"website": "cookexperts.com", "summary": "Professional Chef",
"location": {
"address": "Vikrant Road, Chatiya Gaun, Bilaspur",
"postal Code": "220020",
"city": "Dehradun",
"country": "India",
},
"profiles": [{

"Twitter": "http://www.twitter.com/kashyap26", "LinkedIn": "http://linkedin.com/in/kashyap26"
}]

},
"work": [{

"company": "Hell's Kitchen",
"position": "Junior assistant chef",
"start Date": "11-05-2016",
"end Date": "07-04-2017",
"summary": "My job at Hell's Kitchen was to help the senior chef",

},
{"company": " Elen Paradise",

"position": "Intern",
"start Date": "01-05-2017",
"end Date": "01-07-2017",
"summary": "Intern under the head chef",
},
],
"volunteer": [{"organization": "Mirchiram",
"position": "mentor", "start Date": "05-10-2016",
"end Date": "05-10-2016",

"summary": "taught basics of cooking to students",
}],

5 Conclusion

In conclusion, the data extraction process was successful. Résumés of multiple formats were converted into standard text. NLP techniques were applied to the text to extract name, age, gender, education, email, experience, skills, phone number, and date of birth fields which were later stored systematically into a CSV file. The .csv file can be downloaded after the extraction process.

6 Future Work

The future is full of possibilities. Keyword scraping can be used to filter job profiles and suitable candidates; machine-learning algorithms can be used to effectively find the best possible employer-employee matches. Fields extracted from the NLP process are potential data point candidates for data analytics as well.

References

1. Machine learning methods for solving complex ranking and sorting issues in human resourcing. Inst Electr Electron Eng (IEEE)
2. Intelligent hiring with resume parser and ranking using natural language processing and machine learning. Int J Innov Res Comput Commun Eng (IJIRCCE)
3. Sonar S (2012) Resume parsing with named entity clustering algorithm. IEEE Res. http://www.slideshare.net/swapnilmsonar/resume-parsing-with-named-entity-clustering-algorithm
4. Sovren Resume/CV Parser. http://www.sovren.com
5. Connectifier. http://www.connectifier.com
6. Rchillies. http://www.rchillies.com
7. Sengupta E, Garg D, Choudhury T, Aggarwal A (2018) Techniques to elimenate human bias in machine learning. In: Proceedings of the 2018 international conference on system modeling and advancement in research trends, SMART 2018. https://doi.org/10.1109/SYSMART.2018.8746946
8. Verma A, Shukla A, Choudhury T, Chauhan A (2020) Recommendation system for prediction of tumour in cells using machine learning approach. Commun Comput Inf Sci (CCIS) 1229. https://doi.org/10.1007/978-981-15-5827-6_18
9. Choudhury T, Tripathi A, Arora B, Aggarwal A (2020) Implementation of common spatial pattern algorithm using EEG in BCILAB. Commun Comput Inf Sci (CCIS) 1229. https://doi.org/10.1007/978-981-15-5827-6_25
10. Choudhury T, Kaur A, Verma US (2017) Agricultural aid to seed cultivation: an agribot. In: Proceeding—IEEE international conference on computing, communication and automation, ICCCA 2016. https://doi.org/10.1109/CCAA.2016.7813860
11. Sharma J, Kaur A, Choudhury T (2017) Method for calculating relative performance of microprocessors for comparison under ideal state. In: 2016 international conference on information technology, InCITe 2016—the next generation IT summit on the theme—internet of things: connect your worlds. https://doi.org/10.1109/INCITE.2016.7857592
12. Kumar V, Choudhury T, Sabitha S (2017) Physical telepresence: growth trends of tangible user interface and its future. In: Proceedings on 2016 2nd international conference on next generation computing technologies, NGCT 2016. https://doi.org/10.1109/NGCT.2016.7877420

13. Soni N, Mohta M, Choudhury T (2017) The looming visible light communication Li-Fi: an edge over Wi-Fi. In: Proceedings of the 5th international conference on system modeling and advancement in research trends, SMART 2016. https://doi.org/10.1109/SYSMART.2016.7894519
14. Joshi T, Badoni P, Choudhury T, Aggarwal A (2019) Modification of Weiler-atherton algorithm to address loose polygons. J Sci Ind Res 78:771–774
15. Kumar V, Choudhury T (2019) Real-time recognition of malignant skin lesions using ensemble modeling. J Sci Ind Res (JSIR) 78:148–153

A Comprehensive Study on Transparent Antennas for Wireless Applications

Ajay Kumar Gangwar

Abstract There is a great demand to design a thin, low-cost, and highly transparent antenna for wireless applications although conventional antennas have not fulfilled this demand. For this purpose, there are several transparent antennas that have been developed, which are based on indium thin oxide, glass and polydimethylsiloxane substrate, acrylic material, elastomer and PDMS material, etc. The main goal of the review paper is to identify thin high transparent antenna design material.

Keywords Antenna · Transparent · Material

1 Introduction

In wireless communication, the antenna transmits and receives electromagnetic waves. The antenna is a device that converts the electrical signal to the electromagnetic wave and vice versa [1]. For the planar surface, a microstrip antenna is a suitable candidate because it can be easy to design, lightweight, low cost, and easy to design for desire applications [2, 3]. The conventional microstrip antennas are non-transparent, but there is a huge demand for low-cost, high transparent antennas for different applications such as mobile phones, smartwatches, and solar panels [4–7].

To date several transparent antennas have been designed using different techniques as described in literature survey Sect. 2. Section 3 covered performance comparison and finally, Sect. 4 covered the conclusion.

A. K. Gangwar (✉)
HMR Institute of Technology and Management, Hamidpur, New Delhi, India

Guru Govind Singh Indraprastha University, New Delhi, India

Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh, India

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

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_42

2 Literature Survey

In this section, papers related to transparent antennas are discussed. In the year 2005, Tai-Lee Chen et al. was designed a comb-shaped dipole antenna on a transparent substrate for wireless communications [4]. The antenna consists of two conducting strips with two comb structures as shown in Fig. 1a. It is designed on transparent Mylar, and the frequency range of the dipole antenna is much broader than an ordinary dipole antenna. Further, in the next year, a transparent monopole antenna was designed by Chen-Tin Lee et al. for WCDMA and WLAN applications [5]. It is designed on polyethylene terephthalate plastic with an indium tin oxide coating layer. The antenna consists square patch with two notches as shown in Fig. 1b. It is resonated at 1.95 GHz and 5.58 GHz with the bandwidth covered from 1.77 GHz to 2.19 GHz and 4.81 GHz to 6.0 GHz, respectively.

In the year 2009, A. Katsounaros et al. designed a transparent antenna for ultra-wideband applications [6]. The antenna was designed on a 2.0 mm thick A4-sized Perspex substrate with AgHT-4 film. The antenna consists of a combined structure of coplanar fed circular patch antenna as shown in Fig. 2a. It covered 1.0–8.5 GHz frequency band and exhibits bi-directional and omnidirectional radiation patterns on the E and H-plane, respectively. Similarly, in the year 2011, a UWB transparent antenna was designed by T. Peter et al. for green wireless applications [7]. The

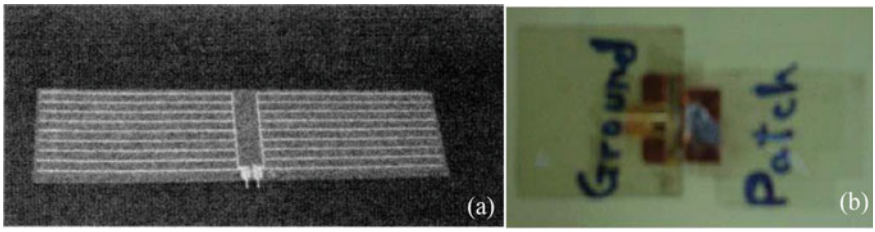


Fig. 1 **a** A comb shape transparent dipole antenna. **b** A monopole transparent antenna

Fig. 2 **a** A ultra-wideband transparent disc antenna. **b** A compact size ultra-wideband transparent antenna

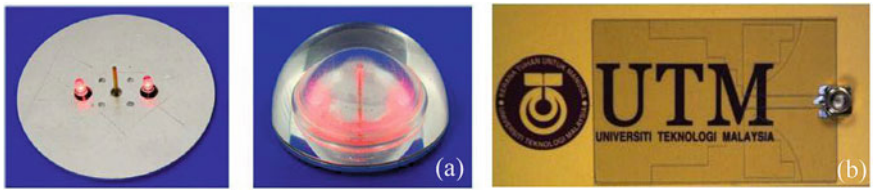


Fig. 3 **a** A hemispherical transparent dielectric resonator antenna. **b** A coplanar fed transparent antenna for ultra-wideband applications

antenna is designed on AgHT-8 conducting layer with polyethylene terephthalate substrate. The antenna is designed by coplanar-fed rectangular patch as shown in Fig. 2b. The antenna covered frequency band from 3.1 to 10.6 GHz with the gain of –6.0 dBi to 4.5 dBi with the UWB. The antenna provides omnidirectional radiation pattern which is suitable for receiving end applications.

In the year 2014, a transparent dielectric resonator antenna was designed by Xiao Sheng Fang et al. for wideband applications [8]. The two-layer hemispherical DRA was designed by using acrylic and glass material in the inner and outer layer, respectively, as shown in Fig. 3a. The antenna covered 3.4–3.7 GHz band for WiMAX applications. Further, in the same year, a CPW-fed transparent antenna was designed by S. Hakimi et al. for ultra-wideband applications [9]. It is designed on 0.175 mm thick silver-coated polyester film with a 3.24 dielectric constant. The antenna is designed by four symmetrical rectangular stubs on the top plane with a quarter circular slot on the ground plane as shown in Fig. 3b. The antenna covered frequency range from 3.15 to 32 GHz with VSWR less than 2.

In the year 2014, a transparent mesh circular monopole antenna was designed by Okan Yurduseven et al. for UWB applications [10]. It is designed by a circular monopole patch on a transparent acrylic substrate placed on Poly-Si solar cell as shown in Fig. 4a. The antenna covered the frequency band from 2.33 to 10.8 GHz with 4.1 dBi gain and 91% transparency. And in the next year, Colan G. M. Ryan et al. was designed and developed a transparent circular polarized metamaterial loaded antenna [11]. It is designed by a wire mesh patch and the same grid cut from the substrate as shown in Fig. 4b. In the antenna for controlling the phase difference, a metamaterial is loaded into the patch of the orthogonal mods. The dual-band circular

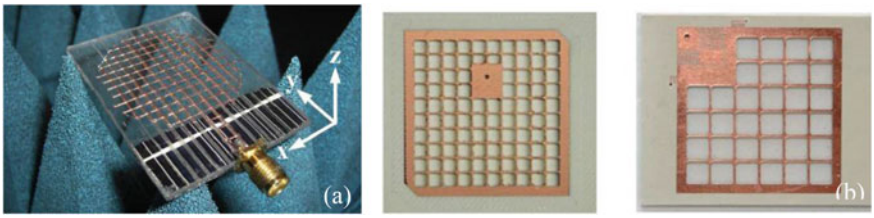


Fig. 4 **a** A transparent mesh circular patch monopole antenna. **b** Metamaterial loaded circular polarized transparent antenna

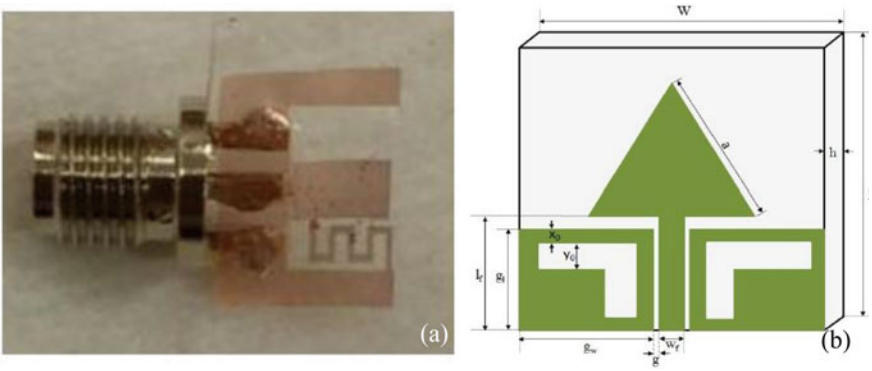


Fig. 5 **a** A transparent mechanical reconfigurable antenna. **b** A coplanar fed circular polarized triangular patch transparent antenna

polarization antenna resonates at 2.35 and 2.73 GHz with the efficiency of 70% and 78%, respectively.

In the year 2015, a transparent mechanical reconfigurable antenna was designed by Jang et al. [12]. The antenna is designed on elastomer and PDMS material, which consists meander line with CPW feed, and re-configurability is achieved by a vertical oriented meandering line as shown in Fig. 5a. The resonance frequency of the antenna is mechanically tuned from 2.94 to 2.46 GHz. And in the next year, a transparent triangular shape antenna was designed by J. P. Shinde et al. for S-band applications [13]. It is designed by indium tin oxide film with a soda line glass substrate. The antenna is designed by CPW fed triangular shape patch as shown in Fig. 5b. The antenna is resonated at 2.5 GHz with the bandwidth covered from 2.12 to 4.31 GHz.

In the year 2016, a flexible transparent antenna was designed by H. A. Elmobarak et al. for UWB applications [14]. It is designed by transparent conductive tissues on the polydimethylsiloxane substrate as shown in Fig. 6a. The UWB antenna was designed by a modified circular disk patch that covered the frequency range from 2.2 to 25 GHz frequency band. The antenna achieved more than 75% efficiency through the frequency band and in the next year, Chi-Yuk Chiu et.al. designed a

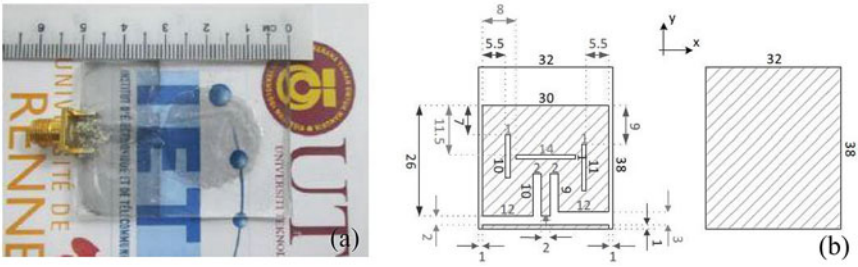


Fig. 6 **a** A flexible transparent ultra-wideband antenna. **b** A transparent dual-band antenna

transparent dual-band antenna for smartwatch applications [15]. The antenna was designed by display glass utilized as a substrate and indium thin oxide (ITO) utilized as a conducting layer as shown in Fig. 6b. The antenna operates at dual-frequency bands 2.4 and 5.0 GHz. The antenna radiates a directional radiation pattern, which minimized interference with the other components.

In the year 2017, Arpar desai et al. designed a dual-band transparent antenna for wireless applications [16]. It is designed on the silver layer with tin oxide layers. The transparent antenna was designed by slot to interconnect ring resonator as shown in Fig. 7a, which is resonated at 2.4 GHz and 3.7 GHz with the gain of 2.26 dBi and 2.11 dBi, respectively for WLAN and WiMAX applications and in the next year, a metal mesh-based transparent antenna was designed by Seok Hyon Kang et al. for smart window, transparent mobile devices [17]. It is designed by copper wire metal mesh with micro metal mesh film use as a transparent patch as shown in Fig. 7b. The antenna operates from 2.4 GHz to 2.5 GHz WLAN frequency band with efficiency and transparency lie from 49.09 to 56.88% and 69.44 to 82.64%, respectively.

In the year 2018, a transparent double folded loop antenna was designed by Yohei Koga et al. for IOT applications [18]. The antenna is designed by 2.0 mm thick transparent polycarbonate with 0.025 mm Indium Tin Oxide transparent conductor as shown in Fig. 8a. The antenna is designed by a transparent conductor as well as a copper conductor and compares the performance. The antenna was designed by a double folded loop for a 2.4 GHz resonance frequency. It achieved –6.5dBi and 4 dBi gain at transparent and copper conductor surfaces. Further, in the next year, a transparent flexible wideband antenna was designed by Weiwei Li et al. for 5G band applications [19]. It was designed by an inkjet printer using silver nanowires. The antenna is designed by a triangular monopole antenna with microstrip line feeding. The transparency of the antenna is increased by introducing slots in the patch and ground plane as shown in Fig. 8b. The antenna achieved 26 GHz (18.0–44.0 GHz) bandwidth with 55% efficiency and 1.45 dBi gain.

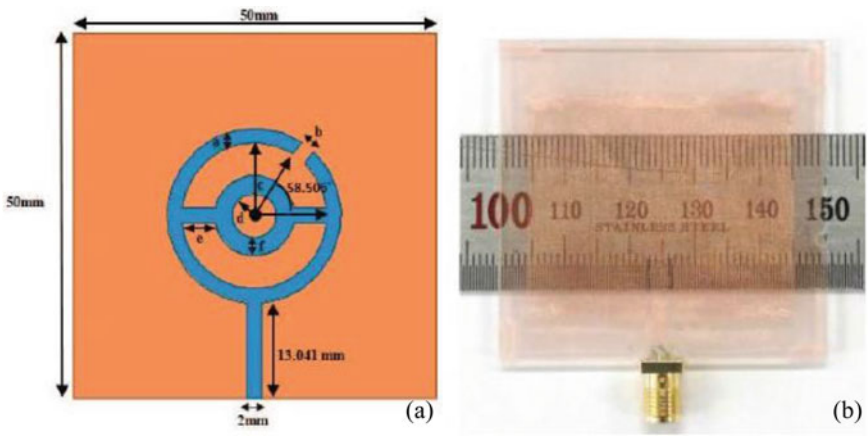


Fig. 7 **a** A transparent dual-band resonator antenna. **b** A metal mesh-based transparent antenna

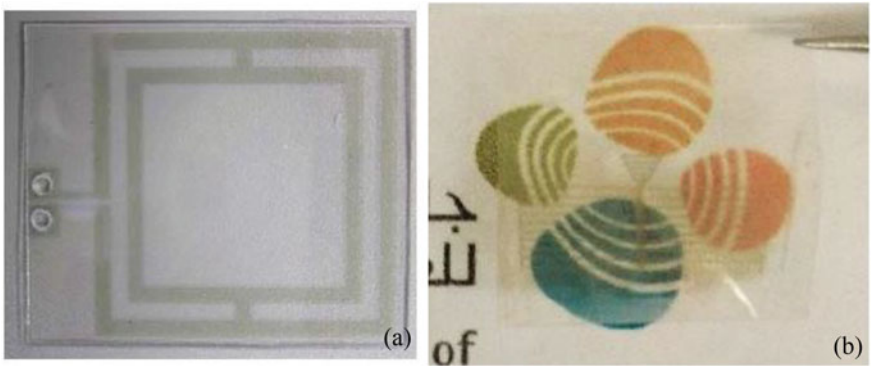


Fig. 8 **a** A transparent double folded loop antenna. **b** A transparent flexible nanowire-based antenna

In the year 2019, a transparent antenna was designed by Jie Sun et al. for wideband applications [20]. It is designed with distilled water with a transparent Plexiglas container. In the antenna, the water patch is excited by a metallic L-shaped feed. The antenna achieved an efficiency of more than 75% with a maximum gain of 6.6 dBi from the frequency of 1.61 GHz to 2.29 GHz and in the same year, a transparent antenna was designed by Rong Zhang et al. for transparent screen and window applications [21]. It is designed by graphene films with transparent glass. The transparent patch antenna operates from the frequency of 5.0 GHz to 20 GHz. The antenna provides light transmission of more than 72%.

In the year 2019, a transparent small radar cross section antenna with absorbing metasurface was designed by Jia-Qi Zhou et al. for X-band applications [22]. It is designed by tin oxide conductive film with transparent glass. The antenna was designed co-planar fed rectangular patch and as an absorber, a square resistance is arranged around the patch. The antenna resonating at 8.88 GHz, and it reduced radar cross section minimum 10.0 dB from 8.0 to 12.0 GHz. Further, in the next year, a transparent millimeter frequency-based antenna was designed by Roc´ıo Rodr´ıguez-Cano et al. for the antenna on display applications [23]. It is designed by a glass display with diamond grid cells. It is designed by two meshed bow-tie-shaped monopoles antennas with a row of mesh patches. The two antennae provide more than 7.0 dBi gain and a row of patches reduce the coupling between them. The antenna covered a band of frequencies from 26.5 to 29.5 GHz with efficiency achieved more than 70%.

3 Summary

In this review article, different transparent antennas have been studies. These antennas are designed for single band, dual-band, wideband, ultra-wideband, and reconfigurable applications. The basic design principle of these transparent antennas is the

same, but only the design material materials have differed. In the text design techniques, design materials, and applications are summarized in Table 1. This article will provide an overview of the researcher that is working on the transparent antenna.

Table 1 Summarized table of transparent antenna

References	Design principle of transparent antenna	Operating frequency	Types of antenna
Chen and Hsieh [4]	Transparent Mylar	557 MHz	Single band
Lee et al. [5]	Polyethylene terephthalate plastic with Indium tin oxide coating layer	1.95 and 5.58 GHz	Dual band
Katsounaro et al. [6]	Perspex substrate with AgHT-4 film	1.0–8.5 GHz	Ultra wide band
Peter et al. [7]	AgHT-8 conducting layer with polyethylene terephthalate substrate	3.1–10.6 GHz	Ultra-wide band
Fang and Leung [8]	Acrylic and glass material	2.87–3.96 GHz	Wide band
Hakimi et al. [9]	Silver coated polyester film	3.15–32.0 GHz	Ultra-wide band
Yurduseven et al. [10]	Transparent acrylic substrate placed on Poly-Si solar cell	2.33–10.8 GHz	Ultra-wide band
Ryan and Eleftheriades [11]	Wire mesh patch and the same grid cut from the substrate	2.35 and 2.73 GHz	Dual band
Jang et al. [12]	Elastomer and PDMS material	2.94–2.46 GHz	Mechanical frequency tunable
Shinde et al. [13]	Indium tin oxide film with soda line glass substrate	2.19 GHz	Single band
Elmobarak et al. [14]	Conductive tissues on the polydimethylsiloxane substrate	2.2–25 GHz	Ultra-wide band
Chiu et al. [15]	Glass utilized as a substrate and indium thin oxide	2.4 and 5.0 GHz	Dual band

(continued)

Table 1 (continued)

References	Design principle of transparent antenna	Operating frequency	Types of antenna
Desai et al. [16]	Silver layer with tin oxide layers	2.42 and 3.71 GHz	Dual band
Kang and Jung [17]	Copper wire metal mesh with micro metal mesh film	2.45 GHz	Single band
Koga and Kai [18]	Transparent polycarbonate with indium tin oxide conductor	2.4 GHz	Single band
Li et al. [19]	Silver nanowires	18.0–44.0 GHz	Wide band
Sun and Luk [20]	Distilled water with transparent Plexiglas container	1.61–2.29 GHz	Wide band
Zhang et al. [21]	Grapheme films with transparent glass	5.0–20 GHz	Wide band
Zhou et al. [22]	Tin oxide conductive film with transparent glass	8.0–12.0 GHz	Wide band
Rodr´ıguez-Cano et al. [23]	Glass display with diamond grid cells	26.5–29.5 GHz	Wide band

4 Conclusion

In this survey paper, various transparent antennas for several wireless applications are studied. The transparent antenna and its applications are suitable for satellite, solar panel, smartwatches, etc. In the study of several papers, it is observed that the author tried to design high transparency with acceptable antenna parameters. This article provides an overview of the transparent antennas and their design materials, which will provide helpful for new researchers/scientists.

References

1. Balanis CA (1997) Antenna theory, analysis and design. John Wiley and Sons, Inc.
2. Gangwar AK, Alam MS (2020) Frequency reconfigurable dual-band filtenna. Int J Electron Commun (AEÜ) 1–12. Elsevier
3. Gangwar AK, Alam MS (2019) A miniaturized quad-band antenna with slotted patch for WiMAX/ WLAN/GSM applications. Int J Electron Commun (AEÜ) 1–9. Elsevier
4. Chen T-L, Hsieh C-W (2005) Comb-shaped dipole antenna on transparent substrate. In: IEEE Antennas propagation society international symposium, pp 610–612. IEEE, Washington, DC, USA

5. Lee C-T, Lee C-M, Luo C-H (2006) The transparent monopole antenna for WCDMA and WLAN. In: IEEE annual wireless and microwave technology conference. IEEE, Clearwater Beach, FL, USA
6. Katsounaros A, Hao Y, Collings N, Crossland WA (2009) Optically transparent ultra-wideband antenna. Electron Lett 722–723. IET
7. Peter T, Sun YY, Yuk TI, AbuTarboush HF, Nilavalan R, Cheung SW (2011) Miniature transparent UWB antenna with tunable notch for green wireless applications, In: International workshop on antenna technology (iWAT). IEEE, Hong Kong, China
8. Fang XS, Leung KW (2014) Design of wideband omnidirectional two-layer transparent hemispherical dielectric resonator antenna. IEEE Trans Antennas Propag 5353–5357. IEEE
9. Hakimi S, Rahim SKA, Abedian M, Noghabaei SM, Khalily M (2014) CPW-fed transparent antenna for extended ultra-wideband applications. IEEE Antennas Wirel Propag Lett 1251–1254. IEEE
10. Yurduseven O, Smith D, Elsdon M (2014) A transparent meshed solar monopole antenna for UWB applications. In: The 8th European conference on antennas and propagation, pp 2145–2149. IEEE
11. Ryan CGM, Eleftheriades GV (2015) Single- and dual-band transparent circularly polarized patch antennaswith metamaterial loading. IEEE Antennas Wirel Propag Lett 470–473. IEEE
12. Jang T, Zhang C, Youn H, Zhou J, Guo LJ (2015) Transparent and mechanically reconfigurable small antenna based on stretchable micromesh. In: IEEE International symposium on antennas and propagation & USNC/URSI national radio science meeting, pp 2403–2404. IEEE
13. Shinde JP, Shinde PN, Gondane NY (2016) Circularly polarized transparent equilateral triangular shaped antenna with defected ground. In: International conference on computing communication control and automation (ICCUBEA). IEEE, Pune, India
14. Elmobarak HA, Rahim SKA, Himdi M, Castel X, Abedian M (2016) A transparent and flexible polymer-fabric tissue UWB antenna for future wireless networks. IEEE Antennas Wirel Propag Lett 1333–1336. IEEE
15. Chiu C-Y, Shen S, Murch RD (2017) Transparent dual-band antenna for smart watch applications. In: 2017 IEEE International symposium on antennas and propagation & USNC/URSI national radio science meeting, pp 191–192. IEEE
16. Desai A, Upadhyaya T, Palandoken M, Patel R, Patel U (2017) Dual band optically transparent antenna for wireless applications, pp 960–963. IEEE, Kuala Lumpar, Malaysia
17. Kang SH, Jung CW (2018) Transparent patch antenna using metal mesh. IEEE Trans Antennas Propag 1–5. IEEE
18. Koga Y, Kai M (2018) A transparent double folded loop antenna for IoT applications. In: IEEE-APS topical conference on antennas and propagation in wireless communications (APWC), pp 762–765. IEEE, Cartagena des Indias, Colombia
19. Li W, Meredov A, Shamim A (2019) Silver nanowire based flexible, transparent, wideband antenna for 5G band application. In: IEEE International symposium on antennas and propagation and USNC-URSI radio science meeting, pp 275–276. IEEE, Atlanta, GA, USA
20. Sun J, Luk K-M (2019) A wideband and optically transparent water patch antenna with broadside radiation pattern. IEEE Antennas Wirel Propag Lett 341–345. IEEE
21. Zhang R, Zhao Y, Li B, Zhang Z, Wang D (2019) A transparent microstrip antenna using grapheme film. In: IEEE 4th advanced information technology, electronic and automation control conference, pp 2625–2627. IEEE, Chengdu, China
22. Zhou J-Q, Zu H-R, Wu B, Wu M-S, Chen L (2019) Design of transparent and low RCS antenna based on absorbing metasurface. In: International applied computational electromagnetics society symposium. IEEE, Nanjing, China
23. Rodr´ıguez-Cano R, Zhang S, Pedersen GF (2020) Transparent mm-wave array on a glass substrate with surface wave reduction. In: 14th European conference on antennas and propagation (EuCAP). IEEE, Copenhagen, Denmark

IoT-Based Framework for Real-Time Heart Disease Prediction Using Machine Learning Techniques

Siddhant Salvi, Ritwik Dhar, and Sunil Karamchandani

Abstract In several disciplines, especially in personalized healthcare, the Internet of Things (IoT) is gaining increasing interest from a technological point of view. However, in both cases, effective diagnosis of heart disease and consultation of a patient for 24 h by a specialist are not accessible due to many factors. In this research, we present a real-time monitoring system based on concepts of IoT and machine learning for the detection of heart diseases through a portable setup accessible even in remote areas. The system integrates various sensor nodes for heartbeat, blood pressure, temperature, and RFID reading onto a Raspberry Pi to maintain and build on a database for optimized results. In order to evaluate the efficiency of heart disease diagnosis, the proposed algorithms are tested using commonly available open-access databases. Including parameter optimization, a comparative study of statistical machine learning models has been applied with a random forest algorithm (0.868) producing best performance followed by SVM (0.852), k-NN (0.836), and decision tree (0.786). The database building and prediction are based on the cloud which further improves on the computational and time efficiency of the system.

Keywords SVM · Random forest · Decision tree · k-NN · sensor · Heart disease · IoT

1 Introduction

Rapid advances in health systems and low-cost cellular connectivity have significantly helped deal with the issue of fewer hospital facilities over the past decade. The introduction of mobile connectivity with wearable sensors has enabled the transition from clinic-centric to patient-centric healthcare facilities and is referred to in the literature as telemedicine [1]. Telemedicine has been introduced, according to recent medical surveys, to take care of patients with coronary disease, asthma,

S. Salvi (✉) · R. Dhar · S. Karamchandani
Department of Electronics & Telecommunication, Dwarkadas J. Sanghvi College of Engineering, Mumbai, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_43

hypotension, hypertension, hyperthermia, and hypothermia [2, 3]. Real-time surveillance of chronic diseases like cardiopulmonary disease and cardiac arrest in patients located outside of medical care facilities by wireless tracking systems is the most promising application [4]. Heart conditions are still one of the leading causes of human deaths; for example, around 2.8 million people die per year as a consequence of becoming overweight as obesity can lead to adverse physiological effects on cholesterol and blood pressure, eventually raising the risk of coronary heart attack, stroke, diabetes, and many types of cancer. According to the WHO, the prevalence of the cardiac disease is expected to rise to 23.3 percent globally by 2030 [5]. The management of such cardiovascular conditions needs constant and long-term care in order to manage the danger. In addition, while wearable technology has made a major contribution to the development of patient monitoring, there are issues that impact the efficiency of health monitoring systems. In applications such as neurological, biological, and motion sensing [6], current experiments of wearable sensors provide tracking. These devices were used for tracking patients' fitness metrics and body positions, as well as for tracking sports and other activities [7]. Due to the fact that these sensors are low-cost, readily available, user-friendly, sensitive, and effective, these wearable sensors are becoming exciting. The therapeutic uses of wearable sensors in respiratory, neurological, asthma, and hypertension disorders have been investigated in various studies [8]. For example, [9] has created a device for tracking congestive heart failure in patients, consisting of a ring-shaped biosensor that tracks heart data. Similarly, systems for monitoring respiratory diseases which capture sound waves by placing a microphone on the patient's neck while breathing have been designed. The system consisted of a band-pass filter to minimize noise and other signal distortions that helped achieve approximately 90% of the precision of the measurement. Study work on the diagnosis of apnea using algorithms was then extended [10]. Wearable technology is also helpful with the use of multiple sensors incorporated on a single chip to solve the problems of motion artifact control. Another example of application is the incorporation of various sensors on the same platform (tight fitting in garments) to track respiratory diseases. It is found that these systems are stronger than spirometry, but they do need advances to eliminate motion artifacts [11]. There is also a lot of medical data available that can be used to build predictive models for cardiovascular diseases due to the advancement of automated healthcare networks. Machine learning or data mining is a method of experimentation for processing big data from a number of viewpoints and encapsulating it into usable information. "Data Mining is a non-trivial retrieval of data information that is tacit, previously unknown and theoretically useful" [12]. A massive amount of knowledge about medical diagnosis, patients, etc. is now created by the healthcare industry. Data mining offers a range of tools that uncover secret data structures or comparisons. A machine learning algorithm is therefore suggested in this paper for the implementation of a prediction method for heart disease that has been tested on two datasets for open-access heart disease prediction. Another link to this article is the definition of a cardiac patient monitoring system using the Internet of Things (IoT) concept with different physiological signal sensors and Arduino microcontrollers. Sensor networks are now using the Internet of Things (IoT) technologies to

collect, store, and transfer information from one node to another. The senors take this data and work on it to trigger the necessary action and create an intelligent cloud-based monitoring, planning, and decision-making platform. IoT-developed devices, such as embedded systems, allow information to be shared between other nodes or over the Internet, and it has been projected that about 8 to 50 billion devices will be connected by 2020 [13].

2 System Architecture

It is considered that enabling the use of IoT in healthcare with interconnected medical sensors, particularly wearable or implantable, is capable of providing smart, precise, and cost-effective personalized healthcare services. The total IoT-enabled healthcare infrastructure designed is seen in Fig. 1.

The sensing layer is designed to observe both visually, mentally, and emotionally the health status of the users; hence, a series of sensors can be embedded, such as wearable sensors such as ECG and blood pressure, to capture biomedical parameters. The model proposed takes into consideration the following sensors for patient health analysis and data prediction purposes:

– The **heartbeat sensor** is a quick way to research the role of the heart, which tracks blood flow through the area of contact. The amount of blood in the region of touch varies over time, as the heart pushes blood into the blood vessels through the skin. The sensor shines through the skin patch with a light lobe (small incandescent

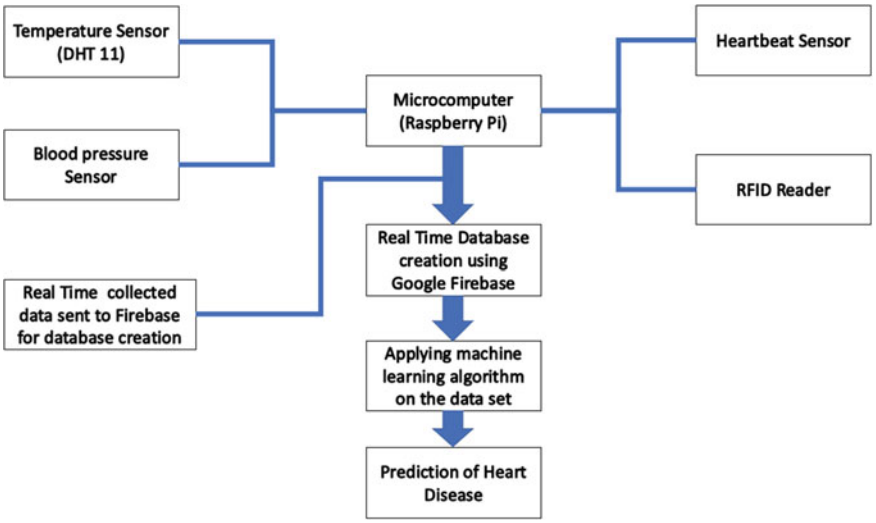


Fig. 1 Real-time heart disease monitoring system

lamp) and tests the light that is transmitted. The clip may be used on the thumb and index finger on a fingertip or on the network of tissue.
- The **temperature sensor** DHT11 is a system that contains a sensor for humidity and temperature. It will then determine the temperature and humidity of the air in a given location, whether indoors or outdoors. Humidity and temperature are the necessary criteria to track the comfort of the patient or to check the physical state of the patient in that position as the patients feel insecure due to sudden shifts in temperature and humidity that can also cause irregular cardiac responses.
- The **blood pressure sensor** is a non-invasive device developed for human blood pressure monitoring purposes. Using the oscillometric form, it tests systolic, diastolic, and mean arterial pressure. All the time, blood pressure does not remain the same. It moves to satisfy the needs of the body. Various variables, including body location, breathing or mental condition, exercise, and sleep, influence it.
- An **RFID reader** is connected to the processor board, which facilitates the identification of different users, thus a RFID tag should be amounted to each individual patient for optimum data logging and easier user-interfacing.

Secondly, the network layer facilitates the reliable and stable transmission of data to the corresponding data process units. A host of short-range communication protocols have been commonly adopted, such as ZigBee. The third layer is the module for data analysis, which gathers valuable information from the sensor data obtained from the first layer. As the processing unit for sensor data aggregation and cloud capabilities, the deployed concept uses a Raspberry Pi. For database building, the device takes the health information of the patient from the sensing layer to Google Firebase. Learning-based approaches have been the most promising technique for data mining. In order to improve machine learning model precision for heart disease prediction, the initial dataset is expanded with patient data integration. Intelligent tools and applications, such as disease identification based on the behavior of the top three tiers, can eventually be given. With improved sensor feedback and real-time monitoring and cloud access capabilities, the system can be very useful across remote locations with heart disease critical patient history.

3 Dataset

The training of our proposed system requires the use of datasets that take into consideration the attributes/symptoms generally seen in patients with cardiovascular diseases. Based on certain parameters and characteristics such as resting blood sugar, cholesterol, maximum heart rate, and exercise-induced angina as shown in Fig. 2, two publicly accessible datasets of heart diseases of the same type and number of characteristics are combined to form a broader dataset and to achieve greater precision. Cleveland Heart Disease dataset along with the Statlog Heart Disease take into consideration 9 characteristics as shown in Table 1.

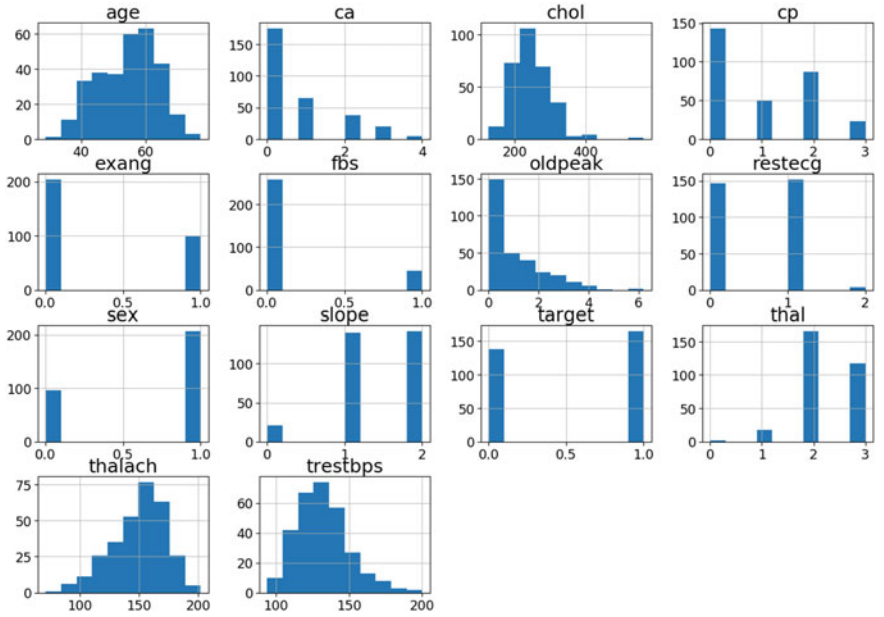


Fig. 2 Dataset histogram

Table 1 Dataset features considered for our model

Features	Description
Age	Patient age in years
Cp	Chest Pain Type (typical angina = 1, atypical angina = 2, non-anginal pain = 3, asymptomatic = 4)
Exang	Stress-induced angina
Oldpeak	Exercise-based ST depression
Restecg	Resting electrocardiographic results (normal = 0, ST-T wave abnormality = 1, left ventricular hypertrophy = 2)
sex	Female = 0, Male = 1
Slope	Slope of the peak exercise ST segment (Upsloping = 1, flat = 2, down sloping = 3)
Thal	Normal = 0, fixed defect = 2, reversible defect = 3
Thalach	Maximum heart rate

4 Machine Learning Algorithms

The performance of four algorithms has been considered here to evaluate the accuracy based on the input variations and evaluation parameters discussed above. We have implemented support vector machines (SVM), random forest algorithm (RF), decision trees, and k-Nearest Neighbor (k-NN) based on the dataset provided.

4.1 Support Vector Machines (SVM)

By constructing an n-dimensional, SVM performs classification and maximizes the margin to achieve the best result in classification. SVMs are based on the concept of hyper plane or linear separability classifiers. Supposing we have n training data points (x_1, y_1), (x_2, y_2), $\ldots.(x_n y_n)$ where $x_i \in R^m$ and $y_i \in \{-1, +1\}$. Considering a hyper plane given by (w, b), where w is weight and b is bias.

Classification can be given by

$$D(x) = sign(w.x + b) = sign(\sum_{i}^{n} a_i y_i (x_i.x) + b) \tag{1}$$

where w represents the hyper plane, and the weight vector direction gives us the class expected. The data points that are similar to the hyper plane, which are called the support vectors, have a minimum distance to the decision boundary as shown in Fig. 3.

SVM has limitations that it experiences lots of computational expenses and produces inconsistent results when dataset is characterized by wide space of features and train dataset is small. This can be solved by adding a kernel function in the feature space, instead of the inner product of two transformed data vectors. A function of

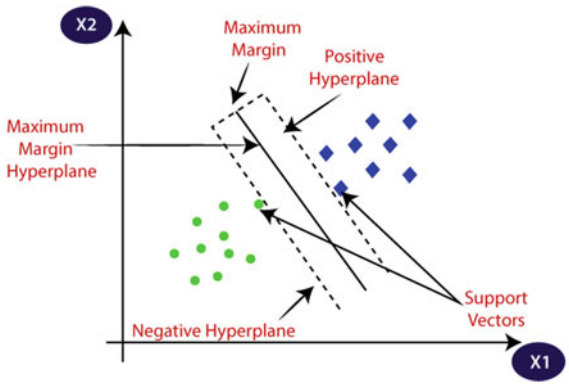


Fig. 3 Decision boundary with hyperplanes

the kernel is set to that corresponds to a dot product of two characteristic vectors in some extended space. In such processes, there are some widely used kernel functions which have been implemented for comparison of accuracy:

– Linear Kernel function:

$$K(x_i, x_j) = x_i.x_j \tag{2}$$

– Sigmoid Kernel function:

$$K(x_i, x_j) = tanh(ax_i.x_j + b) \tag{3}$$

– Polynomial Kernel function:

$$K(x_i, x_j) = \left(x_i.x_j + 1\right)^p \tag{4}$$

– RBF Kernel function:

$$K(x_i, x_j) = exp[-\gamma||x_i - x_j||^2] \tag{5}$$

4.2 Random Forest Algorithm (RF)

Random Forest is an ensemble of learning algorithms based on methods. RF consists of a series of classifiers for tree. Every tree is composed of nodes and edges. The received group classifies new data points through a majority within each classification model's predictions, as shown in Fig. 4.

This approach incorporates a bagging cycle (bootstrap aggregation) and a set of random splits. Each tree is extracted from the dataset from a separate bootstrap

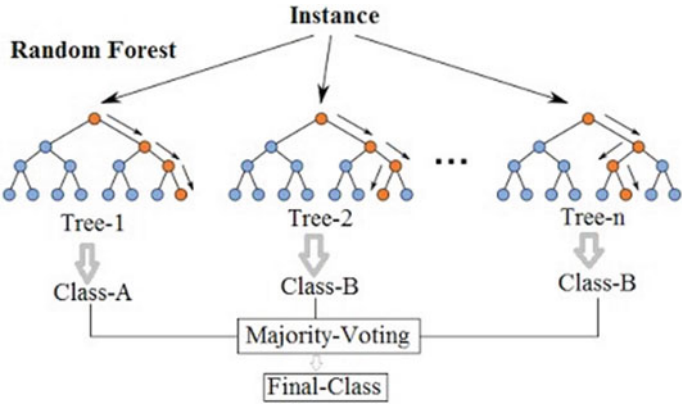


Fig. 4 Random forest classifier

sample, and each tree categorizes the data. The final outcome is a majority vote between the trees. The random forest algorithm is defined by the following steps:

- Construct samples of the data from k trees bootstrap.
- For each of the bootstrap samples, grow an unpruned tree.
- Randomly sample n-try of the predictors at each node and pick the best split among those factors.
- Predict new data through a combination of the k tree predictions.

4.3 K-Nearest Neighbor

The k-nearest neighbor algorithm (k-NN) is a method for the classification of an object among its k-nearest neighbors based on the majority class. k-NN is a slow learning algorithm where the function is approximated locally and all calculations are postponed until classification. The Euclidean distance is commonly used by the k-NN algorithm. However, it is also possible to use some other points, such as the Chebyshev norm or the Mahalanobis distance. Euclidean distance is being used in this experiment. Suppose the question example of coordinates (a, b) and the training sample coordinate is (c, d), then the Euclidean distance square is

$$x^2 = (c - a)^2 + (d - b)^2 \tag{6}$$

4.4 Decision Tree

The tree is a system of data that consists of nodes and edges. Sub-trees on the main tree path can be split into three, namely the root node, branch/internal nodes, and leaf nodes. For a finite number of levels, the decision tree is a basic representation of a classification strategy. The internal node and the root node are marked with an attribute name; the edges are marked with potential attribute values; and the leaf node is marked with various classes. The decision tree is one of the most common classification models because the results obtained are easy to explain and easier to understand (Fig. 5).

In order to represent the decision to make the decision linked to the decision tree, trees are used. The decision tree method is to adjust the form of the data (label) into the model of the tree, transforming the model tree into a law. The biggest advantage of using the decision tree is that it becomes possible to break down complicated decision-making mechanisms so that decision-makers perceive the problem-solving solution. The decision tree is also useful for exploring the details, discovering hidden relationships with the input variables of the target variable with certain candidates.

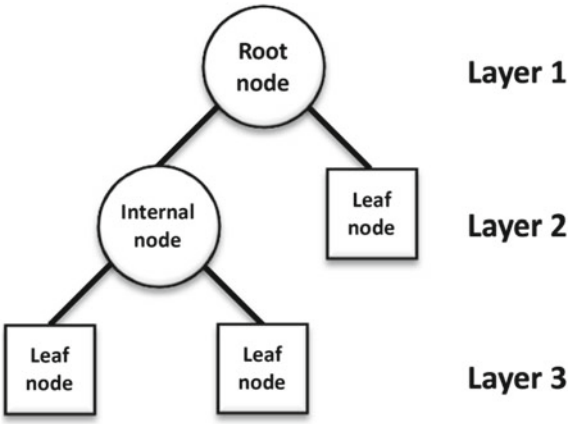


Fig. 5 Decision tree layers

5 Results

The results of the proposed methodology for selecting the right machine learning algorithm and the operation of the cardiac monitoring system are discussed in this section.

5.1 Parameter Optimization of Machine Learning Models

Various machine learning algorithms have been implemented to get the best performance as shown in the above section. To further analyze and optimize the performance before conducting a comparative analysis, the algorithm parameters are iterated across to find the best match. As shown in Fig. 6a, four kernel functions are used and trained with the dataset for the heart disease monitoring system. The performance of each kernel function, i.e. polynomial (0.78), linear (0.81), radial basis (0.85), and sigmoid (0.85), is presented from which we go forward with sigmoid kernel function for support vector machine algorithm. Next, we consider the k-Nearest Neighbor algorithm across various values of k, i.e. number of neighbors to analyze the variation in performance and choose the optimum value for comparative analysis across algorithms for the dataset. As shown in Fig. 6b, the accuracy is highest on k = 8, 12, 14 which will be implemented.

Similarly, the optimum number of features have been computed for the decision tree as shown in Fig. 7a, with 10 features. The number of estimators shows variation in the performance of random forest algorithm although minimally. The highest scores are noticed on 200 estimators with 100 estimators being a close second as shown in Fig. 7b.

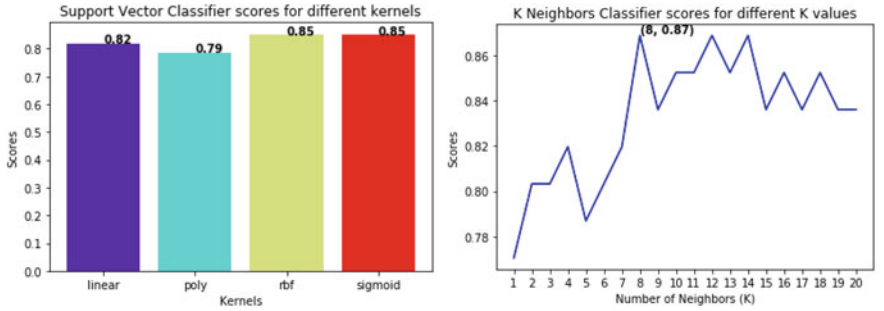


Fig. 6 **a** Support Vector Classifier scores for different kernels. **b** K Neighbors classifier scores for different k values

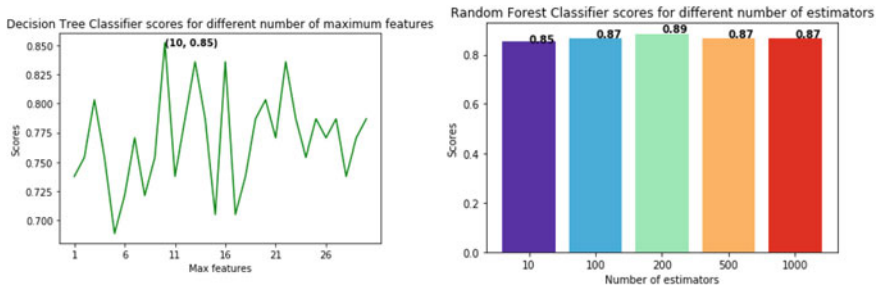


Fig. 7 **a** Decision Tree classifier scores for number of maximum features. **b** Random forest classifier scores for different number of estimators

5.2 Implementation

The optimum parameters are chosen for each algorithm and the dataset is training using these specifications. The training and testing data is segregated by a k-cross fold validation of 80 (train) : 20 (test) on the combined dataset mentioned in Sect. 3. The algorithms are further evaluated on the basis of F1 scores, recall, and precision to finalize on the algorithm to be implemented as shown in Table 2. Out of the four algorithms implemented, Random Forest (0.868) gives the best performance in terms of accuracy followed by SVM (0.852), k-NN (0.836), and decision tree (0.786) as shown in Fig. 8. With the random forest as the the choice for implementation, the time complexity of the model with the initial record being 303 and 9 variables is given by the big-O notation as O(6766).

The dataset is further built upon with the help of real-time sensor monitoring which is updated into the original dataset on which machine learning models are trained to further optimize the overall accuracy as well as give health and cardiac activity support to the user. The output on the heartbeat sensor, temperature sensor, and blood pressure sensor is extracted and pre-processed to satisfy the original dataset

Table 2 Performance evaluation

Algorithms	Accuracy	F1	Recall	Precision
k-NN	0.836	0.75	0.75	0.75
RF	0.868	0.85	0.79	0.92
SVM	0.852	0.798	0.77	0.783
DT	0.786	0.696	0.73	0.67

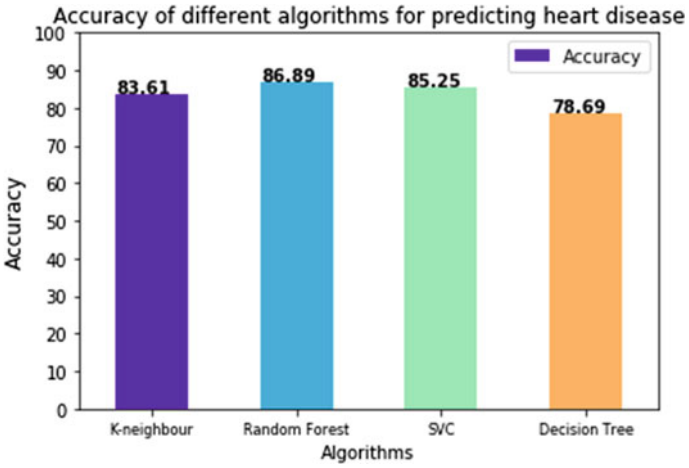


Fig. 8 Accuracy of k-NN, RF, SVM, and DT algorithm (left to right)

features and add to the generalization of the system. The RFID tag is additionally implemented to keep a log of the patient's health with a unique identification number which keeps the system user-friendly.

6 Conclusion

For assistive paradigms, a need for real-time health and behavior detection with sensor input is a prerequisite. This paper provides a concise summary of developments focused on artificial learning and IoT innovations in wellness and behavior-monitoring methods. Secondly, it demonstrates a modern technology architecture for health tracking, which requires real-time surveillance of patients or elderly users and requires access to data from the cloud along with the creation of a network to further enhance outcomes.

References

1. Ping W, Jin-Gang W, Xiao-Bo S, Wei H (2006) The research of telemedicine system based on embedded computer. In: Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS' 05), pp 114–117, IEEE, Shanghai, China, January 2006
2. Gani A, Gribok AV, Lu Y, Ward WK, Vigersky RA, Reifman J (2010) Universal glucose models for predicting subcutaneous glucose concentration in humans. IEEE Transactions on Information Technology in Biomedicine, vol 14, no 1, pp 157–165
3. Pervez MA, Silva G, Masrur S et al (2010) Remote supervision of IV-tPA for acute ischemic stroke by telemedicine or telephone before transfer to a regional stroke center is feasible and safe. Stroke, vol 41, no 1, pp e18–e24
4. World Health Organization (2012) Health Statistics and Health Information Systems. World Health Organization
5. Pierleoni P, Pernini L, Belli A, Palma L (2014) An android based heart monitoring system for the elderly and for patients with heart disease. Int J Telemed Appl 2014 Article ID 625156, 10 p
6. Singh M, Agarwal A, Sinha V, Manoj Kumar R, Jaiswal N, Jindal I, Pant P, Kumar M (2014) Application of handheld Tele-ECG for health care delivery in rural India. Int J Telemed Appl 2014, Article ID 981806, 12 p
7. Morris D, Schazmann B, Wu Y et al (2008) Wearable sensors for monitoring sports performance and training. In: Proceedings of the 5th International Summer School and Symposium on Medical Devices and Biosensors (ISSS-MDBS' 08), pp 121–124, IEEE
8. Patel S, Park H, Bonato P, Chan L, Rodgers M (2012) A review of wearable sensors and systems with application in rehabilitation. J NeuroEng Rehabil vol 9, no 1, article 2
9. Taplidou SA, Hadjileontiadis LJ (2007) Wheeze detection based on time-frequency analysis of breath sounds. Computers in Biology and Medicine, vol 37, no 8, pp 1073–1083
10. Corbishley P, Rodriguez-Villegas E (2008) Breathing detection: towards a miniaturized, wearable, battery-operated monitoring system. IEEE Transactions on Biomedical Engineering, vol. 55, no. 1, pp. 196–204
11. Lanat A, Scilingo EP, Nardini E, Loriga G, Paradiso R, De-Rossi D (2010) Comparative evaluation of susceptibility to motion artifact in different wearable systems for monitoring respiratory rate. IEEE Transactions on Information Technology in Biomedicine, vol 14, no 2, pp 378–386
12. Patel J, Upadhyay P, Patel D (2016) Heart Disease Prediction Using Machine learning and Data Mining Technique. Journals of Computer Science and Electronics 7:129–137
13. Chavan Patil AB, Sonawane P (2017) To Predict Heart Disease Risk and Medications Using Data Mining Techniques with an IoT Based Monitoring System for Post-Operative Heart Disease Patients. International Journal on Emerging Trends in Technology (IJETT) 4:8274–8281

Efficiency Enhancement in a Piezoelectric Energy Harvester for a Low-Frequency System Using Synchronous Charge Inversion

Parul Chaudhary and Puneet Azad

Abstract This article incorporates an efficient hybrid technology to improve the efficiency of a piezoelectric energy harvesting system. The power of the system improved by a staggering 50-fold over conventional full-bridge rectifiers. This nonlinear approach is an integration of an active rectifier and synchronized switch harvesting on the inductor (SSHI) circuit. The power extraction capability of this circuit is much better as compared to conventional full-bridge and voltage doubler circuits. Maximum power of 753.8 μW power is obtained while walking at a speed of 6 km/h.

Keywords Piezoelectric · Energy harvesting · Bias-flip · Active rectifier · Inductor sharing

1 Introduction

As the increasing energy demand has escalated the burden on exhaustible resources, viable energy sources surrounding us are becoming more and more relevant. Harvesters extracting energy from the sun [1], piezoelectric [2–4], thermoelectric [5], triboelectric [6–11] and pyroelectric [12–14] devices are in huge demand. These devices can provide usable power ranging from 10 to 1000s of μW. Piezoelectric energy harvesters have an advantage of a simple device structure along with good voltage and high energy density. Human motion energy has been harnessed to provide an uninterrupted power supply using piezoelectric harvesters in multiple

P. Chaudhary (✉) · P. Azad
Department of Electronics and Communication Engineering, Maharaja Surajmal Institute of Technology, New Delhi 110058, India
e-mail: parul@msit.in

P. Azad
e-mail: puneet.azad@msit.in

P. Chaudhary
University School of Information, Communication and Technology, GGSIP University, Delhi 110078, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_44

self-powered devices [3, 4]. This energy extraction, however, suffers from a big challenge to find a proper interface circuitry for maximum power transfer. Conventional interface circuits include bridge rectifier and voltage doubler circuits. The rectifier fairly converts the alternating input to a usable dc value. Still, its output power is comparatively low, and hence it is not an efficient technique for direct interface with the piezoelectric harvester [15].

Another method is by using an active rectifier, which significantly improves the rectification efficiency of the system by reducing the amount of voltage drop across the diodes [16]. Synchronized switch harvesting on inductor (SSHI) has also shown promising results when used with piezoelectric energy harvester [17]. This method is based on synchronous switching in the circuit while using a shared inductor to form a resonant circuit with the internal capacitance of the transducer. This technique has the advantage of harnessing the energy spent to charge the capacitance, which was getting wasted in any conventional method. Considerable research has been done to include modifications in this method to improve efficiency further [18–22]. In this paper, we have utilized a combination of SSHI and active rectifier to serve as an efficient interface, thereby delivering 50 times more power than these circuits [23].

A wedge-heeled shoe with a sandwich sole containing one single PZT drum transducer at the toe region is used as an energy harvester. This transducer is harnessing walking energy in D33mode and transforming this energy into an electrical signal [2]. The active rectifier−SSHI interface circuitry is added to it, and comparative results with conventional rectifier and voltage doubler have been analyzed to see the improvement in power delivered. The experimental output of the PZT harvester is put to theoretical simulations and the output power increased by 50%.

2 Experimental Section

This system integrates a perovskite crystal of PZT (lead zirconate titanate) with shoe sole at toe region utilizing piezoelectric D33 mode for energy extraction as shown in Fig. 1a. The PZT transducer converts the biomechanical motion energy into electrical energy generating an ac voltage and current of 38 V and 33.9 μA. The system operates at a frequency of 2 Hz while walking in a wedge-heeled shoe with a force of 900 N. Figure 1b shows an equivalent model of the PZT transducer having a current source of 33.9 μA with a parallel capacitance and an internal resistance. This model is used in Psim simulation with various interface circuitry. The values for internal capacitance and resistance are 10 pF and 150 KΩ. The standard circuits for bridge rectifier and voltage doubler are demonstrated in Fig. 2.

Figure 2a shows a conventional bridge rectifier circuit generally used for interfacing the harvester with the load. Four BAT45 Schottky diodes are used for this purpose at the output of the harvester. For each half value of the input cycle from the PZT harvester two diodes conduct and a pulsating unidirectional current is achieved at the rectifier output. The capacitor at the output is a considerably large value to hold the charge for the load. A rectifier can convert the alternating current into a

Fig. 1 **a** Schematic diagram of the experiment. **b** Equivalent circuit of PZT harvester

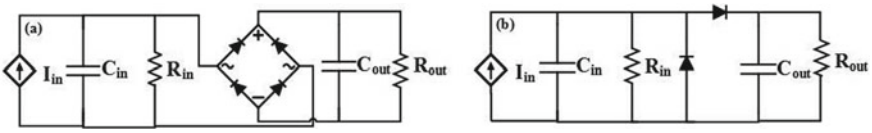


Fig. 2 **a** Full-wave rectifier. **b** Voltage doubler circuit

usable DC value. The voltage doubler circuit, as shown in Fig. 2b, is also an AC–DC converter. The maximum output power for the voltage doubler is almost the same as that of the rectifier but at 2× voltage at the output. The only improvement is obtained because of the lesser number of diodes involved in this circuit as compared to four diodes in a rectifier. The problem with these interface circuits is that most of the energy is used up in charging and discharging the output capacitance. This limits the output power, and hence lesser energy is available at the output.

Figure 3 illustrates an active rectifier–SSHI circuit with the PZT harvester. The circuit consists of two switches S1 and S2 connected with the harvester using an inductor L. Two active diodes D1 and D2 are associated with output capacitor C_{out} through these switches. The clocking frequency of S1 and S2 are kept equal and

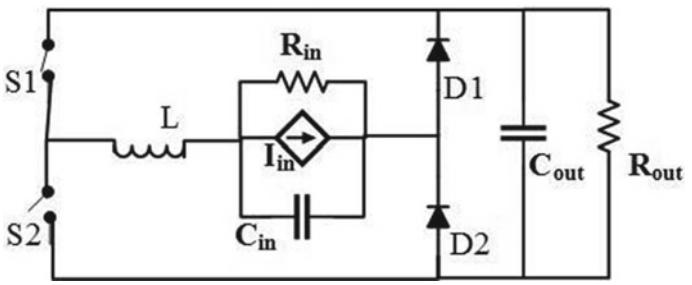


Fig. 3 Active rectifier–SSHI circuit

opposite to each other. An output resistor Rout is used across the capacitor to act as a load. A switching frequency of 1000 Hz is applied to S1 and S2. For the operation of the circuit, switch S2 is closed for a positive half cycle of the current from the piezoelectric harvester, and switch S1 is open. The current through an inductor is almost zero, and the two diodes are cut off to block the current through the load, thereby charging the internal capacitor Cin.

As the capacitor voltage increases beyond the diode threshold voltage, diode D1 starts conducting, and the transducer current starts to flow in the load resistance. This causes the inductor current to grow exponentially with diode D1 acting as a rectifier. As the inductor current reaches its peak value and becomes negative, the capacitor Cin starts discharging blocking diode D1. Switch S1 is closed and switch S2 is opened turning on diode D1. Since S1 and D1 are both turned on, a resonant loop is created to perform the SSHI operation. The energy stored in Cin is transferred to the inductor, and inductor energy is transferred back to the capacitor, thereby flipping the capacitor voltage to a negative level. Diode D1 functions to restrict the flow of current only from piezoelectric to the inductor through the diode. The flipping of voltage gets completed as the current through this loop becomes zero, also reducing the inductor voltage to 0 V. A similar process occurs for the negative half cycle as well achieving the full cycle for the SSHI operation. Diode D2 acts as a rectifier for the negative half cycle charging the input capacitance Cin. The process continues to operate till the harvester is extracting piezoelectric energy from human biomechanical motion. The output is measured across variable resistance using Fluke 287 true-RMS electronics logging multimeter, and power extracted is measured for analysis.

3 Results

The peak open-circuit current and voltage for the piezoelectric harvester are presented in Fig. 4. The graph shows a maximum of 38 V ac voltage and 33.9 μA current. Power

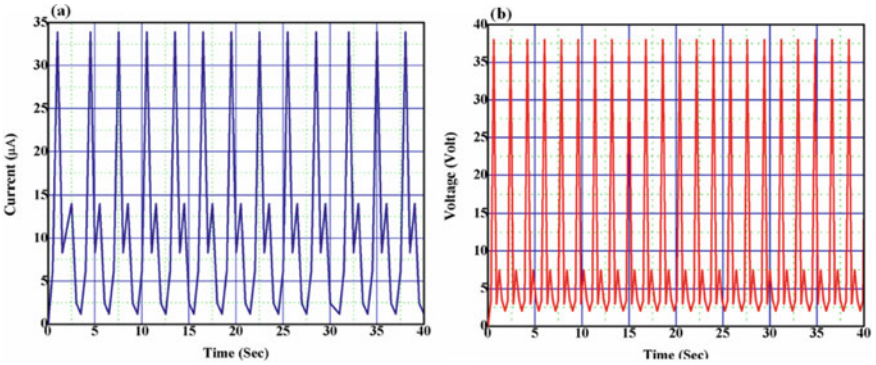


Fig. 4 **a** AC current graph. **b** AC voltage graph

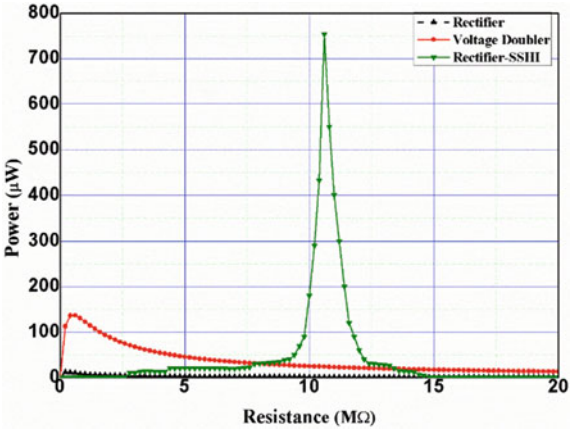


Fig. 5 Output power

for the three discussed circuits is measured and analyzed using Fluke 287 true-RMS electronics logging multimeter.

A comparison of maximum power delivered from a rectifier, voltage doubler and rectifier–SSHI circuit is presented in Fig. 5. The power provided for the load resistance increases as the load increases initially for the rectifier only until it reaches a maximum value of 11.9 μW at a resistance of 2 MΩ, as shown in Fig. 5. Piezoelectric transducer, when interfaced with a voltage doubler, showed an increase in maximum power transferred with 136 μW at a resistance of 6 MΩ. Rectifier–SSHI interface circuit reports an increase in shared power by 62% with a value of 753 μW at a resistance of 10.6 MΩ. This increase in power is contributed to synchronous switching used along with an active rectifier in the circuit.

4 Conclusion

This paper compares extractable power using different rectifier schemes. Active rectifier operated with synchronous switching can enhance this power by greater than 50× compared to conventional circuits used with the harvester. The active diode used in the resonant circuit ensures that the capacitor voltage is flipped at optimal times, making it even more efficient. This harvester design with an improved interface circuit gives better output power for the same input energy.

References

1. Azad P, Vaish R (2019) Solar energy harvesting using pyroelectric effect associated with piezoelectric buzzer. Phys Status Solidi (a) 216(20):1900440
2. Chaudhary P, Azad P (2020) Energy harvesting using shoe embedded with piezoelectric material. J Electron Mater. https://doi.org/10.1007/s11664-020-08401-6
3. Sodano HA, Inman DJ, Park G (2005) Comparison of piezoelectric energy harvesting devices for recharging batteries. J Intell Mater Syst Struct 16(10):799–807
4. Covaci C, Gontean A (2020) Piezoelectric energy harvesting solutions: a review. Sensors 20(12):3512
5. Azad P (2019) Temperature controlled voltage regulated boost converter for thermoelectric energy harvesting. IETE J Res. https://doi.org/10.1080/03772063.2019.1651678
6. Azad P, Vaish R (2017) Portable triboelectric based wind energy harvester for low power applications. Eur Phys J Plus 132(6):1–7
7. Chaudhary P, Azad P (2017) Demonstration of double electrode vertical-sliding triboelectric generator. In: International conference on computing, communication and automation (ICCCA), pp 1483–1487. IEEE, Noida
8. Yadav D, Azad P (2018) Low-cost triboelectric sensor for speed measurement and weight estimation of vehicles. IET Intell Transport Syst 12(8):958–964
9. Chaudhary P, Azad P (2018) Energy harvesting from human biomechanical energy for health-monitoring devices. IETE J Res. https://doi.org/10.1080/03772063.2018.1530074
10. Azad P (2019) A triboelectric energy harvester using human biomechanical motion for low power electronics. Bull Mater Sci 42(3):12
11. Azad P (2020) Design and analysis of a synchronized interface circuit for triboelectric energy harvesting. J Electron Mater 49(4):2491–2501
12. Sharma M, Singh VP, Singh S, Azad P, Ilahi B, Madhar NA (2018) Porous Ba 0.85 Ca 0.15 Zr 0.1 Ti 0.9 O 3 ceramics for pyroelectric applications. J Electron Mater 47(8):4882–4891
13. Azad P, Singh VP, Vaish R (2018) Candle soot-driven performance enhancement in pyroelectric energy conversion. J Electron Mater 47(8):4721–4730
14. Azad P, Sharma M, Vaish R (2019) Diesel exhaust emission soot coated pyroelectric materials for improved thermal energy harvesting. Glob Chall 3(6):1800089
15. Le TT, Han J, Jouanne AV, Mayaram K, Fiez TS (2006) Piezoelectric micro-power generation interface circuits. IEEE J Solid-State Circuits 41(6):1411–1420
16. Shim M, Kim J, Jeong J, Park S, Kim C (2015) Self-powered 30 μW to 10 mW piezoelectric energy harvesting system with 9.09 ms/V maximum power point tracking time. IEEE J Solid-State Circuits 50(10):2367–2379
17. Lefeuvre E, Badel A, Benayad A, Lebrun L, Richard C, Guyomar D (2005) A comparison between several approaches of piezoelectric energy harvesting. J de Phys IV (Proceedings) 128:177–186. EDP Sciences
18. Garbuio L, Lallart M, Guyomar D, Richard C, Audigier D (2009) Mechanical energy harvester with ultralow threshold rectification based on SSHI nonlinear technique. IEEE Trans Industr Electron 56(4):1048–1056
19. Xu SW, Ngo KDT, Nishida T, Chung GB, Sharma A (2007) Low frequency pulsed resonant converter for energy harvesting. IEEE Trans Power Electron 22(1):63–68
20. Liao W, Do XD, Lee SG, Ha DS (2016) A self-powered and optimal SSHI circuit integrated with an active rectifier for piezoelectric energy harvesting. IEEE Trans Circuits Syst I Regul Pap 64(3):537–549
21. Sanchez DA, Leicht J, Jodka E, Fazel E, Manoli Y (2016) A 4μW-to-1mW parallel-SSHI rectifier for piezoelectric energy harvesting of periodic and shock excitations with inductor sharing, cold start-up and up to 681% power extraction improvement. In: Proceedings IEEE international solid-state circuits conference (ISSCC), pp 366–367. IEEE, San Francisco

22. Ramadass YK, Anantha PC (2009) An efficient piezoelectric energy harvesting interface circuit using a bias-flip rectifier and shared inductor. IEEE J Solid-State Circuits 45(1):189–204
23. Wu L, Do XD, Lee SG, Ha DS (2016) A self-powered and optimal SSHI circuit integrated with an active rectifier for piezoelectric energy harvesting. IEEE Trans Circuits Syst I Regul Pap 64(3)5:37–549

Identifying Indicators of Sustainable Smart Agriculture Driven by Big Data Using Modified Total Interpretive Structural Modeling (mTISM)

Nikhil Dogra and Mohd. Adil

Abstract It is widely acknowledged that adoption of sustainable policies is a tool for the inclusive growth of the country. Although, manufacturing industry has received a considerable attention as compared to service industry, sustainability in agriculture is the newest approach. The purpose of this study is to explore key indicators of sustainable smart agriculture driven by big data. First, we identify the key indicators that affect sustainability and then investigate the contextual relationships among them. Modified total interpretive structural (m-TISM) is employed to investigate the interrelationships amongst the identified indicators. The findings indicate that the weather prediction and big data learning are the key indicators with high driving power. It implies that any change in these indicators would bring a significant change in other key indicators as well. Therefore, managers are cautioned to continuously monitor and deal with them with utmost care.

Keywords Smart agriculture · Big data · Internet of thing · Sustainability · TISM

1 Introduction

Nowadays rapid increase in population is prime challenge. This challenge of growing population can only be controlled by increasing production of agricultural products. An effective way to increase the quantity and quality production of agricultural products is technology i.e. Big data analytics. It is an appropriate tool for sustainable smart agriculture driven by big data. production. According to FAO [1], farmers should increase the food production upto 70% against 9.6 billion population by 2050.Along with this challenge of increasing population there are some other challenges hindering the increasing production such as climate change, rapid urbanization, scarcity of cultivable land, and insufficient fresh water. To overcome these challenges, smart technology like big data, internet of thing and cloud computing acts as a key to enhance the agricultural productivity. Big data technologies are the instruments used

N. Dogra · Mohd. Adil (✉)
Department of Management Studies, NIT Hamirpur, Hamirpur, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_45

to gather large amount of agricultural data at low cost which stimulates effective agriculture production. Advance smart technologies are deliberate to establish and drive artificial intelligence in agriculture sector through big data analytics and Internet of Things (IoT). The utilization of smart technologies leads to precision and smart agriculture [2] and maximize crop yield in a sustainable manner. Smart agriculture plays a prominent role to ensure sustainability and addresses the risk and challenges related to pest attack and climate change [3]. The clear understanding of big data concept is essential to respond the challenges associated with the farming management and crop cultivation. Because the concepts of big data technology and smart agriculture are almost new, therefore this concept need more research in agriculture sector to explore its applicability. Big-data driven smart sustainable agriculture is also a novel approach. Scholars have reported both positive and doubtful reviews on the usage of smart technology in agriculture [4–6]. This study contributes significantly to the existing literature (a) by identifying key indicators of sustainable smart agriculture driven by big data in agriculture sector that have been previously not explored by researchers and (b) classifying them based on their driving or dependence power. Therefore, with this backdrop, the current study aims at answering the following research questions—

RQ1. What are the indicators for sustainable smart agriculture in the context of an emerging economy?
RQ2. Are these indicators interrelated and how?
RQ3. Which technique enables us to determine the interactions between these indicators?
RQ4. What would be the outcome and its managerial implications in the agriculture sector?

To answer these research questions, we have scanned through a number of literature primarily for the following purposes—(a) to identify the indicators of sustainable smart agriculture driven by big data; (b) to determine the inter-relationships amongst the indicators by applying the modified total interpretive structural modeling (m-TISM) model; and (c) to delineate managerial implications by proposing the linkages between the identified indicators.

After the introduction, which sets the background of sustainable smart agriculture, the remainder of the paper is structured as follows: next section presents the literature review that presents a brief of big data technology, smart agriculture and indicators of sustainable smart agriculture driven by big data. Next, the paper proceeds to discuss the method used in the study followed by another section on analysis and discussion. The final section includes the conclusion and explicates directions for future research.

2 Literature Review

2.1 Big Data in Agriculture

Data is a potential tool for analyzing situations and making decisions accordingly. Agricultural big data signifies creation of large amount of data from stage of seed sowing to harvesting. Big data technology helps to renew the agricultural practices, those contributing in all dimensions of agriculture. It is an instrument to collect data from various stages and sources of agriculture. Collection of data should be done by using smart technologies such as aerial vehicles, smart phone applications, and special cameras. This data is then stored in form of database and analyzed by effective algorithms. Some already initiated applications for big data management are AgriPrice Book developed by North America Strategic Institute in 2014 for helping farmers to leverage data for improving transparency in agricultural supply chain and Citizen Science developed by CGIAR for promoting climate change and food security management.

2.2 Smart Agriculture

Smart agriculture also known as third green revolution represents the application of information communication technology (ICT) into agricultural practices. Smart agriculture act as a key and has potential to achieve sustainable agriculture by integrating technology into agriculture [7]. İt should provide the farmers with added value in the form of better decision making by collection real time data related to harvesting, soil, pest management and future prediction. İt is estimated that big data driven smart agriculture increase the crop yield by 20% and reduce the cost by 50%. Smart agriculture makes the agriculture sector smart with the support of technologies like big data and IoT and helps in tracking, monitoring, automating, and analyzing operations. Smart agriculture also include precision agriculture that help to increase food quality and productivity. This type of agriculture is based on environment friendly approach that applies the practices of various disciplines like biological and social sciences in a holistic way. Big data technology and Iot are appropriate instruments to drive the smart agriculture in a more sustainable way.

3 Methodology

The qualitative approach m-TISM is used in this study. The TISM [8] is developed from the modification of ISM as it enables researchers to interpret the "what" and "how" of research whereas m-TISM also provide information about the third question, i.e. "why" it is needed to interpret. m-TISM methodology is approved

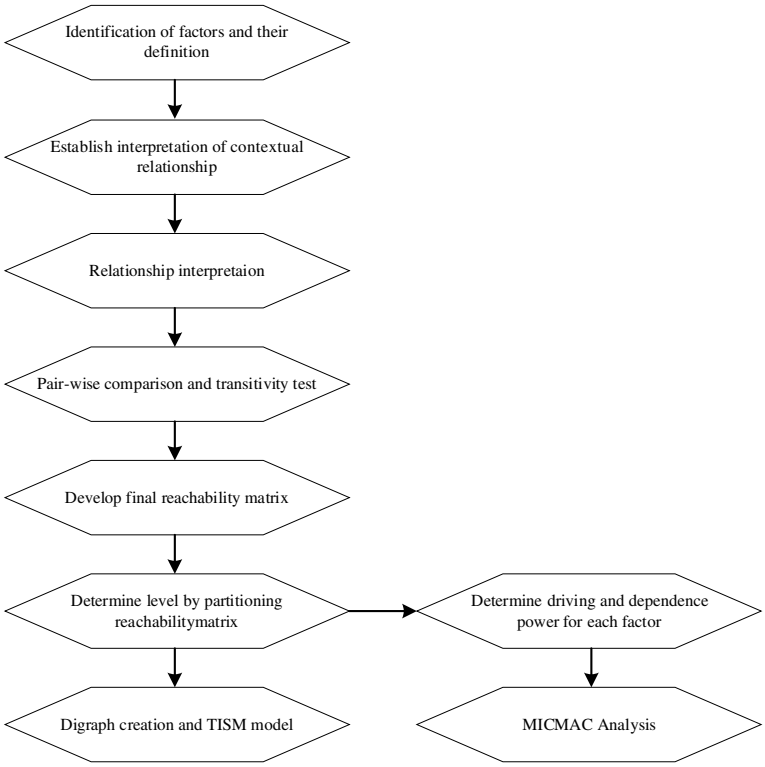


Fig. 1 Flow chart of m-TISM step

to examine the indicators of sustainable insurance and then to establish a stratified structure among the indicators.

To construct an m-TISM model, the reachability matrix and its level partitioning [9, 10] are used. The flowchart exhibited in Fig. 1 briefly explains the procedures used in the research methodology.

4 Analysis and Discussion

The findings of the present study extend our understanding by examining the driving and dependence power of indicators to drive sustainability at smart agriculture. Using m-TISM, we established theoretical argumentation and a qualitative framework with transitivity, indicating that a combination of such associations would evoke positive response towards adoption of sustainability. This study would serve as the point of reference in establishing the theoretical plausibility for sustainable framework.

Table 1 presents the reachability matrix (with transitivity), which enables the

Table 1 Reachability matrix (with transitivity)

S. no.	ER1	ER2	ER3	ER4	ER5	ER6	ER7	ER8	ER9	Driving force
ER1	I	I	I	I	I*	I*	I*	I*	I*	9
ER2	ɵ	I	ɵ	ɵ	ɵ	I	I*	I*	I*	5
ER3	ɵ	I	I	ɵ	I	I*	I*	I*	I*	7
ER4	I	I*	I	I	I	I*	I*	I*	I*	9
ER5	ɵ	I*	I	ɵ	I	I	I*	I*	I*	7
ER6	ɵ	I	ɵ	ɵ	ɵ	I	I	I*	I*	5
ER7	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ	I	I	I*	3
ER8	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ	I	I	I	3
ER9	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ	I	1
Dependence power	2	6	4	2	4	6	8	8	9	

identification of contextual relationships amongst the indicators. Table 2 presents the partitioning matrix result for all the indicators considered under the present study. This enables the researcher in developing different levels of indicators by matching sets of reachability and intersection using multiple iterations. The iteration process was repeated until each indicator found its level resulting into formation of an m-TISM model (as shown in Fig. 2).

In m-TISM model, indicator 9 (i.e. smart monitoring) is at the topmost level of the framework, which highlights that this indicator does not hold much importance as compared to other indicators that are at the lower level of the m-TISM framework. However, at the same time, one should be cautious that this finding does not undermine the role of technology in achieving sustainability.

Indicator 7 (policy formation) and indicator 8 (market information) are at the second level of the framework, highlighting their importance to that of the top level. indicator 2 (crop disease and pest management) and indicator 6 (irrigation management) is placed at the third level, which is even more important than those of the second and the first level. Similarly, rest of the indicators were levelled. Bottom level indicators are considered to be the most significant and important and therefore are strong drivers in implementing sustainability in smart agriculture. Any changes in indicators at the bottom level, would bring a significant impact on both the middle to top level indicators. Collectively, the modified m-TISM model suggests that the most important indicators for achieving sustainability in agriculture are: (a) weather prediction and (b) Big data learning.

Table 2 Partitioning matrix

S. no.	Reachability	Antecedent	Intersection	Level
Iteration 1				
ER 1	1, 2, 3, 4, 5, 6, 7, 8, 9	1, 4	1, 4	
ER 2	2, 6, 7, 8, 9	1, 2, 3, 4, 5, 6	2, 6	
ER 3	2, 3, 5, 6, 7, 8, 9	1, 3, 4, 5	3	
ER 4	1, 2, 3, 4, 5, 6, 7, 8, 9	1, 4	1, 4	
ER 5	2, 3, 5, 6, 7, 8, 9	1, 3, 4, 5	3, 5	
ER 6	2, 6, 7, 8, 9	1, 2, 3, 4, 5, 6	2, 6	
ER 7	7, 8, 9	1, 2, 3, 4, 5, 6, 7, 8,	7, 8	
ER 8	7, 8, 9	1, 2, 3, 4, 5, 6, 7, 8,	7, 8	
ER 9	9	1, 2, 3, 4, 5, 6, 7, 8, 9	9	1st
Iteration 2				
ER 1	1, 2, 3, 4, 5, 6, 7, 8	1, 4	1, 4	
ER 2	2, 6, 7, 8	1, 2, 3, 4, 5, 6	2, 6	
ER 3	2, 3, 5, 6, 7, 8	1, 3, 4, 5	3, 5	
ER 4	1, 2, 3, 4, 5, 6, 7, 8	1, 4	1, 4	
ER 5	2, 3, 5, 6, 7, 8	1, 3, 4, 5	3, 5	
ER 6	2, 6, 7, 8	1, 2, 3, 4, 5, 6	2, 6	
ER 7	7, 8	1, 2, 3, 4, 5, 6, 7, 8,	7, 8	2nd
ER 8	7, 8	1, 2, 3, 4, 5, 6, 7, 8,	7, 8	2nd
Iteration 3				
ER 1	1, 2, 3, 4, 5, 6,	1, 4	1, 4	
ER 2	2, 6	1, 2, 3, 4, 5, 6	2, 6	3rd
ER 3	2, 3, 5, 6	1, 3, 4, 5	3, 5	
ER 4	1, 2, 3, 4, 5, 6	1, 4	1, 4	
ER 5	2, 3, 5, 6	1, 3, 4, 5	3, 5	
ER 6	2, 6	1, 2, 3, 4, 5, 6	2, 6	3rd
Iteration 4				
ER 1	1, 3, 4, 5	1, 4	1, 4	
ER 3	3, 5	1, 3, 4, 5	3, 5	4th
ER 4	1, 3, 4, 5	1, 4	1, 4	
ER 5	3, 5	1, 3, 4, 5	3, 5	4th
Iteration 5				
ER 1	1, 4	1, 4	1, 4	5th
ER 4	1, 4	1, 4	1, 4	5th

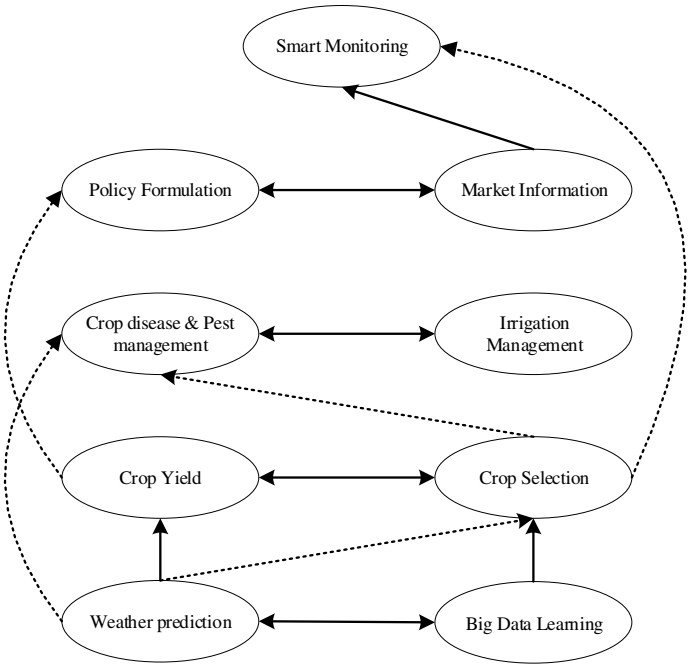


Fig. 2 m-TISM model with significant transitive links

5 Conclusions

The contribution of this research is multifold. This study investigates the indicators of sustainability in the area of the smart agriculture driven by big data, which is unique in the literature and attempts to fill the research gaps in the following ways:

(a) It identifies the key indicators of sustainability and proposed a well-defined qualitative framework.
(b) It also analyses the interactions amongst the identified indicators of sustainability, which has been a gap in the literature until now.
(c) The m-TISM methodology helps researchers to organize the indicators as per their dependence and driving power.
(d) The m-TISM methodology also helps in answering the key questions of theory building such as "what, " "how" and "why" and offers a better insight for researchers, practitioners, and academicians.

The findings of the study would also help farmers to identify grey areas in smart agriculture that require a more concentrated effort based on the importance and hierarchical level of the indicators in the conceptualized framework.

The study, based on interrelationships, classified all the nine indicators into the dependent and the driving force. This would assist managers and farmers in defining indicators that require continuous supervision and help management in channelizing its energy towards enhancing these indicators for the effective implementation

of sustainability in agriculture. Monitoring and evaluation of indicators of model must be supervised on regular basis and the policies must have enough flexibility to cater the deviating needs of the agriculture sector. The continuous monitoring of recent trends in the big technology and sustainable smart agriculture would help to improve the sustainable services. In addition, to achieve the intended results mandatory for sustainable adoption, the managers need to give more attention toward the indicators connected through transitivity links indicated by the m-TISM model. Further, sustainability indicators may help to formulate sustainable related policies and strategic plans more precisely and accurately. This may enable them to determine the priority of sustainable indicators so that positive measures can be taken to address them. Policymakers and management need to be cautious and give extra care towards the indicators lying at the bottom of the m-TISM framework, which reflects high driving power for successful implementation of sustainability. The proposed hierarchal model may enable the decision makers in understanding the interactions amongst the indicators and their impact before implementing sustainable practices at agricultural practices.

The findings of the study would supervise the management of big data to respond the current demands of farmers with special consideration to sustainability. The flow of direction of each indicator guide the management to take appropriate change in processes, which led to achieve competitive advantage and superior quality of service. Lastly, this study gives a new direction to create new sustainable projects strengthening the socio-economic and environment condition of countries.

References

1. FAO (2017) E-agriculture in action. Rome, Italy,
2. Kamilaris A, Kartakoullis A, Prenafeta-Boldú FX (2017) A review on the practice of big data analysis in agriculture. Comput Electron Agric 143:23–37
3. Singh NP (2016) Application of data warehouse and big data technology in agriculture in India, pp 1–8
4. Huang Y, Chen ZX, Tao YU, Huang XZ, Gu XF (2018) Agricultural remote sensing big data: Management and applications. J Integr Agric 17(9):1915–1931
5. Bendre MR, Thool RC, Thool VR (2016) Big data in precision agriculture: Weather forecasting for future farming. In: International conference on next generation computing technologies, NGCT 2015, pp 744–750
6. Channe H, Kothari S, Kadam D, Professors A (2015) Multidisciplinary model for smart agriculture using internet-of-things (IoT), sensors, cloud—computing, mobile—computing and big—data analysis keywords. Int J Comput Technol Appl 6:374–382
7. Belaud JP, Prioux N, Vialle C, Sablayrolles C (2019) Big data for agri-food 4.0: application to sustainability management for by-products supply chain. Comput Ind 111:41–50
8. Sushil S (2012) Interpreting the interpretive structural model. Glob J Flex Syst Manage 13(2):87–106
9. Ajmera P, Jain V (2019) Modeling the factors affecting the quality of life in diabetic patients in India using total interpretive structural modelling. Benchmarking Int J 26(3):951–970
10. Sindhwani R, Malhotra V (2017) A framework to enhance agile manufacturing system. Benchmarking Int J 24(2):467–487

Comparative Investigation of Machine Learning Algorithms for Wind Power Forecasting

Ayush Kumar, Neeraj Kumar, Bharat Singh, Aditya Chaudhary, Karan Dikshit, and Akash Sharma

Abstract Wind power unlike power generated from conventional sources is not constant. There are many factors that influence the power generated from wind energy, like wind speed, location, climate change etc. Owing to this, there is always uncertainty in wind power output. Thus, for proper load scheduling and better integration of wind power with the grid, it becomes essential to develop a robust wind power forecasting system. For developing a reliable forecasting system, it is essential to factor in all the possible factors that affect the wind power output and analyze a huge amount of data set for a higher accuracy rate. This paper proposes the use of two machine learning techniques, namely LASSO and XGBoost classifier, and a comparison is made between the two to find which technique is better for our task. For training and validation of this model, wind power data of the Kolkata region is taken. The result shows that XGBoost is better than LASSO for forecasting wind power accurately with a MAPE value of 1.121 for XGBoost and 62.1476 for LASSO.

Keywords Short-term wind power forecasting · Machine learning · LASSO · XGBoost · RMSE · MAPE

1 Introduction

Owing to the overexploitation of natural resources in today's world, the conventional resources deployed for generating energy are exhausting at an unprecedented rate. As a result of this, modern society demands resources that are sustainable and less polluting to the environment. This is where the concept of generating energy from the flow of wind comes into the picture as wind energy can cater to the needs of modern world consumers [1].

At the same time, it should be taken into consideration that wind energy is highly intermittent in nature and thus, for the optimum operation of the wind energy conversion systems (WECS), we need to devise modern methods which would help in

A. Kumar · N. Kumar (✉) · B. Singh · A. Chaudhary · K. Dikshit (✉) · A. Sharma
Electrical and Electronics Engineering Department, Bharati Vidyapeeth's College of Engineering, New Delhi, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_46

predicting the nature of wind in advance [1, 2. This would potentially lead to better integration of the WECS systems with the nearest power system grid and would also help in the uniform distribution of the energy being generated [3]. Thus, the impact of wind energy integration directly depends upon the accuracy of these prediction or forecasting models [1].

As of now, several forecasting models have been developed which cater to different fields involving multiple applications. The most common ones include the physical model, traditional model and advanced forecasting models based on machine learning algorithms [4–6]. The application of these models in different situations depends upon the time scale of forecasting. Based on this parameter, forecasting trends are classified into [7–9]:

- Ultra short-term forecasting (1 h)
- Short-term forecasting (14–15 h)
- Medium-term forecasting (1 week)
- Long-term forecasting (1 year or more).

The statistical mode incorporates several physical parameters (such as precipitation patterns, air density, real-time variations in wind speed etc.) as input data on which the magnitude of the generated power is dependent, and makes a prediction based on the historical trends in the available dataset [10]. On the other hand, the physical model [6] deploys an analytical approach by evaluating past trends in the wind patterns to make predictions of wind power. These models are considered to be fairly accurate when it comes to forecasting short-term wind power. But for long-term wind power forecasting (LTWPF), these models become highly inconsistent. That is where the need for devising new models based on modern machine learning algorithms comes into the picture. These models are based on the concept of ensemble learning, which facilitates them to develop and utilize multiple statistical models in order to solve a given computational problem such as long-term wind power forecasting (LTWPF) [11].

The given paper proposes a machine learning model meant to forecast wind power in advance. To ensure the highest possible accuracy of the proposed model, a large input dataset comprising meteorological data and wind power data is taken into consideration. The model tries to establish a relationship between the different input features and the magnitude of the generated power. The predictive modelling can be performed using several methods, including artificial neural networks [12, 13], decision trees [14], random forests [15] etc.

The paper aims to study wind power forecasting using two different machine learning algorithms, namely XGBoost classification and the LASSO technique.

XGBoost Boost [16] is a machine learning algorithm that builds on the gradient boost algorithm to give even more accurate results. It works similar to the random forest technique by implementing ensemble learning [17] where some weak learning models like the decision trees are taken and then their output clubbed together either by boosting or by bagging for accurate results.

LASSO is [18] a machine learning technique based on regularization. It is used for prediction purposes in the case of very complex models having a large number of

features. Regularization is needed to reduce a very annoying problem called overfitting. Overfitting increases in the case of complex models. LASSO regression works by performing L1 regularization and penalizes the magnitude of the coefficients with the help of a tuning parameter called alpha (α).

2 Methodology

2.1 Generation of Wind Power and Forecasting

The use of aerodynamic forces on the rotor blades for the generation of electricity has helped us in solving many problems that mankind is facing today. The unethical and extensive use of fossil fuels has taken an enormous toll on our nature, landscapes and ecosystems and had forced us to switch to alternate sources, with wind power being one of them. Wind power is generated by harnessing the mechanical aerodynamic forces present in the air by converting it into rotational power to drive the alternators [5] (Fig. 1).

Wind power generation differs from conventional power sources not only in terms of the use of renewable sources instead of conventional ones but also in terms of their availability due to their stochastic nature. Wind power generation is dependent on many factors such as air density, precipitation, temperature and primarily wind speed, and hence today's power operators are dealing with challenges of balancing supply and demand due to its variable and unpredictable nature [19].

Thus, accurate forecasting of wind power plays a major role in ensuring reliable large-scale integration with the existing electrical network. Short-term and very short-term wind power forecasting methods can thus be used for planning power scheduling, unit commitment and dispatch by the power operator and to minimize the cost and maximize the profit by electricity traders [7].

Here, we have used two machine learning techniques for wind power forecasting, namely XGBoost and LASSO. In both these techniques, we have taken four parameters as independent values along with actual output as dependent values. The dataset

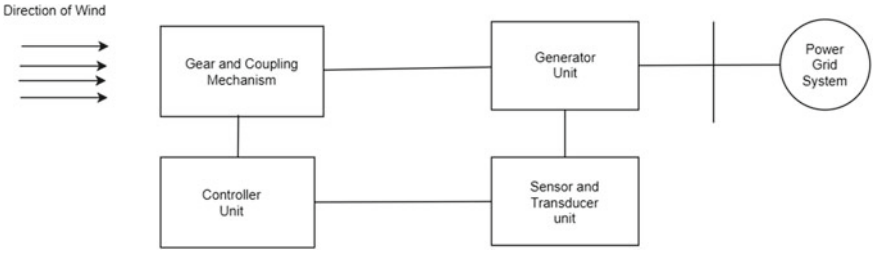


Fig. 1 Block diagram of wind energy conversion systems (WECS)

has been obtained from the website https://www.renewables.ninja for the Kolkata region of India and the same dataset is used for training and testing the algorithm.

2.2 *Extreme Gradient Boost (XGBoost Classifier)*

XGBoost is yet another method that can be used to implement classification and regression. This technique is known by different names when implemented, with the most common being gradient boosting machine [20] (GBM) and XGBoost, XGBoost being the most viable due to its versatility.

XGBoost Boosting resembles random forest to a great extent. That is because they both follow a technique known as ensemble learning. Ensemble learning works by utilizing a large number of weak classifiers [21] and clubbing them together for a better result. It implies that a final model is derived using a collection of singular models. The accuracy of prediction is weak for the singular models with chances of overfitting [22] but the result is improved manifold by combining them together in an ensemble. Like the random forest, the weak model used most commonly in XGBoost Boosting is also decision trees (Fig. 2).

When implementing deep learning and neural networks, gradient descent is used to find the minimum of the loss function [23], that is the parameters or weights for which a singular model gives the least error of prediction. Due to the clubbing of multiple weak models, the model parameters need not be optimized directly as the final boosted parameter can be optimized at the end. Hence the gradients can be added spontaneously to the training process by fitting the new tree also to these values (Fig. 3).

As per the gradient descent [24], we find the learning rate, shrinkage and loss function as hyperparameters, similar to neural networks. Some other hyperparameters of XGBoost are as follows:

1. No. of iterating trees in the ensemble,
2. No. of observations in each leaf,
3. Depth and complexity of the trees,
4. The fraction of samples and
5. The fraction of features selected for training.

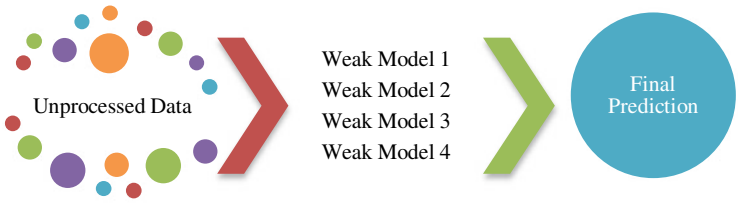


Fig. 2 Diagrammatic representation of ensemble learning

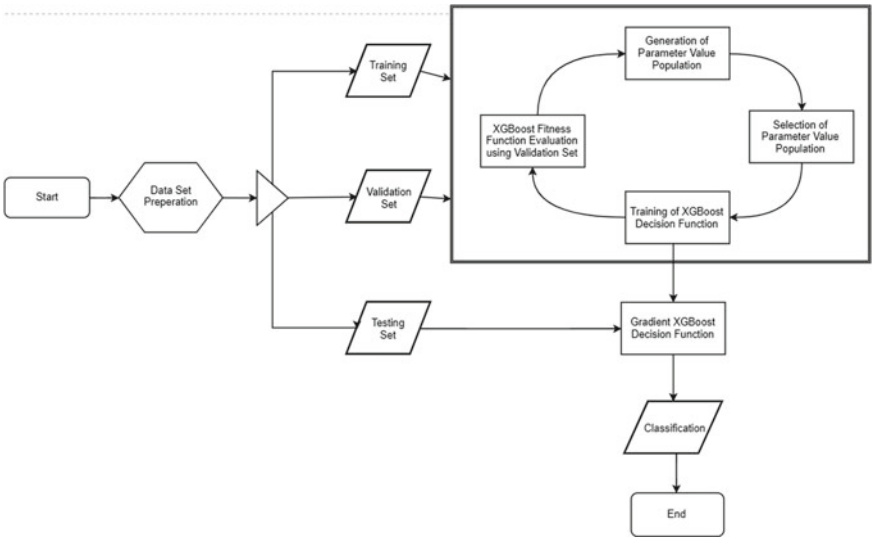


Fig. 3 Flowchart for the generalized algorithm

2.3 LASSO (Least Absolute Shrinkage and Selection Operator)

In machine learning there are seven types of regression techniques. One of them is LASSO, which stands for least absolute shrinkage and selection operator. It is a very powerful prediction technique used for very complex models, i.e., models having large number of features. It works on the principle of regularization. LASSO regression is very similar to [25] ridge regression as both techniques are based on regularization. Both these techniques work by penalizing the coefficients magnitude. The only difference lies in how the penalty is assigned to the coefficients.

LASSO regression [26] performs L1 regularization by adding a bias term or penalty which is equal to sum of absolute value of magnitude of coefficients (Fig. 4).

Minimization objective = (Least square objective) + α * (sum of absolute value of coefficients).

Here α (alpha) is used to vary the magnitude of coefficients as follows [27]:

$\alpha = 0$; no change in magnitude of coefficients.
$\alpha = \infty$; all coefficients become zero.
$0 < \alpha < \infty$; all coefficients between 0 and previous value of linear regression.

Here, $\beta\hat{}$ represents the pair of two coefficients, $\beta 1$ and $\beta 2$. They minimize the least square objective of unregularized model.

The size of constraint region shown is governed by tuning parameter α.

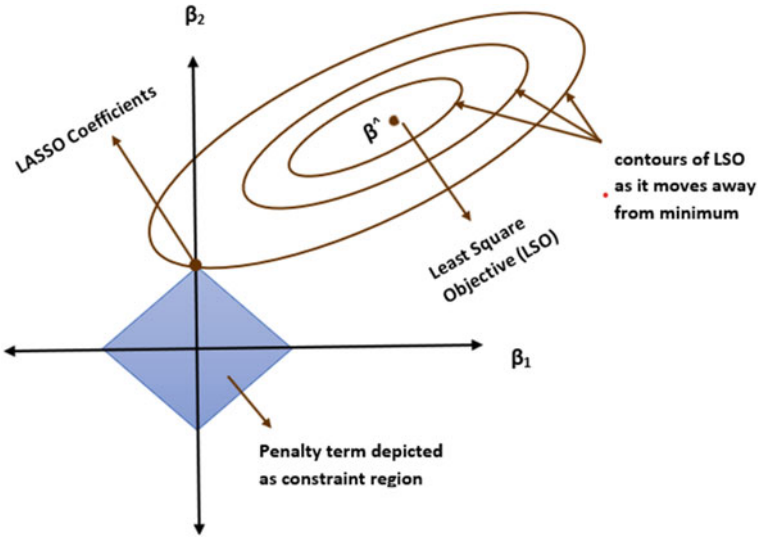


Fig. 4 LASSO regression visual representation

- When the α value is 0 the blue region becomes infinitely large, i.e., no constraint on the magnitude of coefficients.
- When α increases, the blue region keeps on decreasing, i.e., magnitude of coefficients get closer to 0.

The need for penalizing [28] the magnitude of coefficients arises because on increasing the magnitude of the coefficients, the model becomes complex and has a large number of features. The large value of the coefficient means putting too much importance on the particular feature for prediction. This leads to the model giving too accurate results for the trained dataset which eventually causes overfitting.

3 Result and Comparison

These techniques can now be demonstrated by implementing them in MATLAB. The version used here is the MATLAB r2018a. A comparison can then be made between all the methods to the actual output and find the accuracy of prediction using a loss function.

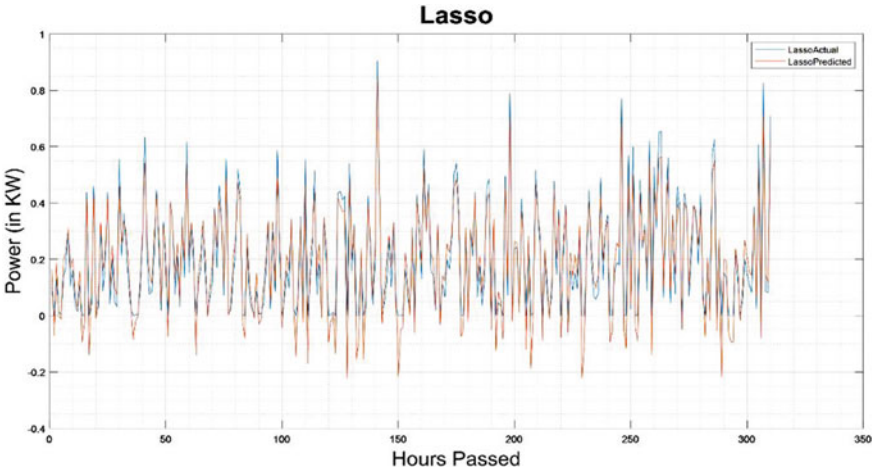


Fig. 5 LASSO prediction plot

3.1 LASSO

Here we can see the output plot of the LASSO technique. As seen, it provides not so accurate results as the predictions go way below zero, which in the case of generated power is impractical. The blue line shows the actual power and the red shows the predicted power (Fig. 5).

3.2 XGBoost Classifier

This is the final plot for the XGBoost technique and has one of the best results of all the techniques discussed here (Fig. 6).

3.3 RMSE and MAPE

The accuracy of the prediction is required to be judged after making the predictions. Several loss functions are deployed for this purpose, including RMSE, MAE, MAPE etc. Out of these, we will be using the RMSE and MAPE. RMSE stands for root mean square error and MAPE stands for mean absolute percentage error. They are two parameters that can be used to observe the accuracy of the predictions.

The RMSE is calculated as:

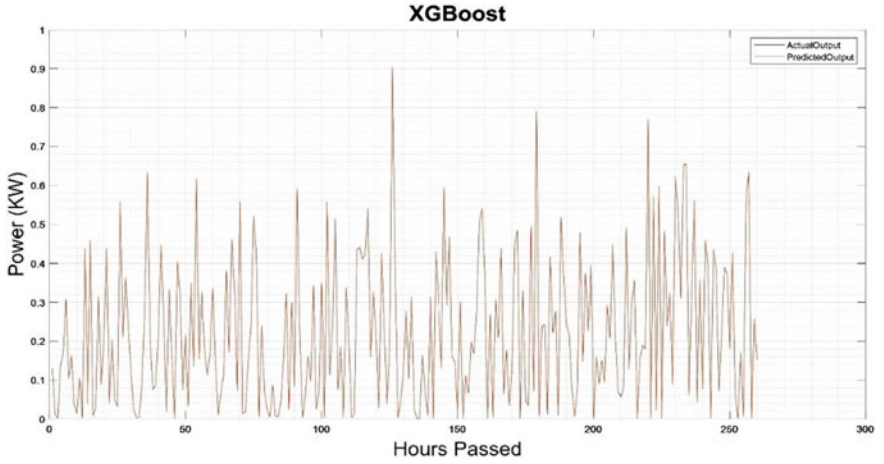


Fig. 6 XGBoost prediction plot

$$\sqrt{\frac{\sum_{i=0}^{n}(Ai - Pi)^2}{n}} \tag{1}$$

The MAPE is calculated as:

$$\frac{\sum_{i=0}^{n}\frac{|Ai-Pi|}{Ai}}{n} \times 100 \tag{2}$$

where Ai tells the actual power generated, Pi tells the predicted power generated, and n tells the number of observations.

Fig. 7 gives the RMSE and MAPE values of the LASSO and XGBoost Boost technique. As seen by the bar graph itself, XGBoost gives better result.

4 Conclusion

The paper exhibits the competence of two different machine learning techniques, i.e., XGBoost Boost and LASSO in predicting the output power generated for short-term wind power forecasting. The accuracy of forecasting is then examined by their respective MAPE and RMSE. The RMSE comes out to be 0.043 and 0.0013 for LASSO and XGBoost, respectively, whereas the MAPE is 62.142 and 1.121 for LASSO and XGBoost, respectively.

Hence, from the output plots and the RMSE and MAPE values, it can be clearly observed that the XGBoost provides way better forecasting for wind power than the LASSO. Overall, the best and most accurate prediction is done by the

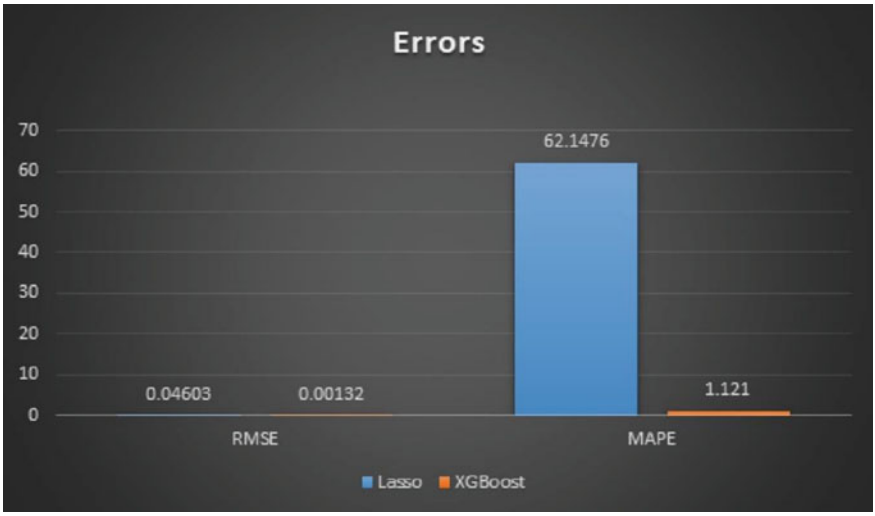


Fig. 7 RMSE and MAPE of LASSO and XGBoost classifier

XGBoost boosting technique and can be successfully used for short-term wind power forecasting.

Wind power forecasting is essential for the successful integration of wind power into the grid and its applications and uses are only going to increase as we progress to the future. Hence optimum research and study in this field are necessary.

References

1. Bhaskar M, Jain A, Venkata Srinath N (2010) Wind speed forecasting: present status. Power system technology (POWERCON), 2010 international conference, pp 1–6, 24–28 October 2010
2. Sharp J, Hodge B, Florita A, Margulis M, Mcreavy D (2010) The value of improved short-term wind power forecasting. In: 9th International proceedings on proceedings, pp 1–2. http://www.nrel.gov/publications.
3. Orwig K, Ahlstrom M, Banunarayanan V, Sharp J, Wilczak J, Freedman J, Haupt S, Cline J, Bartholomy O, Hamann H, Hodge B-M, Finley C, Nakafuji D, Peterson J, Maggio D, Marquis M (2015) Recent trends in variable generation forecasting and its value to the power system. IEEE Trans Sustain Energy 6:924–933
4. Manwell JF, McGowan JG, Rogers AL (2010) wind energy explained: theory, design and application . John Wiley & Sons, Hoboken, NJ, USA
5. Gasch R, Twele J (2011) Wind power plants: fundamentals, design, construction and operation. Springer, Berlin, Germany
6. Foley AM, Leahy PG, Marvuglia A, McKeogh SJ (2012) Current methods and advances in forecasting of wind power generation. Renew Energy 37:1–8
7. Wang X, Guo P, Huang X (2011) A review of wind power forecasting models. Energy Proc 12:770–778

8. Zhao X, Wang S, Li T (2011) Review of evaluation criteria and main methods of wind power forecasting. Energy Proc 12:761–769
9. Dongmei Z, Yuchen Z, Xu Z (2011) Research on wind power forecasting in wind farms. In: Proceedings of the 2011 IEEE power engineering and automation conference (PEAM), Wuhan, China, 8–9 September 2011
10. Buhan S, Cadirci I (2015) Multistage wind-electric power forecast by using a combination of advanced statistical methods. IEEE Trans Ind Inf 11(5):1231–1242
11. Lu HJ, Chang GW (2018) Wind power forecast by using improved radial basis function neural network. In: 2018 IEEE power & energy society general meeting (PESGM), Portland, OR pp 1–5
12. Palomares-Salas JE, de I a Rosa JJG, Ramiro JG, Melgar J et al (2009) ARlMA vs. neural networks for wind speed forecasting. In: Proceedings of IEEE international conference on computational intelligence for measurement systems and applications, pp 129–133
13. Tesfaye A, Zhang JH, Zheng DH, Shiferaw D (2016) Short-term wind power forecasting using artificial neural networks for resource scheduling in microgrids. Int J Sci Eng Appl (iJSEA) 5(3)
14. Song YY, Lu Y (2015) Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry 27(2):130–135. https://doi.org/10.11919/j.issn.1002-0829.215044
15. Sammut C, Webb GI (eds) (2011) Random forests. In: Encyclopedia of machine learning. Springer, Boston, MA
16. Friedman JH. IMS 1999 greedy function approximation: a gradient boosting machine Reitz lecture
17. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. arXiv:1603.02754
18. Gauraha N (2018) Introduction to the LASSO. Reson 23:439–464. https://doi.org/10.1007/s12045-018-0635-x
19. Foley AM, Leahy PG, Marvuglia A, McKeogh EJ (2012) Current methods and advances in forecasting of wind power generation. Renew Energy 37(1):1–8
20. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery & data mining, San Francisco, CA, USA, 13–17 August 2016
21. Martinez-De-Pison FJ, Fraile-Garcia E, Ferreiro-Cabello J, Gonzalez R, Pernia A (2016) Searching parsimonious solutions with GA-PARSIMONY and XGBoost in high-dimensional databases. In: Grana M, LopezGuede JM, Etxaniz O, Herrero A, Quintian H, Corchado E (eds) International joint conference Soco'16-CISIS'16-ICEUTE'16, San Sebastián, Spain, 19–21 October 2016, vol 527, pp 201–210. Springer, Cham, Switzerland
22. Song RW, Chen SD, Deng BL, Li L (2016) XGBoost boosting for identifying individual users across different digital devices. In: Cui B, Zhang N, Xu J, Lian X, Liu D (eds) Proceedings of the web-age information management: 17th International conference, WAIM 2016, Nanchang, China, 3–5 June 2016, vol 9658, pp 43–54. Springer, Cham, Switzerland
23. Sheridan RP, Wang WM, Liaw A, Ma JS, Gifford EM (2016) XGBoost boosting as a method for quantitative structure-activity relationships. J Chem Inf Model 56:2353–2360
24. Ye J, Chow J-H, Chen J, Zheng Z. Stochastic gradient boosted distributed decision trees. In: Proceedings of the 18th ACM conference on information and knowledge management, CIKM '09
25. Rodríguez O (2013) A generalization of ridge, lasso and elastic net regression to interval data. https://doi.org/10.13140/2.1.3753.0883
26. Arribas-Gil A, Bertin K, Meza C, Rivoirard V. Lasso-type estimators for semiparametric nonlinear mixed-effects models estimation. Stat Comput 24(3):443–460
27. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. Stanford University, USA
28. Boulesteix A-L, De Bin R, Jiang X, Fuchs M (2017) IPF-LASSO: integrative-penalized regression with penalty factors for prediction based on multi-omics data, vol 2017, Computational and mathematical methods in medicine, Hindawi, pp 1748–670X. https://doi.org/10.1155/2017/7691937

Voltage Conditioning and Harmonic Mitigation Using UPQC: A Review

Manish Kumar Singh and Vivek Saxena

Abstract The quality of delivered power at the user terminals is the major area of concern in modern electric power systems. It has nowadays become very important and essential due to the drastic increase in the use of sophisticated electrical devices, whose performance parameters are highly dependent and sensitive to the quality of power being delivered. The major problem is linked to the quality of supply voltage which often deteriorates due to common sag in voltage amplitude and also swells in voltage. UPQC is used to enhance the quality of supplied power and the performance parameters are dependent on the coordination of a series dynamic channel and shunt dynamic channel. The subsequent topology has attributes of the two channels and can regulate voltage and unsettling influences and enhances the client's energy quality. In this paper how various control strategies can be employed in different UPQC topologies to regulate load voltage under conditions of fast disturbances in voltage profile while also mitigating current harmonics has been discussed.

Keywords UPQC · Series Active Filter · Shunt Active Filter

1 Introduction

Nowadays, consumers of electric power whether individual clients or bulk power users, i.e., industrial and commercial users utilize numerous electronic equipments having in-built converters that draw the harmonic currents from the supply mains, which in turn influences the network's quality and stability. Power quality is one of the vital issues to be addressed by the electrical engineers which results due to extreme usage of loads working on solid-state devices and variable loads. Many of the FACTS devices were tested to improve the delivered end-user power quality. The sophisticated equipments are more sensitive equipment and are the prime cause for the injection of harmonics in the distribution system. Such disturbance in power can cause the microprocessor-based or programmable logic controller or microcontroller-based control system to malfunction. The communication lines running parallel to

M. K. Singh (✉) · V. Saxena
ABES Engineering College Ghaziabad, Ghaziabad, Uttar Pradesh, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_47

power transmission lines are disturbed by electric and magnetic interference. The life of electrical equipment is also decreased due to poor power quality. False tripping of circuit breakers and flickering of lights are very severe problems in daily life. A lot of equipments with nonlinear characteristics and whose operation requires the usage of power electronic devices inject voltage distortion and current harmonics. A momentary change in loads and switching of equipments creates a transient voltage waveform, and owing to this, malfunction of relay and sophisticated equipments occur. At distribution stations, the voltage is not at its appropriate value, which means voltage dip and voltage swell problems may occur at end users. Owing to increased utilization of indirect loads and power converters-based electronic equipments in industries as well as by retail consumers, it has become necessary to improve the voltage profile to a minimum desired value while also keeping the current harmonics at an acceptable limit. Keeping in mind the end goal to give quality energy to the clients, different regulating bodies and participating utilities are gearing to work together to achieve the goal of quality supplied power.

Many custom devices for power quality enhancement are used to minimize the dip in voltage and problems arising out of a swell of voltage. The power supplied by generating power plants must be up to mark and standard so that end customers receive quality power. For transferring a good quality of power, custom power devices, namely UPFC, IPFC, SSSC, STATCOM, D-STATCOM, DVR etc., were used at the distribution level. UPQC is one of the modern devices for controlling simultaneously end voltage and end current harmonics by decoupled control of supply voltage and load current. UPQC is basically a combination of a series as well as a shunt control channel associated with a normal common dc interface. In the most recent years, investigations and researches have been done to find the optimum control strategy for controlling various attributes of UPQC and diminish the intricacy of the topology of converters. In this way, a series converter fills in as a coordinated source of voltage and a shunt converter fills in as a current source. The schematic connection diagram of one of the typical UPQC has been represented in Fig. 1.

2 Literature Review

In [1] the author provides an exhaustive review of UPQC using 12 acronyms being listed, particularly UPQC-DG, UPQC-VA, UPQC-D, UPQC-S, UPQC-I, UPQC-R, UPQC-L, UPQC-Q, UPQC-MC, UPQC-ML, UPQC-MD and UPQC-P.

In [2] the author manages control quality change by using a combination of UPQC with passive filters in the power system. The proposed plan can reduce voltage unsettling influences and current harmonics in spite of the regular coupling of control circuitry networks.

In [3] the author introduces a series along with a parallel real power compensation technique to create a harmonic current reference value connected with the series control converter applied in single-phase UPQC.

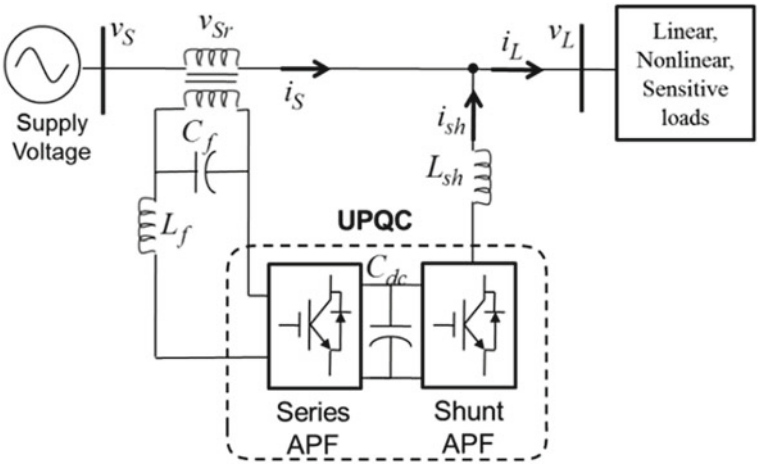


Fig. 1 Schematic representation of UPQC

In [4] the proposed technique involves using together with the shunt as well as APF in UPQC incorporating fuel cell model in order to get better results. The model described has been simulated with MATLAB and the results are confirmed by utilizing FFT analysis.

In [5] the d-q technique and p-q idea are utilized for control of UPQC. The UPQC-based mathematic approach to deal with moderate power quality problems had been incorporated by using MATLAB/Simulink.

In [6] the author presents a strategy for control of the series filter in unbalanced and mutilated load conditions. The proposed strategy depends on PLL. For providing a unified approach for UPQC, the shunt filter controller is also described.

In [7] two different modulation modes were optimized along with decreased switching losses and harmonic distortion so as to show the adaptability obtained through space vector modulation of UPQC.

In [8] an approach for the UPQC, known as a sag-based topology for phase regulation, i.e., UPQC-SPAC has been proposed. A proper allocation for a UPQC-SPAC has also been observed to be advantageous for the systems with distributed generators.

In [9] ANN-oriented controller is intended for the control of shunt APF and has been trained offline by the conventional PI controller. The digital controller has been used in real-time simulations.

In [10] the paper depicts a novel topology of one-phase UPQC-DG for networks and was examined by means of reenactment with PSCAD/EMTDC investigation program programming. The outcomes are exhibited to affirm the legitimacy of the proposed approach.

In [11] hysteresis control technique is utilized for exchanging power across PAF and SAF which allows zero-voltage exchanging (ZVS) for its switches during turn ON.

In [12] the paper exhibits a switching technique for multilevel inverters on the basis of space-vector-oriented theory. This switching procedure creates a vector of voltage with low harmonics distortion and a reduction in switching frequency.

In [13] the execution examination of a UPQC and SSSC to help fault in the ride-through ability of 2 MW FSIG has been completed and UPQC is observed to be much more practical in relation to the device rating.

In [14] the proposed SERC has the ability to share lagging reactive demand of load with shunt compensator while reducing its loading as well. The UPQC utilizing this kind of quadrature injection of voltage in series has been named UPQC-Q.

In [15] the paper applies UPQC to upgrade the LVRT ability of the DFIG-based WECS. The UPQC is connected to shield the framework from the ground faults, permits quick restoration of generation characteristics, enhances the system p.f. and keeps the network from the rotor over-currents and the dc-interface overvoltage.

In [16] to expand the compensation capacity of UPQC large supercapacitors have been combined with bidirectional DC to DC converters to create energy storage systems in parallel to the DC-link while keeping DC voltage steady and that to be utilized additionally as an extra UPS.

In [17] an MC-UPQC network is proposed that can control the load voltage along with parallel distribution system during the balanced as well as distortion in loading conditions and acquire state-space modeling for MC-UPQC. The simulation results using the SRF-oriented control technique are displayed in the paper using the MATLAB/Simulink software.

In [18] a UPQC model based on control of phase (UPQC-PAC) is demonstrated. In this series an inverter infuses a voltage having a regulated phase so that the magnitude of voltage is unchanged at the load end.

In [19] a UPQC model based on control of phase (UPQC-PAC) is utilized. Within the proposed model, optimal location, reactive power demand to compensate at the location and parameters of the UPQC are controlled by limiting three targets: ratings of the UPQC, losses in the network and the percent of a node with a low voltage issue.

In [20] real power regulation approach is utilized to compensate for swell/sag and has been integrated along with PAC theory to control load's reactive power within two inverters. Although series inverter at the same time conveys real and reactive power, this idea is called UPQC-S.

In [21] OPEN UPQC has been modeled for an LT grid rated for 400-kVA as a test system to measure steady-state execution. The results obtained during steady-state conditions justify the better device performance.

In [22] the paper exhibits another system (MC-UPQC), in which all converters must be connected back-to-back over the DC terminal and offer a typical DC interface capacitor. The result obtained on PSCAD/EMTDC for two-bus and two-feeder systems demonstrates the robustness of the proposed design.

In [4], the Simulink model for shunt, series and fuel cell along with the UPQC are designed in MATLAB. Also, a hysteretic current controller along with the hysteretic voltage controller has been proposed as the control method.

In [23] the generalized scheme has been proposed to reduce the fundamental rating of a parallel converter, the series converter as well as series transformers. Data-driven controller (DDC) is produced utilizing variable PAC technique to understand the execution of the outlined UPQC system under different compensation conditions.

In [24] the control system of PAF depends on a mixture of p-q theory as well as the symmetrical components whereas the control circuit for the SAF depends on only symmetrical components.

In [25] the mathematical model of UPQC (as a combination of a shunt-APF along with series-APF) is being discussed in the distribution power system. It is further expanded by including a series of VSC with each feeder. The p-q and d-q concepts are utilized for control of UPQC.

In [26] a power converter topology where UPQC is joined with the photo-voltaic generators is displayed. This topology depends on a double structure of the established three-phase two-level inverters. The current and voltage controllers are executed in the synchronous pivoting reference (dq0-hub).

In [27] the execution investigation of left shunted UPQC is examined in MATLAB/SIMULINK along with OPAL-RT for voltage drop compensation, the current harmonic compensation and the reactive compensation.

In [28] the paper shows a straightforward technique for control of series filter during unbalancing and the distorted load condition by utilizing a three-phase PLL control. Two strategies to control shunt filters have been analyzed. The improvements in power quality have been examined by comparing THD on the source side after and before installing UPQC.

In [29] the paper proposes the use of SAPF and UPQC to reduce power quality issues. SAPF is in action during the normal conditions yet when an unbalance is observed in the network. The network with DER and SAPF/UPQC has been modeled in MATLAB, the performance analysis of SAPF and UPQC is done and the results are compared.

In [30] yet another topology called Open UPQC is shown to be more adaptable toward field applications since the power quality problem is controlled by the power bills of clients.

In [31] a dynamic mathematical model of UPQC is created in MATLAB/SIMULINK and the controller has been designed for a combination of direct and non-direct loads.

In [32] the paper examines the plan and the reenactment of the Open UPQC, concentrating on the series part. The improvements in power quality have been examined by comparing THD on the source side after and before installing UPQC.

In [33] for UPQC control, SAF has been controlled through the d-q-o method-ology. This control technique requires a diminished number of estimations for source voltage as well as load current.

In [34] an improved control methodology has been proposed for three-phase UPQC under the mutilated source as well as nonlinear load condition. The proposed controller comprises a relatively indispensable and dull controller (RC) that is created on both series as well as shunt APF of UPQC for adjusting $(6n \pm 1)$ harmonics

voltages within distorted source voltages and harmonic current produced by the nonlinear loads.

In [35] a technique for prioritizing sag and swell compensations over the active filtering has been proposed. The unutilized power compensation capability of UPQC (post prioritizing sag and swell compensation) has been utilized for the harmonic filtering on the source side. Adaptive control of harmonic compensating current allows prioritizing various harmonic to be required to being compensated.

In [36] an SRF-oriented modified PAC scheme has been proposed for control of UPQC. The SRF-oriented PAC method is more robust, whereas existing ones provide only a small value of the angle and thus limit improvement in the unbalancing of the VA loadings for the series as well as shunt APF.

In [37] an original hybrid UPQC scheme has been proposed as the alternative solution toward power quality issues. Besides the novelty of this topology, reactive compensation ability has been realized within HUPQC scheme by achieving adaptive controlled DC terminal voltage. To validate the effectiveness of this topology, an experimental study has been carried out.

3 Results and Discussions

A comparative study between various topologies and their control structure has been discussed and presented in Table 1.

Table 1 Comparative analysis of different proposed techniques in the different supply system

Ref. No	Topology and Control Methodology used	Results and Conclusions
1	Twelve different topologies of UPQC were used for review	With the help of UPQC it has become possible to maintain power quality within acceptable limits
2	UPQC with passive filters	A hybrid solution was proposed to compensate voltages and currents using UPQC
3	ANPC was controlled by ANN having AI	Quality of electric power improved with the help of single-phase UPQC system
4	D-q theory with hysteretic current–voltage control for series and shunt APF	UPQC maintains the voltage constant at 700 V during occurrence of swell, sag and interruptions. The results have been verified using FFT analyses
5	D-q and p-q technique are used	Voltage and current harmonics have been reduced by introduction of UPQC

(continued)

Table 1 (continued)

Ref. No	Topology and Control Methodology used	Results and Conclusions
6	PLL controller for series and shunt APF	Three-phase PLL to control SF for unbalanced and distorted load condition
7	Novel space vector modulating technique has been used	Two optimized modulating modes along with reduction in switching losses and harmonic distortions obtained
8	Sag-based topology for phase regulation of UPQC-SPAC	Power-loss reduced, voltage sag mitigation, enhancement of the voltage stability margins were obtained
9	ANN control trained by PI is used for parallel APF	With the use of ANN controller there was substantial improvement within the response time for control of DC-link current
10	UPQC-DG having low voltage was connected at the DC-link	UPQC DC-link capacitor and the bulky power frequency transformer are eliminated in this configuration
11	Hysteretic control is used for SAF that allows ZVS	The better static as well as dynamic behavior for UPQC topology for single phase as well as for three phases has been proved
12	Space-vector switching which creates a vector of voltage to be regulated	Voltage vector had very low harmonics distortion and reduced frequency of switching. Also it has better operating performance than that of carrier-oriented PWM methods
13	Fault ride-through technique is used for wind-driven FSIG	Around 100% active power was transferred for lower levels of voltage sag
14	The PC hybrid control is used for UPQC-Q along with SERC and SHUC	For long durations voltage sag VA consumption for UPQC-Q is minimal. It is also capable in maintaining harmonics isolation within utility and the load
15	UPQC is used to upgrade the LVRT ability of the DFIG-based WECS	The results depicted fast recovery for rotor, stator, and the DC-link current to their final steady-state value. Results depicted better performance and validation of WECS
16	The low-frequency model d-q state-space technology is used	The control strategy for voltages and currents compensation as well as energy management scheme of super capacitors is given
17	SRFcontrol is used for state-space modeling of MC-UPQC	The power quality is improved in three-phase system operating under various load considerations

(continued)

Table 1 (continued)

Ref. No	Topology and Control Methodology used	Results and Conclusions
18	UPQC model involving on control of phase (UPQC-PAC) has been used	The optimal allocation for UPQC-PAC resulted in significant reduction in power loss, during voltage mitigation as well as enhancement in stability margin of voltage
19	Control of phase (UPQC-PAC) is utilized	Load flow solution including UPQC-PAC has been devised. Performance with proposed solution is validated by different cases
20	Extended UPQC-PAC scheme known as UPQC-S has been used	Multifunctionality of the series inverter for compensating voltage changes, utilization of ratings of series converter and reduction in ratings of shunt converter was obtained
21	A new scheme known as OPEN-UPQC has been demonstrated	OPEN UPQC improves power quality and cost reduction in case of need of better quality power with a good device performance
22	MC-UPQC is used with one shunt VSC, at least two and more than two series VSC	Protects critical and sensitive load for distortion, swell, sag and interruptions in two-feeder system and compensation of interruption without having requirement of a BESS
23	UPQC with fuel cell using “d-q theory” and hysteretic current controller is used	Fuel-cell-powered UPQC was able to maintain voltage constant at 700 V when swell, sag and interruption occurred
24	Data-driven controller (DDC) is used as a variable in the PAC technique	DDC controller was developed using variable phase control in the PAC method, which enhances the overall power quality in the UPQC system
25	PAF control uses p-q theory as well as the symmetrical components while for SAF control symmetrical components are used	The issue of generating three-phase unbalance reference current for expanded p-q theory had been solved. The validity of control strategy has been studied by means of simulation results
26	Shunt-APF with a series-APF is used by including series VSC with each feeder	By adding the number of APFs to each feeder advanced UPQC was able to achieve different mitigating techniques
27	UPQC combined with PV generators is used	This allowed fine accuracy for extracting fundamental component in load current instead of compromising dynamic performance

(continued)

Table 1 (continued)

Ref. No	Topology and Control Methodology used	Results and Conclusions
28	UPQC with left shunted UPQC has been used	Left shunted UPQC rating was much less when compared with a right shunted UPQC
29	Three-phase PLL control for shunt APF and series APF has been used	Improvements in power quality were analyzed through comparison of THD of the source side after and before installing UPQC
30	SAPF with UPQC has been used in combined topology	Eliminates harmonics to a greater extent, and mitigates the voltage unbalance problems when a disturbance occurs in the system
31	A new scheme known as OPEN-UPQC has been demonstrated	Open UPQC possesses better compensating characteristics for both steady state as well as transient condition of system disturbance
32	UPQC topology has been applied to non-direct loads	Results obtained depicted effectiveness of UPQC. THD of system reduced to 5.25% from 21.40%
33	A new scheme known as OPEN-UPQC has been demonstrated	Series control logic has been presented and different simulations for three power quality issues have been discussed
34	UPQC with SAF controlled through the d-q-o methodology	Different UPQC topologies enhanced performance and capabilities. This system is effective to address power quality problems
35	Relatively robust and a dull controller (RC) that is created on both shunt as well as series APF of UPQC	The proposed RC gives much faster real-time response compared with traditional RC and far better steady-state characteristics while also providing fast dynamic results against load variation
36	Adaptive control of harmonic compensating current allowed prioritizing various harmonic to be compensated	A technique for prioritizing sag and swell compensation over the active filtering has been proposed. It helps in the case where passively tuned filter is installed for the harmonic frequencies against which UPQC need not respond
37	SRF-oriented modified PAC scheme has been proposed for control of UPQC	The PAC method decreased unbalance in the VA loadings of series as well as shunt APF of UPQC through sharing the reactive power among two APF
38	HUPQC which consists of shunt hybrid APF (SHAPF), the DVR, and bidirectional DC converter	SHAPF enables decrease in voltage ratings of DC capacitor, reduction in price and size and rating of DC capacitance, and thus reduces the switching loss in VSC. The DC-link voltage has been adaptively controlled

4 Discussions

The paper discussed and compared the various strategies in different supply systems used by many researchers to reduce sag in voltage as well as swell in voltage and to reduce harmonics. This review paper also explained different control techniques and topologies to control voltage and current injections by UPQC and further compare with the other series or shunt FACTS power devices. The paper presents the use of UPQC to mitigate simultaneously voltage and current-related issues in the power system. UPQC makes use of the series and shunt active filters, where the series active filter is used to meet voltage-linked issues and the shunt active filter is used to meet the current-related issues.

5 Conclusions

UPQC has been concluded as a very powerful device for series compensation as well as shunt compensation in the power system. The control strategies discussed for UPQC was analyzed for various kinds of power distribution systems, such as low-voltage distribution and medium-voltage-level distribution system. It has been concluded that UPQC is one of the very powerful, robust and reliable compensating FACTS power devices among all other compensating devices. The UPQC has been found to have a superb capability in improving the voltage profile and thus reducing the overall harmonic content at the end-user level.

References

1. Khadkikar V (2012) Enhancing Electric Power Quality Using UPQC: A Comprehensive Overview. IEEE Trans Power Electron 27(5):2284–2297
2. Naderipour, Voltage and current compensation in dispersed generation systems A (2010) First Power Quality Conferance. Tehran 2010:1–4
3. R. Barriviera, E.J. Acordi and R. Q. Machado, "Selective compensation of harmonics applied in a multilevel single-phase UPQC system," 2017 Brazilian Power Electronics Conference (COBEP), Juiz de Fora, 2017, pp. 1–6
4. Samal S, Hota PK, Barik PK, Fuel cell integrated UPQC System for power quality improvement (2017) IEEE Calcutta Conference (CALCON). Kolkata 2017:325–330
5. Krishna D, Sasikala M, Ganesh V (2017) Mathematical modeling and simulation of UPQC in distributed power systems," 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE), Karur, 2017, pp.1–5.
6. S. Ivanov, M. Ciontu, D.Sacerdotianu and A. Radu, "Simple control strategy of the series filter within a unified power quality conditioner," 2017 International Conference on Electromechanical and Power Systems (SIELMEN), Iasi, 2017, pp.427–430.
7. Lu Y, Xiao G, Wang X, Blaabjerg F, Lu D (April 2016) Control Strategy for Single-Phase Transformerless Three-Leg Unified Power Quality Conditioner Based on Space Vector Modulation. IEEE Trans Power Electron 31(4):2840–2849

8. Ganguly S (Aug. 2014) Impact of Unified Power-Quality Conditioner Allocation on Line Loading, Losses, and Voltage Stability of Radial Distribution Systems. IEEE Trans Power Delivery 29(4):1859–1867
9. Kinhal VG, Agarwal P, Gupta HO (Jan. 2011) Performance Investigation of Neural-Network-Based Unified Power-Quality Conditioner. IEEE Trans Power Delivery 26(1):431–437
10. A. Mokhtarpour, M. Bathaee, H. A. Shayanfar, "Power quality compensation in smart grids with a single phase UPQC-DG," in Smart Grids (ICSG), 2012 2nd Iranian Conference on Smart Grid.
11. da Silva SAO, Negrão FA (May 2018) Single-Phase to Three- Phase Unified Power Quality Conditioner Applied in Single- Wire Earth Return Electric Power Distribution Grids. IEEE Trans Power Electron 33(5):3950–3960
12. Rodriguez J, Moran L, Correa P, Silva C (Aug 2002) A vector control technique for medium-voltage multilevel inverters. IEEE Trans Industr Electron 49(4):882–888
13. Jayanti NG, Basu M, Conlon MF, Gaughan K (June 2009) Rating requirements of the unified power quality conditioner to integrate the fixedspeed induction generator-type wind generation to the grid. IET Renew Power Gener 3(2):133–143
14. Basu M, Das SP, Dubey GK (May 2008) Investigation on the performance of UPQC-Q for voltage sag mitigation and power quality improvement at a critical load point. IET Gener Transm Distrib 2(3):414–423
15. Saleh MA, Eskander MN, Amer S, Nashed Enhancing the LVRT capability of grid connected wind energy conversion system using Unified Power Quality Controller MNF (2014) IEEE Conference and Expo Transportation Electrification Asia-Pacific (ITEC Asia-Pacific). Beijing 2014:1–5
16. Z. Hailiang, W. Jianru, Y. Chenhu and L. Shuchao, "Analytical research on unified power quality conditioner based on super capacitors energy storage system," 2011 4th International Conference on Power Electronics Systems and Applications, Hong Kong, 2011, pp.1–5.
17. Naidu PV, Basavaraja B, Design of a SRF based MC UPQC used for load voltage control in Parallel distribution systems (2012) IEEE Fifth Power India Conference. Murthal 2012:1–6
18. S. Ganguly, "Unified power quality conditioner allocation for reactive power compensation of radial distribution networks," in IET Generation, Transmission & Distribution, vol. 8, no. 8, pp. 1418–1429, 8 2014.
19. Ganguly S (July 2014) Multi-Objective Planning for Reactive Power Compensation of Radial Distribution Networks With Unified Power Quality Conditioner Allocation Using Particle Swarm Optimization. IEEE Trans Power Syst 29(4):1801–1810
20. Khadkikar V, Chandra A (Sept. 2011) UPQC-S: A Novel Concept of Simultaneous Voltage Sag/Swell and Load Reactive Power Compensations Utilizing Series Inverter of UPQC. IEEE Trans Power Electron 26(9):2414–2425
21. Brenna M, Faranda R, Tironi E (Oct. 2009) A New Proposal for Power Quality and Custom Power Improvement: OPEN UPQC. IEEE Trans Power Delivery 24(4):2107–2116
22. Mohammadi HR, Varjani AY, Mokhtari H (July 2009) Multiconverter Unified Power-Quality Conditioning System: MC-UPQC. IEEE Trans Power Delivery 24(3):1679–1686
23. J. Ye and H. Beng Gooi, ", Optimization of the size of UPQC system based on data-driven control design,," (2017) IEEE Power & Energy Society General Meeting. Chicago, IL 2017:1–1
24. M. T. Haque, T. Ise and S. H. Hosseini, "A novel control strategy for unified power quality conditioner (UPQC)," 2002 IEEE 33rd Annual IEEE Power Electronics Specialists Conference. Proceedings (Cat. No.02CH37289), 2002, pp. 94–98 vol.1.
25. D. Krishna, M. Sasikala and V. Ganesh, "Mathematical modeling and simulation of UPQC in distributed power systems," 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE), Karur,2017,pp.1–5.
26. S. Devassy and B. Singh, "Control of solar photovoltaic integrated UPQC operating in polluted utility conditions," in IET Power Electronics, vol. 10, no. 12, pp. 1413–1421, 10 6 2017.
27. N. Patnaik, A. K. Panda and P. R. Mohanty, "Performance and comparative rating evaluation of single phase left shunt UPQC," 2016 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Trivandrum, 2016, pp. 1–6.

28. S. Ivanov, M. Ciontu, D. Sacerdotianu and A. Radu, "Simple control strategies of the active filters within a unified power quality conditioner (UPQC)," 2017 International Conference on Modern Power Systems (MPS), Cluj-Napoca, 2017,pp.1–4.
29. P. K. Nair and P. Reji, "Power quality improvement in distribution networks containing distributed energy resources using unified power quality conditioner," 2015 International Conference on Power, Instrumentation, Control and Computing (PICC),Thrissur, 2015, pp. 1–7.
30. Kotturu J, Agarwal P, Comparative performance analysis of UPQC and Open UPQC (2015) Annual IEEE India Conference (INDICON). NewDelhi 2015:1–6
31. A. Shrivstava and P. Nene, "Power Quality Enhancement Using UPQC connected with PV Arrays," 2015 Fifth International Conference on Communication Systems and Network Technologies,Gwalior,2015,pp.1232–1237.
32. D'Antona G, Faranda R, Hafezi H, Accetta G, Della Giustina D, Open UPQC: A possible solution for power quality. Series unit analysis (2014) International Symposium on Power Electronics. Electrical Drives, Automation and Motion, Ischia 2014:1104–1109
33. Kwan KH, Tan KT, So PL, Kwan KH, An unified power quality conditioner for load sharing and power quality improvement (2012) Asia-Pacific Symposium on Electromagnetic Compatibility. Singapore 2012:963–967
34. Trinh QN, Lee HH, A repetitive control scheme to improve performance of UPQC under distorted source and nonlinear load conditions (2012) 10th International Power & Energy Conference (IPEC). Ho Chi Minh City 2012:418–423
35. S. R. Choudhury, A. Das, S. Anand, S. Tungare and Y. Sonawane, "Adaptive shunt filtering control of UPQC for increased nonlinear loads," in IET Power Electronics, vol. 12, no. 2, pp. 330–336, 20 2 2019.
36. A. Patel, H. D. Mathur and S. Bhanot, "Enhancing VA sharing between the shunt and series APFs of UPQC with a modified SRF-PAC method," in IET Power Electronics, vol. 13, no. 2, pp. 275–285, 5 2 2020.
37. Koroglu T, Tan A, Savrun MM, Cuma MU, Bayindir KC, Tumay M (2020) Implementation of a Novel Hybrid UPQC Topology Endowed With an Isolated Bidirectional DC–DC Converter at DC link. IEEE Journal of Emerging and Selected Topics in Power Electronics 8(3):2733–2746

Multi Linear Regression Model to Detect Distributed Denial of Service Attacks in Cloud Environments

Swathi Sambangi and Lakshmeeswari Gondi

Abstract DDoS attacks are the attacks carried in distributed environments. For this to be achieved attackers usually stagger the network bandwidth, halt system resources thus causing denial of access for legitimate users. The problem of identifying Distributed Denial of Service attacks in cloud environments is a significantly challenging research problem because of the computational complexity issues that are to be addressed. In this paper, the research objective is to study the problem of DDoS attack detection in cloud environment by considering CICIDS 2017 benchmark dataset and build a multiple regression-based machine learning model. In addition to evaluating the MAPE and prediction accuracy values of the multiple regression model, visualizations of residual plots and fit charts of the model are also studied.

Keywords DDoS attack · Fit chart · Traffic packet · Classification · Prediction

1 Introduction

DDoS attacks are the attacks carried in distributed environments. For this to be achieved attackers usually stagger the network bandwidth, halt system resources, thus causing denial of access to legitimate users. The problem of identifying Distributed Denial of Service attacks in cloud environments is a significantly challenging research problem because of the computational complexity issues that are to be addressed. In this paper, the research objective is to study the problem of DDoS attack detection in cloud environment by considering CICIDS 2017 benchmark dataset and

S. Sambangi · L. Gondi (✉)
Gitam (Deemed To Be University), GITAM Institute of Technology, Vishakhapatnam, India
e-mail: lgondi@gitam.edu

S. Sambangi
e-mail: swathswathi_s@vnrvjiet.in

S. Sambangi
VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_48

build a multiple regression-based machine learning model. In addition to evaluating the MAPE and prediction accuracy values of the multiple regression model, visualizations of residual plots and fit charts of the model are also studied.

Some of the recent research interests from cloud computing that have received prime importance are DDoS attack prevention, detection and mitigation. Among these issues, DDoS attack detection had gained prominent importance from research community and industry. Researchers across the globe are working on coming out with various methods and approaches to address DDoS attack detection in computationally optimal way. But these contributions could not put a stop to DDoS attacks. Unfortunately, even today the deployment of these available methods could not resist DDoS attacks thus affecting the cloud environments. In fact, DDoS attacks are increasing with each passing day, both interms of attack frequency and attack size. One of the most common reasons for the failure to address this issue is from having no consensus among various end points in a distributed internet network as one cannot enforce cooperation globally. Other reasons include socioeconomic factors that makes the global cooperation enforcement difficult and nature of DDoS attacks. Hence, there is an emerging immediate need to properly study, identify and figure out the reasons of failure of the available methods in the research literature by revisiting the fundamental techniques, by carrying detailed analysis for failure of existing approaches.

This paper is outlined as follows. Section 2 outlines the need for DDoS attack detection and few related works. Section 3 outlines the proposed multiple linear regression model. Section 4 outlines the experimental analysis in terms of MAPE and Accuracy of the prediction model and it also shows the visualization of parameters considered in terms of residuals, residual plot for the model and fit chart of the prediction model. Section 5 concludes the research contributions of this paper.

2 Related Literature

In this section, we first discuss some of the recent significant contributions that have addressed detection of DoS, U2R, R2L, Probe attacks in a network using KDD and NSL-KDD datasets and then outline some of the contributions that addressed DDoS attacks.

Gunupudi et al. [1–4] proposed gaussian based measure for anomaly detection and achieved promising results with the experimentation over NSL KDD and KDD datasets. Kumar et.al [5] presents a novel membership function, and achieved very good performance over the detection of low frequency attacks such as User2Root and Remote2Local along with DoS attacks. The authors achieved 99.8% for DoS, with a threshold of 0.9999 and 99.5% accuracy with a threshold of 0.999999. The authors in their research compared and correlated the results with CLAPP [6]. In CLAPP the accuracy of DoS attacks was 99.8 and the same detection accuracy of 99.8% is achieved with this approach [6] along with excellent performance of low frequency attack detection. Mangathayaru et al. [8], Nagaraja et al. [8] Gunupudi et al.

[10] and Aljawarneh [11] also contributed in this research and achieved promising results. Arun et al. [9] performed experimentation over the NSL KDD and KDD, considering 19 attributes and 41 attributes that includes DoS Attacks along with four other attributes. KDD data set has 4,94,021 instances out of which 3,91,418 instances were DoS attacks. The authors achieved 99.9% detection accuracy for DoS attacks using J48 classifier. The dataset used in this research is motivated from the research by Iman Sharafaldin et al. [12]. It consisted of 5-day traffic log files which has normal and attack instances. One of the most recent significant works in detection of DDoS attacks is by Kurniabudi et al. [13] which carried dataset feature analysis for the CICIDS 2017 dataset by applying information gain concept for retrieval of significant features for anomaly detection. The present research is motivated from this contribution [13].

3 Multiple Linear Regression Model—DDoS Attack Detection

The basic idea behind this research study is to show the importance of regression model-based prediction. The idea is to build a prediction model by applying the multiple linear regression technique which is one of the important statistical modeling and machine learning technique. Figure 1 shows the prediction model which is based on regression analysis. The first element of the prediction model is the traffic data which is preferably captured in a distributed environment. We can capture the traffic data either using a virtual environment setup or from a distributed real time environment. In this paper, we restrict our research study to analysis of regression model through using publicly available CICIDS 2017 benchmark dataset.

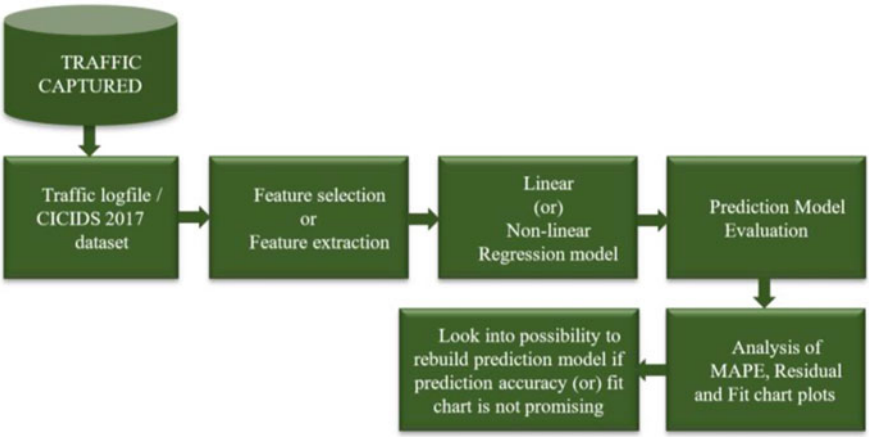


Fig. 1 Multiple linear regression model for DDoS attack detection

The benchmark dataset which is considered has 78 attributes and 1 class label. We choose to restrict the regression model by preferring the top 16 attributes out of the 22 chosen attributes of the dataset [12]. For obtaining top 16 attributes, we apply feature selection technique such as Information Gain based feature selection and then list out the top 16 attributes with higher information gain when compared to information gain of remaining attributes. By using these 16 attributes, we build a multiple linear regression (MLR) model. The model is studied by comparing R^2, adjusted R^2 values. The prediction model obtained is used to study the output evaluation parameters such as MAPE, ACCURACY and visualize residual plots for each attribute, residual plot of the model along with the fit chart of the prediction model. Data visualization of fit chart helps us to know the performance of the model irrespective of prediction accuracy. A prediction model is considered to be a better model, if fit chart of the prediction model is such that the prediction label and actual label cannot be differentiated visually. Residual plots of each attribute can also be plotted to visually study the behavior of attributes. The overall residual plot also helps to understand the behavior of the prediction model. Analytically, we can evaluate MAPE, ACCURACY values of the prediction model and figure out possibility of redesigning a better regression model based on adjusted R^2 value.

4 Experimental Analysis and Key Findings

In this section, the benchmark dataset used for experimental analysis and development of prediction model is described. Also, the experimental findings of the multiple linear regression model are studied and presented.

4.1 Dataset Description Used for Experimental Analysis

CICIDS 2017 Dataset consists of different attack profiles generated in a distributed environment. In this paper, the DDoS attack profiles that were captured on Friday afternoon are considered. The pcap file of Friday afternoon DDoS attack profile consisted of 78 attributes for which we have to apply feature extraction so that high mutual information attributes are preferred. Information gain can be applied to the generated log file so that, we can select the highest ranked attributes to consider as input for designing MLR model. But some of the attributes namely Flow Bytes/s, Flow Packets/s have non numeric values like ‘infinity’, ‘NaN’. Other set of attributes such as Init_Win_bytes_backward have negative values. As information gain cannot be applied on negative values and non-numeric values, we replaced the mentioned records with suitable values during preprocessing.

After the replacement, we carried feature extraction using information gain approach that results ranked attributes. All the 78 attributes are ranked according to their significance. We selected the top 16 Attributes as input for design of MLR

Table 1 List of selected attributes obtained using information gain computation

S. No	Selected attribute name	S. No	Selected attribute name
1	Flow Duration	9	Fwd IAT Min
2	Total Length of Fwd Packets	10	Bwd IAT Mean
3	Bwd Packet Length Min	11	Fwd PSH Flags
4	Bwd Packet Length Std	12	Average Packet Size
5	Flow IAT Mean	13	Subflow Fwd Bytes
6	Flow IAT Std	14	Init_Win_bytes_forward
7	Flow IAT Min	15	Init_Win_bytes_backward
8	Fwd. IAT Mean	16	Active Min

model. The list of 16 attributes considered for experimentation and building model are as depicted in Table 1

4.2 Result Analysis

As mentioned in the previous section the dataset chosen for experimentation consisted of five-day log records from Monday to Friday in csv format. For experiment analysis, we have considered the log fie of Friday afternoon which also consisted of two class labels. The class labels are Benign (Normal) and DDoS (attack). The total number of traffic packets in the log file considered for experimental analysis included 2,25,745 traffic packets out of which normal or benign traffic traces are 97,718 and DDoS Attack traces are 1,28,027. Figure 2 depicts the Regression statistics for CICIDS 2017 Friday Afternoon DDoS Traffic Log file.

From Fig. 2, it can be observed that the R2 value is 0.819265 and adjusted R2 value is 0.819249. This means that incase if we think of reducing the number of attributes for obtaining a better prediction model then, it may not help in improving the current prediction accuracy of the model. Also, the p-value of 14 out of 16 attributes is less than 0.05 which means that the attribute coefficients obtained are statistically significant. However, if we see the p-value for second and third attributes (Total length of Fwd packets and Bwd packet length min) they are non-numeric. This means that there is some degree of insignificance w.r.t these attributes. Although this is not evident directly from the p-value of these attributes, it is possible to understand the insignificance of these attributes if we can refer to the residual plot (or) line fit plot of these two attributes. For example, consider the residual plot shown in Fig. 3 which depicts the individual residual plot for each of the 16 attributes considered to design the multiple linear regression (MLR) model. In the residual plot shown in the Fig. 3, it is evident that there is some pattern showing heteroscedasticity property, i.e. the residual plot is showing say, an increasing trend pattern.

Regression Statistics	
Multiple R	0.905132929
R Square	0.819265619
Adjusted R Square	0.819249179
Standard Error	0.210647209
Observations	225745

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	16	45402.80959	2837.675599	68215.02648	0
Residual	225729	10016.10284	0.044372247		
Total	225745	55418.91243			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.050678541	0.001382296	760.0966481	0	1.047969276	1.053387806	1.047969276	1.053387806
Flow Duration	-3.56778E-10	3.82364E-11	-9.330832248	1.05948E-20	-4.3172E-10	-2.81835E-10	-4.3172E-10	-2.81835E-10
Total Length of Fwd Packets	0	0	65535	#NUM!	0	0	0	0
Bwd Packet Length Min	-0.000227078	1.21457E-05	-18.69612355	#NUM!	-0.000250883	-0.000203273	-0.000250883	-0.000203273
Bwd Packet Length Std	2.13372E-05	6.36171E-07	33.53995321	5.1124E-246	2.00903E-05	2.2584E-05	2.00903E-05	2.2584E-05
Flow IAT Mean	3.24331E-08	6.36847E-10	50.92756079	0	3.11849E-08	3.36813E-08	3.11849E-08	3.36813E-08
Flow IAT Std	1.23309E-08	2.66791E-10	46.21939437	0	1.1808E-08	1.28538E-08	1.1808E-08	1.28538E-08
Flow IAT Min	-3.64306E-09	2.23887E-10	-16.27189384	1.68953E-59	-4.08187E-09	-3.20425E-09	-4.08187E-09	-3.20425E-09
Fwd IAT Mean	-1.23141E-08	4.01923E-10	-30.63791119	1.0154E-205	-1.31018E-08	-1.15263E-08	-1.31018E-08	-1.15263E-08
Fwd IAT Min	1.80723E-09	3.05072E-10	5.923966733	3.1472E-09	1.2093E-09	2.40517E-09	1.2093E-09	2.40517E-09
Bwd IAT Mean	-8.29991E-09	2.1754E-10	-38.15343847	0	-8.72629E-09	-7.87354E-09	-8.72629E-09	-7.87354E-09
Fwd PSH Flags	-0.290295787	0.002602345	-111.5516254	0	-0.295396316	-0.285195258	-0.295396316	-0.285195258
Average Packet Size	0.000638376	1.88848E-06	338.0375055	0	0.000634674	0.000642077	0.000634674	0.000642077
Subflow Fwd Bytes	-4.89843E-05	2.0511E-07	-238.8198078	0	-4.93863E-05	-4.85823E-05	-4.93863E-05	-4.85823E-05
Init_Win_bytes_forward	-1.20854E-05	2.93645E-08	-411.5651522	0	-1.21429E-05	-1.20278E-05	-1.21429E-05	-1.20278E-05
Init_Win_bytes_backward	1.11243E-05	2.39641E-08	464.2064772	0	1.10773E-05	1.11713E-05	1.10773E-05	1.11713E-05
Active Min	-5.57051E-09	6.41432E-10	-8.684492926	3.82903E-18	-6.8277E-09	-4.31332E-09	-6.8277E-09	-4.31332E-09

Fig. 2 Regression Statistics for CICIDS 2017 Friday Afternoon DDoS Traffic Log file

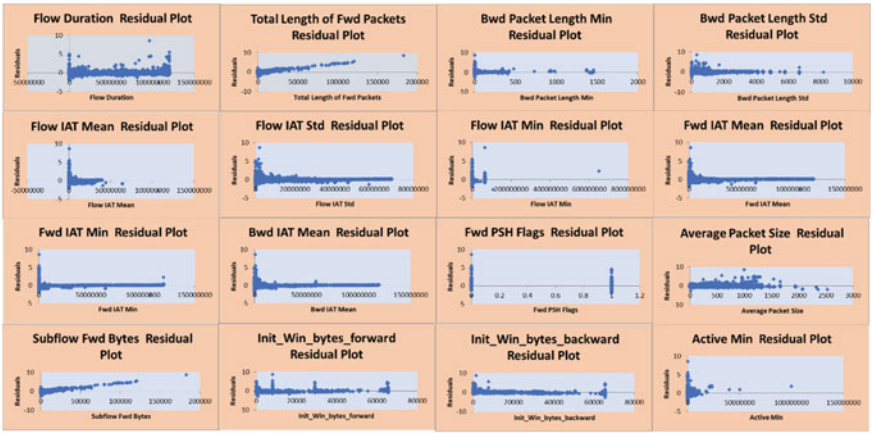


Fig. 3 Individual Residual plot for each of 16 Attributes considered to design MLR Model

Similarly, another finding which can be made is through considering the line fit plot of the Fwd Packets Line fit plot shown in Fig. 4. Here, Fig. 4 depicts individual Line Fit plot of 12 attributes out of 16 attributes which are considered to design MLR Model. It is visible that the predicted label and actual label can be easily differentiated. If the actual label and predicted label both overlap each other then, this means that the prediction is a better one. But in this case, there is some difference visible due to differentiable overlap. This means that this attribute is sure to create

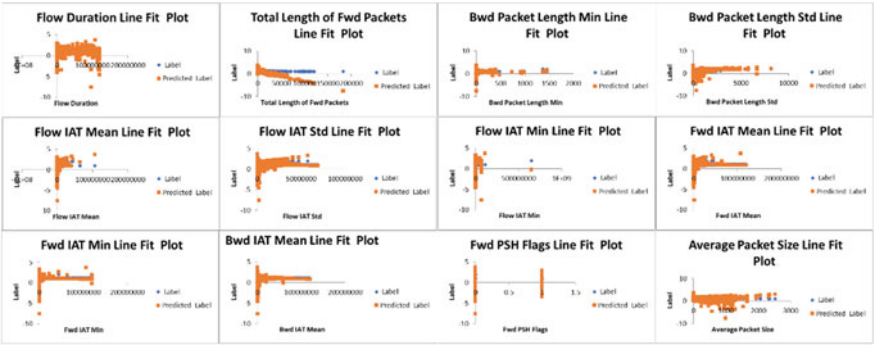


Fig. 4 Individual Line Fit plot for each of 12 Attributes considered to design MLR Model

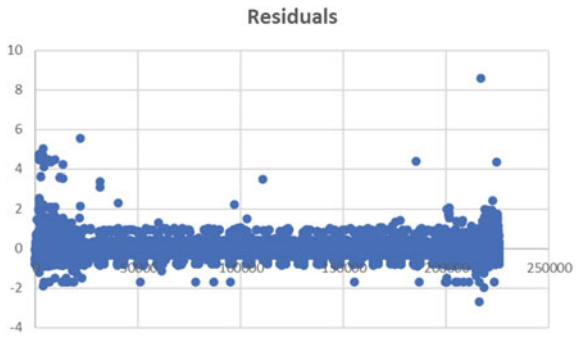


Fig. 5 Overall Residual plot for 16 Attributes considered to design MLR Model

some problem during the prediction task. Similar case is w.r.t the other attribute Bwd Packet Length Min.

Figure 5 depicts the residual plot of the overall prediction model by considering all the sixteen attributes mentioned in the Table 1. It is visible from the residual plot shown in Fig. 5 that there is no specific pattern w.r.t plotted residuals. So, we may consider this as a good model. This can be evident if we visualize the fit chart for the prediction model.

Figure 6 shows visualization of the fit chart for the MLR prediction model. It can be seen that the predicted label and actual label are almost overlapped. There are minor instances of non-overlapping of actual and predicted labels of the obtained prediction model. In fact, fit chart is one of the visualization techniques which we may use to decide the importance of the prediction model.

In this case, we may consider this model as a good model as there is sufficient overlap. In contrast to visualization, alternately we can analytically obtain the mean absolute percentage error and prediction accuracy by considering the regression equation which is function of statistically significant regression coefficients and the intercept.

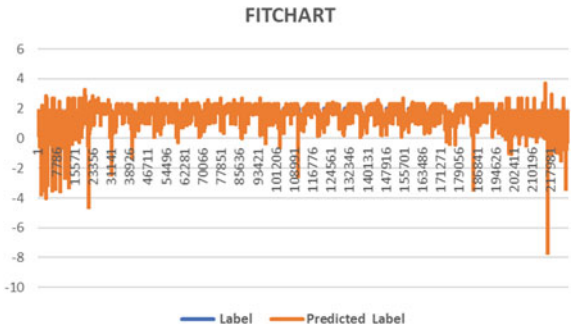


Fig. 6 Overall Fit chart plot for 16 Attributes considered to design MLR Model

In the present case, by considering the regression coefficients and the intercept values, the prediction accuracy is obtained as 86.38%. The mean absolute percentage error (MAPE) is obtained as 0.1362, i.e. 13.62%. This means that the model is a good model if the acceptable prediction accuracy is 87% with an error of 13%. In this case, the error is 13%. This is how, we can apply the regression analysis to study the behavior of the prediction model. From regression statistics of Fig. 7, it can be observed that the R2 value is 0.6445 and adjusted R2 value is nearly the same. This value is less than the R2 value obtained after preprocessing. This shows that the initial preprocessing of handling negative values appropriately has improved the accuracy.

Regression Statistics	
Multiple R	0.802858184
R Square	0.644581263
Adjusted R Square	0.644556071
Standard Error	0.295397246
Observations	225745

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	16	35721.99259	2232.624537	25586.0244	0
Residual	225728	19696.91983	0.087259533		
Total	225744	55418.91243			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.357865487	0.001260334	283.9450126	0	0.355395265	0.360335709	0.355395265	0.360335709
Flow Duration	-6.02702E-09	5.53794E-11	-108.8315552	0	-6.13557E-09	-5.91848E-09	-6.13557E-09	-5.91848E-09
Total Length of Fwd Packet	-4.62881E-05	2.93871E-07	-157.5116968	0	-4.6864E-05	-4.57121E-05	-4.6864E-05	-4.57121E-05
Bwd Packet Length Min	-0.002203188	1.34332E-05	-164.0111516	0	-0.002229517	-0.002176859	-0.002229517	-0.002176859
Bwd Packet Length Std	2.2365E-05	8.93075E-07	25.04263145	3.2487E-138	2.06145E-05	2.41154E-05	2.06145E-05	2.41154E-05
Flow IAT Mean	2.69597E-07	1.37021E-09	196.7557348	0	2.66911E-07	2.72282E-07	2.66911E-07	2.72282E-07
Flow IAT Std	-6.87091E-08	5.49112E-10	-125.127628	0	-6.97854E-08	-6.76329E-08	-6.97854E-08	-6.76329E-08
Flow IAT Min	-2.50531E-07	1.49868E-09	-167.1674085	0	-2.53468E-07	-2.47593E-07	-2.53468E-07	-2.47593E-07
Fwd IAT Mean	3.48228E-08	5.69508E-10	61.14533371	0	3.37065E-08	3.5939E-08	3.37065E-08	3.5939E-08
Fwd IAT Min	-4.72463E-08	4.77705E-10	-98.90255913	0	-4.81825E-08	-4.631E-08	-4.81825E-08	-4.631E-08
Bwd IAT Mean	-3.45909E-08	3.15393E-10	-109.6754278	0	-3.5209E-08	-3.39727E-08	-3.5209E-08	-3.39727E-08
Fwd PSH Flags	-0.332739117	0.003631252	-91.63206837	0	-0.339856278	-0.325621956	-0.339856278	-0.325621956
Average Packet Size	0.000397867	2.44607E-06	162.6555401	0	0.000393072	0.000402661	0.000393072	0.000402661
Subflow Fwd Packets	0.001787293	4.38643E-05	40.74597807	0	0.00170132	0.001873266	0.00170132	0.001873266
Init_Win_bytes_forward	-8.47002E-06	8.98975E-08	-94.2186516	0	-8.64622E-06	-8.29383E-06	-8.64622E-06	-8.29383E-06
Init_Win_bytes_backward	-7.20996E-06	1.47641E-07	-48.83426762	0	-7.49933E-06	-6.92059E-06	-7.49933E-06	-6.92059E-06
Active Min	5.38464E-09	9.10778E-10	5.912135864	3.38183E-09	3.59954E-09	7.16974E-09	3.59954E-09	7.16974E-09

Fig. 7 Regression statistics for friday afternoon log file without preprocessing

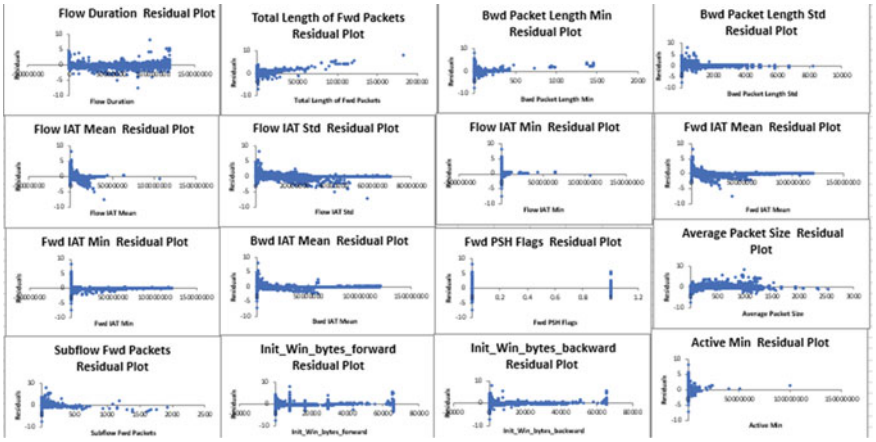
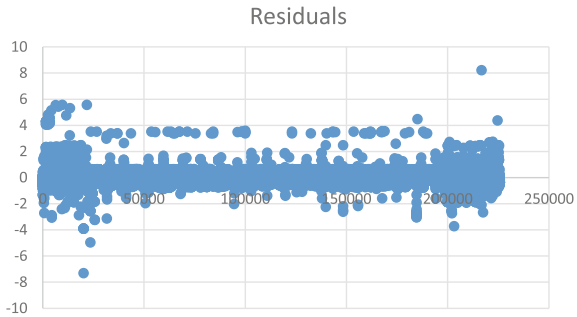


Fig. 8 Individual residual plot for each of 16 Attributes (without preprocessing)

Fig. 9 Overall residual plot for 16 attributes (without processing)



This is one of the important finding in this contribution. Figures 8 and 9 shows the individual residual plot for each of 16 Attributes (without preprocessing) obtained for the MLR Model and overall residual plot for prediction model respectively. Figure 10 shows the fit chart of the MLR model obtained without handling negative values in the dataset appropriately. It is visible from the fit chart that there is a clear distinction between actual and predicted labels and hence the reduce in prediction accuracy is

Fig. 10 Overall fit chart plot for 16 Attributes (without preprocessing)

seen. For this case, the prediction accuracy is obtained as 77% which is lesser than the former. This shows the importance of initial preprocessing required for feature handling in obtaining a better accuracy rate. The importance of fit chart can be understood in making performance analysis of the prediction model.

5 Conclusions

In a distributed environment like cloud, DDOS attacks have become more common, this makes it as an essential need to early detect these attacks which shall otherwise cause the service unavailability in cloud. In order to identify these attacks, we may apply machine learning techniques. Alternately, this task can also be achieved by using regression analysis which is one of the important statistical analysis-based prediction models. The main research objective behind the present study is not only to build a regression-based prediction model that is ensemble of feature selection using information gain and regression analysis but also to highlight and show the importance of data visualization in analysis of prediction model. For analyzing, the prediction model performance, we have used the most popularly known CICIDS 2017 dataset that is publicly available for researchers. In this dataset, we restricted the study to specifically Friday afternoon logfile. It has been observed that through this ensemble model for Friday afternoon log file the prediction accuracy is obtained as 86.38% for 16 attributes obtained through information gain-based feature selection and regression analysis-based ML model. This research thus shows importance of visualization and regression analysis in building a ML model. We have considered visualizations such as fit chart and residual plots to study the performance of the model. In this work, we have limited our analysis for one day log file and in future this research may be extended to consider all traffic log files of 5 days and come out with a consensus-based machine learning model. The important finding from this contribution is that there is a wide scope for improving prediction accuracy of the regression model by suitably carrying feature analysis and extraction. In future, this research can be extended to propose novel feature extraction or selection approaches to improve the prediction accuracy of the prediction model to detect DDoS attacks in cloud environments.

References

1. Kumar GR, Mangathayaru N, Narsimha G (2016) An approach for intrusion detection using fuzzy feature clustering," 2016 International Conference on Engineering & MIS (ICEMIS), Agadir. pp. 1–8. https://doi.org/10.1109/ICEMIS.2016.7745345
2. Gunupudi Rajesh Kumar N, Mangathayaru N, Narasimha G (2015) An approach for Intrusion Detection using Text Mining Techniques. In Proceedings of the The International Conference on Engineering & MIS 2015 (ICEMIS '15). ACM, New York, NY, USA, Article 63 , 6 pages. https://doi.org/10.1145/2832987.2833076

3. Gunupudi Rajesh K, Nimmala Mangathayaru, Gugulothu Narsimha, and Aravind Cheruvu (2018) Feature Clustering for Anomaly Detection Using Improved Fuzzy Membership Function. In Proceedings of the Fourth International Conference on Engineering & MIS 2018 (ICEMIS '18). ACM, New York, NY, USA, Article 35, 9 p
4. Kumar G, Mangathayaru N, Narsimha G, Reddy GK (2017) Evolutionary approach for intrusion detection. In: 2017 International Conference on Engineering & MIS (ICEMIS), Monastir, 2017, pp. 1–6. https://doi.org/10.1109/ICEMIS.2017.8273116
5. Kumar GR, Gugulothu N, Nimmala M (2019) An Evolutionary Feature Clustering Approach for Anomaly Detection Using Improved Fuzzy Membership Function: Feature Clustering Approach for Anomaly Detection. International Journal of Information Technology and Web Engineering (IJITWE) 14(4):19–49. https://doi.org/10.4018/IJITWE.2019100102
6. Rajesh Kumar G (2017) Mangathayaru Nimmala, Narsimha Gugulothu, Suresh Reddy Gali, CLAPP: A self constructing feature clustering approach for anomaly detection. Futur Gener Comput Syst 74:417–429
7. Mangathayaru N, Kumar GR, Narsimha G (2016) Text mining based approach for intrusion detection. In: 2016 International Conference on Engineering & MIS (ICEMIS), Agadir, 2016, pp. 1–5. https://doi.org/10.1109/ICEMIS.2016.7745351
8. Nagaraja A, Uma B, Gunupudi RK (2019) UTTAMA: An Intrusion Detection System Based on Feature Clustering and Feature Transformation. Found Sci
9. Nagaraja U, Boregowda K, Khatatneh R, Vangipuram, Nuvvusetty R, Sravan Kiran V (2020) Similarity Based Feature Transformation for Network Anomaly Detection. In IEEE Access, vol. 8, pp. 39184–39196. https://doi.org/10.1109/ACCESS.2020.2975716.
10. Rajesh Kumar G, Mangathayaru N, G. Narasimha (2019) Similarity function for intrusion detection. In Proceedings of the 5th International Conference on Engineering and MIS (ICEMIS '19). Association for Computing Machinery, New York, NY, USA, Article 28, 1–4. https://doi.org/10.1145/3330431.3330460
11. Aljawarneh SA, Vangipuram R (2020) GARUDA: Gaussian dissimilarity measure for feature representation and anomaly detection in Internet of things. J Supercomput 76:4376–4413. https://doi.org/10.1007/s11227-018-2397-3
12. Iman S, Arash Habibi Lashkari, Ghorbani AA (2018) Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization", 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018
13. Kurniabudi, D. Stiawan, Darmawijoyo, M. Y. Bin Idris, A. M. Bamhdi and R. Budiarto, "CICIDS-2017 Dataset Feature Analysis With Information Gain for Anomaly Detection," in *IEEE Access*, vol. 8, pp. 132911–132921, 2020, doi: https://doi.org/10.1109/ACCESS.2020.3009843.
14. Swathi S, Lakshmeeswari G (2020) A Machine learning approach for DDoS attack detection using multiple linear regression. In: Proceedings at the 14th International Conference INTER-ENG 2020 Interdisciplinarity in Engineering, 8–9 October 2020

Credit Card Fraud Detection Using Machine Learning Algorithms

Mudita Sharma, Harshita Sharma, Pawan Bhutani, and Ira Sharma

Abstract **Fraud**: Fraud is any activity that causes financial loss to any other person. With the advancement of technology in the modern era, fraud related to credit card is continuously increasing. Credit card frauds are costing a huge amount of dollars to the customers. Nowadays, scammers are continuously finding new ways to do the scams. Also, to minimize the credit card fraud losses, different types of systems have been developed which detect frauds. We have to make a highly sensitive and powerful fraud detection system up to such an extent that it detects the fraud even before it takes place and in a sequential and exact manner. In this paper, we have mentioned three important algorithms, namely Isolation Forest, Support Vector Machine and Local Outlier Factor for credit card fraud detection. Also, we have discussed methodologies used in all three algorithms as mentioned above. We have also discussed the applications of the fraud detection system and how it will be implemented in the future.

Keywords Fraud · Local Outlier Factor · Support factor machine · Isolation Forest · Credit card fraud detection

1 Introduction

Credit card fraud: This fraud occurs when someone uses one's credit card or credit card information for making transactions without his permission or which he doesn't authorize. The fraudster (one who does the fraud) can steal one's credit card number, PIN number, etc. Now, a credit card can be defined as a card that helps a customer by giving a pre-set credit limit so that he can use it for making the payment. There are many benefits of having a credit card. Some are listed as follows:

- Chance to gain credit points.
- Earn rewards such as cashback and miles point.
- Protection against **credit card fraud**.
- Free information about credit score.

Mudita S (✉) · Harshita S · Pawan B · Ira S
HMR Institute of Technology and Management, GGSIPU, Delhi, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_49

- No foreign transaction fees.
- Increased purchasing power.
- Not linked to checking and savings account.

Because of advancements in electronic technology, the use of credit card transactions has been increased. There are basically two types of transaction:

(i) *Virtual Transaction:* In a virtual transaction, only some important information is necessary to make the payment.
(ii) *Physical Transaction:* In the physical transaction, the cardholder physically presents the card for making payment and the fraud is discovered after a few days the transaction has been made.

Various credit card frauds are:

Application frauds: By entering the details or information of the user such as password, pin number, etc. scammers gain control over the application system of the user.

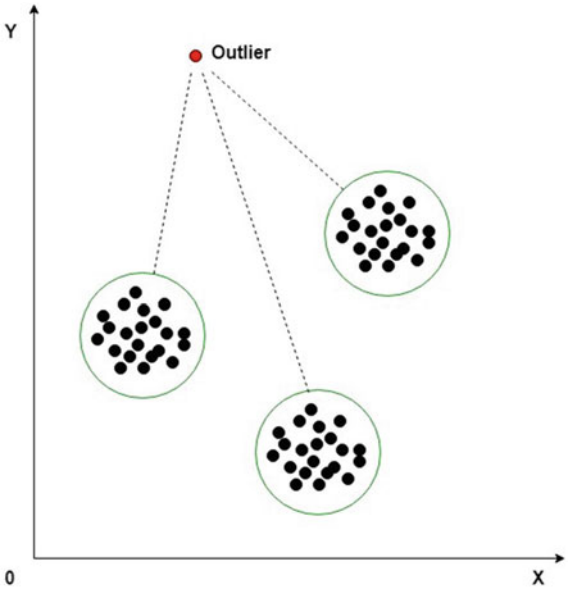
Electronic credit card imprints: Scammers can steal the information that is present on the magnetic stripe of the card. The information present on the magnetic stripe of the card is confidential, so one should handle this cautiously.

Card Not Present (CNP): In some cases, the card can also be used without a card number. This can be possible when the scammer has information about the expiry date and account number.

Card ID theft: In this type of fraud the scammer gets all the details about the credit card and uses that information to open another new (fake) account.

Credit card fraud occurs when one person uses the credit card information of the other person. We have to minimize the presently available limit on the credit card as a solution to the present problem statement of fraud-related credit card transactions. The best way to reduce the frequency of fraud is to analyze the present or normal purchase data of cardholders. Every person follows a particular pattern in the transactions and when there is a deviation in that pattern it indicates fraud. **There are three basic types of algorithms, i.e., Isolation Forest, Local Outlier Factor and Support Vector Machine, which are important ways to solve the frauds.** There are various advantages that came to light after practicing these algorithms in the detection of frauds caused by credit card transactions. These are after using these algorithms; first, there is a drastic change that can be seen very easily in the frauds. The second advantage is, based on the previous pattern followed by the user, fraud can be easily detected. Next is that we can add more updates in our system, i.e., the fraud detection system will block the user after certain attempts which are formerly set by the system organization (Fig. 1).

Fig. 1 Outliers



2 Dataset

We have taken the dataset from Kaggle. It contains information on transactions made by European Cardholders in September 2013. The total number of records is 2,84,807 out of which there are 492 fraud transactions. There are a total of 31 features out of which we know only three features, i.e., time, amount and class. Rest 28 features are not disclosed due to the confidentiality issue and are named as V1, V2, …, V28. Time represents the seconds elapsed between the first transaction and the consequent transaction. Amount refers to the amount that has been transacted. Class is the final prediction. Class 0 represents the normal transaction and Class 1 represents fraud transactions.

3 Related Works

Electronic commerce technology has reached new heights in the past few years and so the use of credit cards, net banking, debit cards and other mobile payment platforms has increased at a fast pace. A credit card, being the most versatile mode of payment for both online and regular purchases, is somewhat insecure too if we do not handle it with care. Hence, the cases of credit card fraud are also surging. To resolve this issue we have various techniques: Isolation Forest, Support Vector Machine, Random Forest, Logistic Regression etc. The paper by Massimiliano and his colleagues [1] has concluded in the result that the score obtained by a standard ANN classification

model can be improved by the features that are extracted from a network-based representation of data and play a vital role in rectifying the score. To improve data mining models, this paper tested a hypothesis that complex networks can be used for the same. These data mining models are specifically used for detecting fraud transactions in credit card transactions. This hypothesis provides a new approach to the combined usage of complex networks and data mining tools for detecting fraud cases (Figs. 2, 3 and 4).

In the paper [2], the authors—Yashvi Jain, Namrata Tiwari, Shripriya Dubey and Sarika Jain—had done a comparative analysis on various fraud detection techniques such as Support Vector Machine, Artificial Neural Network, Bayesian network, K-nearest neighbor, fuzzy logic-based system, decision tree and Logistic Regression. They concluded that we have major gaps in the techniques and methods of fraud detection. Artificial Neural Network (ANN) and Naïve Bayesian network have high accuracy and better detection rates, but they are costly to train because they need GPUs for training purposes, otherwise they'll work very slowly and can take a lot of time to train. Some algorithms like KNN and SVM give tremendous results with small datasets but large datasets cannot be trained properly to use them. Decision tree and SVM, being supervised machine learning algorithms, work well on sampled and pre-processed data. While raw data can be used efficiently with Logistic Regression and fuzzy systems, the authors have suggested using the hybrid of various techniques. According to them, to develop a good hybrid model there is a need to pair an expensive model with an optimization technique that will take care of the cost of the model. The application and environment of the fraud detection system play an important role in choosing the right algorithm. Some examples of hybrid techniques that have been mentioned in the paper are: a hybrid of decision tree and neural network; a combination of neural network and genetic algorithm; a hybrid of fuzzy clustering and neural network; Hybrid Bayesian Network and Artificial Neural Network and the combination of Support Vector Machines and decision trees.

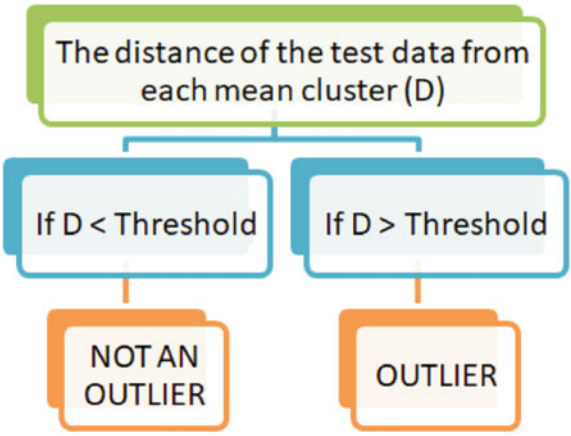


Fig. 2 Detection of outliers

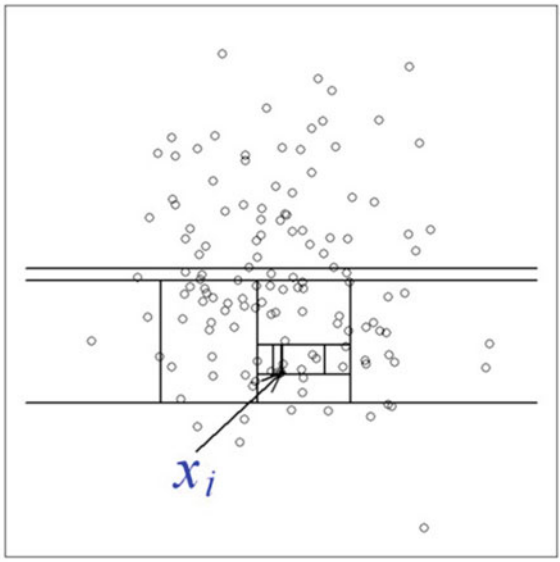


Fig. 3 Isolating normal point

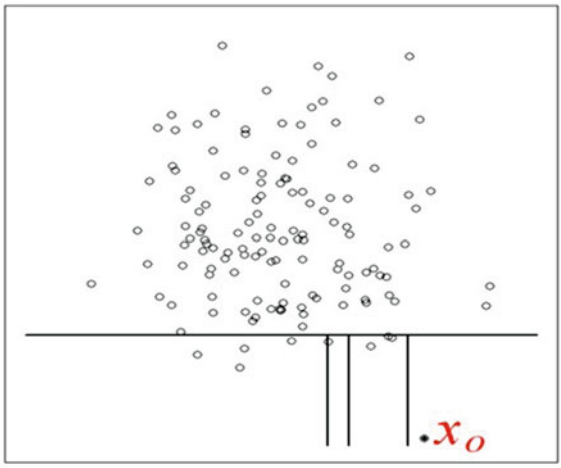


Fig. 4 Isolating an outlier

The authors Ishu Trivedi, Monika and Mrigya Mridushi of the paper [3] found that to lower the number of false alerts and finding out fraudulent transactions, the genetic algorithm method gives more accurate results. This algorithm requires a very short span of time after the transactions have been made, to detect whether the transaction is fraud or not, which will ultimately prevent the banks and customers from huge losses and also will reduce risks.

The paper [4] by Linda Delamaire (UK), Hussein Abdou (UK) and John Pointon (UK) gives information about various types of frauds, such as theft fraud, bankruptcy

fraud, application fraud, counterfeit fraud, theft fraud and behavioral fraud. In today's world, it is important to know about various frauds in order to be careful while making any transaction. The methods that they have used for detecting fraud transactions are genetic algorithms, clustering techniques, pair-wise matching, decision trees, and neural networks.

The authors Samaneh Sorournejad, Zahra Zojaji, Reza Ebrahimi Atani, and Amir Hassan Monadjemi have stated some difficulties of credit card detection techniques in the paper [5]. These are imbalanced data, different misclassification importance, overlapping data, lack of adaptability, fraud detection cost and lack of standard metrics. They have also given the complete classification of credit card fraud detection techniques. Their paper has shown that fraud detection techniques are of two types: 1. Supervised/user's behavior/misuses, 2. Unsupervised/transaction analysis/anomaly detection, in which supervised techniques are inductive logic programming, artificial immune system, expert system, Artificial Neural Network (BP), and the case-based reasoning, Support Vector Machines, Bayesian network, rule base, and fuzzy system. And unsupervised techniques are hidden Markov model, Support Vector Machines, artificial immune system, Artificial Neural Network (self-organizing map), the fuzzy system and Bayesian network. They have also described the complete classification of the dataset's attribute. Credit card transaction datasets are usually divided into two types, i.e., numerical and categorical attributes. The numerical attributes may consider features such as time, amount, etc. while the categorical attributes can be the class which is the final output. They identified decision tree and rule induction algorithms as categorical-based, whereas SVM, KNN, AIS (NSA), GA and ANN algorithms are numerical-based. Naive Bayes, CBR, AIS (DCA) and fuzzy system are some algorithms that come under both types. The paper has also stated about the open issues, like nonexistence of extensive credit card benchmark dataset, nonexistence of standard algorithm, nonexistence of suitable metrics and the deficiency of adaptive credit card fraud detection systems.

The authors Navanshu Khare and Saad Yunus Sait have shown in their research paper [6] that Random Forest will give the highest accuracy whereas Logistic Regression, SVM, decision tree gives less accuracy than Random Forest. Though Random Forest gives the highest accuracy, but it does not hold well in case of speed during testing. In the case of SVM, datasets require pre-processing in order to get better results.

In the research paper [7], the authors Masoumeh Zareapoor, Pourya Shamsolmoali compared various standard classifiers with bagging ensemble classifier, which is an interestingly new technique based on decision tree. They also identified that it works very well in the detection of fraudulent transactions. Bagging ensemble has the highest fraud catching rate and a relatively low false alarm rate. They concluded that the bagging ensemble classifier has performed stably during training, testing and evaluation. Also, it is independent of the rate of fraud. It also has the capability to handle imbalanced datasets.

The authors of the research paper [8] have identified that the isolation technique gets high AUC score when compared to K-means, OCSVM and LOF. It is more accurate, performs much better than others in detecting fraud transactions and detects

Table 1 Comparison of algorithm

Classifier	Advantages	Disadvantages
Local Outlier Factor	• LOF identifies outliers in a data set locally • It works well for unlabeled data	• It works better for larger datasets rather than smaller ones
Isolation Forest	• Values need not be scaled in feature space • Value distributions need not be assumed • It is robust and easy to optimize because of fewer parameters	• The result visualization is complicated • Training time can be very long and computationally expensive if it's not correctly optimized
Support Vector Machine	• Accuracy is good • Works well on smaller and cleaner datasets • It can use a subset of data points, hence making it efficient	• It is less efficient for noisier datasets with overlapping classes • Training time is large for larger data sets

few errors than the other three methods. Isolation Forest is proficient in detecting anomalies of credit card fraud transactions.

In the research paper [9] the authors clearly stated that the Local Outlier Factor has the highest accuracy than Isolation Forest. LOF performed much better and gives the best results when compared with Isolation Forest (Table 1).

4 Algorithms Used

Credit card fraud is a matter of great concern. Everyone wants a secure transaction and hence people take care of various things to keep their cards safe but still sometimes can get trapped in a situation that can lead to a great loss. Hence, the total numbers of fraud transactions are very less as compared to genuine transactions. So, we can say that in any dataset, the fraud transactions act like an outlier.

"*Observation which deviates so much from other observations as to arouse suspicion it was generated by a different mechanism"—Hawkins (1980).*

An object which deviates remarkably from the rest of the objects can be defined as an ***Outlier****.*

These points represented within a circle form a cluster.

First, we have to calculate the mean for each cluster and then initialize a threshold value. Hence, an outlier can be identified in the following way:

In our project, we have used three methods to find these outliers or frauds.

(I) ***Local Outlier Factor***

It is an unsupervised outlier detection technique which works on the principle of density measurement and comparison. It measures the density of different data points, or their clusters and compares them with one another. The anomaly or the outlier will have less density in comparison with the other data points

or clusters. So, by this technique, we can differentiate the outliers with normal data points, and hence detect them.
There are certain terms associated with LOF:

(a) *k*
"k" is defined as the number of neighbors. It can be considered as a threshold value for calculations of the Local Outlier Factor. Its value should not be either too small or too large. If k is very small, then it would be able to consider very few points. If k is very large, then it can miss some of the outliers.
(b) *k-distance*
It is the distance which is based on the value of k. A point's distance to its kth neighbor is called k-distance.
(c) *Reachability distance*
If '**a**' and '**b**' are two points, then reachability distance can be described using the formula:
(d) *Local reachability distance (LRD)*
It can be found out using the reachability distance by using the formula:
(e) *Local Outlier Factor (LOF)*
The ratio of the LRD of the neighbors of point **'a'** to the LRD of point '**a'** is known as Local Outlier Factor.

(II) ***Isolation Forest Algorithm***
Isolation Forest algorithm is also a type of anomaly detection algorithm. It is very much similar to decision trees. The basic difference between the two algorithms is decision trees start with a parent node and then keep on partitioning on the basis of information gain, whereas in Isolation Forest algorithm, the partitioning occurs randomly on the basis of the value of the selected feature. The selected feature will have a maximum and minimum value which perhaps gives the range of the split value. We continue splitting until we isolate the outlier from other data points.

(a) *Isolating a normal point*
A normal point will be much clustered, i.e., it would be surrounded by many data points. Many data points will be there in close proximity to a normal point. So, more splits would be required to isolate a normal point from other data points.
(b) *Isolating an outlier*
While if we want to isolate an outlier or an anomalous point which would not be in the vicinity of other normal data points, less number of splits would be required. An outlier is a point which lies away from other normal points. There will be no clusters as well as the density will be very less around an outlier. So, it would be easier to find the anomalous point or the outlier.

(III) ***Support Vector Machine***

Support Vector Machine is a supervised learning classifier. It is used for dividing the dataset into two or more classes, but it depends on a number of factors such as kernel, hyperplane, support vectors etc.

(a) *Hyperplane*

Support Vector Machine or SVM can be thought of as a classifying machine. Suppose we are working in two-dimensional space, having two features X1 and X2, then these two features can be separated using a plane known as hyperplane which is also known as maximum margin hyperplane. This plane separates the two features linearly. The main aim is to choose a hyperplane with maximum margins. The plane which lies on the right of hyperplane is known as positive hyperplane, whereas the plane which lies on the left is known as negative hyperplane.

(b) *Margins*

The distance between the two lines on either side of the features from the hyperplane (center line) is known as margins. And the combination of the margins of both side of the hyperplane is known as maximum margin.

(c) Support Vectors

These are the points which help in making the hyperplane. It supports the positive and negative hyperplanes on either side of the features.

(d) Kernel

Kernels are the tools for analyzing the different types of patterns in SVMs and hence partition the data. They are useful in separating different features from one another.

In our project, we have specifically used ***One Class SVM,*** which is an unsupervised machine learning algorithm. It trains on normal data and its boundary and hence identifies all the points that lie outside this boundary. It is suitable for outlier detection.

In Table 1, we have compared these three algorithms on the basis of their advantages and disadvantages. The table shows a theoretical comparison between these algorithms, i.e., Local Outlier Factor, Isolation Forest and Support Vector Machine. The Support Vector Machine algorithm works well with small dataset but here, in credit card fraud detection, the dataset is humungous. It contains around 284,000 records. So, the SVM method took more than 14 h for its execution and its result was also not appreciable which we will see in the result section.

This comparison table will help the users to unleash the basic concepts of these algorithms as well as will enhance the understanding of readers.

Table 2 Confusion matrix

Actual class	Predicted class		
		Negative	Positive
	Negative	True negative	False positive
	Positive	False positive	True positive

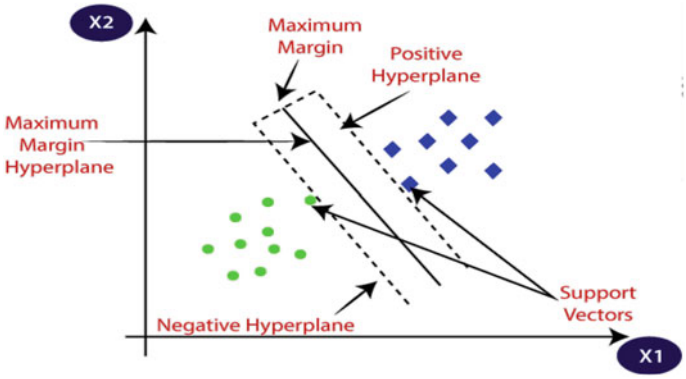


Fig. 5 Support Vector Machine

5 Methodology

We have used three algorithms to find the fraud transactions in our dataset which contains around 284,315 transactions. We proceeded according to the given flowchart. For data visualization, we have used four different plots. The first plot gives us the total number of fraud as well as normal transactions. The second plot is a histogram between the amount and number of transactions for both fraud and normal transactions. The third plot is a scatter plot between the time of the transaction and the amount for both fraud and normal transactions. Also, we have a heatmap, which clearly explains the total transactions. The fraud transactions are classified as "1", while normal transactions are classified as "0". We have used three classifiers, i.e., Local Outlier Factor, Isolation Forest and Support Vector Machine. We fit the model with the dataset and find the result. Isolation Forest has the highest accuracy of 99.74%, whereas SVM has the lowest accuracy. Also, SVM takes a large amount of time as compared to the other two methods.

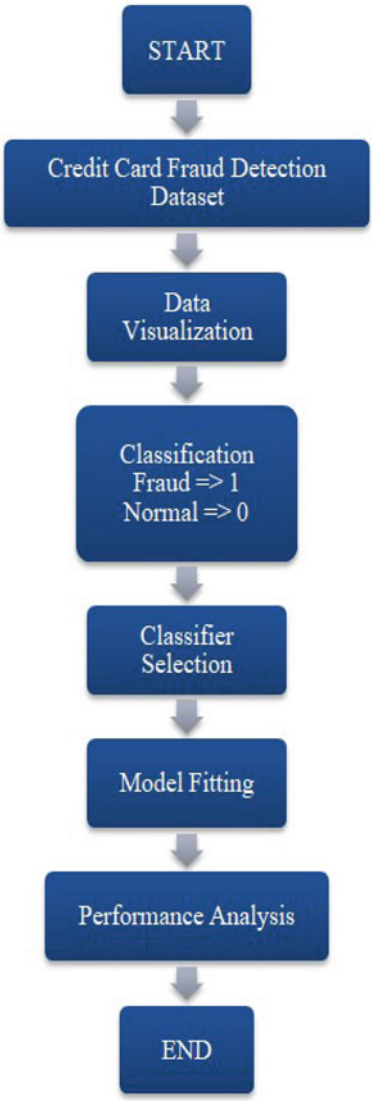
The flowchart given in Fig. 5 explains the process flow very efficiently and in a simple and readable format (Fig. 6).

6 Results

The quality of predictions from a classification algorithm can be measured using a **classification report**.

The classification report uses four different metrics, namely **precision, recall, F1-score and support**. These metrics are calculated by using some parameters called true positive, false positive, true negative and false negative. The values of these four parameters can be found out on the basis of true class and predicted class using the confusion matrix shown in Table 2.

Fig. 6 Process flow



1. *True Positive (TP):* These are the values which have been predicted correctly. It includes only the positive values which belong to the class "True or Yes".
2. *True Negative (TN)*: These are the values which have also been predicted correctly. It includes only the negative values which belong to the class "False or No".
3. *False Positive (FP)*: It signifies the wrong prediction. The value of the actual class is NO and the value of the predicted class is YES.

4. *False Negative (FN)*: It also signifies the wrong prediction. Here, the value of the actual class is YES and the value of the predicted class is NO.

False positive and false negative classes occur in the case of class mismatch between the actual and the predicted class.

Classification report metrics are as follows:

1. Precision
2. Recall
3. F1-score
4. Support

Precision: Precision in simple words can be defined as the accuracy of positive predictions. It is the number of correct predictions made by the classifier.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$$

Recall: It is the fraction of positive predictions made by the classifier that are correctly identified. It can be used to find the true positive rate.

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN})$$

F1-score: It is the mean of precision and recall.

$$\text{F1} - \text{Score} = 2 * (\text{Recall} * \text{Pre-cision}) / (\text{Recall} + \text{Precision})$$

Support: It is the number of occurrences of each class in correct target values.

Table 3 shows the accuracy as well as the number of wrong predictions made by that particular algorithm.

As described above, we have used the method of the classification report which gives us the result on the basis of four parameters, i.e., precision, recall, F1-score and support. **Here, Class 1 denotes fraudulent transactions, while Class 0 represents the normal transactions.** We have shown the classification report in Table 4.

The Isolation Forest method performs the best out of all the algorithms with an accuracy of 99.76%. Out of 284,315 data entities, the Support Vector Machine shows the poorest result. It has just the accuracy of 75.52% and it predicts 71,078 records wrongly. Local Outlier Factor also shows the good result, and its result is close to that

Table 3 Accuracy and wrong predictions

Algorithm	No. of wrong predictions	Accuracy (%)
Isolation Forest	683	99.76
Local Outlier Factor	935	99.67
Support Vector Machine	71,078	75.52

Table 4 Classification report

Algorithm	Class	Precision	Recall	F1-score	Support
Isolation Forest	0	1.00	1.00	1.00	284,315
	1	0.31	0.31	0.31	492
Local Outlier Factor	0	1.00	1.00	1.00	284,315
	1	0.05	0.05	0.05	492
Support Vector Machine	0	1.00	0.75	0.85	284,315
	1	0.00	0.32	0.00	492

of Isolation Forest. It has an accuracy of 99.67% and misclassifies only 935 values. Both the methods, Isolation Forest and Local Outlier Factor, also show a good result and are close to that of the Isolation Forest whereas the Support Vector Machine is not that accurate. Also, SVM is a very slow algorithm. It took more than 14 h to run which is not at all beneficial.

7 Conclusion and Future Scope

Fraud transaction cases can be eliminated to a greater extent with the help of machine learning algorithms. In this paper, we have compared three fraud detection techniques, i.e., Support Vector Machine, Local Outlier Factor and Isolation Forest. In all these fraud detection techniques, the Isolation Forest is proficient.

Isolation Forest gives 99.7% accuracy in results when applied on the complete dataset, whereas LOF gives 99.6% accuracy and SVM being the slowest one gives 99.7% accuracy. During the execution of code, SVM takes the longest time to give results as compared to other techniques. Its accuracy is also very low as compared to the other two techniques.

We can conclude here that to get the highest accuracy and fast results, use the Isolation Forest fraud detection technique. It will give you a precise and efficient result.

Despite these outstanding techniques, there will always be a percentage of fraud prevailing, until we integrate good security software/hardware with present software/hardware available in the market. There is always a possibility of improvement. So, for the future scope, we can add a fingerprint sensor in ATM machines to avoid fraudulent transactions. Nowadays, we all use mobile phones or laptops for making transactions. The inbuilt camera in these devices can act as the most useful aid for security. We can also add iris detection system in the bank's software or mobile application. Whether the user uses a mobile phone or laptop for the transactions, it will be a safe transaction if the bank's software is integrated with iris detection system. We can also add a face detection system with the bank's software.

References

1. Massimiliano Zanin, Miguel Romance, Santiago Moral and Regino Criado “Credit Card Fraud Detection through Parenclitic Network Analysis”, Volume 2018, Article ID 5764370.
2. Yashvi Jain, Namrata Tiwari, Shripriya Dubey, Sarika Jain, “A Comparative Analysis of Various Credit Card Fraud Detection Techniques”, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277–3878, Volume-7 Issue-5S2, January 2019.
3. Ishu Trivedi, Monika, Mrigya Mridushi, “Credit Card Fraud Detection”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016.
4. Linda Delamaire (UK), Hussein Abdou (UK), John Pointon (UK), “Credit card fraud and detection techniques: a review”, Banks and Bank Systems, Volume 4, Issue 2, 2009.
5. Samaneh Sorournejad, Zahra Zojaji, Reza Ebrahimi Atani, Amir Hassan Monadjemi, “A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective.”
6. Navanshu Khare and Saad Yunus Sait (2018) Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models. International Journal of Pure and Applied Mathematics 118(20):825–838
7. Masoumeh Zareapoor, Pourya Shamsolmoali, “Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier”, ICCC-2015 Procedia Computer Science 48 (2015) 679 – 685.
8. Soumaya Ounacer, Hicham Ait El Bour, Younes Oubrahim, Mohamed Yassine Ghoumari1 and Mohamed Azzouazi, “Using Isolation Forest in anomaly detection: the case of credit card transactions” Periodicals of Engineering and Natural Sciences ISSN 2303–4521 Vol.6, No.2, December 2018, pp.394–400.
9. Hyder J, Sameena Naaz, “Credit Card Fraud Detection using Local Outlier Factor and Isolation Forest”, Journal of Computer Sciences and Engineering Vol.-7, Issue-4, April 2019 E-ISSN: 2347–2693.

Identification of Parameters for Noise Pollution Based Electrical Energy Converter

Arunesh Kumar Singh, Shahida Khatoon, and Kriti

Abstract This paper presents identification of parameters for noise pollution based electrical energy converter. Parameters involved in traffic noise model and their effects on the performance of noise pollution based electrical energy converter are also discussed.

Keywords Noise pollution · Electrical energy · Traffic noise · Energy converter

1 Introduction

Noise, the unwanted sound is one of the most important constituents of environmental pollution. Different forms of noise found in our daily life are railway noise (sound of the train wheels and its horn), aircraft noise(sound of air-plane at the time of takeoff and landing), industrial noise (due to heavy machines and other industrial activities), construction noise and traffic noise [1]. Among all these types of noise pollution, traffic noise is the most significant as it contributes approximately 55% of total noise pollution [2]. Increase in the noise pollution leads to the various health hazards among human beings, such as lack of sleep, stress, depression, high blood pressure etc. Scientists and researchers are trying to achieve reduction of environmental noise pollution with various innovative techniques [3].

At present [4], different forms of machines have been inherent part of our daily life, such as use of electricity based home appliances, use of laptops, smart watches, mobile phones and other digital devices for individual purposes and use of electrical machines for industrial purposes etc. In this technology driven time, use of electricity has increased more than ever. For electricity generation, conventional energy resources are falling short of meeting this huge demand. So it becomes inevitable to

A. K. Singh · S. Khatoon · Kriti (✉)
Department of Electrical Engineering, Jamia Millia Islamia, New Delhi 110025, India

A. K. Singh
e-mail: asingh1@jmi.ac.in

S. Khatoon
e-mail: skhatoon@jmi.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_50

look for alternative methods of electricity generation. The idea is to utilize the noise pollution for electricity generation. Noise (sound) energy based energy conversion method provides the clean electrical energy which does not pollute the environment unlike the coal and petroleum does.

2 Noise Pollution Based Electrical Energy Converter

A simple noise pollution based energy conversion scheme is shown in Fig. 1.

The noise pollution based energy converter transforms the noise signal into low level electrical signal. The energy converter can be based on any of the multiple principles of energy conversion from acoustic signal. Converted low level electrical signal goes to suitable rectifier and amplifier which provide necessary rectification and amplification. Super capacitors can be used for energy storage and regulator provides regulated voltage at the output [5].

Noise signal generated from noise pollution is a very low level signal. Hence for this application most suitable technique is based on the concept of piezoelectricity [6, 7]. Li et al. has reviewed various piezoelectric converters for low frequency as well as high frequency applications. Piezoelectric material is of four types namely ceramics, single crystal, polymers and composite. Their uses depend upon the application. Lead-zirconate-titanate, the ceramic type piezoelectric (PZT) material can be used for high frequency application [8]. For low frequency application polymers and composites can be used. He also discussed about the uni-morph and bi-morph structure of piezoelectric converter. Uni-morph has only one layer of piezoelectric on cantilever beam whereas bi-morph has two layers of PZT on metal strip which gives high output power. Et al L. Jiang proposed a piezoelectric cantilever beam which produces an electric charge when placed under mechanical stress as shown in Fig. 2. Output of piezoelectric converter is in current microamperes which is proportional to the input vibration. It acts as current source [9].

As the piezoelectric materials have ability to generate maximum output at resonance frequency so resonator can also be integrated with the energy converter to increase its efficiency of energy conversion. When the frequency of incoming signal matches with the resonance frequency it produce large signal which further can be

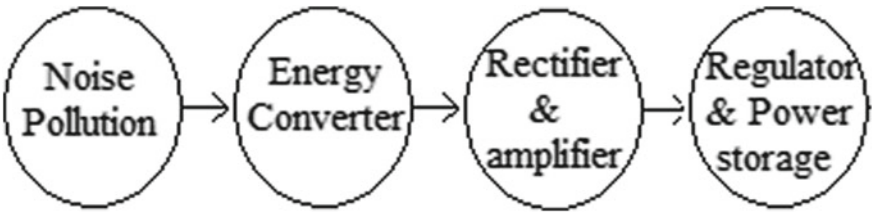


Fig. 1 Noise pollution based electrical energy converter

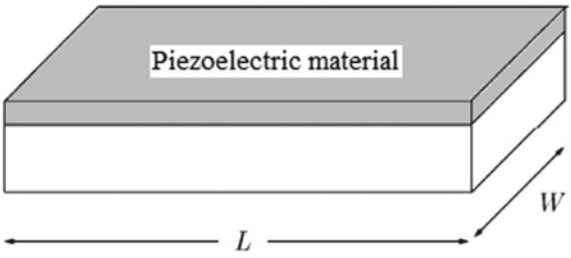


Fig. 2 Piezoelectric cantilever beam

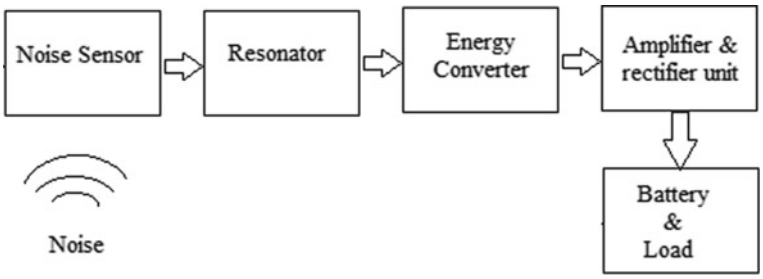


Fig. 3 Noise pollution based energy converter with resonator

amplified. In case of traffic noise pollution the frequency of noise (sound) signal is very unstable [10]. Use of resonator [11, 12] will definitely help to get good results. Figure 3 shows the Noise pollution based energy converter with resonator [13]. In the event of resonance it provides large signal to the energy converter. After being converted this electrical energy gets amplified and rectified. Then it is supplied to the load.

3 Effect of Different Parameters on the Performance of Noise Pollution Based Energy Converter

Many parameters can have an effect on the performance [14] of noise pollution based energy converter. We have enlisted few parameters that may have effect on the signal converted from noise pollution.

(a) Variation in local pressure from the ambient atmospheric pressure due to sound is called as sound pressure. It is basic parameter of noise pollution. It is measured by using microphone.

(b) SPL (Sound pressure level) is the logarithmic ratio of sound pressure to the reference pressure. It is given by

$$\mathrm{SPL} = 10\log_{10}\left(\mathrm{P}^2/\mathrm{P}_{\mathrm{ref}}^2\right) \tag{1}$$

(c) Intensity of sound represents the sound power carried by the sound wave per unit area in the perpendicular direction.

$$I = P \times v \tag{2}$$

where particle velocity is the velocity of sound particle by which it transmit the wave in a medium.

$$v = \frac{\partial \delta}{\partial t} \tag{3}$$

where δ the displacement cause by sound wave and t is is the time.
Sound intensity level is given by

$$L_I = 10 \log_{10}\left(\frac{I}{I_O}\right) dB \tag{4}$$

where reference intensity, $I_o = 10^{-12}$ W/m^2.

(d) Amount of emitted sound energy from a source in one second is called as power of sound. Sound power does not depend on the distance from the measurement location whereas the sound pressure does.
Sound Power (PW) is given by.

$$P_W = I \times A \tag{5}$$

where A is the surface area and I is sound intensity. Sound intensity or sound power is the property of noise/sound source [15] whereas it's effect in the environment is sound pressure.

(e) Electric Charge (Q) develops when piezoelectric sensor deforms due to applied sound pressure. When an acoustic signal imparts on the sensor, a displacement occurs at opposite ends and thus an electric charge develops. This charge get stored in super capacitors [16] and is responsible for the current flow through circuit. Equation (6) shows the relation between vibration energy force (F) and electric charge (Q) [17]

$$Q = k \cdot F \tag{6}$$

Where k is the piezoelectric sensor constant.

(f) The capacity of a capacitor to store electric charge is capacitance(C) and defines in terms of Farad. The relation between electric charge (Q) and output voltage (V_o) of piezoelectric sensor is given by

$$Q = C \cdot V_o \tag{7}$$

(g) Distance of the measurement point from the noise source is an important factor in energy conversion. As the distance (D) increases the sound pressure level decreases significantly.

(h) Equivalent SPL is measure by using SPL meter. A-weighted frequency filter provides the frequency response similar to the human ear. L_{eq} is the measure of noise pollution. Equivalent SPL (L_{eq}) is the Continuous A- Weighted SPL measured over a time period (T). For some measurements [18, 19] L_{50}, L_{70}, L_{90} can be calculated respectively for SPL beyond 50, 70 and 90 percentile of time period T.

(i) Output voltage of a piezo-electric sensor (Vo) depends upon the electric charge stored by the capacitor. This voltage can be enhanced by using suitable amplifier.

(j) Human ear is sensitive to the audio frequencies (20 Hz to 20,000 Hz). Frequency of the sound pressure signal is defined as total number of pressure vibrations per second. [20] Larger pressure vibrations refer to are called as high amplitude or louder sound whereas high frequency refers to high pitch of sound

(k) Energy conversion efficiency of the energy converter can be given by

$$\eta = \frac{\text{Power delivered to load}}{\text{Power harvested from Q} - \text{V cycle}} \tag{8}$$

4 Traffic Noise Model

Single largest component of noise pollution is transportation noise as it contributes almost 55% to the total noise pollution. To regulate the traffic noise, traffic noise models have been adapted by various countries. Equation (9) shows the simple traffic noise model. It shows the relationship between sound pressure level (Leq) and noise pollution [21].

$$L_{eq} = X + Y\log_{10}W \tag{9}$$

In above equation, W is the traffic volume, X and Y are the traffic model constants.

Reference [22] in traffic noise modeling, sound pressure level depends upon various factors including type of vehicle, Number of vehicles, vehicle's average speed, traffic volume, distant between noise source and measurement location, sound power of noise source etc. Spike in the level of sound pressure has also been detected due to honking and sudden brake application. We have enlisted few parameters of traffic noise model in Table 1 which may impact on the converted electrical signal from transportation noise.

Table 1 List of parameters involved in noise pollution based energy converter

S. No	Parameter	Symbol	Unit
1	Sound Pressure	P	Pascal or dB
2	Sound power level	L_W	dB
3	Equivalent Sound Pressure level(A-weighted)	L_{eq}	dB(A)
4	Input sound intensity level	L_I	Wm-2 or dB
5	Input sound signal frequency	f	Hz
6	Resonant frequency	fr	Hz
7	Electric charge	Q	Columb
8	Capacitance	C	F
9	Distance from source	D	meter
10	Time interval	T	second
11	Velocity of source	V	Km/h
12	No. of vehicles or machines	N	–
13	Location of noise measurement	$()_x$	–
14	Traffic volume	W	Vehicle/hours
15	Type of vehicle	a,b,c	–
16	Output power	Po	Watts
17	Output voltage	Vo	Volts
18	Efficiency	η	–

5 Conclusion

In this paper, identification of various parameters related to noise pollution based electrical energy converter and their effects on the performance are discussed. The presented noise pollution based electrical energy converter can fulfil the energy requirement for low power applications such as powering the LEDs roadside at night and other small applications.

References

1. Bhatnagar SR (2012) Converting Sound Energy To Electric. Int. J. Emerg. Technol. Adv. Eng. 2(10):267–270
2. Vijay R, Sharma A, Chakrabarti T, Gupta R (2015) Assessment of honking impact on traffic noise in urban traffic environment of Nagpur, India. J. Environ. Heal. Sci. Eng., vol. 13, no. 1. https://doi.org/10.1186/s40201-015-0164-4.
3. Goines L, Hagler L (2007) Noise pollution: a modern plague. South Med J 100(3):287–294. https://doi.org/10.1097/SMJ.0b013e3180318be5
4. Gieras JF, Wang F, Lai FC (2005) Noise of polyphase electric motors. Noise Polyph. Electr. Mot., no. January, pp. 1–368. https://doi.org/10.1201/9781420027730.

5. Liew HF, Rahim RA, Isa M, Ismail B, Hassan SIS (2018) Analysis of batteries or supercapacitor as energy storage device for a sound energy harvester system. IEEJ Trans Electr Electron Eng 13(12):1699–1708. https://doi.org/10.1002/tee.22733
6. Ahmad S, Abdul Mujeebu M, Farooqi MQ (2019) Energy harvesting from pavements and roadways: a comprehensive review of technologies, materials, and challenges. Int. J. Energy Res 43(6):1974–2015. https://doi.org/10.1002/er.4350.
7. Cha SN et al (2010) Sound-driven piezoelectric nanowire-based nanogenerators. Adv Mater 22(42):4726–4730. https://doi.org/10.1002/adma.201001169
8. Li T, Tian C, Deng Z (2014) Energy harvesting from low frequency applications using piezoelectric materials. Appl Phys Rev. https://doi.org/10.1063/1.4900845.
9. Jiang L, Li Y, Marvin C (2012) Analysis of piezoelectric energy harvesting device with adjustable resonance frequency. J Electron 29(3–4):310–318. https://doi.org/10.1007/s11767-012-0810-2
10. Ali Khalil M, Hamad K, Shanableh K (2019) Developing machine learning models to predict roadway traffic noise: an opportunity to escape conventional Techniques. Transp. Res. Rec 2673(4):158–172. https://doi.org/10.1177/0361198119838514.
11. Mishra S, Unnikrishnan L, Nayak SK, Mohanty S (2019) Advances in Piezoelectric Polymer Composites for Energy Harvesting Applications: A Systematic Review. Macromol Mater Eng 304(1):1–25. https://doi.org/10.1002/mame.201800463
12. Wischke M, Masur S, Kröner M, Woias P (2011) Vibration harvesting in traffic tunnels to power wireless sensor nodes. Smart Mater. Struct. 20(8). https://doi.org/10.1088/0964-1726/20/8/085014.
13. Ge Q (2017) Prospect of electric generation using sound. AIP Conf. Proc., vol. 020050, no. May, 2017. https://doi.org/10.1063/1.4982415.
14. Fang LH, Hassan SIS, Rahim RH, Isa M, Bin Ismail B (20177) Exploring Piezoelectric for Sound Wave as Energy Harvester. Energy Procedia, 2017, vol. 105. https://doi.org/10.1016/j.egypro.2017.03.341.
15. Sakamoto S (2015) Road traffic noise prediction model 'ASJ RTN-Model 2013': Report of the Research Committee on Road Traffic Noise. Acoust Sci Technol 36(2):49–108. https://doi.org/10.1250/ast.36.49
16. Arnab MMB, Ullah SMR, Alam MA, Nondy RK, Alam ASMF, Mishu AP (2014) "Generation of electrical energy using piezoelectric material from train wheels: Bangladesh perspective", 2014 9th Int. Forum Strateg. Technol. IFOST 2014:300–303. https://doi.org/10.1109/IFOST.2014.6991126
17. Selim A, Haggag HM, Yehia FZ. Amer El-Garhy AM (2016) Acoustic energy conversion into useful electric energy from Disk Jockey by using piezoelectric transducers," 2016 18th Int. Middle-East Power Syst. Conf. MEPCON 2016 - Proc., pp. 871–876. https://doi.org/10.1109/MEPCON.2016.7836998.
18. Bijay Kumar Swain SG (2013) Integration and comparison of assessment and modeling of road traffic noise in Baripada town, India. Int. J. Energy Environ., vol. 4, no. 2, pp. 303–310
19. Kumar K, Jain VK (1994) A study of noise in various modes of transport in Delhi. Appl Acoust 43(1):57–65. https://doi.org/10.1016/0003-682X(94)90041-8
20. Piercy JE, Embleton TFW, Sutherland LC (1977) Review of noise propagation in the atmosphere. J Acoust Soc Am 61(6):1403–1418. https://doi.org/10.1121/1.381455
21. Verghese V, Subramanian SC, Vanajakshi L (2013) Model based Traffic Control in Indian Conditions. Procedia - Soc. Behav. Sci. 104:516–525. https://doi.org/10.1016/j.sbspro.2013.11.145
22. Maruyama M (2020) Dynamic properties of peak levels of road traffic noise along a freeway. Appl. Acoust 160, p. 107095. https://doi.org/10.1016/j.apacoust.2019.107095
23. Singh AK, Ibraheem AK, Sharma R (2014) Parameter identification of eddy current braking system for various applications," Proc. Int. Conf. Innov. Appl. Comput. Intell. Power, Energy Control with their Impact on Humanity. CIPECH 2014, November, pp. 191–195. https://doi.org/10.1109/CIPECH.2014.7019113.

24. Kalawapudi K, Singh T, Dey J, Vijay R, Kumar R (2020) Noise pollution in Mumbai Metropolitan Region (MMR): An emerging environmental threat. Environ Monit Assess 192(2):1–20. https://doi.org/10.1007/s10661-020-8121-9

Reconfiguration of Micro-grid for Future Forecasting to Maintain High Reliability

Shivani Mishra and R. K. Viral

Abstract The grid unbalance is a common phenomenon in the northern region of India. To maintain frequency is very difficult where the population is around 1.3 billion. To maintain grid frequency, demand and supply should be matched. Micro-grid plays a major role in demand and supply equalization. Micro-grid is a feasible solution for remote location where electrification is very costly. To develop remote location of India like Ladakh, Bhuj etc. micro-grid provides better solution to increase the economy of country. In micro-grid, input cost is totally depending on the availability of region natural resources. Solar is the main availability in countries like India, which can be supported by bio-mass, wind, bio-gas, storage system, etc. to prepare hybrid micro-grid for 24 h. reliability. After electricity act amendment of new rule is implemented, which is ABT, (availability-based tariff) in which penalty and reward are applicable. To implement ABT, reconfiguration of micro-grid is required. Reconfiguration is the adoption of new technology of previous electrical models where implementing of ring-main unit, smart metering, feeder segregation and forecasting of the load is done. ABT (availability-based tariff) gives better reliability to a system, which can be utilized with micro-grid solution. The paper identifies future updated scenarios of micro-grid as per the latest regulation of India.

Keywords Micro-grid · Reconfiguration · Reliability · Forecasting

1 Introduction

For exchange of electricity in India, there are many components like PPA (power purchase agreement), energy exchange, REC (renewable energy certificate) but every generation and distribution companies are responsible for maintaining grid

S. Mishra (✉) · R. K. Viral
Department of Electrical & Electronics Engineering, Amity School Of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

R. K. Viral
e-mail: rviral@amity.edu

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_51

frequency, for this, three-part tariff procedure is introduced called ABT (availability-based tariff) and every Discom has to purchase 1% REC of its demand [1]. The task of harmonizing the load demand with the generation is significant. Extra generation provides high frequency and less generation provides low frequency. As the current scenario micro-grid with the inclusion of forecasting methodology will help to improve reliability as well as profitability. In India, maintaining frequency is a major task where combination of multiple EHV lines as well as multiple generations, where some phenomena, are required for penalty and reward system, which will maintain or distract the grid frequency.

2 ABT (Availability-Based Tariff)

The trading of power is moderately diverse from on article of trade to the alternative one for electricity supply because it cannot be stored and subjected to the demand–supply sense of balance also aspects lots of transmission constriction. Fixed charge, variable charge and UI (unscheduled power interchange) for penalty and incentives are pricing schemes for ABT (availability-based tariff) [2]. The authorized power from the generation station is in proportion to the imposed-on heirs to the fixed cost. The plant capacity is proportionate to fixed cost, which is shared by beneficiaries. So, the fixed charges are called as the capacity charges. Generating station cannot due to a little amount as per the availability of electricity irrespective of its fixed cost [3]. The generating station is dependent on the plant availability by which it is payable for the fixed charge reimbursement. The generating station has to pay more if the plant accessibility is moreover the year as per the set standards and the vice-versa is applied when the plant accessibility is low. That's why it (tariff) is called as ABT (availability-based tariff). Previously, tariff for the fixed charges depends on the plant load factor but now ABT depends on the plant accessibility [4]. Generating station acquired variable charges day by day as per day MW production of electricity, this is also called as energy charges. It includes the fuel charges like nuclear power plant for fuel nuclear bundle, fuel for gas power plant, coal fuel for thermal power plant, etc. UI (unscheduled interchange) charges depict deviation from plants' scheduled generation. Presume that the scheduled power to deliver by the generating station is 500 MW but its supplying 600 MW on a day, then the energy charges paid by the station for the scheduled generation, which is for 500 MW. For the excess 100 MW, the energy charge rate depends predominantly on the grid frequency of that time. The charges of energy for the surplus supply of 100 MW as per presumption are called UI charges, which are linked with grid frequency. The idle frequency rate is 50 Hz, if the grid frequency increases up to or more than 50.2 Hz then there will be no UI charges that means the excess 100 MW generation is paid by the generating station. So, to maintain grid frequency, station involuntary is to reduce generation. Same as for lower grid frequency, the excess UI rates paid for excess generation incentives to the generating station at UI rates, assume that the frequency at that time is 49.4Hz. Micro-grid work as a single controllable entity in comparison to the grid within defined

electrical boundaries because it is a group of interconnected loads and distributed energy resources (DER). Micro-grid work as both as grid-connected as well as in isolated mode. As per CEA (central electricity authority), maintaining frequency in the grid is the primary responsibility of all generators and distributors [1]. Micro-grid is also a generator which can help to maintain frequency at the emergency time as well as after new rule of government of India distribution company will purchase 1% of its demand of electricity will purchase from renewable energy sources. Micro-grid can help to utilize the rule of government of India. At the off-peak time, micro-grid can sale extra electricity through participating in scheduling.

3 Micro-grid Reconfiguration

Currently, micro-grid has no tools for scheduling and no specific feeders for specific consumers, which cannot help power scheduling and finding the actual load [5]. In conventional micro-grid direct connection from generation to the distribution, only single switches were installed for breaking the load, which is also shown in Fig. 1.

Figure 1 shows the conventional micro-grid technology in which single switch was used and no feeder segregation was available. It is the full radial system at the time of fault restoration time will be high and all consumers will be affected till fault restoration. There are many changes required in conventional micro-grid to perform scheduling of power maintain high reliability specified consumer's feeder. As per the current scenario for scheduling, 96 slots do prepare for electricity sale. 96 slots are subdivided into 15 min. each. This report is further given to SLDC (state load dispatch centre) and communicates to RLDC (Regional load dispatch center) further. At the time of emergency when the frequency dips, micro-grid can help to maintain grid frequency and can get profit through availability-based tariff [6]. In the reconfiguration of micro-grid, installation of smart meters for gathering the actual and correct load as per specified slots as well as these smart meters will be installed at the starting of each feeder, which can be differentiated like industrial feeder, residential feeder and commercial feeder. This is also shown in Fig. 2.

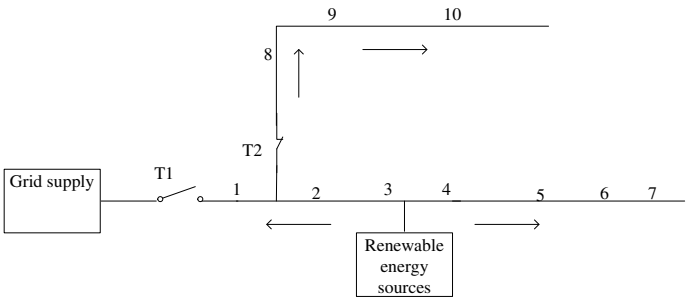


Fig. 1 Conventional micro-grid technologies

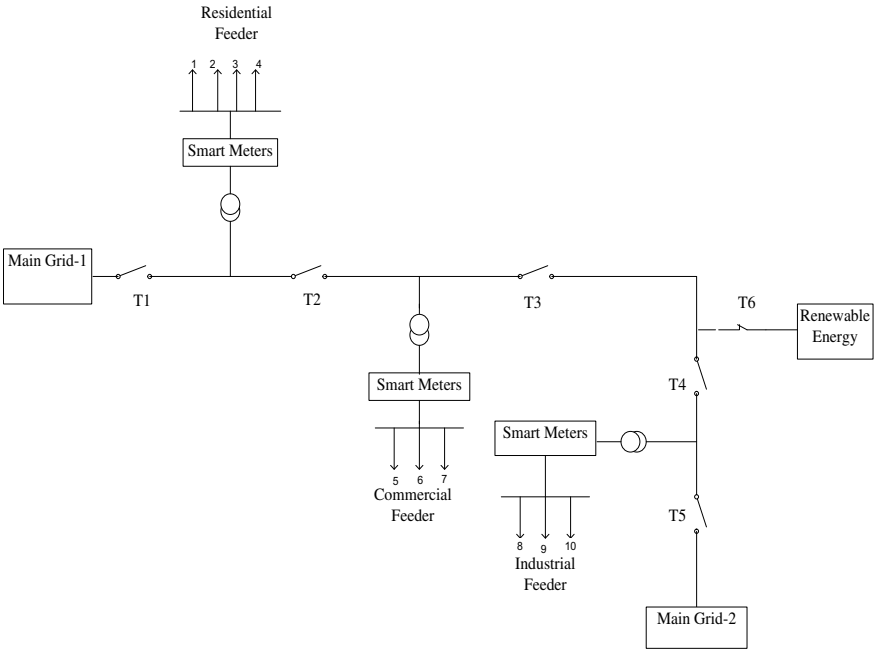


Fig. 2 Micro-grid reconfiguration with smart meters

In Fig. 2 adoption of some changes in conventional micro-grid, in which multiple online switches install with smart meters and specified feeder segregation which will help to early restoration of electricity [7, 15]. At the time of fault, only some consumers will be affected. Through smart meters, actual loading of feeders will get for better forecasting and planning.

4 Benefits of Micro-grid Reconfiguration

After reconfiguration of micro-grid, there are several benefits [8]. These are:

Loading

In conventional micro-grid, no procedure was available to gather data of actual loading in the reconfigured micro-grid using of smart meter identification of actual loading through modem will be very easy [9]. The actual loading will get the current scenario of consumers.

Table 1 Major benefits of DERs [39]

S. no	Factors	Conventional micro-grid	Reconfigured micro-grid
1	Reliability	Low	High
2	Forecasting	No	Yes
3	Planning for maintenance	Low	High
4	Frequency management	No	Yes
5	Sale of electricity	Low	High
6	Planning for electricity trading	No	Yes

Planning

In the conventional micro-grid, single switches were used but in reconfigured micro-grid, multiple switches are used, which can help in planning with better maintenance without affecting whole reliability, which improves SAIDI (system average interruption duration index).

Frequency

In the conventional micro-grid, no methodology was available to provide help to the grid at the time of frequency disturbance. After reconfiguration, micro-grid can provide help to the main grid at the time of frequency disturbance with proper loading and planning.

5 Comparison Between Conventional Micro-grid Reconfiguration

There are many ways to compare micro-grid and reconfigured micro-grid. Comparison graph is in Table 1: in which factors are considered, i.e. reliability, forecasting, planning and maintenance, frequency management, sale of electricity and planning for electricity trading. This comparison table shows the effectiveness of the grid with the factors used.

6 Forecasting Methodology

There are some factors that are affecting electricity forecasting. These are:

Factor of Time

It is an imperative factor in load forecasting that's why it highly influences consumer load.

Economic Factor

The economic factor has the function as load pattern, which is the growth of population, development of industries, cost of electricity and GDP (gross domestic products).

Condition of Weather

For considering the forecasting of load various weather variables are taken, which are humidity, speed of wind, temperature and cover of cloud.

Customer Factor

It defines the customer consumption of electricity as a factor of number, size and type while its electrical equipment and their installation depend on various customer priorities.

In conventional methodology forecasting is done by the following methods:

- Load taken by consumer
- Total load capacity taken by transformer.

In both methodologies, real-time analysis could not be done for real-time analysis and actual consumption of consumer or feeder [10, 16]. This will help for better and future forecasting. In electricity, trading forecasting is the foremost device to classify the demand and supply of grid along with regions. As per regulatory guidelines [1], it is mandatory that the formation of 96 slots is 15 minutes each for the power exchange. As per government policies, 96 slots to be provided before 24 hrs. At that time, the reconfigured micro-grid will help to forecast the consumption of electricity as well as to sale of electricity if required. At the time of emergency, sale of electricity in ABT regime will be beneficial. In the ABT regime to interrupt grid disturbance, UI charges will be extra benefit.

7 Reliability

Reliability designates the functionality of the system and equipment (connected through various components or devices) with specific period of time. In the power system, the following reliability areas are described:

- Power interruption
- Connected load
- Number of interruptions
- Consumer counts
- Duration of interruption.

Reliability defines the network capability of handling the consumer demand means the consumer can get 24 hours supply of electricity [11, 12, 14]. In the reconfigured micro-grid model, segregation of feeder is done for higher consumer satisfaction.

Reliability can be defined by various manners: SAIDI (system average interruption duration index), CAIDI (customer average interruption duration index), SAIFI (system average interruption frequency index) and MAIFI (momentary average interruption frequency index) [13]. In the reconfigured model, multiple switches are used to reduce consumer interruption count, which will increase next-generation reliability.

SAIDI (*System Average Interruption Duration Index*)

This index measures an interruption for the total duration for average customer through a given period of time.

$$\text{SAIDI} = \sum (r_i * N_i)/N_T \tag{1}$$

where, $\sum$ = function of summation, r_i = restoration time, N_i = total number of customers interrupted, N_T = total number of customers served.

CAIDI (*Customer Average Interruption Duration Index*)

It is the restoration time due to fault or outages occur in systems.

$$\text{CAIDI} = \sum (r_i * N_i) \Big/ \sum (N_i) \tag{2}$$

where, $\sum$ = function of summation, ri = restoration time, Ni = total number of customers interrupted.

SAIFI (*System Average Interruption Frequency Index*)

The system consumer experiences an average number of times, outages during the year

$$\text{SAIFI} = \sum (N_i)/N_T \tag{3}$$

where, $\sum$ = function of summation, N_i = total number of customers interrupted, N_T = total number of customers served.

MAIFI (*Momentary Average Interruption Frequency Index*)

The customer experiences the average number of momentary interruptions at a given period of time.

$$\text{MAIFI} = \sum (ID_i * N_i) \Big/ N_T \tag{4}$$

where, $\sum$ = function of summation, $ID_{i\,=}$ number of interrupting device operations, N_i = total number of interrupted customers, N_T = total number of served customers.

8 Recommendations

In India, Din Dayal Gramin VidyutiKaran Yojna (DDGVY) is going on where isolated and tribunal areas still have some problems for electrification. The above reconfiguration solution will give a reliable solution for maintaining the grid frequency. As per above description, the following points to be recommended:

i. This reconfigured model provides peak load demand.
ii. Proposed model is time-dependent model where it is helpful for power trading.
iii. The proposed model provides the long-term planning for economic growth.
iv. The proposed model provides the annual growth rate to upgrade the system simultaneously.
v. The model provides the essential data for increasing the reliability.

9 Conclusion

The reconfigured micro-grid will give better reliability due to the use of multiple switches as compared with conventional micro-grid where a single switch is used. Reconfigured micro-grid will also provide as the profitability solution at non-peak session, which was not available at conventional micro-gird. Reconfiguration in the micro-grid is a continuous process. In this reconfiguration using of latest problem-solving techniques in the combination of smart meters with Wi-Fi. Segregation of multiple feeders and upgraded online multiple switches enhance the capability of conventional micro-gird to the increase of reliability, online forecasting, enhance power quality, frequency maintenance etc., reconfigured micro-grid model is presented to perform various objectives like operating cost, restoration timing, etc. One of the major constraints will be the radial system and generation limit of micro-grid.

References

1. Central Electricity Authority (2016) Load generation balance report 2015–16. CEA, New Delhi, India. http://cea.nic.in/god/gmd/lgbr_report.pdf
2. Satre J, Deshmukh S (2018) Deviation settlement mechanism linked with market price in Indian power sector. In: IEEE international WIE conference on electrical and computer engineering (WIECON-ECE), pp 168–171. https://doi.org/10.1109/WIECON-ECE.2018.8783024
3. Priya GS, Sivakumar P, Krishnan TS (2017) Survey on combinations of load frequency control, ABT and economic dispatch control in a regulated environment. In: International conference

on computation of power, energy, information and communication (ICCPEIC), pp. 819–828. https://doi.org/10.1109/ICCPEIC.2017.8290480
4. Murali V, Sudha KR (2018) Performance analysis of price based AGC for Indian power system under ABT mechanism. In: 15th IEEE India council international conference (INDICON). https://doi.org/10.1109/INDICON45594.2018.8986980
5. Kirthiga MV, Gurunathan S, Daniel SA (2011) Optimal re-configuration of micro-grids based on ranking of buses. In: IEEE PES innovative smart grid technologies. https://doi.org/10.1109/ISET-India.2011.6145370
6. Kanimozhi K, Pallath D, Kirthiga MV (2016) Adaptive re-configuration and islanding of micro-grids based on frequency deviations. In: IEEE international conference on power electronics, intelligent control and energy systems, pp 1–6. https://doi.org/10.1109/ICPEICES.2016.7853198
7. Gazijahani FS, Hosseinzadeh H, Tagizadeghan N, Salehi J (2017) A new point estimate method for stochastic optimal operation of smart distribution systems considering demand response programs. In: Electrical Power Distribution Network Conference, pp 39–44. https://doi.org/10.1109/EPDC.2017.8012738
8. Gazijahani FS, Hosseinzadeh H, Abadi AA, Salehi J (2017) Optimal day ahead power scheduling of micro-grids considering demand and generation uncertainties. In: 25th Iranian conference on electrical engineering (ICEE), pp 943–948. https://doi.org/10.1109/IranianCEE.2017.7985174
9. Sarkar SJ, Kundu PK (2017) Performance analysis of resumable load data compression algorithm (RLDA) for power system operational data. IEEE Calcutta conference (CALCON), pp 16–20. https://doi.org/10.1109/CALCON.2017.8280687
10. Kumar K, Alam M, Dutta V (2019) Techno-economic analysis of metal hydride-based energy storage system in micro-grid willey. Energy Storage 1–14. https://doi.org/10.1002/est2.62
11. Raol KV, Parvatikar AG, Gokul S, Nitish N, Rao P (2016) A novel fault diagnostic strategy for PV micro grid to achieve reliability centered maintenance. In: IEEE international conference on power electronics. Intelligent control and energy systems, pp 1–4. https://doi.org/10.1109/ICPEICES.2016.7853546
12. Yingcong G, Hanmei P, Yi H, Shuaihu L, Ning W (2018) Short-term reliability assessment method based on time-varying probability ordered tree selection algorithm of islanded micro-grid. In: Proceedings of the 37th Chinese control conference, pp 8914–8919. https://doi.org/10.23919/ChiCC.2018.8482735
13. Abdulgalil MA, Khalid M, Alshehri J (2019) Microgrid reliability evaluation using distributed energy storage systems. In: IEEE PES innovative smart grid technologies Asia. https://doi.org/10.1109/ISGT-Asia.2019.8881068
14. Cheng S, Liu Z (2018) Reliability evaluation of generation system with micro-grid based on glowworm swarm optimization. In: International conference on intelligent green building and smart grid (IGBSG), pp 1–4. https://doi.org/10.1109/IGBSG.2018.8393551
15. Amleh N, Al-Muhaini M, Djokic S (2018) Smart restoration for improved reliability of micro-grids with renewable energy sources. In: IEEE PES innovative smart grid technologies conference Europe (ISGT-Europe). https://doi.org/10.1109/ISGTEurope.2018.8571715
16. Paghdar S, Sipai U, Ambasana K, Chauhan PJ (2017) Active and reactive power control of grid connected distributed generation system. In: Second international conference on electrical, computer and communication technologies (ICECCT). https://doi.org/10.1109/ICECCT.2017.8118019

Wireless Sensor Network in Agriculture: Needs, Challenges and Solutions

Neha Gupta, Preeti Singh, and Pardeep Kaur

Abstract Wireless communication technologies are increasingly growing in today's era, which are providing great research opportunities in the networking area. Wireless sensor network (WSN) is one such example of wireless communication technology. WSNs are widely used in agricultural field in order to help farmers cut down their expenses and increase the profit margin. Precision agriculture (PA) is a management strategy that helps to improve the quality as well as the quantity of the production. In this paper, sensor networks are classified on the basis of different parameters, the various issues and the challenges that are faced while deploying WSNs are also reviewed for improved farming. In this review paper, the comparison of different wireless communication protocols and energy-efficient protocols is analyzed comprehensively.

Keywords Base station · Routing protocol · Sensor nodes · Wireless sensor networks

1 Introduction

With the increasing population, the demand for agricultural crops is also increasing at a rapid pace. As per the reports of UN world population prospects 2019, the population of the world is estimated to rise by 2 billion in the next 30 years, i.e. from 77 billion cur recently to 9.7 billion in 2050 [1]. Also the agricultural land has reduced to 37% according to the food and agriculture organizations. The reasons for this decrease comprise of urbanization, global warming, natural disasters and the shortage of water availability. Wireless sensor network (WSN) technology can alleviate the problem of less agricultural output by implementing the precision agriculture. WSN is a system that comprises of sensors, processing unit, radio frequency transceiver unit and power unit as shown in Fig. 1.

N. Gupta (✉) · P. Singh · P. Kaur
UIET, Panjab University, Chandigarh, India
e-mail: nehagupta101005@gmail.com

P. Singh
e-mail: preeti_singh@pu.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_52

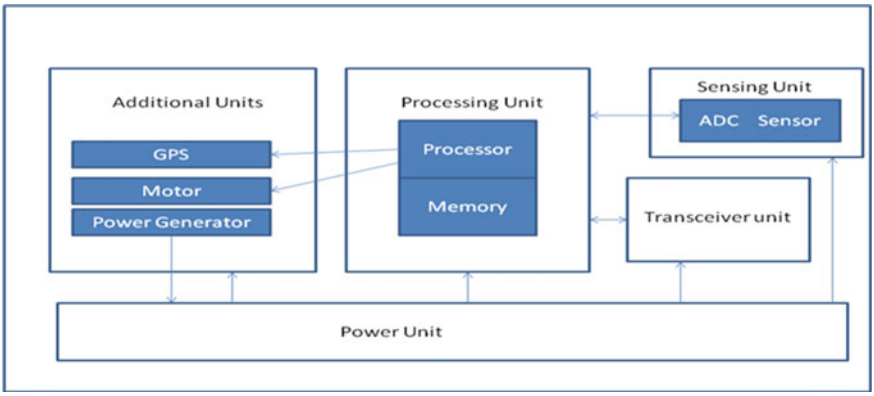


Fig. 1 Wireless sensor node architecture

Wireless sensor networks can be used for various applications in agriculture like precision agriculture, irrigation scheduling and optimization of plant growth, farmland monitoring, agricultural production, process management and security in crops [2]. Precision agriculture is based on the phenomena of collecting the information from the soil using the sensors and sending the information to the sink node either by single-hop or multi-hop communication. The sink node then transmits the collected information to the base station and then the base station (BS) forwards it to the system where the actual decision is taken and the information is finally sent to the farmer regarding the crop as depicted in Fig. 2. There are different concerns that need to be taken into mind while working in the field of wireless sensor networks like quality of service, energy consumption, data processing and compressing techniques [3]. Among all energy consumption in WSNs is a major concern in agricultural field. When the base station and the source nodes are located far away from each other and if there is direct communication between them then a lot of energy gets wasted in

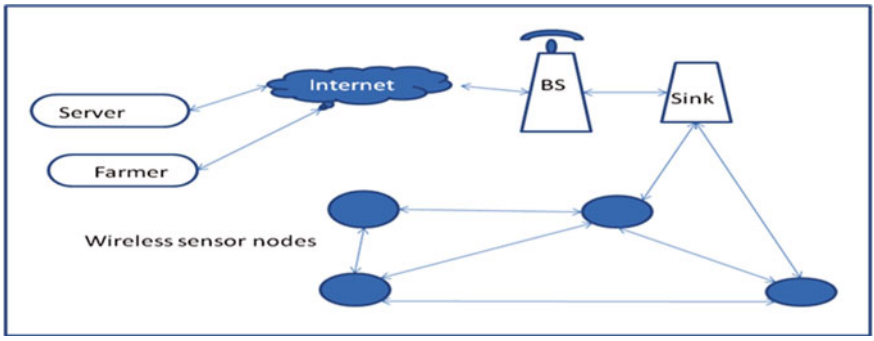


Fig. 2 Deployment of WSN in field

this process. Thus an efficient routing algorithm is needed that minimizes the energy consumption.

1.1 Motivation

In today's era, agriculture sector is a very important factor for economic growth of a country so it should be modernized with technologies and should replace the traditional way of farming. With the availability of micro and cheap sensors in the market, it has enabled advancement in traditional agricultural practices [4]. But there are also few challenges associated with WSN, i.e. energy consumption, less memory of sensors, fault tolerance and computational complexity, which need to be addressed [5]. WSN technology can be applied in different areas like irrigation scheduling by monitoring the moisture of the soil, early warning system of plant health by predicting any invasive species or disease likely to happen. The solutions need to be cost-effective so that the farmers can use these technologies. There is a lot of potential of WSN in agriculture since very limited work is done till date. In today's era of artificial intelligence, wireless sensor networks are widely used in agricultural fields like animal census is being carried out on a large scale. Drone monitoring of the agricultural fields, spraying of pesticides, fertilizers is nowadays a very common trend.

1.2 Organization of the Paper

In this paper, the different techniques that are associated with the WSNs are discussed along with their potential scope for the advancement in agricultural field. The sensor networks are classified on the basis of different parameters, which are used in the agricultural field. Section 2 presents the basics of WSNs and sensors network classification is also discussed. Issues and challenges faced while deploying the sensor nodes are presented in Sect. 3. The standards, technologies that are used in agricultural applications are discussed in Sect. 4. Section 5 discussed the implementation of WSNs in the agricultural field. The existing routing algorithms are discussed and analyzed in Sect. 6. In Sect. 7, the paper is concluded with the scope of future work.

2 Types of Sensor Networks

The sensor networks are classified on the basis of different parameters, which are discussed below. The type of sensor network is selected based upon the requirement of the application. Figure 3 shows the classification of sensor networks with respect to different parameters.

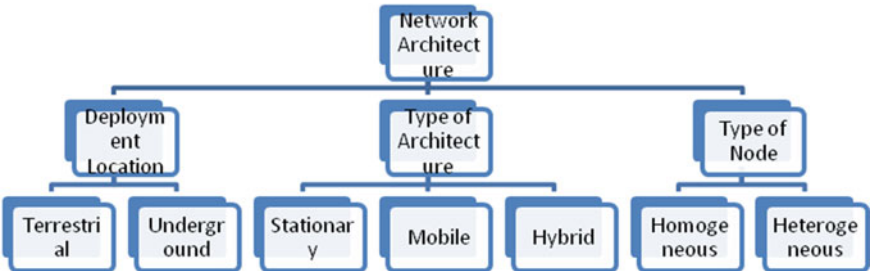


Fig. 3 Classification of sensor networks

2.1 Based on the Deployment Location

The location of the deployment of the sensor nodes is a very important parameter in wireless sensor networks as it determines the viability of the network. The sensor nodes can be deployed below the ground or above the agricultural ground depending upon the requirements of the application.

2.1.1 Terrestrial Wireless Sensor Networks (TWSN)

In such type of sensor networks, sensor nodes are deployed above the ground to create an intelligent network by the means of small and cheap sensors. In precision agriculture field, this type of sensor network is used in soil moisture, temperature and humidity measurement [6]. It is widely used in WSN applications because of less cost, less energy consumption and higher communication range and frequency.

2.1.2 Wireless Underground Sensor Networks (WUSN)

In this type of sensor networks, the sensor nodes are deployed inside the soil in the agricultural field. It is used to check the quality index of the soil [7]. One of the major limitations of these networks is the attenuation of signals at higher frequencies and comparatively lower frequency signals are able to penetrate inside the soil. As a result of which the communication area gets limited and thus more number of nodes are required when compared with TWSN thus it is not widely used in wireless sensor applications.

2.2 Based on the Characteristics of Sensors

Wireless sensor networks have different architectures depending upon the nature of sensor nodes, i.e. whether the autonomous nodes are stationary or moving. Sensing the data input from the nodes plays an important part in an effective communication so sensor nodes should be chosen carefully.

2.2.1 Stationary Architecture

In this type of architecture, the sensor nodes are always stationary and the routing is fixed. Applications like irrigation management, optimizing the use of pesticides, groundwater monitoring require stationary architectures [8].

2.2.2 Mobile Architecture

In this type of architecture, the sensor nodes are continuously moving in order to collect the data from different nodes and give it to the base station or sink. In mobile wireless sensor networks, the routing is dynamic [9]. The major limitations are self-organization of nodes, navigation and control, maintenance, localization.

2.2.3 Hybrid Architecture

In this type of architecture, the sensor nodes are both stationary as well as mobile in nature. It is one of the best architectures when compared with other two architectures in terms of energy and self-organization capability. This type of architecture can be used for tracking the movement of animals, which enters the field and sending the message to the farmer.

2.3 Based on the Sensing and Transmission Power

2.3.1 Homogeneous Network

In this type of network, all the sensor nodes in the network collect the same level of the transmission power with the same parameters of the data, and the collected data are sent to the base station. This type of network is best suited for in situ monitoring of the agricultural field. One of the major drawbacks of this network is the lack of variety in communication hardware.

2.3.2 Heterogeneous Network

In this type of network, all the sensor nodes sense the different parameters of the data having different transmission ranges. These networks use multiple hoping techniques to reach the cluster head of the network [10]. One of the limitations is non-uniform energy drainage because every node requires different amount of energy to send the data to the cluster head based on its location. But the major advantage is the increase in the lifetime of the network.

Thus, it is concluded that uniform energy drainage and lower hardware cost are the two parameters that need to be kept in mind while deploying sensor nodes. On one hand, homogeneous networks have uniform energy drainage and on the other hand, heterogeneous networks have lower hardware cost. The solution can be proposed to include both features in a single network.

2.4 Layout of the Sensor Nodes

Layout means the physical arrangement of the sensor nodes in the agricultural field. Placing the sensor nodes in the field is a very important part of precision agriculture, which should be given proper attention. The agricultural output will eventually grow with time and thus can affect the output of the system. So care should be taken while placing the nodes in the field. The nodes can be placed both horizontally and vertically.

2.4.1 Horizontal Layout

In this type of layout, the nodes are placed in the grid pattern, i.e. nodes are placed in rows and columns forming a grid pattern. Although this layout looks like easy implantation but overlapping of the sensor nodes is a serious problem. Jing et al. have proposed to implement this layout in orchid farms to measure soil moisture, humidity, temperature parameters [11].

2.4.2 Vertical Layout

In this type of layout, the nodes are not placed on the ground, i.e. they are placed on the plants. This is because of the fact that all the plants grow either in upward or downward direction. Hence, this type of layout is better than the horizontal layout. Harris et al. (2016) have placed the nodes at seven different heights of the tomato plant to measure the temperature–humidity parameters [12]. Also, Akkas et al. (2017) have fixed the height of the sensor nodes at one point and monitored the humidity, temperature and pressure parameters [13].

3 Challenges Associated While Deploying Sensor Nodes

Deploying sensor nodes in an organized manner ensures better monitoring of the agricultural field. Although there are various challenges that are faced while placing sensor nodes in an agricultural field. These challenges are discussed as follows:

3.1 Node Size

The size of a sensor is a very crucial factor that needs to be addressed properly. With the advancement of MEMS technology, it has been made possible that small size sensors are available for different parameters like temperature sensors, humidity sensors, wind sensors etc., so appropriate sensors should be chosen as per the requirements of the application [14].

3.2 Energy Consumption

Energy is consumed while performing different tasks like sensing the data, processing it and transmitting the data to the base station. So a lot of energy is required for day-to-day monitoring of the agricultural field. The proposed solution can be the solar energy. It is one of the potential alternatives to battery-powered wireless sensor networks. Greenhouse monitoring using solar panels is one such application.

3.3 Fault Tolerant

The system needs to be fault-tolerant as the nodes are deployed in an open field so there are lots of physical barriers experienced like with the increase in height of crops interference in signals occur, lack of communication between different sensor nodes, blockage by animals, etc. [15]. So in order to address these problems, fault-tolerant techniques need to be designed and addressed. To the best of author's knowledge not much work has been done in this field so there is a lot of potentials to work in this field.

The proposed solution can be the consideration of the coverage area. It is one of the most important aspects to consider while deploying sensor nodes in a network that shows how a particular area or barrier is being monitored. There are three coverage problems, which need to be addressed:

I. Target coverage: To cover only particular points in the monitored area is termed as targets.
II. Area coverage: To cover a particular area of the sensor field.

III. Barrier coverage: It is a round area in which any person, animal can be detected by the sensors, which are deployed inside the circle.

3.4 Transmission Range

Transmission range is also a very important factor since the sensors need to be placed very close to one other for effective communication. But in an open field due to ecological imbalances like rainfall, wind etc. the communication between the sensor nodes gets affected [16]. So one alternative is indoor agriculture, i.e. greenhouse cultivation in which the crops are cultivated in a closed environment and monitored continuously [17]. Another way is to increase the communication range by using different architecture topologies like mesh architecture, multi-tier architecture [18].

4 Communication Technologies Used in Wireless Sensor Network

The data that have been sensed by the sensors need to be communicated to the system where the decision can be taken and conveyed same to the farmer. In this, sensors need to communicate the data to the base station. For efficient way of data communication, the technology should be chosen based on the requirements of the application [19]. A survey of various communication technologies is discussed in this section.

4.1 Zigbee

The origin of Zigbee technology originated in 1991 but it was standardized in 2000 by the Zigbee alliance and came into use [20]. It operates in 2.4 GHz ISM band and is used in communication from layer 3 onwards [21]. This standard operates as IEEE 802.15.4; its function is to provide the network with routing and networking functionalities [22]. It supports various network topologies like master to master or master to slave. In this technology, the system comprises of three different devices: router, Zigbee coordinator and Zigbee end device [23]. The network has one coordinator, which functions as a bridge by handling and storing the data while doing the receiving and transmission tasks [24]. It has the communication range of 10–20 m. Due to its low power consumption and low-cost capabilities, it is a widely used communication technology.

4.2 Bluetooth

The Bluetooth technology was invented in 1994. This standard operates as IEEE 802.15.1 with frequency band 2.45 GHz [25]. It can connect eight different devices within a range of 10 m. This network comprises of a single master and up to seven slaves. This type of arrangement is termed as piconet and when these piconets combine together this arrangement is called scatternet [26]. Hseih et al. (2016) have used this configuration to design an irrigation network. Master initiates the communication and the slaves respond to it [27]. The Bluetooth module is active in the first three layers whereas the host is active in the last two layers. This type of interfacing between these two groups is termed as host controller interface [28].

4.3 Wi-Fi

Wi-Fi nowadays is the trending most used wireless network technology. It stands for "Wireless Fidelity". The origin of Wi-Fi dates back to 1991 and was invented by AT&T in the Netherlands. This standard operates as IEEE 802.11 and uses radio frequency band [29]. It has the range of 20–100 m with a high speed 2–55 Mbps. Thakur et al. (2018), Guo et al. (2018) have used the Wi-Fi module to communicate the data to the farmers. This technology is widely used to connect many devices over a long range [30].

4.4 GPRS

General Packet Radio Service (GPRS) was introduced in the market in 2000 by the European Telecommunications Standards Institute (ETSI) but is now maintained by the third generation partnership project (3GPP) [31]. It is one of the best technologies in terms of latency and throughput. It has a data speed ranging from 56 to 114 Kb/s [32]. When 2G technology combines with GPRS, it is termed as 2.5G, i.e. a technology that is between 2 and 3G Technology. Humberto et al. analyzed this technology as a gateway between Zigbee wireless sensor networks and internet for precision agriculture. Joaquin et al. used the GPRS module [33] to communicate the soil moisture and temperature parameters to the system, which resulted in water savings of up to 90% when compared with traditional irrigation practices.

4.5 WiMAX

Worldwide Interoperability for Microwave Access was introduced in 2001 by the WIMAX forum. This standard is based on the IEEE 802.16 and was introduced as an alternative to cables and DSL [34]. It has data speed up to 1Gb/s. Musha et al. (2014) have used this technology [35] to monitor and control the agricultural field for irrigation system using sensors. However, it is an expensive technology in terms of installation and operational cost. Also, it has poor quality service. This is because in situations when more number of people are accessing the same tower at the same time, it becomes very tough to maintain high quality [36].

4.6 Comparison of Different Communication Technologies

As per Table 1, the comparison of different communication technologies on the basis of different parameters is shown. If the application requires data transmission of longer distance with medium power consumption then WiMAX is the best option. Based upon the requirements and applications, suitable technology can be selected to decrease the cost.

Table 1 Comparison of technologies [38]

Parameters	Zigbee	Bluetooth	WIFI	GPRS	WiMAX
Frequency band	2.5 GHz	2.4 GHz	2.4 GHz	900–1800 MHz	2.5–3 GHz
Range	10–100 m	1–100 m	20–100 m	Gprs coverage area	1–30 m
Data rate	20–250 Kbps	1–24 Mbps	2–54 Mbps	32–48 Kbps	1 Gbps
Cost	Low	Low	High	High	High
Power consumption	Low	Medium	High	High	Medium
Network size	64,000 nodes	8 nodes in piconet	30 nodes	1000 nodes	Varies
Mode of communication	Peer to peer	Master and slave, Peer to Peer	From one point to all devices	From Base station to all devices	From one point to multipoint
Security (bits)	128	128	128	128	128

5 Wireless Sensor Network Applications in Agriculture

Wireless sensor network technology can be applied in the agricultural field to increase the production of the crops and thus increase the profit margin of the farmers. The various applications include.

(1) Monitoring the environmental conditions.
(2) Precision agriculture.

5.1 Monitoring the Environmental Conditions

The environment can be monitored through stationary sensors, which are deployed in the fields. These sensors can sense the water quality, soil quality and send the information to the system where the data can be analyzed and the corresponding action can be taken [37].

Crossbow Technology Inc. (2004) developed a wireless sensor network [39] based on solar power that collected the atmospheric weather condition like moisture in the air through the use of sensors and the collected data were conveyed to the end-users. **Soil** Tsvetelina et al. (2016) developed a wireless sensor network [40] to monitor the soil quality parameters. But, however, there were few limitations in this research work while dealing with an appropriate management technology. There are very few research papers in the monitoring of soli parameters so work can be done in the future in this area. S.N. Shylaja et al. (2017) developed a mobile application [41] for the farmers to help them gather real-time information about the fertility of soil and suggested the right time to put fertilizers in the field.

5.2 Precision Agriculture

Wireless sensors are used in precision agriculture to help in data collection, precision irrigation and technology for communication. In data collection, J.John (2018) developed the shortest path tree-based wireless sensor network [42] to collect the different parameters like relative humidity, atmospheric pressure. The system was composed of different sensors deployed in the field that include atmospheric temperature sensor, relative humidity sensor, soil moisture sensor, soil temperature sensor. In precision irrigation, Y.Hamouda et al. (2018) developed an optimal heterogeneous irrigation system for measuring the content of water in the soil. Kalman filter was used to differentiate the sensed soil moisture and the temperature from the surrounding environment.

6 Different Routing Protocols

Routing protocol reduces the power consumption by minimizing the path between the sensor nodes and the base station. This section analyzes the different routing algorithms used for WSN**.** Haider et al. (2013) have proposed REECH-ME [43] routing protocol in which a node having maximum energy became the cluster head that resulted in a fixed number of cluster head in each round. In this protocol, network lifetime and throughput improved significantly. It has better results when compared with LEACH routing protocol. Amjad et al. (2013) have proposed DREEM–ME [44] routing protocol in which the network area was divided into three concentric circles with center as origin that resulted in a reduction in distance between cluster heads and base station. This protocol has improved network stability and lifetime of the network significantly when compared with LEACH protocol. Nadeem et al. (2013) have proposed gateway-based M-GEAR [45] protocol for WSN in which the area was divided into four logical regions. The base station was placed outside the sensing region and the gateway node was placed at the center. Arati et al. (2001) have proposed TEEN [46] routing protocol for reactive networks. This protocol reduced the transmissions by using hard and soft threshold values. It is best for time-critical applications where lifetime of the network is a crucial issue but if the threshold values are not reached in time then it becomes the drawback of this protocol. Then to overcome this problem, the author proposed a new protocol called APTEEN in which the problem of periodic data collection was resolved. The major limitations of both these protocols are overhead and time constraints. Parul et al. (2010) have proposed TDEEC [47] routing algorithm for a heterogeneous WSN to increase the energy efficiency of the system.

6.1 Comparison of Different Routing Protocols

As per Table 2, the comparison of different routing protocols on the basis of different parameters is shown. Among all the routing protocols, Regional Energy-Efficient

Table 2 Routing protocol comparison

Routing protocol in comparison With LEACH protocol	Network stability	Network lifetime in terms of alive nodes
REECH-ME [42]	79% more	66% more
DREEM-ME [43]	40% more	36% more
MGEAR [44]	50% more	64% more
TEEN [45]	50% more	40% more
APTEEN [46]	70% more	60% more

Cluster Heads based on Maximum Energy (REECH-ME)-based protocol is the best in terms of network stability and network lifetime.

7 Conclusion and Future Work

The agricultural field has a lot of potential for a country's economic growth so proper attention should be given to this area and this can be achieved with the help of wireless sensor networks. In this paper, a variety of communication technologies and power reduction techniques are reviewed. The major issues involved while deploying wireless sensor networks are power consumption, cost and complexity. Power consumption is a very crucial factor in today's era in every field so an efficient routing protocol should be used in order to minimize the energy consumption. While deploying sensors in the field, the cost factor should be kept in mind. The deployment of sensors, communication protocols are very complex, so the generalized structure needs to be designed. In this paper, different communication technologies like Zigbee, GPRS, Wi-Fi and more technologies are reviewed. Among all these technologies, Zigbee is the best technology in terms of transmission data rate. Various routing protocols like MGEAR, TDEEC are also reviewed. Among those, REECH-ME is the best routing protocol in terms of network lifetime and stability. Also, sensor deployment challenges are discussed in this paper. Deployment of wireless sensor nodes in agriculture can increase the profit margin, agricultural output for a farmer and thus boosts the economy of a country.Wireless sensor technologies have a lot of potential in agriculture, which is not yet explored. There are many challenges of wireless sensor networks that are not much explored yet. These include throughput, network lifetime and delay. Balancing the tradeoff between power consumption and network lifetime is an area of concern. Digital signal processing field can be collaborated with this field to solve this issue.

References

1. Akyildiz IF, Kasimoglu IH (2004) Wireless sensor and actor networks: research challenges. Ad Hoc Netw 2(4):351–367
2. Wang N, Zhang N, Wang M (2006) Wireless sensors in agriculture and food industry—recent development and future perspective. Comput Electron Agric 50(1):1–14
3. Ruiz-Garcia L, Lunadei L, Barreiro P, Robla I (2009) A review of wireless sensor technologies and applications in agriculture and food industry: state of the art and current trends. Sensors 9(6):4728–4750
4. Srinivasan A (2006) Handbook of precision agriculture: principles and applications. CRC
5. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2009) A survey on sensor networks. IEEE Commun Mag 40(8):102–114
6. Abowd G, Dey AK, Brown P, Davies N, Smith M, Steggles P (1999) Towards a better understanding of context and context-awareness. In: The workshop on the what, who, where, when,

and how of context-awareness as part of the 2000 conference on human factors in computing systems (CHI 2000), pp 304–307. Springer, The Netherlands
7. Schilit BN, Theimer MM (1994) Disseminating active map information to mobile hosts. IEEE Netw 8(5):22–32
8. Morais R, Valente A, Serôdio C (2005) A wireless sensor network for smart irrigation and environmental monitoring. In: EFITA/WCCA joint congress on IT in agriculture, Portugal, pp 845–850
9. Wang C, Zhao C, Qiao X, Zhang X, Zhang Y (2008) The design of wireless sensor networks node for measuring the greenhouse's environment parameters. In: Computer and computing technologies in agriculture, vol 2, pp 1037–1046. Springer, Boston
10. Kolokotsa D, Saridakis G, Dalamagkidis K, Dolianitis S, Kaliakatsos I (2010) Development of an intelligent indoor environment and energy management system for greenhouses. Energy Convers Manage 51(1):155–168
11. Burrell J, Brooke T, Beckwith R (2004) Vineyard computing: sensor networks in agricultural production. IEEE Pervasive Comput 3(1):38–45
12. Beckwith R, Teibel D, Bowen P (2004) Report from the field: results from an agricultural wireless sensor network. In: 29th annual IEEE international conference on local computer networks, Tampa, FL, USA, pp 471–478
13. Shaikh AZA (2008) Towards design of context-aware sensor grid framework for agriculture. In: Fifth international conference on information technology, XXVIII-WASET conference, Rome, Italy, pp 244–247
14. Goumopoulos C, Christopoulou E, Drossos N, Kameas A (2004) The PLANTS system: enabling mixed societies of communicating plants and artefacts. In: Ambient intelligence, pp 184–195. Springer, Berlin/Heidelberg
15. Kaur P, Sohi BS, Singh P (2018) Recent advances in MAC Protocols for the energy harvesting based WSN: a comprehensive review. Wirel Personal Commun. Springer
16. Basu T, Thool MVR, Thool RC, Birajdar AC (2006) Computer based drip irrigation control system with remote data acquisition system. In: 4th world congress of computers in agriculture and natural resources, USA
17. Escobar C, Galindo J (2004) Fuzzy control in agriculture: simulation software. In: Industrial simulation conference, pp 45–49
18. Ferentinos KP (2018) Deep learning models for plant disease detection and diagnosis. Comput Electron Agric 145:311–318
19. Kim Y, Evans RG, Iversen WM (2008) Remote sensing and control of an irrigation system using a distributed wireless sensor network. IEEE Trans Instrum Meas 57(7):1379–1387
20. Kim Y, Evans RG (2009) Software design for wireless sensor-based site-specific irrigation. Comput Electron Agric 66(2):159–165
21. Cugati S, Miller W, Schueller J (2003) Automation concepts for the variable rate fertilizer applicator for tree farming. In: The proceedings of the 4th European conference in precision agriculture, Berlin, Germany, pp 14–19
22. Ehlert D, Schmerler J, Voelker U (2004) Variable rate nitrogen fertilisation of winter wheat based on a crop density sensor. Precis Agric 5(3):263–273
23. He J, Wang J, He D, Dong J, Wang Y. The design and implementation of a integrated optimal fertilization decision support system. Mathematical and Computer Modelling (in press)
24. Chen X, Zhang F (2006) The establishment of fertilization technology index system based on "3414" fertilizer experiment. China Agricult Technol Extens 22(4):36–39
25. Yanlin H, Shoulun C (2004) Summarization of fertilization model research. Chin J Soil Sci 35(4):493–501
26. Dammer KH (2010) Variable rate application of fungicides, Precision crop protection the challenge and use of heterogeneity, pp 351–365. Springer Science and Business Media
27. Butler Z, Corke P, Peterson R, Rus D (2004) Virtual fences for controlling cows. In: IEEE international conference on robotics and automation (ICRA), New Orleans, LA, pp 4429–4436
28. Radenkovic M, Wietrzyk B (2006) Wireless mobile ad-hoc sensor networks for very large scale cattle monitoring. In: 6th International workshop applications and services in wireless networks (ASWN 06), pp 47–58

29. Andonovic I, Michie C, Gilroy M, Goh HG, Kwong KH, Sasloglou K, Wu T (2010) Wireless sensor networks for cattle health monitoring. In: ICT innovations 2009, pp 1–31. Springer, Berlin Heidelberg
30. Zhang W, Kantor G, Singh S (2004) Integrated wireless sensor/actuator networks in an agricultural application. In: 2nd ACM international conference on embedded networked sensor systems, p 317
31. Aqeel-ur-Rehman ZAS, Yousuf H, Nawaz F, Kirmani M, Kiran S (2010) Crop irrigation control using wireless sensor and actuator network (WSAN), 2nd IEEE International conference on information and emerging technologies (ICIET-2010), Karachi, Pakistan, pp 1–5
32. Mizunuma M, Katoh T, Hata S (2003) Applying IT to farm fields—a wireless LAN. NTT Tech Rev 1(2):56–60
33. Gutiérrez J, Medina JFV, Garibay AN, Gándara MAP (2003) Automated irrigation system using a wireless sensor network and GPRS module. IEEE Trans Instrum Meas 63(1):1–11
34. Hernandez-Perez JA, Garcıa-Alvarado MA, Trystram G, Heyd B (2004) Neural networks for the heat and mass transfer prediction during drying of cassava and mango. Innov Food Sci Emerg Technol 5:57–64
35. Hinnell AC, Lazarovitch N, Furman A, Poulton M, Warrick AW (2010) Neuro-drip: estimation of subsurface wetting patterns for drip irrigation using neural networks. Irrig Sci 28:535–544
36. Kalaivani T, Allirani A, Priya (2011) A survey on Zigbee based wireless sensor networks in agriculture, pp 85–89. IEEE
37. Katariya SS, Gundal SS, Kanawade MT, Mazhar K (2015) Automation in agriculture. Int J Rec Sci Res 6(6):4453–4456
38. Kamilaris A, Prenafeta-Boldú FX (2018) Deep learning in agriculture: a survey. Comput Electron Agric 147:70–90
39. Kavdir S, Guyer DE (2003) Apple grading using fuzzy logic. Turk J Agric 27:375–382
40. John J, Kasbekar GS, Sharma DK, Ramulu V, Baghini MS (2018) Design and implementation of a wireless sensor network for agricultural applications. EAI Endorsed Transactions on Internet of Things, vol 4, issue 16
41. Abbasi AAZ, Shaikh ZA (2008) Building a smart university using RFID technology. In: 2008 International conference on computer science and software engineering (CSSE 2008), Wuhan, China, pp 641–644
42. Haider A, Javaid N, Amjad N, Awan AA, Khan A, Khan N (2013) REECH-ME: regional energy efficient cluster heads based on maximum energy routing protocol for WSNs. In: International conference on broadband and wireless computing, Communication and applications
43. Amjad N, Javaid N, Haider A, Awan AA, Rahman M (2013) DREEM-ME: distributed regional energy efficient multi-hop routing protocol based on maximum energy in WSNs. In: 8th International conference on broadband and wireless computing ,Communication and applications
44. Nadeem Q, Rasheed MB, Javaid N, Khan ZA, Maqsood Y, Din A (2013) M-GEAR: gateway-based energy-aware multi-hop routing protocol for WSNs. In: 8th International conference on broadband and wireless computing ,communication and applications
45. Manjeshwar A, Agrawal DP (2001) TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. In: Proceedings of 15th international parallel and distributed processing symposium
46. Manjeshwar A, Agrawal DP (2002) APTEEN: a hybrid protocol for fficient routing and comprehensive information retrieval in wireless. In: Proceedings of 16th International parallel and distributed processing symposium, Lauderdale, USA
47. Heinzelman W, Chanrakasan A, Balakrishnan H (2000) Energy efficient communication protocol for wireless microsensor networks. In: Proceedings of 33rd Hawaii conference on system sciences

Spin Device-Based Self-Calibrated Comparator

Yogendra Kumar Upadhyaya and Ajay Kumar Kushwaha

Abstract The comparator is the basic block for data converter circuits, designing modern analog and mixed-signal systems, for example, telecommunications that interface with sensing circuits. Several analog to digital converters (ADCs) require low-power, high-speed comparators through less area of chip. The spin device-based self-calibrated comparators are accustomed to design high-speed, low-offset and power ADCs. This paper presents two spin devices based on self-calibrated comparator, which is an area and power efficient compared with other comparators. Our proposed self-calibrated comparator consumes 1.3 μW power and 0.0272 fJ energy at a supply voltage of 0.75 V. The spin devices (MTJ and a Strip), SPICE compatible model-based Verilog-A language and PTM CMOS 45 nm model are used to simulate its performance, like the power and delay, and compared it with dynamic comparators.

Keywords Self-calibrated · Magnetic strip · MTJ · Spin devices · Comparator

1 Introduction

A high-speed, low-offset and power comparator is very attractive for various applications such as storage, sensing circuits, data converters [1, 2]. But, the scaling of MOS transistors supports high-speed and low-power operation, the offset voltage in comparator will be increased because of mismatching in transistors. In conventional designs, pre-amplifiers are employed to decrease offset voltage [3]. Nevertheless, self-calibrated technique requires a high voltage gain to reduce the offset voltage and maintaining effectiveness because of transistors scaling. Moreover, large power consumption of the amplifier is inevitable for realizing a wide bandwidth amplifier [4]. In this technique, the identical input signal of each comparator such as the reference voltage is turned over in flash type ADC, and output node capacitances are used digitally (ideal value). Nevertheless, the calibration time depends on resolution of ADCs. Furthermore, the speed of the comparator is decreased due to the increase in

Y. K. Upadhyaya (✉) · A. K. Kushwaha
Government Engineering College, Raipur, Chhattisgarh, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_53

nodes' capacitances. Additionally, the considerably large size circuits to hold in the calibration are necessary for each comparator. In addition, a noise in the comparator decides the SNR of ADCs. The high-level resolution in SAR ('successive approximation ratio') needs very less noise comparator to stop errors during conversions [5, 6]. A new tactic to diminish the design complication and enhance the measurement quality of a comparator is presented in this paper by using calibration technique, a strip and MTJ with CMOS sense amplifier for its actualization. This report uses a strip [7, 8], reference MTJ [9, 10] along with a CMOS sense amplifier to make a dynamic comparator. The right and left layers in a magnetic strip consist of fixed magnetizations but are opposite in nature. These are called 'pinned/fixed' layers. One layer is in the middle of both is called 'free' layer. Third layer is fixed on upper of 'free' layer in Fig. 3. Here, a MTJ is formed according to the magnetization of 'free' layer and third 'fixed/pinned' layer for sensing state. Applied voltages at both ends decide the propagation of domains in strip from one to other ends.

In this paper, the proposed dynamic self-calibrated magnetic comparator is compared with CMOS dynamic comparator' in terms of area, delay, 'power' and energy.

The paper is arranged in this manner: Sect. 2 illustrates the 'double-tail' comparator. Section 3 contains the proposed spin device-based self-calibrated comparator. Section 4 contains the results and discussion. The conclusion is described in Sect. 5.

2 CMOS Comparator

The symmetrical structure of dynamic comparator [11–13] is displayed in Fig. 1. In pre-charge' phase (CLK' is low), transistor P3 and N7 are OFF' and transistors' P7 and P6 are ON that result in the pre-charging of both nodes f_n/f_p to V_{dd} and discharging of the outputs Out− and Out + nodes to ground. If the voltage IN + is greater than IN− in the evaluation phase (CLK is high) then fp drops faster than fn due to the larger current flow in N6 than N5. In the meantime, f_p is continuously falling and one p-MOS transistor P5 starts turning on pulling the node f_n to V_{dd}. The transistor P5 remains off that allows the discharge of f_p. Therefore, latch regeneration time is reduced due to voltage' difference between f_p and f_n, which rises in an exponential manner (Fig. 2).

3 Proposed Spin Device-Based Self-Calibrated Comparator

The spin devices are used in the design of comparator as follows:

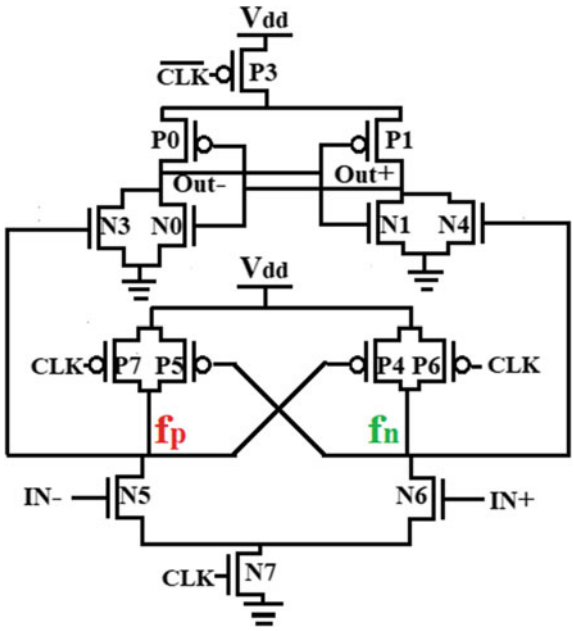


Fig. 1 CMOS-based dynamic comparator

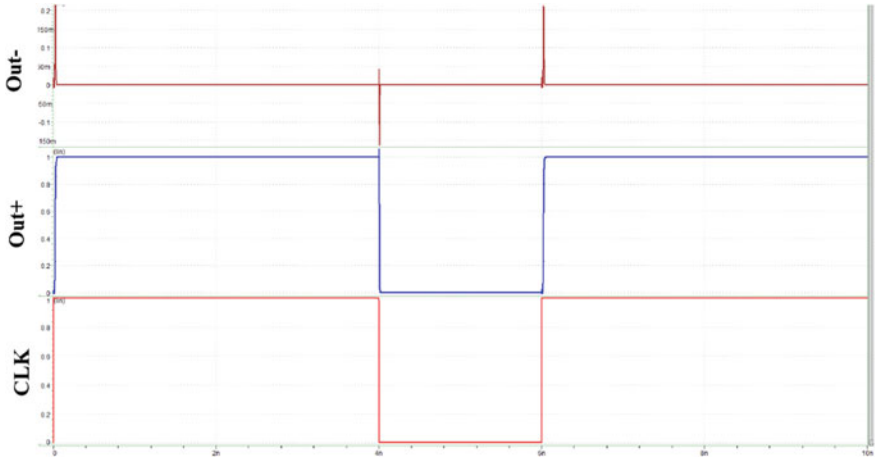


Fig. 2 Simulation of CMOS-based dynamic comparator

3.1 Magnetic Tunnel Junction (MTJ)

Three layers (one free, fixed and one thin layers) is made MTJ by inserting a MgO thin layer between a free and fixed layer. The upper and bottom layers are free (can

change magnetization) and pinned (fixed magnetization) layers, respectively. If the free and the bottom fixed (pinned) layers' magnetization is identical then MTJ is in parallel state ('minimum resistance') and the free layer and bottom pinned layer's magnetization is in reverse manner, then MTJ [10] is in antiparallel/high resistance state.

3.2 Magnetic Strip

A magnetic strip [14, 15] is built by using three layers (two pinned with fixed opposite magnetization) and one free layer between both pinned layer and a third 'fixed/pinned' above the 'free' layer. A detecting MTJ is formed between 'fixed/pinned' layers3 and 'free' layer. Voltages IN− and IN+ are applied at both the ends of the magnetic strip. Domains flow from left to right end (IN− > IN+) of the magnetic strip, and the magnetization of the free layer becomes a left fixed layer. It means that the magnetization of both layers (free layer and top fixed layer 3) is equal, and the sensing MTJ will be in parallel ('minimal resistance') state. Domains flow from right to left end (IN+ > IN−) into the magnetic strip resulting in antiparallel ('high resistive') state of the sensing MTJ. It is easily fabricated above CMOS device with the same area. Parameters that are used for calculating power, delay are shown in Table 1.

Table 1 Parameters for 'power, delay' and modeling of 'spin device-based strip'

Parameter	Description	Values
Technology	Model	45 nm
PMOS	Width of PMOS	240 nm
NMOS	Width of All NMOS	120 nm
V_{dd}	Supply voltage	0.75
T_{CLK}	Time period	2 ns
T_R	CLK's rising time	1 ps
T_F	CLK's falling time	1 ps
TMR	MTJs' tunnel magneto resistance	400%
W	Domain's 'width	65 nm
Area	MTJ' surface	$(65\ nm)^2$
V	Free layer's volume	Area × 1.3 nm
R. A	Resistance area product	$10\ \Omega\mu m^2$
$J_{C\ nucleation}$	Critical current density for domain wall nucleation	$5.7 \times 10^6\ A/cm^2$
$J_{C\ Propagation}$	Critical current density domain wall propagation	$6.2 \times 10^8\ A/cm^2$

3.3 *Proposed Self-Calibrated Comparator with Spin Devices*

The proposed self-calibrated comparator is designed by using spin devices (like magnetic strip and MTJ) with a sense amplifier as shown in Fig. 3. It is arranged in three manners (i) sense amplifiers, (ii) magnetic strip ('based on domain wall motion') and reference MTJ and (iii) self-calibrated circuit. The discharging of fn/fp nodes must be fast for the better performance of dynamic comparator architecture. It is achieved by using spin devices as a comparator instead of CMOS.

A fixed reference MTJ resistor in the dynamic comparator is occupied as the average of the parallel' and anti-parallel resistance of the magnetic strip-based sense MTJ as follows for proper operation [16, 17]:

$$R_{ref} = (R_P + R_{AP})/2$$

A reference MTJ is used as a fixed resistor to minimize the area as well as to track process-based changes in the two arms of the dynamic comparator. The value of the resistance of reference MTJ' will depend upon tunnel magneto resistance ('TMR = $(R_{AP}\text{-}R_p)/ R_p$') of the sense MTJ of the strip.

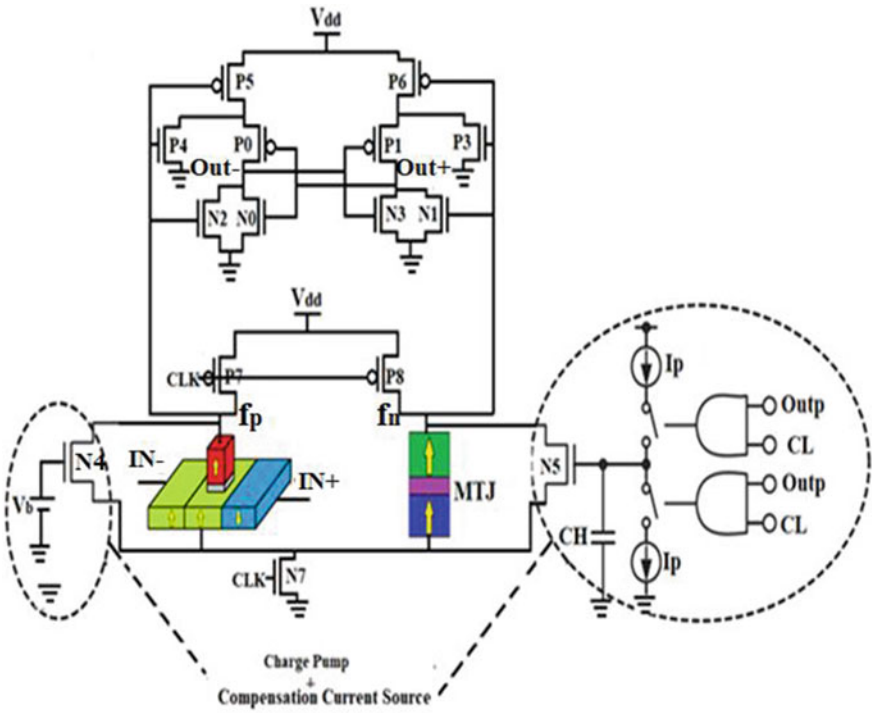


Fig. 3 Proposed spin device-based self-calibrated comparator

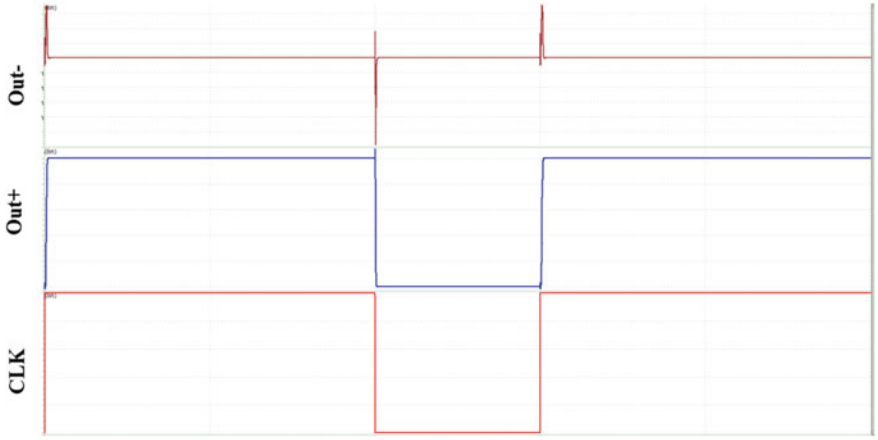


Fig. 4 Simulation of spin device-based self-calibrated comparator

The compensation current is generated through transistors N4 and N5, which are attached to the comparator's inner output node. The transistor N4 is attached to V_b to establish voltage through charge pump. The transistor N5 and capacitor CH are connected to each other and are charged to V_b initially. If the comparator contains the offset voltage, the output is high, and draws charge from CH through current source Icp. The current source's control voltage ('Vc') drops, as result the comparator's offset voltage moves toward zero. While Vc surpasses offset voltage, the comparator will adapt zero and the outputs will vary in terms of high and low alternately (Fig. 3). CH retains offset value in conversion' mode. Thus, the offset voltage is canceled in this way [3, 18] (Fig. 4).

Input voltage IN− is more than IN+ then low resistance sensing MTJ is formed. Reference MTJ has more resistance then sensing MTJ of magnetic strip, node fp is discharged faster as compared to node fn. Out+ is discharged to ground. In case IN+ is greater than IN− then high resistance sensing MTJ is formed. Reference MTJ has less resistance than sensing MTJ of magnetic strip, node fn is discharged faster than fp. Out− is discharged to ground [12, 19].

3.4 Monte-Carlo Simulation

Figure 5 shows Monte Carlo results of the proposed spin device-based self-calibrated dynamic comparator using 45 nm PTM and Verilog-A models. The variations are taken to be 10% in each parameter with 500, 1000 and 10,000 samples. The samples are mismatched during simulation 23, 39 and 282 out of 500, 1000, 10,000 samples, respectively, for the proposed dynamic comparator.

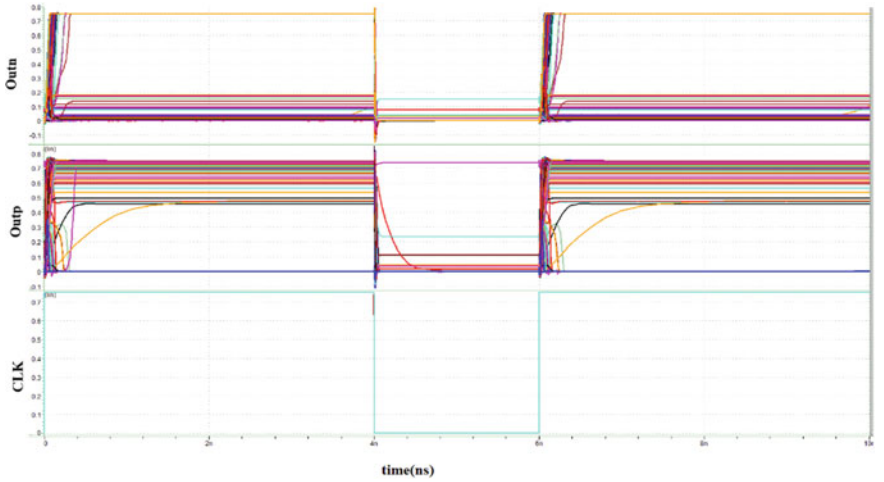


Fig. 5 Monte Carlo simulation of a proposed spin device-based self-calibrated comparator for 1000 samples

4 Comparison Between CMOS Comparator and Spin Device-Based Self Calibrated Comparator

The simulation results are obtained using H-Spice simulation and Verilog-A compact model-based strip (based domain wall motion). Reference MTJ and strip's TMR is 400% and 120%, respectively. Here, it is assumed that parallel ('minimal resistance, R_P') state of 2 KΩ and an antiparallel state ('large resistance, R_{AP}') of 10 KΩ for strip's forming sense MTJ and a parallel state ('low resistance, R_P') of 1 KΩ and 2.2 KΩ anti-parallel ('large resistance, R_{AP}') state for reference resistance. The resistance of reference MTJ depends upon its TMR and its desired value (average of 'R_{AP}' and 'R_P' of forming sense MTJ of the strip) is adjusted by properly choosing its TMR. By including the MTJ and strip-based self-calibrated comparator along with CMOS sense amplifier in the design of a dynamic comparator, its area and power can be diminished maintaining its speed. The MTJ and strip reduce area because of fabrication over CMOS logic. The performance of both (proposed self-calibrated and CMOS dynamic double-tail) comparators is shown in Table 2.

5 Conclusion

This paper successfully proposed spin device-based self-calibrated low power, high-speed, low-offset and high-density comparator. It offers much better performance in terms of 'area' delay, power and compared with the CMOS-based dynamic double-tail comparator.

Table 2 Power consumption and delay for proposed spin device-based self-calibrated comparator

Comparator structure	CMOS comparator	Proposed spin device-based self-calibrated comparator
Technology	45 nm	45 nm
Supply voltage (V)	0.75	0.75
Maximum sampling frequency	500 MHz	500 MHz
Delay (ps)	23.64	20.91
Power consumed (μW)	76.385	1.30
Energy (fJ)	1.806	0.0272

Acknowledgements This work is supported by the 'Third Phase of Technical Education Quality Improvement Program' (TEQIP-III) under the 'TEQIP Collaborative Research Scheme', 'National Project Implementation Unit' (NPIU), a unit of the 'Ministry of Human Resource Development' (MHRD), Government of India for implementation of 'World Bank'-assisted projects in 'technical education'.

References

1. Upadhyaya YK, Gupta MG, Hasan M, Maheshwari S (2016) High density magnetic flash ADC using domain wall motion and pre-charge sense amplifiers. IEEE Trans Magn 52(6)
2. Gall B, Zimmermann H (2009) A comparator with reduced delay time in 65-nm CMOS for supply voltages down to 0.65. IEEE Trans Circuits Syst II, Exp Briefs 56(11):810–814
3. Miyahara M, Asada Y, Paik D, Matsuzawa A (2018) A low-noise self-calibrating dynamic comparator for high-speed ADCs. In: 2008 IEEE Asian solid-state circuits conference, pp 269–271
4. Goll B, Zimmermann H (2007) Low power 600MHz comparator for 0.5 V supply voltage in 0.12 μm CMOS. IEEE Electron Lett 43(7):388–390
5. Gupta MK, Hasan M (2015) Robust high-speed ternary magnetic content addressable memory. IEEE Trans Electron Devices 62(4):1163–1169
6. Gupta MK, Hasan M (2015) Design of high-speed energy efficient masking error immune Penta-MTJ based TCAM. IEEE Trans Magn 51(2):3400209
7. Zhang Y, Zhao WS (2012) Perpendicular-magnetic-anisotropy CoFeB racetrack memory. J App Phys 111(5)
8. Atoofian E (2015) Reducing shift penalty in domain wall memory through register locality. In: Proceedings of the 2015 international conference on compilers, architecture and synthesis for embedded systems (CASES '15). IEEE Press, Piscataway, NJ, USA, pp 177–186
9. Sharad M, Roy K (2013) Spintronic switches for ultra low energy on-chip and interchip current-mode interconnects. IEEE Electron Device Lett 34(8)
10. Zhang Y, Zhao W, Lakys Y et al (2012) Compact modeling of perpendicular-anisotropy CoFeB/MgO magnetic tunnel junctions. IEEE Trans Electron Devices 59(3):169–177
11. Babayan-Mashhadi S, Lotfi R (2012) An offset cancellation technique for comparators using body-voltage trimming. J Analog Integr Circuits Signal Process 73(3):673–682
12. Babayan-Mashhadi S, Lotfi R (2014) Analysis and design of low power double tail comparator. IEEE Trans Very Large-Scale Integr (VLSI) Syst 22(2):343–352
13. Lin Y, Jyh Y, Liu Y, Liu C, Huang G (2009) A 5b 800 MS/s 2mW asynchronous binary-search ADC in 65nm CMOS. IEEE Int Solid-State Circuits Conf (ISSCC) Dig Tech Papers 80–81

14. Upadhyaya YK, Hasan M (2016) High speed area efficient 5-bit hybrid Flash ADC using spintronic devices and CMOS. Int J Emerg Adv Electron Electr Eng (IJEAEEE) 1(10)
15. Upadhyaya YK, Hasan M, Maheshwari S (2017) Low power and high density magnetic flash analog to digital converter using spintronic devices and CMOS. In 2017 international conference on multimedia, signal processing and communication technologies (IMPACT), pp 237–241
16. Upadhyaya YK, Hasan M, Maheshwari S (2018) Low power, high speed and high-density magnetic comparator using spintronic devices. Nano Trends J Nano Technol Appl 20(1)
17. Upadhyaya, YK, Hasan M, Maheshwari S (2018) High speed and low power sensing amplifier based on MTJ. In: 2018 5th IEEE Uttar Pradesh section international conference on electrical, electronics and computer engineering (UPCON), pp 1–3
18. Sharma P, Upadhyaya YK, Singh K (2018) Design of ultra low-power robust SRAM cell. Nano Trends J Nanotechnol Its Appl 21(2):7–18. ISSN:0973–418X (online)
19. Upadhyaya YK, Hasan M (2019) Energy efficient robust pre-charge sense amplifier. In: 2019 international conference on electrical, electronics and computer engineering (UPCON), pp 1–5

Performance Analysis of Graphene-Based Field-Effect Transistors in Ternary Logic: A Review

Badugu Divya Madhuri and S. Sunithamani

Abstract Considering the development of carbon nanotube field-effect transistors (CNTFETs) and graphene nanoribbon FETs (GNRFETs), this paper presents a review of previous work. The ternary logic design is promising than the conventional binary logic as it offers reduced interconnects, fast computations, and reduced area. Hence, the ternary logic is introduced in both CNTFET and GNRFET-based circuits. Using the basic gates such as standard ternary inverter (STI), positive ternary inverter (PTI), negative ternary inverter (NTI), NAND gate, and NOR gate, the complex circuits are designed. The complex circuits such as half adder, one-bit multiplier, memory cell, and multiplier are designed. All the designs are simulated in HSPICE to verify the correctness. From the observations, it is noticed that the GNRFET-based ternary circuits provide high performance for future integrated circuits (ICs) over the CNTFET circuits.

Keywords Carbon nanotubes · Graphene nanoribbons · VLSI circuits · CNTFET · GNRFET · Ternary logic

1 Introduction

Due to the continuous development in the semiconductor industry, the electronic gadgets become a primary need for human life. According to the Moore's law, the FETs count on a chip doubles for every 2 years with reduced cost [1–6]. Traditionally, metal oxide semiconductor FETs (MOSFETs) are used to design the integrated circuits [7–9]. A 10-nanometer feature size can allow a tera-scale chip with a trillion number of transistors according to the ITRS [10]. This extraordinary growth has been attained by downscaling the size of digital integrated circuit (IC). However, scaling the traditional transistor size in nanometer range leads to problems such as large

B. D. Madhuri (✉)
Department of ECE, RGM College of Engineering and Technology, Nandyal, India

S. Sunithamani
Koneru Lakshmaiah Education Foundation, Guntur 522502, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_54

leakage current, complex interconnect structures, and power density. These difficulties can reduce the performance of the ICs. Therefore, the semiconductor industry introduced the graphene materials such as CNTs and GNRs. Over past decades, these materials attracted researchers because of their excellent physical and chemical properties [11–20]. Recently, the use of CNTs and GNRs as channel material is being analyzed to develop new devices called CNTFET and GNRFET [21, 22]. Because of their outstanding physical properties and ballistic transport, the CNTFETs and GNRFETs are considered as the most promising candidate for designing future ICs.

In recent days, ternary logics are considered as best because of their benefits than the binary logics for implementing high-performance digital system [23–25]. The ternary logic systems allow three logic states, while binary logic system is carried out with two logic states. However, the disadvantages of binary logic are the interconnection and pin-out problems. Moreover, many applications such as decision support system, robotics, and process control are not designed with two logic states. To avoid these defects, the ternary logic can be used instead of the binary logic. The three-value logic can degrade the circuit complexity and chip area and enhances the rate of serial and parallel communication. Additionally, the arithmetic operations can be evaluated faster with reduced computation steps in three-value logic. Among three-value logic design, utilizing e base ($e = 2.718\ldots.$) gives the well-organized implementation. The most efficient three-value logic offers low cost and complexity over the binary logic [26]. Thus, introducing the ternary logic provides the high performance in transistors.

In recent days, a lot of articles studied to design, model, and integration of CNT and GNR-based transistors [27–31]. In [27–31], the CNTFET and GNRFET ternary arithmetic circuits are presented. In particular, this paper focuses on the design of ternary logic circuits using CNTFETs and GNRFETs for future digital ICs.

2 Carbon Nanotubes and Graphene Nanoribbon Transistors

In this section, the review of the design of CNTFETs and GNRFETs is presented. The mathematical equations to calculate the bandgap and threshold voltage of both CNTFET and GNRFET are also described. Furthermore, the suitability of CNTFET and GNRFET in ternary logic is discussed.

2.1 Carbon Nanotube Field-Effect Transistor

CNTFETs use semiconductive SWCNTs to develop electronic gadgets [27–29]. A SWCNT has only one CNT, which makes the device auspicious than conventional Si transistors. Based on the chirality vector, the SWCNT can act either as semiconductor

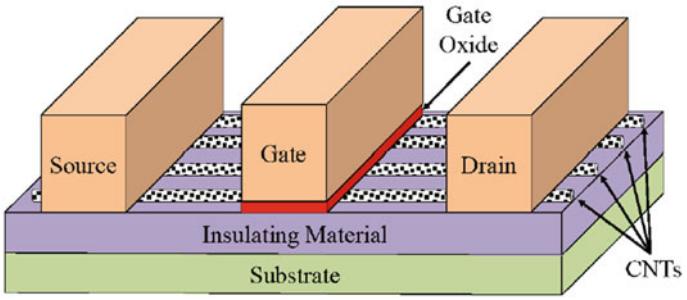


Fig. 1 Schematic diagram of a CNTFET

or metal. The chirality vector is characterized by the integer pair (n, m). An efficient technique to govern whether the CNT is semiconductor or conductor is to take its indexes (n, m): the CNT acts as conductor for $n = m$ or $n - m = 3i$, where i is an integer. Else, the CNT acts as semiconductor. The CNT diameter is evaluated as [27–29]:

$$D_{\text{CNT}} = \frac{\sqrt{3}a_0}{\pi}\sqrt{n^2 + m^2 + nm} \tag{1}$$

where $a_0 = 0.142$ nm. The circuit of CNTFET is shown in Fig. 1. The threshold voltage is the minimum voltage value needed to ON the FET. The CNTFET threshold voltage can be stated as [27–29].

$$V_{\text{th}} \approx \frac{E_g}{2e} = \frac{\sqrt{3}}{3}\frac{aV_\pi}{eD_{CNT}} \tag{2}$$

where $a = 2.49°\text{A}$, $V\pi = 3.033$ eV, e is the electron charge, and D_{CNT} is the diameter of CNT. Figure 2 shows both PCNTFET and NCNTFET threshold voltage values [27]. From Eq. (2), it is noticed that the threshold voltage is varied as CNT diameter is varied and thus it is altered to an essential value by selecting the CNT chiral vector. This makes the CNTFETs promising to be utilized for three value logics. The CNTFET I–V curves are shown in Fig. 3 [28], which looks same as the conventional transistor. Due to obtaining the various threshold voltage and conventional like (I–V) curves, CNTFET is compatible with three value logic implementations.

2.2 Graphene Nanoribbon Field-Effect Transistor

In contrast to the CNT, the GNR can be more compatible [32]. The GNR also shows the superior properties to the CNT for designing the FETs. Moreover, the GNRs

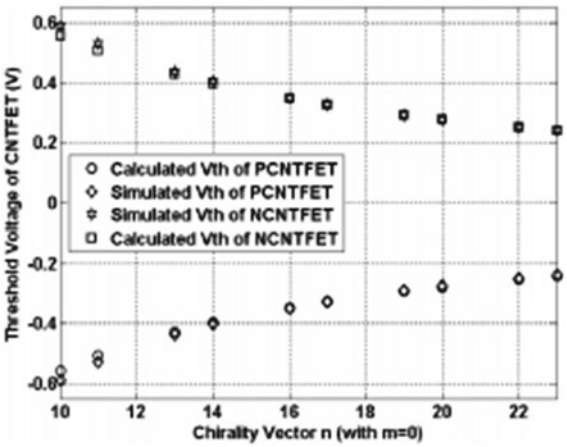


Fig. 2 CNTFET threshold voltage for various n [27]

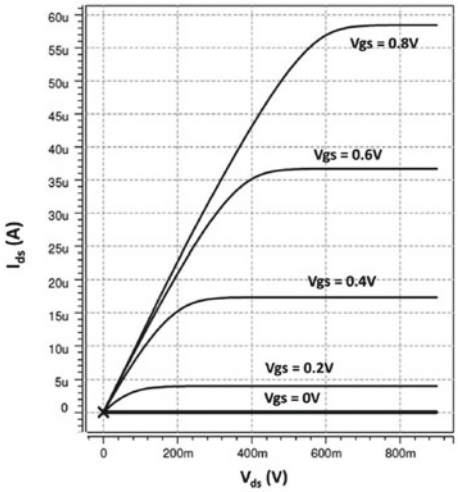


Fig. 3 I–V curve of CNTFET [28]

provide the effective performance over CNTs as they have smooth edges and fewer defects [33]. The multi-threshold devices can be implemented by utilizing GNRs with various widths. Moreover, the transistor current driving capacity can be improved by increasing the graphene channels. Thus, GNRFET is the best alternative to the CNTFET as it provides high throughput.

The GNRFETs use semiconducting armchair GNR as channel materials. The structure of GNRFET is shown in Fig. 4. The mathematical equation of GNRFET threshold voltage is as follows [30, 31].

$$V_{\text{th}} \approx \frac{E_g}{3e} \tag{3}$$

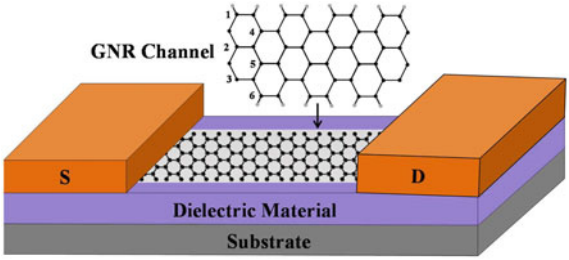


Fig. 4 Structure of GNRFET

The bandgap of the GNR (E_g) can be stated as

$$E_g = 2|\alpha|\Delta E \tag{4}$$

$$\Delta E = \frac{h\upsilon_f \pi}{W} \tag{5}$$

$$W = (n+1)\frac{\sqrt{3}}{2}a \tag{6}$$

where E_g is the bandgap, W is the GNR width, h is the Planck's constant, a is the C–C bond distance, v_f is the Fermi velocity, and, n is the dimer line number. Figure 5 shows the threshold voltage of GNRFET for various dimer lines [30]. It is investigated that the threshold voltage value is reduced as the dimer line value is increased. Due to the possibility of different threshold voltage, GNRFET is also compatible to develop the circuits in ternary logic.

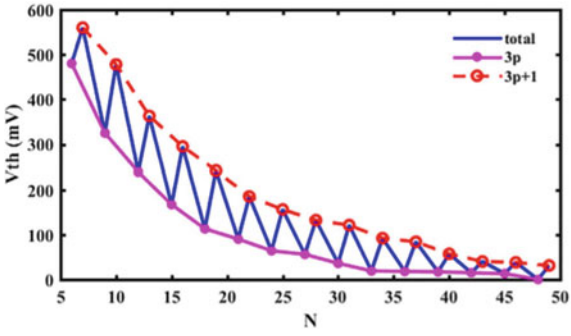


Fig. 5 Threshold voltage of GNRFET versus dimer lines (from [30])

3 Carbon Nanotube and Graphene Nanoribbon-Based Ternary Logic Designs

The design of three value logic arithmetic circuits using CNTFETs and GNRFETs is presented in Refs. [27–29]. The review on designing the complex circuits using basic logic gates is presented. All the CNTFET and GNRFET-based circuits are simulated in industry standard HSPICE to verify correctness.

3.1 Ternary Half Adder Using CNTFETs

The ternary logic gates' design utilizing CNTFETs is developed for the future integrated circuits. Traditionally, a resistive-load CNTFET circuit has been demonstrated to design ternary arithmetic circuits [34]. A method for implementing the CNTFET-based ternary logic gates is discussed in [27]. The basic gates such STI, PTI, NTI inverters are designed to develop the complex circuits. The design of ternary logic gate method is combined with the design method of two valued logic gate to obtain the high speed and reduced power consumption. Figure 6a shows the ternary half-adder circuit diagram. The adder is simulated in the industry standard HSPICE to obtain the output response using the CNTFET model presented in [35]. The obtained response of the half adder is given in Fig. 6b. Furthermore, one-bit multiplier is developed and investigated the output response in Fig. 6c. From the analysis, it is noticed that the presented circuits consume less power and delay compared with the resistive-load circuit.

In [28], a pair of ternary half-adder circuits are developed utilizing CNTFETs. In these designs, the innovative ternary and traditional binary logic schemes are combined. One of the developed designs for ternary to binary decoder shortens the schematic design and delivers tremendous power and delay benefits in adder circuit. The half adder circuit and its output response are shown in Fig. 7. The performance of circuits is associated with traditional designs presented in published articles. One of the developed adders established that power and power delay product reduction up to 63% and 66%, respectively, with reduced FET number. Hence, the utilization of the presented adders in future VLSI designs can be profitable.

3.2 Ternary Memory Using CNTFETs

The ternary memory cell implementation utilizing the CNTFETs is presented in [29]. For writing operation, this memory cell utilizes the transmission gate and uses a buffer for read operation. The CNTFET threshold voltage is operated by the CNT chirality. This avoids the utilization of extra power supply. Figure 8a shows the CNTFET-based ternary memory cell. The memory cells' write and read operations

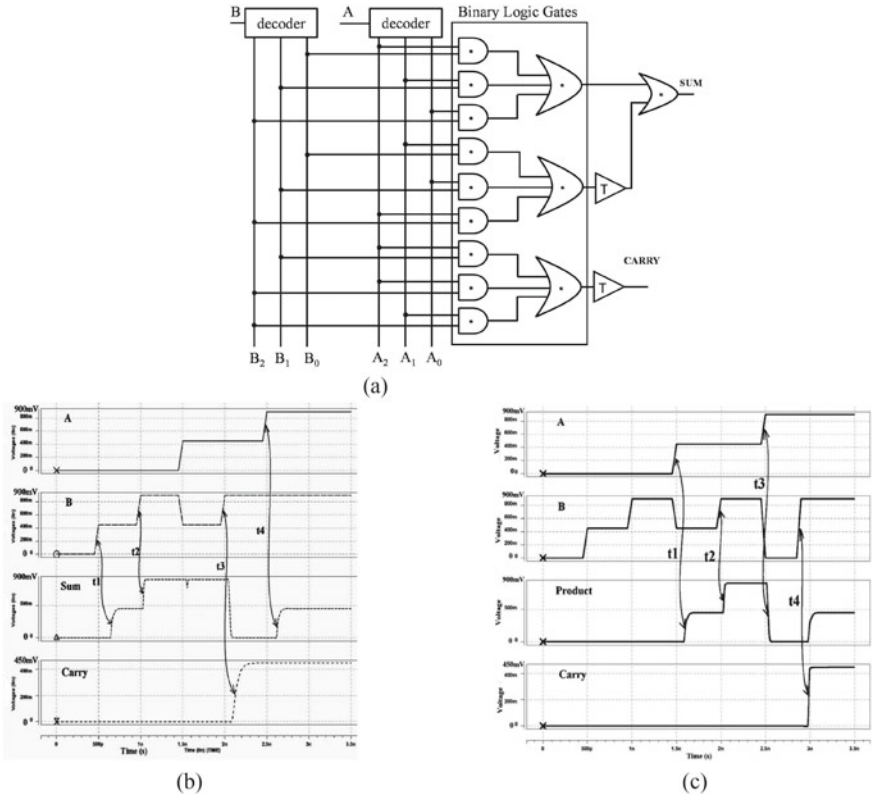


Fig. 6 CNTFET-based circuit **a** half adder, **b** output response of half adder, and **c** output response of multiplier (Reproduced with permission from [27])

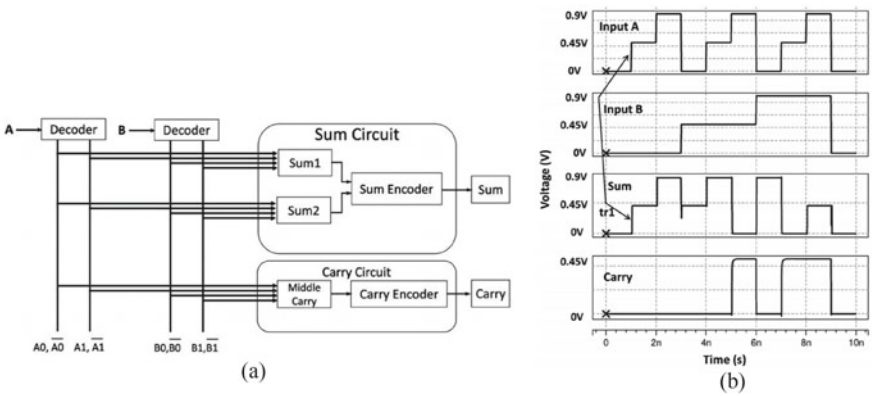


Fig. 7 Half adder **a** circuit diagram and **b** transient response (Reproduced with permission from [28])

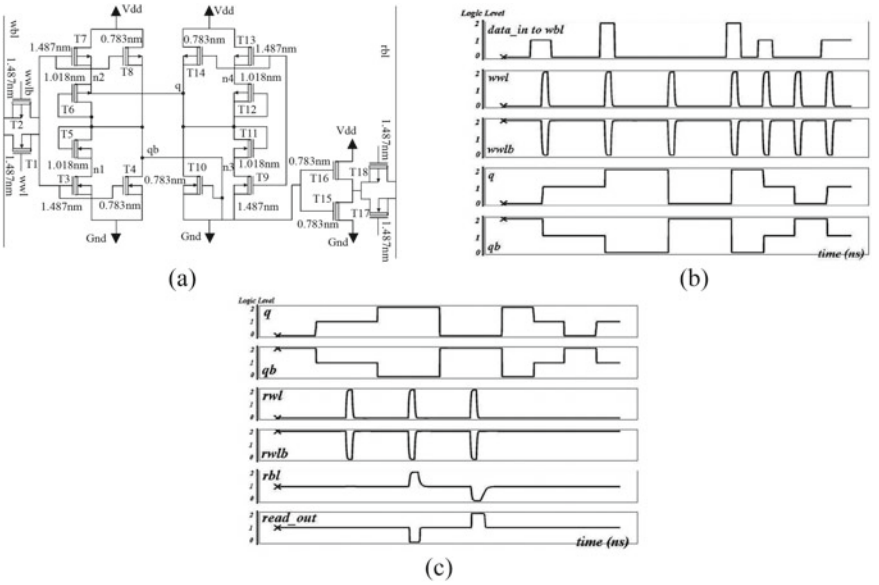


Fig. 8 CNTFET-based memory cell **a** circuit diagram, **b** write operation, and **c** read operation (Reproduced with permission from [29])

are shown in Fig. 8b. The read/write delay and static noise margin of the cell are promising, and it is noticed that memory cell attains a major saving in area (41.6%) than the Si-based ternary circuit.

3.3 GNRFET-Based Ternary Half Adder

The GNRFET-based logic and arithmetic circuits in ternary logic are discussed in [30, 31]. The half adder and 3:1 based multiplexer are developed. Figure 9a shows the ternary half adder circuit representation [30]. The SPICE model of GNRFET mentioned in [36] is used to investigate the simulations. The simulation results of the half adder from reference [30, 31] are given in Fig. 9b, c. Furthermore, the circuit diagram of 3:1 multiplexer and transient response is shown in Fig. 10.

3.4 Comparative Study

The performance comparison is analyzed between GNRFET and CNTFET-based ternary half adders to show the effectiveness of the approach. The power delay product of proposed GNRFET ternary circuits is compared with the CNTFET and placed in

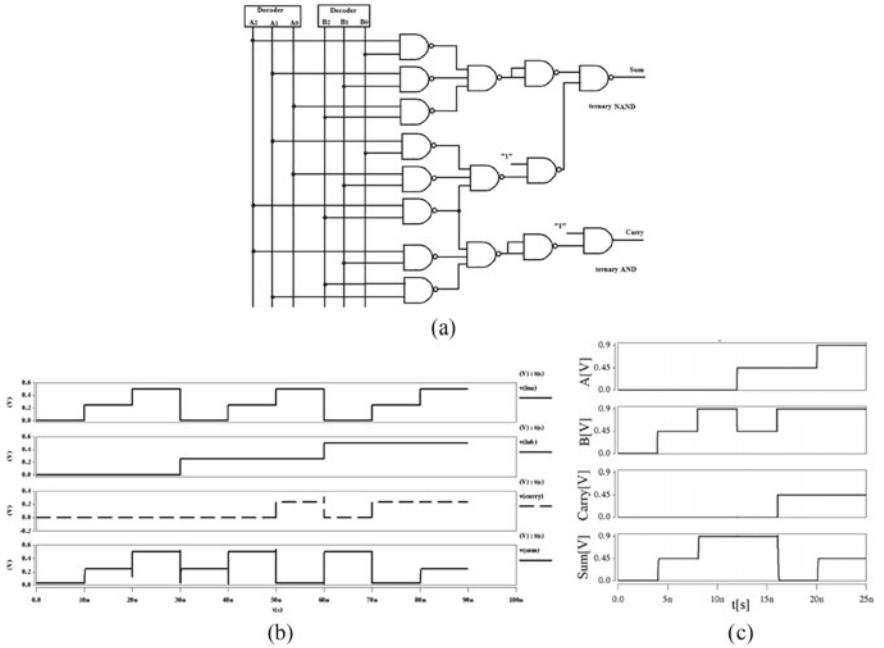


Fig. 9 Circuit diagram **a** half adder [30], **b** output response [30], **c** output response [31]

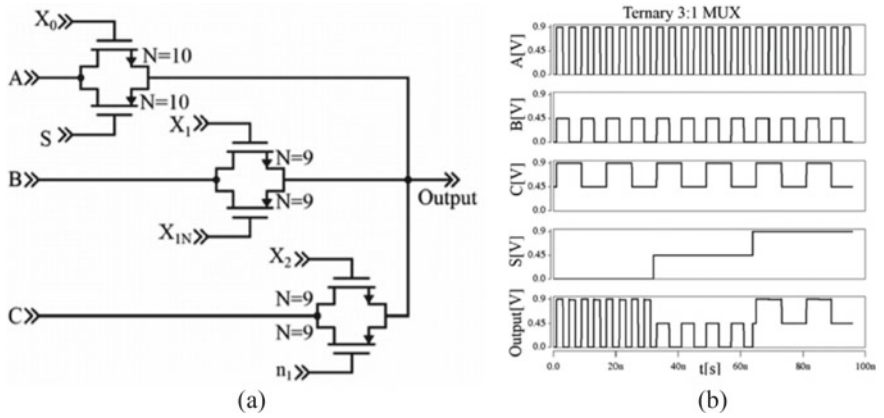


Fig. 10 Circuit diagram **a** 3:1 multiplexer, and **b** transient response (Reproduced with permission from [23])

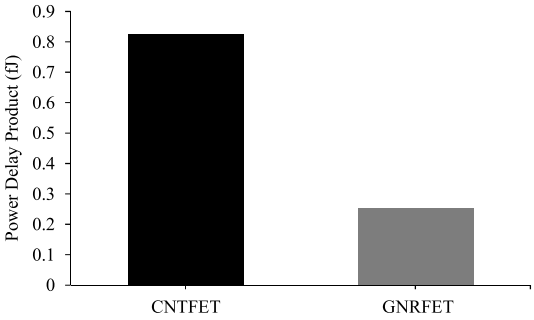


Fig. 11 Comparisons of power delay product for CNTFET and GNRFET-based half adders

Fig. 11. The comparative results specify that the GNRFET-based half adders have low power delay products compared with CNTFET-based with an average reduction is 69%. It is investigated that the GNRFET circuits dissipate high power over the CNTFETs because of the large current driving capability. The average increase in power consumption is observed to be 8%. But still, the GNRFETs are considered to be a promising device for the implementation of ternary logics because of fabrication compatibility due to planar structure, offer high tensile strength, and provide better endurance compared with the CNTFETs.

4 Conclusion

The current status of research on CNTFET and GNRFET-based transistors for developing the complex circuits has been reviewed in this paper. Numerous typical CNTFET and GNRFET-based ternary circuits have been evaluated and discussed. Traditionally, it is demonstrated that the silicon-based ternary circuits are utilized for developing the ICs. However, due to scaling the channel in MOSFETs, the performance of the ICs is reduced. The ternary logic-based transistors using CNTFETs and GNRFETs are the practical solutions to meet the future ICs. It is discussed that GNRFETs have superior properties over the CNTFET, which provide high performance. It is concluded that the GNRFET ternary circuits can be used in long-term digital applications.

References

1. Moore GE (2006) Cramming more components onto integrated circuits, reprinted from electronics, vol 38, number 8, april 19, 1965, pp 114 ff. IEEE Solid-State Circuits Soc Newsl 11(3):33–35
2. Siva Kumar M, Tulasi SK, Arunkanth R, Nandini M, Sudheeer Kumar G (2017) A third order sigma delta modulator in 45nm CMOS technology. J Theor Appl Inf Technol 95(10):2139–2146

3. Patjoshi RK, Suvarsha C, Ali SKI, Basha SKM, Anjum D (2017) Design and analysis of novel high performance CMOS domino-logic for high speed applications. ARPN J Eng Appl Sci 12(17):5109–5114
4. Sarath Chandra K, Hari Kishore K (2018) Physical design implementation of high performance CMOS dynamic latch comparator. J Adv Res Dyn Control Syst 10(9):323–332
5. Ramakrishna P, Kishore HK (2018) Design of low power 10GS/s 6-Bit DAC using CMOS technology. Int J Eng Technol (UAE) 7(1.5):226–229
6. Prasanna JL, Sahiti V, Raghuveera E, Kumar MR (2018) CMOS based power efficient digital comparator with parallel prefix tree structure. Int J Eng Technol (UAE) 7(2.7):647–651
7. Paradhasaradhi D, Sai Kumar Reddy, GLV, Manideep G, Amar Kumar YLNDV (2019) High efficient CMOS rectifier with reduced leakage for low powered bio-implantable devices. Int J Innov Technol Explor Eng 8(7):112–116
8. Prasanna JL, Lasya KNJ, Abbas MS, Sushmanth S (2019) CMOS inverter with second function. Int J Innov Technol Explor Eng 8(7):2812–2816
9. Pavan Kumar KVKVLP, Prabhakar VSV, Bhavani MD, Geetha K, Venkatesh MK, Kishore H (2019) Design and analysis of CMOS schmitt trigger. Int J Innov Technol Explor Eng 7
10. International technology roadmap for semiconductors (2013). http://www.itrs2.net/. Last Accessed 1 Aug 2017
11. Kumar A, Kumar VR, Kaushik BK (2019) Transient analysis of crosstalk induced effects in mixed CNT bundle interconnects using FDTD technique. IEEE Trans Electromagn Compat 61(5):1621–1629
12. Kumar VR, Kaushik BK, Patnaik A (2014) Modeling of crosstalk effects in coupled MLGNR interconnects based on FDTD method. In: 2014 IEEE 64th electronic components and technology conference (ECTC), Orlando, FL, pp 1091–1097
13. Basha, S. J., V. R. Kumar.: Design of MWCNT based Through Silicon Vias with Polymer Liners to Reduce the Crosstalk Effects. ECS Journal of Solid State Science and Technology 9(4), 041002 (2020).
14. Kolanti TNJ, Vemu S, Kumar VR, Rao SK (2018) Modeling of CMS-based nonuniform interconnects using FDTD technique. Int J Circuit Theory Appl 47(1):1–12
15. Kumar VR, Lohith MS, Basha SJ, Reddy RM (2020) Graphene nanoribbon tunnel FET for low-power nanoscale IC design. In: Dhiman R, Chandel R (eds) Nanoscale VLSI. Energy Systems in Electrical Engineering. Springer, Singapore
16. Kumar VR, Kaushik BK, Patnaik A (2015) Improved crosstalk noise modeling of MWCNT interconnects using FDTD technique. Microelectron J 46(12):1263–1268
17. Majumder MK, Das PK, Kumar VR, Kaushik BK (2015) Crosstalk induced delay analysis of randomly distributed mixed CNT bundle interconnect. J Circuits Syst Comput 24(10):1500145
18. Kumar VR, Kaushik BK, Patnaik A (2015) Crosstalk noise modeling of multiwall carbon nanotube (MWCNT) interconnects using finite-difference time-domain (FDTD) technique. Microelectron Reliab 55(1):155–163
19. Kaushik VR, Majumder MK, Kumar VR (2014) Carbon nanotube based 3-D interconnects-A reality or A distant dream. IEEE Circuits Syst Mag 14(4):16–35
20. Kumar VR, Kaushik BK, Patnaik A (2014) An accurate model for dynamic crosstalk analysis of CMOS gate driven on-chip interconnects using FDTD method. Microelectron J 45(4):441–448
21. Gadgil S, Vudadha C (2020) Design of CNTFET-based ternary ALU using 2:1 multiplexer based approach. IEEE Trans Nanotechnol 19:661–671
22. Madhuri BD, Sunithamani S (2019) Design of ternary D-latch using graphene nanoribbon field effect transistor. In: 2019 International conference on vision towards emerging trends in communication and networking (ViTECoN), Vellore, India, pp 1–4
23. Rezaei Khezeli M, Moaiyeri MH, Jalali A (2017) Analysis of crosstalk effects for multiwalled carbon nanotube bundle interconnects in ternary logic and comparison with Cu interconnects. IEEE Trans Nanotechnol 16(1):107–117
24. Rezaei Khezeli M, Hossein MM, Jalali A (2019) Active shielding of MWCNT bundle interconnects: an efficient approach to cancellation of crosstalk-induced functional failures in ternary logic. IEEE Trans Electromagn Compat 61(1):100–110

25. Madhuri BD, Sunithamani S (2019) Crosstalk noise analysis of on-chip interconnects for ternary logic applications using FDTD. Microelectron J 93:104633
26. Madhuri BD, Sunithamani S, Basha SJ, Kumar VR (2020) Design of hardened flip-flop using schmitt trigger-based SEM latch in CNTFET technology. Circuit World. https://doi.org/10.1108/CW-10-2019-0141
27. Lin S, Kim Y, Lombardi F (2011) CNTFET-based design of ternary logic gates and arithmetic circuits. IEEE Trans Nanotechnol 10(2):217–225
28. Sahoo SK, Akhilesh G, Sahoo R, Muglikar M (2017) High-performance ternary adder using CNTFET. IEEE Trans Nanotechnol 16(3):368–374
29. Lin S, Kim Y, Lombardi F (2012) Design of a ternary memory cell using CNTFETs. IEEE Trans Nanotechnol 11(5):1019–1025
30. Nayeri M, Keshavarzian P, Nayeri M (2019) Approach for MVL design based on armchair graphene nanoribbon field effect transistor and arithmetic circuits design. Microelectron J 92 (2019).
31. Sandhie ZT, Ahmed FU, Chowdhury MH (2020) Design of ternary logic and arithmetic circuits using GNRFET. IEEE Open J Nanotechnol 1:77–87
32. Kosynkin DV, Higginbotham AL, Sinitskii A, Lomeda JR, Dimiev A, Price BK, Tour JM (2009) Longitudinal unzipping of carbon nanotubes to form graphene nanoribbon. Nature 458:872–876
33. Kaushik BK, Kumar VR, Majumder MK, Alam A (2016) Through silicon vias: materials, models, design, and performance. CRC press, Inc. Subs. of Times Mirror 2000 Corporate Blvd. NW Boca Raton, FL, United States
34. Raychowdhury A, Roy K (2005) Carbon-nanotube-based voltage-mode multiple-valued logic design. IEEE Trans Nanotechnol 4(2):168–179
35. Deng J, Wong H-P (2007) A compact SPICE model for carbon-nanotube field-effect transistors including nonidealities and its application—part II: full device model and circuit performance benchmarking. IEEE Trans Electron Devices 54(12):3195–3205
36. Chen Y et al (2015) A SPICE-compatible model of MOS-type graphene nano-ribbon field-effect transistors enabling gate- and circuit-level delay and power analysis under process variation. IEEE Trans Nanotechnol 14(6):1068–1082

Automated Optimization Strategy and Usage of RoBERTa for Humor Identification

P. Hemant, Pramod Kumar, and C. R. Nirmala

Abstract A lot of work has been done in the past to identify abstracts like humor, sarcasm, and even irony in sentences. This work tries to figure out whether the text is humorous or not by using a base, large, and base-openai-detector models of RoBERTa, which are trained on ColBERT dataset having 200 k formal texts with 100 k humorous and 100 k non-humorous text in it. This is the first time implementation of humor detection using RoBERTa and its configurations. This paper compares the performance in terms of accuracy, precision, F1 score of the models of RoBERTa by taking AdamW, Adabound, and Adafactor as optimizers for the model. Experimental results show that accuracy of 99% is established by RoBERTa-large with Adafactor as its optimizer, 1% improvement compared with BERT model and 1.2% improvement compared with ALBERT models. With this, extensive experiments are done on automated loss function identification and also the comparison of Adam, Adafactor, and Adabound by varying epsilon values via cosine annealing function.

Keywords RoBERTa · Humor detection · Optimizer · Adam · Adabound · Adafactor · Learning rate

1 Introduction

Humor plays a significant role in communication, aside from making the person comfortable it also creates a cozy environment. The sense of humor can be used in every field of our life like in normal talks, films, television, books, etc., humor can

P. Hemant (✉)
Department of Computer Science and Engineering, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, India

P. Kumar
Department of Computer Science and Engineering, Krishna Engineering College, Ghaziabad, India

C. R. Nirmala
Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_55

be defined as one of the characteristics of human. Why don't we train machines to detect humor as we do? And even train them to reply in a humorous way as in the movie Interstellar, the robot—TARS is used to detect and reply, but in the current century, this can be used in chatbots and personal assistants. Humor is not easy to understand. For some persons, the text might be humorous and for others, it might not, it all depends on the person's state of mind and understanding. Understanding the text is a challenge for machines, only after understanding the text the machine or the model can detect humor in it. This can be achieved using language models, which are meant to understand the language. With advancement in NLP techniques, many researchers have contributed their work on humor detection wherein some used CNNs, RNNs, decision trees, highway network, etc. Generative chatbots is one of an advanced use-case where humor detection can be implemented where the system should be able to detect humor in the text and provide appropriate responses, thus making the conversation appealing. The language models can attain any state of the art task whether it is question answering or humor detection. To detect humor RoBERTa and its different configurations are used like base, large, base-openai-detector. To demonstrate this experiment, we used ColBERT dataset that consists of 200,000 labeled short texts, which are in turn distributed into 100,000 humorous texts and 100,000 non-humorous texts.

2 Literature Review

With the advancement in NLP tasks, identification or detection kind of problems attracted researchers' interest that involves deep understanding of the data. The humor detection also involves deep understanding. Humor detection and understanding is challenging because the level of humor by a person depends upon many conditions like his/her personal background, state of mind, funniness, sometimes humor might be serious. Natural language processing, in short NLP, is a field in artificial intelligence (AI) in which a machine understands the natural language, manipulates it, and analyzes the data in an intelligent and useful way. It is widely used in organizing the knowledge in the structured form so that it can be used in automatic text translation, summarization, relationship extraction, sentiment analysis, speech recognition, text summarization, and topic segmentation. Since the popularity in transfer learning, many pre-trained models started to emerge like ELMO, Googles' BERT, which mainly focuses on storing the learning information on a specific task and then apply the same on any other relevant task. BERT [1] is a bi-directional encoder that uses several stalked encoders for learning. Then came RoBERTa, it is one of the language models, which is an improved version of BERT. RoBERTa is both a robust and optimized approach to BERT. This is mainly pre-trained for English natural language understanding. It is trained a bit longer when compared with BERT and with large data, big sequences and it has a dynamically masking operation on trained data. Some of the previous works done on humor detection involve the use of deep

learning techniques like CNN, RNNs, use of highway networks, etc. Humor detection in code-mixed social media texts [2] where code mixed is a text that contains two or more languages in it, in this case, it was Hindi and English. In this, they have considered the problem as classification problem and applied some machine learning algorithms like Naive Bayes, Support Vector Machines, Random Forest to classify humor. These algorithms are also used for humor detection in Spanish tweets [3] along with decision trees, nearest neighbors. Generating unsupervised humor labels on videos and predicting it using deep learning models [4] have archived a high correlation with human annotations and obtained AUC of 0.751 on speech and text features. The author also showcases the accuracy results to be around 98.1% using BERT. An experiment on humor prosody analysis and humor recognition [5] wherein they used the spoken conversations from the famous television show FRIENDS and labeled the speech as humorous when it was followed by laughter and non-humorous when there was no laughter, they analyzed the data on pitch, energy and temporal information. In the work [6], they have come up with a way of using machine learning models that is mainly transformers to discern what jokes are funny and what jokes are not. This model does not require human interactions and can predict the level of humor for a specific data. Understanding humor is divided into two subtasks [7] as recognition of humor and humor anchor extraction wherein they constructed different computational classifiers to recognize humor, which is on the basis of designed sets of features and associated with each structure and also given a simple and effective maximal decrement method so that we can automatically extract anchors that enable humor in a sentence. This work detected humor in text samples of about 200 k using different pre-trained models of RoBERTa with two different optimizers and compared to results.

3 Parameter and Configuration of RoBERTa

RoBERTa [8] was proposed around 2019 and it is based on BERT with few modifications in hyperparameters and no next sentence prediction. Roberta works on dynamica masking and full sentence prediction methodology. It has multiple large mini batch groups and large byte-pair encoding byte level as compared to BERT. RoBERTa is a pre-trained model on English language using masked language modeling as an objective. It is case-sensitive and is trained longer than BERT.

This work used three configurations of RoBERTa, which are stored on RoBERTa model, namely:

i. RoBERTa-base
ii. RoBERTa-base-openai-detector
iii. RoBERTa-large

These three configurations are different in the number of hidden layers, attention heads, hidden size, and parameters. Hidden layers are located between the input and output layers. Attention [9] head is for a token to attend another token as an

Table 1 Pre-trained language model—RoBERTa

RoBERTa configurations	Hidden layers	Attention heads	Hidden size	Parameters in million
RoBERTa-base	12	12	768	125
RoBERTa-base-openai-detetor	12	12	768	125
RoBERTa-large	24	16	1024	355

attention mechanism. Suppose, there are 12 attention heads then a token can focus on 12 different aspects of other tokens. In simple terms, we can consider the depth of the model as the number of hidden layers and width as hidden size. The table gives configuration of RoBERTa-base, RoBERTa-large, and RoBERTa-base-openai-detector (Table 1).

4 Experiments and Results

This experiment has taken into account the combination of learning rate, different configurations of RoBERTa, scheduler, and two types of optimizers. This work considered the learning rate to be 1e-5 that is 0.00001. In simple terms, the learning rate is the rate at which the model learns. If the learning rate is too high then the model will be trained faster but results won't be accurate and if it is too low then the model might overfit and take a longer time to be trained, so always the learning rate should be in the range of 0.1–0.00001. As mentioned above, we take three different configurations of RoBERTa, those are—RoBERTa-base, RoBERTa-base-openai, and RoBERTa-large.

Optimization is done to identify the error and minimize it. Optimizer is the extended class that contains some information for training specific models and by this, the speed and performance of the model. In this experiment, AdamW and Adafactor optimizers were used for training the model.

i. AdamW implements Adam algorithm with weight decay fix, Adam is an adaptive gradient algorithm. The parameters in AdamW are model parameters, learning rate, Adam's betas parameters, Adam's Epsilon for numerical stability, weight decay, and correct bias.
ii. Adabound is a variant of Adam that changes the learning rate dynamically to achieve gradual and smooth transition. The parameters to define in Adabound are model parameters, learning rate and the final learning rate at which the Adabound performs as SGD optimizer, this is optional.
iii. Adafactor can be used as a replacement for Adam. This enables training of significantly larger models on the same memory-constrained hardware. Parameters should be passed are model parameters, learning rate, threshold of root mean square of final gradient update, decay rate, relative step, weight decay, scale parameter.

Scheduler is to schedule or adjust the learning rate by reducing it during training of the model. Here, this work used a linear schedule with warm-up, which increments the learning rate with the warm-up step specified. Initially, the learning rate will be zero and it will be constantly incremented till the learning rate given in parameters. Linear schedule with warm-up takes optimizer, warm-up steps and the last epoch as its parameters. Loss functions [10] are for assessing how well explicit calculation models will perform with the given information. We considered binary cross-entropy loss for our model to detect error and minimize it with the help of optimizer.

With these, the training was performed using three EPOCHS.

We import transformers, pandas, matplotlib and seaborn, and many more machine learning libraries for visualization and processing. We use RoBERT and its configuration as pre-trained models. For preprocessing of the text, we use Robertas' Tokenizer to get tokens.

Consider an example:

Sentence—Guys, i'm not eating

Tokens - 'Guy', 's', ',', '_i', "'", 'm', '_not', '_eating'

Token ids- [19979, 18, 15, 31, 22, 79, 52, 946]

We then separate token ids and tokens, encode then to get input ids and attention masks. The input ids and attention masks are sent to the model for training with a learning rate of 0.00001, wherein the loss function in our case is binary cross entropy and considers the predicted values by the model and real values for calculating loss score. The loss score is sent to optimizer for optimization of the weights. Then the final prediction is sent to classifier to classify whether the text is humorous or not by outputting true if it is humorous and false if not. This work compared the results in terms of accuracy, precision, recall, F1 score, and the total time taken by the model to get trained. The combination is as follows:

i. Learning Rate—0.00001
ii. Language Model—RoBERTa-base/RoBERTa-base-openai-detector/ RoBERTa-large
iii. Optimizer—AdamW/Adabound/Adafactor
iv. Scheduler—get_linear_schedule_with_warmup

With optimizer as AdamW shown in Fig. 1, we obtained 92.7% accuracy, 82% precision, 88% recall, and F1 in both base and base-openai-detector models with a total time of 3 h and 2 min. The highest accuracy we obtained is 95%, 94% precision, 93.5% recall and 95% F1 score in RoBERTa-large with a total time of 3 h 20 min.

With Adabound shown in Fig. 2, the base model achieved 98% accuracy, 97% precision, 99% recall and 98% F1 score in a total time of 42 min 17secs. The openai-detector model obtained 97.2% accuracy, 95% precision, 97% recall, and 97% F1 score with a total time of 42 min 3 secs. RoBERTa-large achieved accuracy of 98.73, 98% precision, 99% recall and F1 score with Adabound. We obtained 98.7, 98% precision, 99% recall and F1 score with pre-trained model RoBERTa-base-openai-detector and optimizer as Adafactor, which took a total time of 21 min and 1 s. This is just 0.1% less than that of RoBERTa-base.

RoBERTa-base with Adafactor shown in Fig. 3 established accuracy of 98.8%

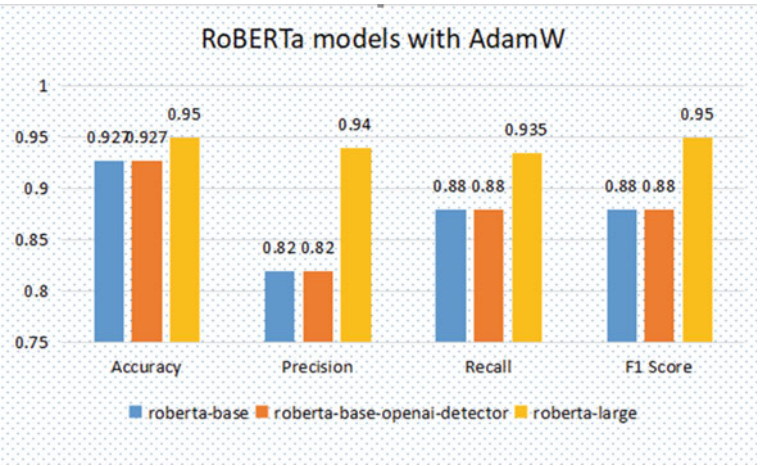


Fig. 1 RoBERTa models with Adam optimizer

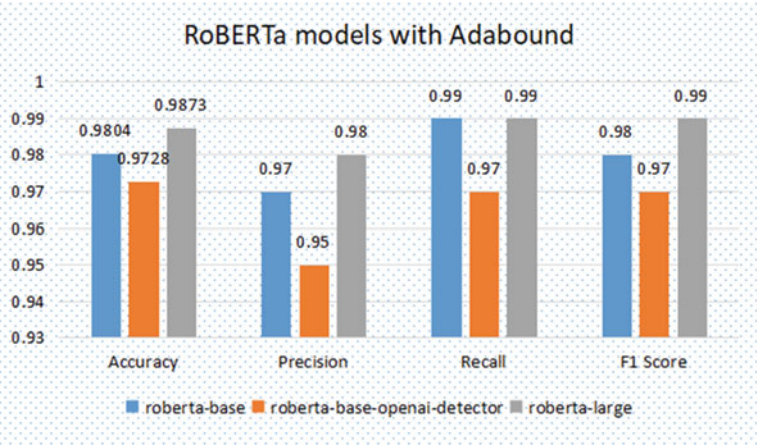


Fig. 2 RoBERTa models with Adabound optimizer

with the same total time, precision, recall, and F1 of Roberta-base-openai detector. The highest accuracy we got was with RoBERTa-large that is 99, 99.1% precision, 99.1% recall, and 99% F1 with a total time 1 h 5 min 31 s, which is one-third the total time taken by base and base-openai-detector models with AdamW optimizer.

The graphical representation of the comparison is given below.

Figure 4 showcases a comparative graph between Adam and Adafactor to showcase the effect of the optimizer on loss for the corresponding iteration and with learning rate as 0.00001 for Adafactor, Adabound, and Adam in Fig. 3. We found that AdamW almost performs in the same way where they lower the loss up to some extent with ongoing iterations and with Adabound the loss is almost constant, but

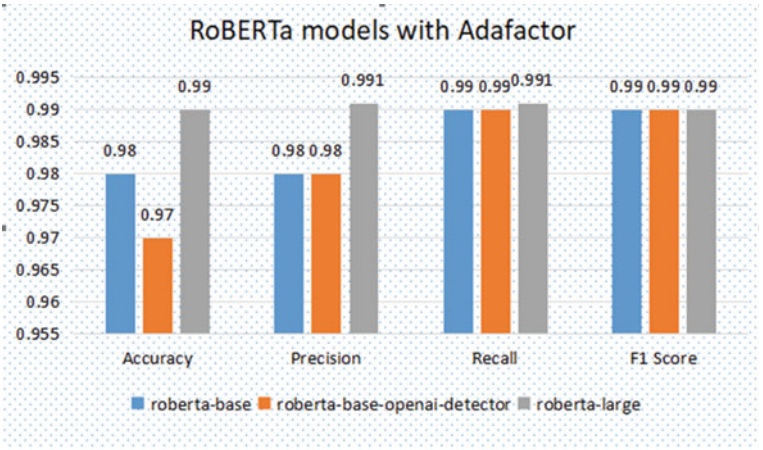


Fig. 3 RoBERTa models with Adafactor optimizer

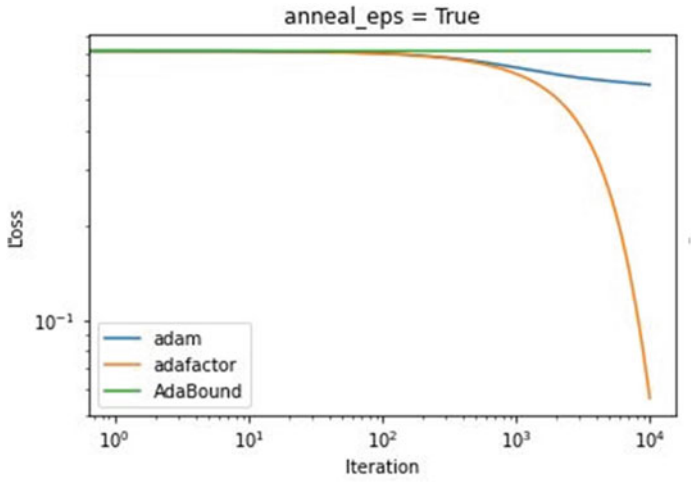


Fig. 4 Comparison between optimizers

Adafactor gradually lowers the loss up to some number of iterations beyond that the loss is exponentially reduced and this is the reason why the model performed well with Adafactor as optimizer by giving 99% accuracy.

We observed that Adafactor not only triggers the performance of the model and gives good accuracy but also reduces the execution time that is it increases the speed of the execution.

5 Results and Discussion

Deep understanding is required for understanding and detecting humor in that given text. In this paper, we demonstrated how hyperparameters like learning rate, epsilon values in optimizers affect the accuracy of the model. This work used RoBERTa and its different configurations that are base, openai-detector, and large for detecting humor from 200 k samples for text data and different factors like optimizers that are Adam, Adafactor, and Adabound. This work compared the outcomes in terms of accuracy, precision, recall, F1 score, and total time taken by the model and observed that the highest accuracy of 99% is obtained by the RoBERTa-large model with Adafactor as its optimizer taking 1 h 5 min as total time. This work also showcases that the model takes more time with Adam as optimizer and moderate time with Adabound, the accuracy for models with these optimizers is a bit accurate when compared with the model with Adafactor. With Adafactor, the model takes almost the same time to get trained no matter what the configuration is. We also demonstrated the effect of these optimizers to know the reason behind the performance of Adafactor.

6 Conclusion and Future Work

İn this paper, we showcase a way to automate the identification of the right optimizer and use the state of art language model RoBERTa to solve the humor detection/identification problem. Based on the existing work that was performed on humor detection, it is very clear and evident that the optimization of existing downstream tasks using state-of-art language models is very important. This paper showcases the accuracy lift of 1% keeping other metrics in line with the existing literature and experiments available on the mentioned dataset. In future studies, the inclusion of impact meter of humor detection in generative chatbots will be taken into account, which can, in turn, be used to see the human behavior and analyze it.

References

1. Yang Y, Abrego GH, Yuan S, Guo M, Shen Q, Cer D, Kurzweil R (2019) Improving multilingual sentence embedding using bi-directional dual encoder with additive margin softmax. arXiv:1902.08564
2. Khandelwal A, Swami S, Akhtar SS, Shrivastava M (2018) Humor detection in english-hindi code-mixed social media content: corpus and baseline system. arXiv:1806.05513
3. Castro S, Cubero M, Garat D, Moncecchi G (2016) Is this a joke? Detecting humor in spanish tweets. In: Ibero-American conference on artificial intelligence. Springer, Cham, pp 139–150
4. Annamoradnejad I (2020) ColBERT: using BERT sentence embedding for humor detection. arXiv:2004.12765
5. Yang D, Lavie A, Dyer C, Hovy E (2015) Humor recognition and humor anchor extraction. In: Proceedings of the 2015 conference on empirical methods in natural language processing, pp 2367–2376

6. Mozafari M, Farahbakhsh R, Crespi N (2019) A BERT-based transfer learning approach for hate speech detection in online social media. In: International conference on complex networks and their applications. Springer, Cham, pp 928–940
7. Cattle A, Ma X (2018) Recognizing humour using word associations and humour anchor extraction. In: Proceedings of the 27th international conference on computational linguistics, pp 1849–1858
8. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) RoBERTa: a robustly optimized BERT pretraining approach. arXiv:1907.11692
9. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems, pp 5998–6008
10. Nie F, Hu Z, Li X (2018) An investigation for loss functions widely used in machine learning. Commun Inf Syst 18(1):37–52
11. Howard J, Ruder S (2018) Universal language model fine-tuning for text classification. arXiv:1801.06146
12. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT press
13. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv:1412.6980
14. Luo L, Xiong Y, Liu Y, Sun X (2019) Adaptive gradient methods with dynamic bound of learning rate. arXiv:1902.09843
15. Shazeer N, Stern M (2018) Adafactor: adaptive learning rates with sublinear memory cost. arXiv:1804.04235
16. Yang D, Lavie A, Dyer C, Hovy E (2015) Humor recognition and humor anchor extraction. In: Proceedings of the 2015 conference on empirical methods in natural language processing
17. Duchi J, Hazan E, Singer Y (2011) Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 12(7)

Chaos-Based Image Encipherment Techniques: Analysis and Comparative Review

Ashish Kumar and N. S. Raghava

Abstract This paper presents an in-depth review of traditional cryptography and chaos-based cryptography schemes. In the last two decades, security services are achieved by advanced methodologies, which are prominent and reliable, and most of them are based on chaos theory. Chaos theory is useful for several applications in various fields. This paper investigates the traditional mechanisms of cryptography and also compares chaotic maps. Chaotic maps and the generated sequences are examined, and their behavior is analyzed on specific parameters to evaluate the performance. The literature review facilitates to find the appropriate system for real-time applications. A general encryption algorithm is applied to obtain PSNR, entropy analysis, SSIM, and time analysis to evaluate the performance of different chaotic maps. Thus, experimental results exhibit maps' performance. This study shows that traditional methods are the foundation of the advanced schemes; however, new algorithms are more suitable for bulky-sized data. Comparative analysis of different algorithms and evaluation of chaotic maps provides a better understanding of nonlinear systems and cryptography.

Keywords Chaos theory · Image encryption schemes · Security services · DES · Pseudorandom numbers

1 Introduction

The chaotic maps are nonlinear systems involved in scientific applications due to their properties, such as ergodicity, dynamicity, randomness, and pseudorandom numbers [1, 2]. Chaotic maps are dependent on initial conditions and parameters that control the map's chaotic behavior, and it is the primary element to make an ideal

A. Kumar (✉)
Department of Information Technology, Delhi Technological University, Delhi, India
e-mail: ashishkumar@dtu.ac.in

N. S. Raghava
Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India
e-mail: nsraghava@dce.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_56

cryptosystem. Researchers and scientists have applied chaotic maps to images and other multimedia in the last two decades, i.e., usually based on confusion diffusion theory [3–5]. In order to achieve the confusion diffusion concept, pseudorandom numbers have been applied to numerous cryptography algorithms based on chaos theory. Shannon introduces the confusion diffusion theory; as per the confusion theory, the algorithm is designed in such a manner that the intruder fails to find the relationship between ciphertext and the key. In diffusion, if a single bit is replaced in ciphertext, the cryptanalyst must fail to develop the original input. It refers to make a more complicated relationship between ciphertext and original input [6]. The entire mechanism, also stated in Fridrich architecture (permutation and substitution), is applied to images to enhance security and make them secured against security attacks. When information is communicated through a nonsecure channel, it is assumed that the information can be intercepted at any point. Therefore, information must be protected, and the cryptography concept arises here, which allows users to encipher the data (convert it into a nonreadable format). Encipherment techniques are classified in two ways: a user can apply either a cryptography scheme or a steganography technique depending upon the security requirements.

In this paper, chaotic maps have been analyzed and the recent approaches for image encryption schemes are studied. Chaotic maps are also applied in the encryption process to evaluate their performance by simulation and test cases. This study shows that all the chaotic maps are different from each other in many aspects and capable of generating significant results and contributing to image encryption schemes.

2 Traditional Security Mechanisms

Traditional algorithms of cryptography were designed for text and written according to 26 distinct symbols, whereas these days, information can be communicated through any mode of multimedia such as images, video, audio. Multimedia covers a considerable number of bits that travel across the world very frequently. Hence, to handle such bulky data's intrinsic properties that hold a huge amount of volume and redundant data, we need such systems that can be implemented efficiently on software and hardware. To achieve confidentiality, Chaos-based schemes follow traditional cryptography schemes' principles, but both work differently in different environments.

2.1 Symmetric Key Cryptography

AES and DES are two popular modern symmetric key cryptography algorithms used to achieve confidentiality and were designed to alter the text [7, 8]. The emphasis is given to encryption stability instead of giving attention to cryptanalysis attacks and statistical property of input data. However, cryptanalysis became more important

these days to check the vulnerability of the system. Stream cipher and block cipher are two techniques, which are designed under the category of symmetric key cryptography. DES, AES, 3DES, and IDEA are popular algorithms used in the commercial sector and the mainstream security policies for privacy. Initially, these algorithms were designed for text, which consists of limited input information and different statistical properties than any other multimedia. So, in a few cases, when these traditional algorithms [9–11] are applied to images, these techniques do not perform well, hence, found unsuitable for image encryption. Also, the theoretical values obtained demonstrate that these techniques are ideal for a small amount of information.

2.2 Asymmetric Key Cryptography

In this scheme, RSA and ECC are popular algorithms for asymmetric key cryptography scheme. Nevertheless, many other algorithms, such as ElGamal, Paillier cryptosystem, are also utilized to fulfill specific service requirements. These techniques are based on a pair of public keys. In [12–14], Chaos theory has emerged with these algorithms for the better performance of cryptosystems. Asymmetric key cryptography techniques are based on exhaustive calculations and it takes a good amount of time to compute the keys and the cipher; also, the primary importance is given to the key generation phase. Cryptography schemes are listed in Table 1.

2.3 Security Issues with Traditional Approaches in Image Cryptosystem

In literature, numerous techniques are available to deal with security issues and vulnerability. Illegitimate users can alter, modify, and delay the true information either actively or passively. This paper attempts to address some major drawbacks present in the existing algorithms and examine the chaos-based cryptosystems. Images are the primary source of information and travel through insured channels. As per usage of images, it has a wide variety of applications in different fields such as internet communication, multimedia systems, medical imaging, telemedicine, military communication, etc. There are many shortcomings of a conventional cryptosystem apart from the computational cost. It is discussed below [15, 16]:

- Traditional encryption algorithms, such as Data Encryption Standard (DES), International Data Encryption Algorithm (IDEA), and Advanced Encryption Standard (AES), etc., were designed for the text. Therefore, these techniques are not suitable for image encryption due to high computational cost and time complexity.
- One is that the image size is practically always larger than the text. Therefore, the old system takes a large amount of time to encrypt image data [17].

Table 1 Comparison between different traditional cryptographic algorithms

	Private key cryptography schemes			Public key cryptography schemes	
Algorithm	DES	3DES	AES	RSA	ECC
Developed	1975	1978	2000	1978	1985
Block round	16	48	10 (128bits key) 12 (192 bits) 14 (256 bits)	01	01
Design	Substitution and permutation	DES is applied three times	Substitution, shift, and bit mixing	Based on modular arithmetic, Euclidean algorithm, and prime numbers	Point addition and point doubling
Key size	56 bits		128,192, 256 bits	>1024 bits	>160 bits
Input blocks size	64 bits	64	128	Depending upon the input	Very tiny 16 Kb
Encryption speed	Moderate	Slow	Faster	Slower	Faster than RSA
Security	Inadequate secured	More secured than DES	Exceptional secured	Least secured	Secured and can be applied to images
Power consumption	Low	High	Low	High	Low

- The second problem is that size of the decrypted text must be equal to the original text. However, this requirement is not necessary for the image data. Thus, to perform the encryption of images, various algorithms are proposed.
- In the traditional cryptography system, it is difficult to secure a large size of multimedia from intruders or attackers and calculation of mathematical equation (built-in encryption technique) is achieved in a high order of complexity.

Thus, the algorithm must be designed in such a manner that it can preserve robustness and reliability against all kinds of statistical attacks with a hard time-bound to operate in real-time applications with consistent performance.

3 Chaos-Based Cryptography

Chaos theory is prevalent in recent times, and it is found that many researchers are working in the same direction using various chaotic maps. Chaos theory has emerged in many fields of science and it has become useful and attractive due to

its properties. A system based on chaos theory is called a chaotic system. There are specific parameters to control the behavior of chaotic maps and also depend on the observations and experiments [18, 19]. The formation of the chaotic map depends on the initial seeds and the parameters; accordingly, the orbit and trajectory of the map are plotted. In order to attain an efficient system, the initial seed (keyspace) must be large over the range of 2^{100}, i.e., the key size must be of at least 100 bits to resist brute force attack, where an attacker tries to find the actual key; it is fulfilled in chaos-based cryptosystems due to the large precision of seeds. Thus, it is highly sensitive to the initial key conditions and deterministic. In [20–23], chaotic systems have collaborated with other existing fields like DNA and wavelet transforms to provide a more realistic and reliable approach to advancing security protocols. Chaotic systems are used to generate the pseudorandom sequence. These sequences are treated as a keystream to obtain a cipher image. Since image encryption techniques are based on Shannon's confusion diffusion strategy, they are more sustainable against statistical attacks. Table 2 exhibits the chaotic map-based cryptography schemes, where several researchers have proposed algorithms based on confusion diffusion.

4 Experimental Results and Observations

In this section, chaotic maps are analyzed and initiated with initial seed value to generate sequences for the encryption scheme as per Algorithm 1. The MATLAB R2015a software has been used for simulation purpose. The proposed framework has been applied similarly to all the chaotic maps described in Table 3. The standard image of 'lena.jpg' of size 204 × 204 is taken as input for the encryption process. Comparative performance matrices are listed in Table 3. Structural similarity index (SSIM) [35], (peak signal to noise ratio) PSNR entropy, and sequence generation time analysis are done to evaluate the performance. In addition, range and cutoff point are also determined that can be used to generate binary sequence on the basis of the threshold value. SSIM is used to quantify the degradation quality of an image after encryption. For the ideal cryptosystem, the SSIM and entropy value must be near zero and eight, respectively. Figure 1 illustrates that all the chaotic maps obtain values that are near zero. Time taken for sequence generation is illustrated in Fig. 2; it shows that all the chaotic maps are fast enough to generate the sequence. Here PSNR values are calculated between original and encrypted images, so low PSNR value indicated an ideal cryptosystem as shown in Table 3.

Table 2 Chaos-based cryptography schemes

Scheme	Parameter for the encryption process	Purpose	Confusion module	Diffusion module	Image type	Test performed	Remark
[24]	Keystream cipher	Confidentiality	2D rectangular transform	Tent map with pseudorandom numbers [0 255]	Color images	Keyspace, key sensitivity, histogram, correlation analysis, entropy, NPCR, UACI, robustness, speed test	It is based on an enhanced version of 2D rectangular transform by adding some variable values
[25]	Bit-level and column level indexing and shuffling (between pixels)	Confidentiality	Henon chaotic map Lorenz map Chua attractor Rossler attractor	–	Color images	Keyspace, correlation analysis FIPS PUB test (monobit test, poker test, runs test, long-run test) NPCR, UACI	Sequences are used to arrange pixels and bits among pixels using chaotic maps
[26]	Blocks pixels	To secure bulky images, i.e., satellite images	Nil	Henon map, Chebyshev, cubic, sine, Tent map	Grayscale images	Maximum deviation, information entropy, Histogram analysis, Keyspace analysis, key sensitivity analysis, performance analysis	Average encryption time is less than 0.31 s

(continued)

Table 2 (continued)

Scheme	Parameter for the encryption process	Purpose	Confusion module	Diffusion module	Image type	Test performed	Remark
[27]	One-time pad scheme	confidentiality	Arnold cat map	Modifies logistic map	Color images	Entropy analysis, Histogram analysis	The maximum Lyapunov exponent is used to check the chaotic behavior of a new chaotic map
[28]	Block-based cipher	Confidentiality with compression	Chebyshev map	6 D chaotic map	Grayscale images	Entropy, Histogram, Correlation, NPCR, UACI, Time complexity Chosen plaintext and ciphertext attack analysis	Compressive sensing is applied using Gauss matrix based on a chaotic map
[29]	2 Round of S Box using cyclically shifted	Confidentiality	Enhanced Rectangular transform, S Box	Double dependent substitution based on logistic map	Grayscale images	Chi-square statistics, Correlation analysis, information entropy, Histogram analysis, NPCR, UACI, Keyspace analysis, speed analysis, robustness against cropping attacks	Rectangular transform scramble nonsquare images, and S Box is applied during the confusion phase on each pixel

(continued)

Table 2 (continued)

Scheme	Parameter for the encryption process	Purpose	Confusion module	Diffusion module	Image type	Test performed	Remark
[30]	One-time pad (XOR)	Confidentiality	Arnold cat map, logistic tent map, logistic sine map	Tent + sine map,	Three gray images simultaneously used as input for the algorithm	Keys space, Histogram, Correlation, entropy, NPCR, UACI, PSNR	A discrete fractional random transform is used after the diffusion process
[31]	All pixels of an image	Confidentiality	5-D multiwing hyper-chaotic system is used to shuffle pixel and bits	5-D multiwing hyper-chaotic system	Grayscale images	Keys space, Histogram, Correlation, Key sensitivity, entropy, NPCR, UACI	keystream is strongly related to plain-image
[32]	Chained block scheme	Compression, integrity and encryption	Chaos-based encryption and Huffman coding	Arbitrarily	Grayscale image	PSNR, entropy, correlation analysis	MAC is used to provide integrity
[33]	One time pad scheme	Confidentiality	Bit level permutation	Hénon chaotic map with Key-value transformation	Color images	Entropy, correlation analysis, mean value analysis, histogram analysis, key sensitivity test	To secure object-oriented images and size is not fixed of the region. It is adaptive as per the region

(continued)

Table 2 (continued)

Scheme	Parameter for the encryption process	Purpose	Confusion module	Diffusion module	Image type	Test performed	Remark
[34]	Multiplication and addition operations	Confidentiality	Logistic map with optimized sequences	Logistic map with optimized sequences	Color images	Decryption error, secret key size analysis, key sensitivity, histogram, correlation, differential, chosen/known plain attack, entropy, time analysis	Image characteristics are added to the key resist known plain image attack

Table 3 Chaotic maps and performances

	Equation	Parameters	Range	Cutoff point	PSNR	Entropy
Logistic map	$x_{n+1} = rx_n(1 - x_n)$	$3.57 \leq r \leq 4$	[0.00001 0.99]	0.551	27.57	7.99
Henon chaotic map	$x_{n+1} = 1 + y_n - ax_n^2$ $y_{n+1} = bx_n$	a = 1.4 b = 0.3	[−1.2846 1.2730]	0.4142	27.58	7.99
Tinker bell chaotic map	$x_{n+1} = x_n^2 - y_n^2 + ax_n + by_n$ $y_{n+1} = 2x_n y_n + cx_n + dy_n$	a = 0.9 b = −0.6013 c = 2, d = 0.5000	[−1.231 0.4600]	−0.1128	27.55	7.99
Tent map	$x_{n+1} = \begin{cases} \mu x_n \quad for\ x < 0.5 \\ \mu(1 - x_n)\ for\ x > 0.5 \end{cases}$	μ = 1.99	[0.00014 0.999]	0.49	27.54	7.99
Sine map	$x_{n+1} = \lambda sin(\pi x_n)$	λ = 0.99	[0.0001,0.99]	0.522	27.59	7.99
Gauss map	$x_{n+1} = exp(-\alpha x_n^2) + \beta$	α = 4.9, β = 0.5	[−0.1434,0.4000]	0.035	27.20	7.81
Duffing map	$x_{n+1} = y_n$ $y_{n+1} = -bx_n + ay_n - y_n^3$	a = 2.75 b = 0.2	[−1.7095, 1.7095]	0.1938	27.59	7.9954

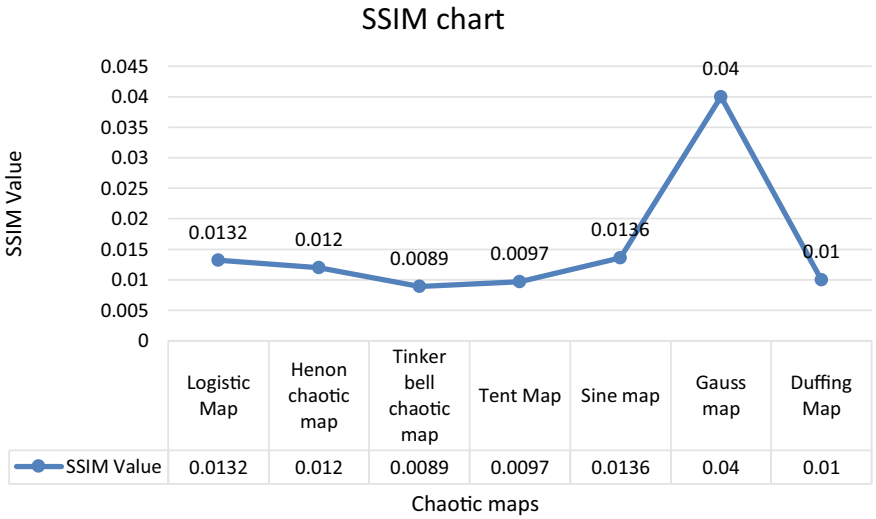


Fig. 1 Structural similarity index measures of chaotic maps

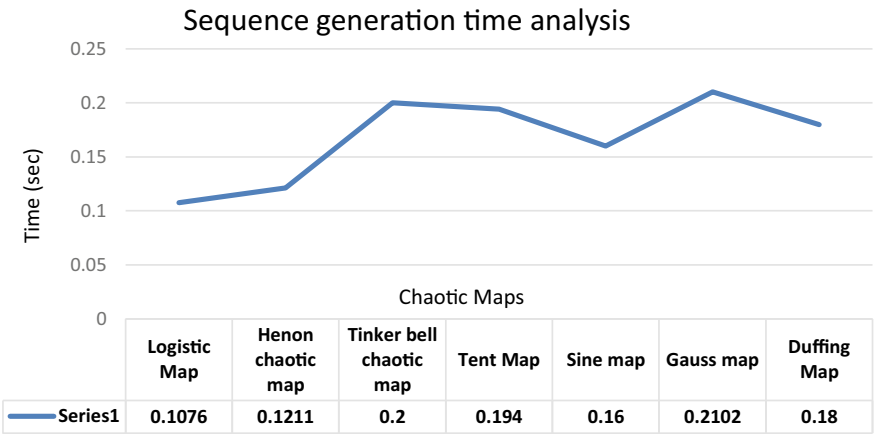


Fig. 2 Sequence generation time analysis of chaotic maps

Algorithm 1: The encryption algorithm

1. **Start**
2. initiate seed and control parameters
3. Read an image with size [m, n]
4. generate sequence [X] of size m×n using a chaotic map
5. Keystream= $(X \times 10^{7)} \bmod 255$
6. Encipher all the pixels of an image using the generated keystream. The procedure applies to all the pixels in a bitwise manner.
7. **End**

5 Conclusion and Discussion

In this paper, chaotic maps are examined and studied, and it is found that chaotic maps are rich in features and can perform cryptographic operations in an insignificant amount of time. These chaotic maps are based on a simple calculation and can be implemented in limited resources. It is also analyzed that all these chaotic maps performed well with images, it is evident through entropy, SSIM, and PSNR tests. In this study, different chaotic maps are studied and it is observed that different parameters are available to evaluate the performance of the cryptosystem. However, the primary goal is only to achieve confidentiality in chaos-based schemes using stream cipher techniques. Pseudorandom sequences are useful in all the fields of science, and thus, sequence generation time can improve the speed of systems based on chaotic maps. Therefore, this paper emphasizes chaos-based schemes and experimental results show that they can be used in real-time applications. Evaluation of chaotic maps and comparative analysis of different algorithms provide a better understanding of nonlinear systems and cryptography.

References

1. Nesa N, Ghosh T, Banerjee I (2019) Design of a chaos-based encryption scheme for sensor data using a novel logarithmic chaotic map. J Inf Secur Appl 47:320–328
2. Ahmad J, Hwang SO (2016) A secure image encryption scheme based on chaotic maps and affine transformation. Multimed Tools Appl 75(21):13951–13976
3. Ye R (2011) A novel chaos-based image encryption scheme with an efficient permutation-diffusion mechanism. Opt Commun 284(22):5290–5298
4. Enayatifar R, Abdullah AH, Isnin IF et al (2017) Image encryption using a synchronous permutation-diffusion technique. Opt Lasers Eng 90:146–154
5. Essaid M, Akharraz I, Saaidi A et al (2018) A new image encryption scheme based on confusion-diffusion using an enhanced Skew Tent map. Procedia Comput Sci 127:539–548

6. AbuTaha M, El Assad S, Farajallah M, Queudet A, Deforge O (2015) Chaos-based cryptosystems using dependent diffusion: an overview. In: 2015 10th international conference for internet technology and secured transactions (ICITST). IEEE, pp 44–49
7. Rihan SD, Khalid A, Osman SE (2015) A performance comparison of encryption algorithms AES and DES. Int J Eng Res Technol (IJERT) 4(12):151–154
8. Sanchez-Avila C, Sanchez-Reillol R (2001) The Rijndael block cipher (AES proposal): a comparison with DES. In: Proceedings IEEE 35th annual 2001 international carnahan conference on security technology (Cat. No. 01CH37186). IEEE, pp 229–234
9. Zhang Y (2018) Test and verification of AES used for image encryption. 3D Res 9(1):3
10. Chen C, Wang T, Kou Y, Chen X, Li X (2013) Improvement of trace-driven I-Cache timing attack on the RSA algorithm. J Syst Softw 86(1):100–107
11. Coppersmith D (1994) The Data Encryption Standard (DES) and its strength against attacks. IBM J Res Dev 38(3):243–250
12. Yue P, Guodong L, Jing Z (2016) "Based on the improved RSA keys and compound chaotic system and design of audio encryption algorithm. In: 2016 International conference on smart city and systems engineering (ICSCSE). IEEE, pp 197–20
13. Nkapkop JD, Effa JY, Toma A, Cociota F, Borda M (2016) Chaos-based image encryption using the RSA keys management for an efficient web communication. In: 2016 12th IEEE international symposium on electronics and telecommunications (ISETC). IEEE, pp 59–62
14. Irfan P, Prayudi Y, Riadi I (2015) Image encryption using combination of chaotic system and rivers shamir adleman (RSA). Int J Comput Appl 123(6)
15. Kansal S, Mittal M (2014) Performance evaluation of various symmetric encryption algorithms. In 2014 International conference on parallel, distributed and grid computing. IEEE, pp 105–109
16. Ahamad MM, Abdullah MI (2016) Comparison of encryption algorithms for multimedia. Rajshahi Univ J Sci Eng 44: 131–139
17. Kushwah K, Shibu S (2013) New image encryption technique based on combination of block displacement and block cipher technique. Int J Comput Sci Inf Technol 4(1):61–65
18. Kocarev L (2001) Chaos-based cryptography: a brief overview. IEEE Circuits Syst Mag 1(3):6–21
19. Holt N (2017) Chaotic cryptography: applications of chaos theory to cryptography
20. ur Rehman A, Liao X, Ashraf R, Ullah S, Wang H (2018) A color image encryption technique using exclusive-OR with DNA complementary rules based on chaos theory and SHA-2. Optik 159:348–367
21. Chai X, Xianglong Fu, Gan Z, Yang Lu, Chen Y (2019) A color image cryptosystem based on dynamic DNA encryption and chaos. Signal Process 155:44–62
22. Chai X, Chen Y, Broyde L (2017) A novel chaos-based image encryption algorithm using DNA sequence operations. Opt Lasers Eng 88:197–213
23. Lisungu TO, Sumbwanyambe M (2019) Image compression-encryption scheme based on 2D DWT, SPIHT and Qi hyper-chaos. In: 2019 Southern African universities power engineering conference/robotics and mechatronics/pattern recognition association of South Africa (SAUPEC/RobMech/PRASA). IEEE, pp 177–182
24. Wu X, Zhu B, Yutong Hu, Ran Y (2017) A novel color image encryption scheme using rectangular transform-enhanced chaotic tent maps. IEEE Access 5:6429–6436
25. Huang CK, Hsiau-Hsian N (2009) Multi chaotic systems-based pixel shuffle for image encryption. Opt Commun 282(11):2123–2127
26. Usama M, Khan MK, Alghathbar K, Lee C (2010) Chaos-based secure satellite imagery cryptosystem. Comput Math Appl 60(2):326–337
27. Zhang X, Cao Y (2014) A novel chaotic map and an improved chaos-based image encryption scheme. Sci World J 2014
28. Zhu S, Zhu C, Fu Y, Zhang W, Wu X (2020) A secure image encryption scheme with compression-confusion-diffusion structure. Multimed Tools Appl 1–24
29. Zhang X, Fan X, Wang J, Zhao Z (2016) A chaos-based image encryption scheme using 2D rectangular transform and dependent substitution. Multimed Tools Appl 75(4):1745–1763

30. Bisht A, Dua M, Dua S (2019) A novel approach to encrypt multiple images using multiple chaotic maps and chaotic discrete fractional random transform. J Ambient Intell Humaniz Comput 10(9):3519–4353
31. Li Y, Wang C, Chen H (2017) A hyper-chaos-based image encryption algorithm using pixel-level permutation and bit-level permutation. Opt Lasers Eng 90:238–246
32. Sankaradass V, Murali P, Tholkapiyan M (2018) Region of interest (ROI) based image encryption with sine map and lorenz system. In: International conference on ISMAC in computational vision and bio-engineering. Springer, Cham, pp 493–502
33. Kumar A, Raghava NS (2020) Selective colour image encryption using Hénon chaotic system with a keyless substitution cipher. Eng Appl Sci Res 47(1):66–76
34. Murillo-Escobar MA, Cruz-Hernández C, Abundiz-Pérez F, López-Gutiérrez RM, Del Campo OA (2015) A RGB image encryption algorithm based on total plain image characteristics and chaos. Signal Process 109:119–13
35. Anwar S, Meghana S (2019) A pixel permutation based image encryption technique using chaotic map. Multimed Tools Appl 78(19):27569–27590

Review of Recent Trends in Additive Manufacturing

Meena Pant, Leeladhar Nagdeve, V. C. Pandey, and Harish Kumar

Abstract Additive manufacturing (AM) emerged as a boon for the manufacturing systems, due to its capability of producing parts through deposition of materials in a layer-by-layer manner and its effects on the global supply chain by making the systems decentralized in nature. AM is also recognized with various names such as direct digital manufacturing, rapid manufacturing, solid-free fabrication process, and rapid prototyping. It has the potential to make complex and internal lattice structures without any need for prior processing tools as compared with the conventional manufacturing process. Currently, this emerging technology is fulfilling customer demand at a smaller scale. Currently, engineers and research institutes are working on AM-based systems to examine the machine capabilities and microstructural features of manufactured parts in order to fully utilize the potential of the AM process in various organization, research departments, and also at the academic level. This paper provides an overview of AM process, its latent application, combating in COVID-19, and discusses the challenges, which are slowing down its full implementation.

Keywords 3D printing · Techniques · Medical industry · Food industry · COVID-19

1 Introduction

Since Stone Age, we have been using tools to remove materials from a block so that the desired shape is achieved and consequently function. This type of manufacturing has not changed significantly for several thousand years. However, over the past few decades, the concept of material addition changes the thinking ability and emerged as a new manufacturing technique known as 3D printing. The term has been known with different names; solid free fabrication, digital manufacturing, solid ground curing,

M. Pant (✉) · L. Nagdeve · H. Kumar
Department of Mechanical Engineering, National Institute of Technology Delhi, Delhi 110 040, India

V. C. Pandey
HMR Institute of Technology and Management, Delhi 110 042, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_57

and rapid manufacturing process. Both 3D printing and additive manufacturing are synonyms. This technology came into the picture in the 1980s which is defined by the American Society of Testing Materials (ASTM) as a material joining process using the concept of layer-by-layer to make 3D physical objects [1, 2]. Figure 1 shows the methodology for conventional manufacturing to the 3D printing process. In early 1990, Kruth [3] characterized different AM processes from three viewpoints: solid-based, liquid-based, and powder-based systems. Thus a change in procedures determined the final accuracy and the mechanical properties of the part. Table 1 shows the AM techniques with their summary.

This way of creating parts has many benefits over traditional manufacturing like; ability to make a custom-based product, infinite design of freedom, single part assembly, no cost for additional tooling, and make complicated and intricate design features. Due to these benefits, AM has increased its domain in the aerospace industry, food industry, construction industry, and medical industry [4]. The remarkable performance of AM has been seen in the recent COVID-19 pandemic. When there is a need of personal protective equipment for healthcare professionals, various AM industries come to forward to combat this crisis. During the pandemic, everything became stagnant, there is a disruption in supply chains, and the demand for essential equipment increases day by day, then digital versatility and rapid prototyping of 3D printing allow a quick mobilization of the technology and response to this crisis. Industries realize the importance of 3D printing. Apart from many benefits of AM, it is not adapt the full focus to the manufacturing industry. There is a need to retrospect the design practices from a technological, economic, and emotional point

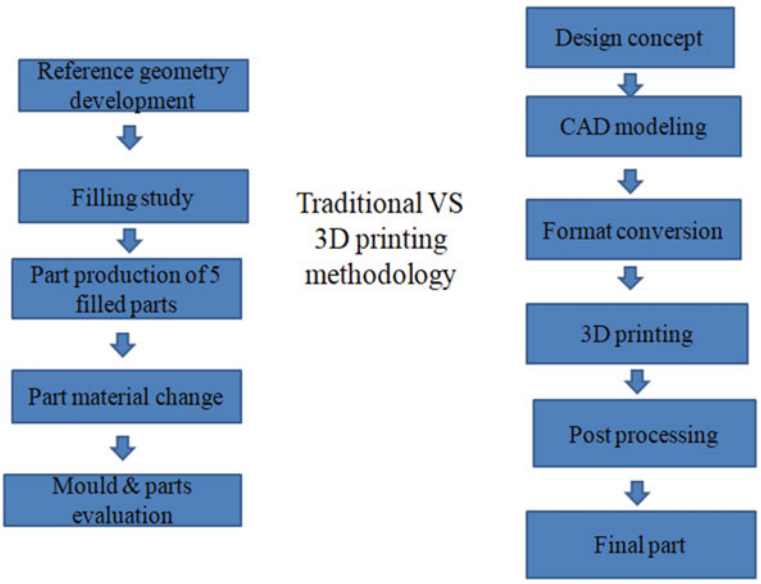


Fig. 1 General trial methodology for traditional manufacturing versus 3D printing

Table 1 Illustration of AM techniques [4, 5]

Process	Techniques	Principle	Characteristics
Material extrusion (ME)	Fused deposition modeling (FDM)	Material is selectively dispensed through a nozzle	Cost is low, high speed, simplicity, and design
Vat photopolymerization	Steriolithograhpy (SLA) Digital light processing	Liquid photopolymer in a vat is selectively cured by light-activated polymerization	High-quality parts < 10 μm fine resolution
Binder jetting (BJ)	3D printing Inkjet printing	Liquid bonding agent is selectively deposited to join powder materials	Reduction in cost incurred in tooling and manufacturing time is low, complex structures are not recommended
Material jetting (MJ)	Polyjet Thermo jet Inkjet	Droplets of build material are selectively deposited	Efficient and fast, flexibility in design, printing complex structures
Powder bed fusion (PBF)	Selective laser sintering(SLS) Selective laser melting (SLM) Electron beam melting (EBM)	Thermal energy selectively fuses regions of a powder bed	Anisotropic microstructural and residual stress effects, powders are fused at the molecular level
Sheet lamination (SL)	Laminate object manufacturing (LOM) Ultrasonic consolidation	Laminated sheets of material are consecutively cut and added to the previous layer	Printing efficiency is high, and fast flow-ability
Direct energy deposition (DED)	Direct metal deposition Electron beam direct melting Laser deposition	Material is fused and added during the production of the part	Controlled microstructure with exceptional mechanical properties, and accurate Composition control

of view toward the AM industry. This paper started with a basic introduction and Sect. 2 discusses the AM application in different fields as well as the role of AM during the COVID-19 crisis. Section 4 discusses the challenges in the AM process and the conclusion of the study.

2 Application of AM

2.1 Healthcare and Medical Industry

Additive manufacturing provides a new strategy, and viewpoint to develop personalized methods with specific and cost-effective requirements for both patients and clinicians. Additive manufacturing can create complex anatomical structures easily through scanned data of magnetic resource imaging and computed tomography imaging, which is difficult to form by conventional manufacturing methods [6]. Tumoi [7] classified AM in medical applications into five main domains and the application of AM in medicine as shown in Fig. 2.

1. Models of healthcare
2. Implant surgery
3. Surgical guides
4. External supports
5. Biofabrication

AM methods can implement new biomedical implants, dental implant, bone implant, controlled drug delivery systems, lab and testing equipment, an instrument for diagnostic and surgery, and engineered tissues and organs. Biomaterials allow for the presence of a cell in the patient's body for the primary reason. Three methods of

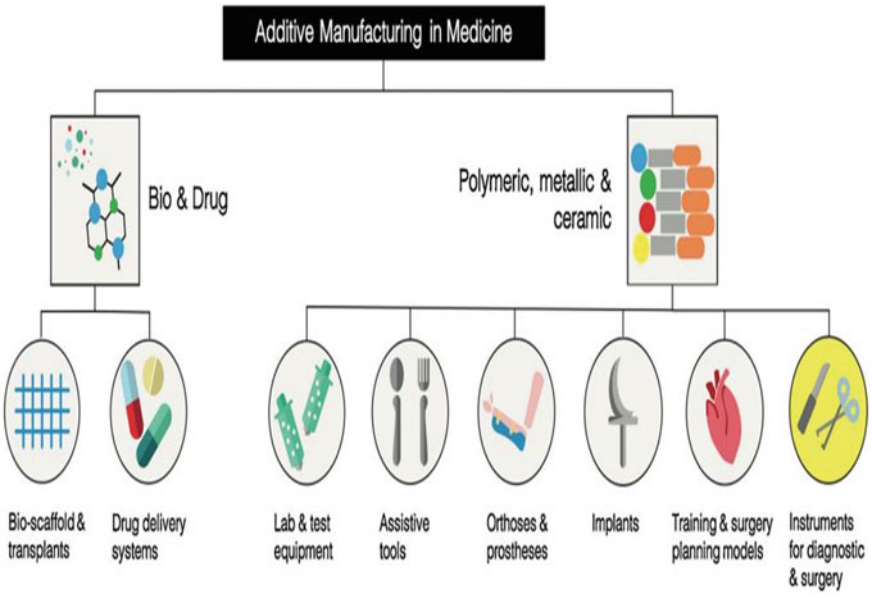


Fig. 2 Application in the medical field [9]

physical change, chemical alteration, and improved compatibility with surface coatings can bind such biomaterials to cells. Scientists and researchers are developing and harnessing AM technologies to construct a practical structure [8].

2.2 Food Industry

Although food printing is a current research and development subject, limited findings have been identified for food printing in 3D format. 3D printing technology is implemented in food processing due to its functionality, flexibility, smoother supply chain, custom nutrition design, etc. Yang et al. [10] reported that manufacturing a complicated 3D cake for the first time. It is based on the extrusion process since 3D printing technology has increasingly grown in the food industry. Currently, the extrusion process, binder jetting, SLS, and inkjet printing are the four most commonly used techniques in food processing because people are now concerned about the higher demand for healthy functional personal food. In the food sector, 3D printing technology has still presented many difficulties due to several reasons, such as less food material availability, compatibility of machines and high cost [11]. Yang et al. [12] used lemon juice gel and potato starch as a composite containing 59.82 g/100 of moisture. The experiment was conducted, taking into account parameters such as viscosity, complex viscoelastic characteristics, and shear stress. The impact was an excellent printout of lemon juice gel (Fig. 3).

2.3 Architecture Building and Construction Industry

AM has a broader application in aerospace, medical, and now is being inspected as a manufacturing method for architecture and construction industries. According to the survey, the total AM market was established to be worth 4 billion dollars in 2014 and is estimated to expand to more than 21 billion dollars by 2020. AM provides the opportunity to exploit different materials for architectural design. It has the ability to freeform fabrication and automates the construction industry directly through digital modeling [15]. Earlier AM process was suitable for factory and small-scale production, now AM application adapted for large building construction with several different solutions such as gantry robot to move the printing head [16]. Traditionally, computer simulations and wood prototypes are used to study designs, but AM introduces practical and unique features and detail to primary treatment. A model's capacity to check saves precious time and money framework. In 2014, the first commercial 3D printed architecture was built in Amsterdam (Fig. 4), using the FFF process. However, AM technology in the construction industry is already in the beginnings, it is seen that using this technology significantly improves the conventional construction process. This technology is projected to be further research and explore, resulting in more possibilities and growing challenges in the construction

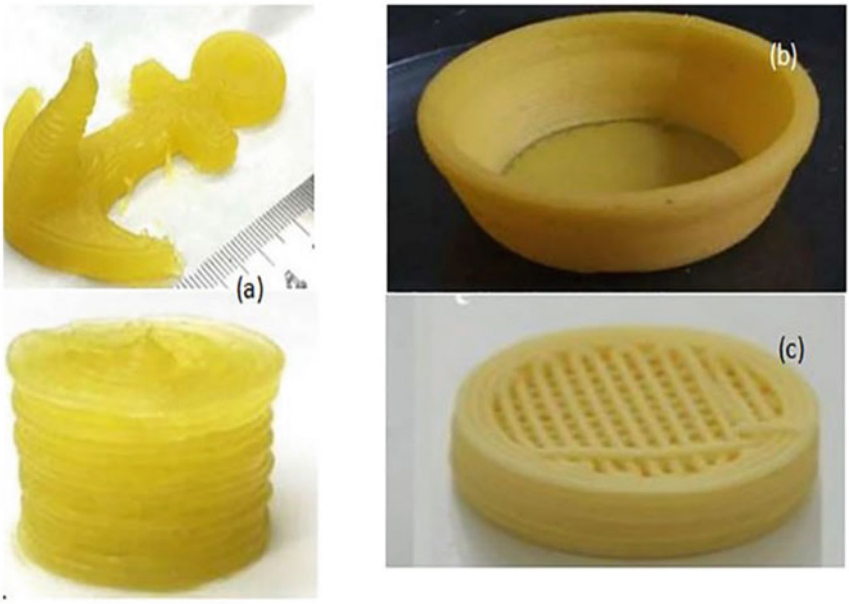


Fig. 3 **a** Product manufactured from lemon juice gel [12] **b** product made by egg and rice [13] **c** Product made from flour, dry congealed mango, and olive oil [14]

Fig. 4 First 3D-printed house by DusArchitects [17]

industry. The incorporation of Wholer's association has been widely recognized as the leading consulting firm and the most prominent expert for AM in the world [15].

2.4 Advanced Alloys

AM technological improvements are closely related to research and innovation of new materials and new systems. The development of new materials such as bioinformatics materials, functionally graded materials (FGMs), non-homogenous materials is of great significance for the survival of AM technology. The material used for AM technology is low, and durability undesirable, disrupting AM molding technology's production. An eco-friendly new material with excellent performance, reduced cost, small shrinkage, and high endurance must be produced to enhance the publication and standardization of AM technology [18]. The FGM paradigm was first suggested in 1987 to create heat-resistant material for the spacecraft propulsion system and airframe. After that, much work has been done to build FGMs for different applications using gradients of chemical, biochemical, physical, and mechanical properties. The FGMs are used in a variety of ways in both the electrical and electronic industries. The microstructural gradient FGMs find their function in components that must have a highly wear-resistant surface and a strong core to withstand the high impact of the process. An example of FGMs material is hardened steel, cams or round gear, bearings or shaft and turbine applications. Advance material elements are those that use their hybrid build-up to incorporate functional capabilities. Less complicated examples include the inclusion of geometrically defined heat conduction paths in polymers or electrical conductors. This form of the process can be further developed at the very complex end of the spectrum to link or even accurately construct electronic components and systems [19, 20]. The use of multi-materials and various complex structures allows static or dynamic functionalities to be introduced, allowing mechanical, thermal, or electrical applications [21]. Shape memory materials (SMMs) do their behavior/actions through external stimuli. A programmable multi-material begins with the original form and seeks one or more transitional shapes to allow the effect of shape memory through step-by-step operation.

2.5 AM in COVID 19

COVID-19 is known as a novel coronavirus caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first case has been seen in Wuhan in China at the end of December 2019. Then it is spread rapidly all around the globe. The World Health Organization declared (WHO) a global pandemic on 11 March 2020 [22, 23]. The common source of spreading this virus is the droplet of saliva and sneezing droplets transferred from an infected person to another person. The presence of this virus shows the common symptom is mild fever, coughing, headache,

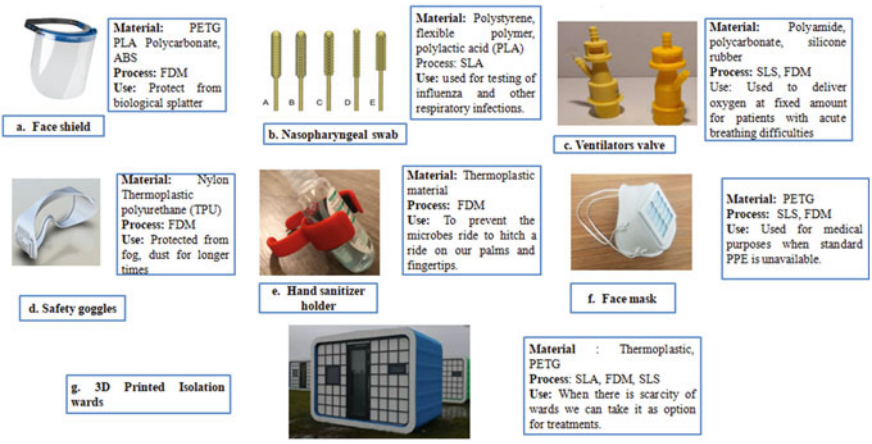


Fig. 5 3D printed medical devices to combat against COVID-19 [25]

and breathing problems so on. Lockdown, social distancing, personal hygiene, and work from home these kinds of policies have been taken to control the spread of disease. Health professionals are also continuously putting their effort to treat the infected patient. While treating the patient, health cares are required personal protective equipment, face shields, swab for testing and ventilator for respiratory-related problems, the rapid increase of this disease creates the shortage of medical devices [24, 25]. Till now, no cure has been found for this disease still research going on the vaccine trial for this virus. During this crisis, AM acts as a boon, which is to respond to emergencies arises in the health care systems. Emergencies are in terms of the high demand of medical devices and personal protective equipment for a health care professional. Following Fig. 5 demonstrates the application of 3D-printed items to fight against this novel coronavirus.

Rapid fabrication of product makes AM systems decentralize in nature and improve the global supply chains. The customer creates design according to their needs send to the nearest 3D printing house and get the custom-based product. This ability of 3D printing reduces the inventory cost, waste and minimizes the lead time. COVID 19 highlighted the need to transform the supply chain.

3 Challenges

Additive manufacturing is a transforming technology for multifaceted applications but there are numerous challenges. The most common challenges are discussed below;

- Preprocessing steps like tessellation affect the surface quality and accuracy of the parts.

- AM techniques in construction fields still in their implementing stage without an interdisciplinary approach.
- AM systems are required high-demand energy and printed parts create toxicity, it has an impact on the environment during the fabrication process.
- Establishing a standard for AM is highly necessary to achieve better quality parts.
- Size restriction is the major issue in AM, limited to the size of printer casing.
- AM is used for mass customization, but when there is need of mass production traditional manufacturing is faster than AM.
- Currently, printer and material cost is too much high.
- Limited simulation software is available to check the real-time behavior of 3D-printed parts.
- 3D-printed parts' design stored in digital platform so there may be chances of data leaked from the company.

4 Conclusion

Additive manufacturing is a well-suited approach for producing a complex and customized product with a fast delivery time to meet customer demand. Although it has many limitations, there are so many factors that are enhancing the popularity of this industry. The recent popularity of the 3D printing industry emerged to fight against the COVID-19 pandemic, where other manufacturing industries have become slow down. The future of additive manufacturing is bright because it is improving every year by utilizing its full potential.

References

1. Mellor S, Hao L, Zhang D (2014) Additive manufacturing: a framework for implementation. Int J Prod Econ 149:194–201
2. Rajaguru K, Karthikeyan T, Vijayan V (2020) Additive manufacturing–state of art. Mater Today Proc 21:628–633
3. Knuth JP (1999) Material increase manufacturing by rapid prototyping technique. Annals CIPP 40(2):603–604
4. Ngo TD, Kashani A, Imbalzano G, Nguyen KTQ, Hui D (2018) Additive manufacturing (3D printing): a review of materials, methods, applications and challenges. Compos B Eng 143:172–196
5. Astm standard f2792, standard terminology for additive manufacturing technologies (2013) http://www.astm.org/Standards/F2792.htm
6. Liaw CY, Guvendiren M (2017) Current and emerging applications of 3D printing in medicine. Biofabrication 9(2):024102
7. Tuomi J, Paloheimo KS, Vehviläinen J, Björkstrand R, Salmi M, Huotilainen E, Mäkitie AA (2014) A novel classification and online platform for planning and documentation of medical applications of additive manufacturing. Surg Innov 21(6):553–559
8. Sears NA, Seshadri DR, Dhavalikar PS, Cosgriff-Hernandez E (2016) A review of three-dimensional printing in tissue engineering. Tissue Eng Part B Rev 22(4):298–310

9. Culmone C, Smit G, Breedveld P (2019) Additive manufacturing of medical instruments: a state-of-the-art review. Addit Manuf
10. Yang J, Wu LW, Liu J (2001) US patent no. 6,280,785. Washington, DC: US Patent and Trademark Office
11. Sun J, Zhou W, Yan L, Huang D, Lin LY (2018) Extrusion-based food printing for digitalized food design and nutrition control. J Food Eng 220:1–11
12. Yang F, Zhang M, Bhandari B, Liu Y (2018) Investigation on lemon juice gel as food material for 3D printing and optimization of printing parameters. Lwt 87:67–76
13. Anukiruthika T, Moses JA, Anandharamakrishnan C (2020) 3D printing of egg yolk and white with rice flour blends. J Food Eng 265:109691
14. Liu Y, Liang X, Saeed A, Lan W, Qin W (2019) Properties of 3D printed dough and optimization of printing parameters. Innov Food Sci Emerg Technol 54:9–18
15. Thompson MK, Moroni G, Vaneker T, Fadel G, Campbell RI, Gibson I, Martina F (2016) Design for additive manufacturing: trends, opportunities, considerations, and constraints. CIRP Ann 65(2):737–760
16. Duballet R, Baverel O, Dirrenberger J (2017) Classification of building systems for concrete 3D printing. Autom Constr 83:247–258
17. 3D print canal house—DUS architects (217) http://houseofdus.com/project/3d-printcanal-house/
18. Gupta N, Weber C, Newsome S (2012) Additive manufacturing: status and opportunities. Science and Technology Policy Institute, Washington
19. Pei YT, Ocelik V, De Hosson JTM (2002) SiCp/Ti6Al4V functionally graded materials produced by laser melt injection. Acta Mater 50(8):2035–2051
20. Mahamood RM, Akinlabi ET (2017) Types of functionally graded materials and their areas of application. In: Functionally graded materials. Springer, Cham, pp 9–21
21. Bosse S, Lehmhus D, Lang W, Busse M (2018) Material-integrated intelligent systems: technology and applications. John Wiley & Sons
22. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al (2019) A novel coronavirus frompatients with pneumonia in China. N Engl J Med 382
23. WHO dashboard coronavirus. https://experience.arcgis.com/experience/685d0ace521648f8a5beeeee1b9125cd. Last Accessed 12 Oct 2020
24. Tarfaoui M, Nachtane M, Goda I, Qureshi Y, Benyahia H (2020) Additive manufacturing in fighting against novel coronavirus COVID-19. Int J Adv Manuf Technol 1–15
25. Choong YYC, Tan HW, Patel DC, Choong WTN, Chen CH, Low HY, Chua CK (2020) The global rise of 3D printing during the COVID-19 pandemic. Nat Rev Mater 5(9):637–639

Performance Analysis of Time Equivalent Space Vector Pulse Width Modulation Scheme for Three-Phase VSI at Inductive Load

Shailesh Kumar Gupta, Mohd. Arif Khan, and D. K. Chauhan

Abstract To get the controlled output from a three-phase voltage source inverter, various pulse width modulation (PWM) controlling schemes are available in the literature but time equivalent space vector pulse width modulation scheme (TESVPWM) is best among all schemes because it has less total harmonic distortion as compared with other. In the article, the time equivalent space vector pulse width modulation scheme is discussed for inductive load. The achievement of above-said scheme is analyzed in accordance with fundamental components and total harmonic distortion. To validate the performance of the three-phase voltage source inverter, simulation results are provided.

Keywords Two-level voltage source inverter (VSI) · Pulse width modulation (PWM) · Total harmonic distortion (THD)

1 Introduction

Inverters are most widely used to convert DC into AC. Inverters are classified as voltage-fed inverters and current-fed inverters. The voltage source inverter develops three-phase voltage having a governable amplitude, phase angle and frequency. To get the controlled output from three-phase voltage source inverter, various pulse width modulation (PWM) controlling schemes are available in the literature but time equivalent space vector pulse width modulation scheme (TESVPWM) is best among all schemes because it has less total harmonic distortion as compared with other. The time duration of conduction of semiconductor device used in inverters depends upon the logic or pulse width modulation techniques, which is used to control the

S. K. Gupta (✉)
Department of Electrical Engineering, Gautam Buddha University, Greater Noida, UP, India

Mohd. A. Khan
Department of Electrical Engineering, Fiji National University, Suva City, Fiji

D. K. Chauhan
Raj Kumar Goel Institute of Technology, Ghaziabad, UP, India
e-mail: drdkcexd@rkgit.edu.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_58

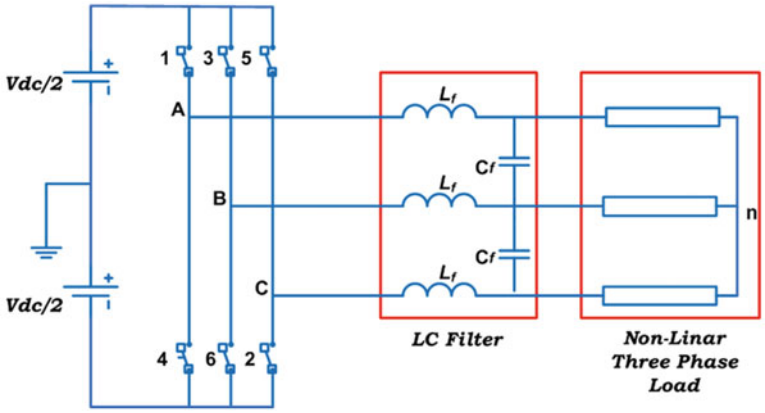


Fig. 1 Power circuit diagram of voltage. Source Inverter at three-phase non-linear inductive load

output. During PWM technique, the output of inverter (voltage and frequency) is controlled by the switching that occurred inside the inverter. The combination of IGBT, MOSFET, etc. along with diode connected antiparallel makes the switch. Three-phase inverters have three legs. Each leg consists of two switches and there are a totally of six switches. Both the switches of the same leg can never be on simultaneously to avoid the short circuit. Figure 1 indicates the power circuit diagram of three-phase voltage source inverter at non-linear three-phase inductive load. Most of the PWM schemes compare three-phase sinusoidal signal with high frequency carrier signal and generate switching signal to turn on the switch. The various PWM schemes used to control the output of three-phase inverters have third harmonic content in output [1].

In 2010, Md. Arif Khan proposed a simple and alternative scheme for three-phase VSI known as time equivalent space vector pulse width modulation scheme (TESVPWM) [2]. This scheme is best among all the schemes because it has less total harmonic distortion as compared with others. In this scheme, the switching time for inverter leg is generated during the sampling interval. An identical scheme has been mentioned in literature [3] for voltage source inverter. To implement this scheme, only sampled peak of phase voltages are utilized. Since sector identifications and look-up table are not used during this scheme, hence, computational time is very much reduced as compared with conventional PWM schemes.

In this article, author first discusses the time equivalent space vector pulse width modulation scheme along with the algorithm. Then presented a result on load and verified the results in MATLAB/SIMULINK environment.

2 Time Equivalent Space Vector Pulse Width Modulation Scheme

The conferred SVPWM referred to as here time equivalent space vector PWM (TESVPWM) utilizes merely the sampled voltages (reference) to come up with the turning time that every electrical converter leg to yield sinusoidal output. This scheme is best among all the schemes because it has less total harmonic distortion (THD) as compared with others. In this scheme, the switching time for inverter leg is generated during the sampling interval. Identical scheme has been mentioned in literature [3] for voltage source inverters. To implement this scheme, only sampled peak of phase voltages are utilized. Since sector identifications and look-up table are not used during this scheme, hence computational time is very much reduced as compared with conventional PWM schemes.

This offset addition-based scheme centers the active switching voltage vectors within the switching period. The algorithm is described herewith, where Vx; x = a, b, c; is the sampled magnitude of reference phase voltages throughout sampling duration and Ts is the converter switching duration. Tp; p = a, b and c; are termed as time equivalents of the sampled magnitudes of reference phase voltages. Tmin and Tmax are the minimum as well as maximum values of Tp throughout sampling duration. To is the time period for which the null voltage vector is fed in the switching duration. Toffset is the offset interval when sum up to time equivalent makes turning time signal or the converter leg switching duration Ttx; x=a, b and c. The switching sequence for sector I is described in Fig. 2. It is clear that the switching pattern is identical to SVPWM.

Algorithm of the TESVPWM:

(a) Reference voltages Va, Vb and Vc are sampled in each switching duration Ts.
(b) Decide the equivalent times duration T_1,$T_2$and T_3 given by equation , Tps = Vps* Ts/ V_d
where p = a, b and c

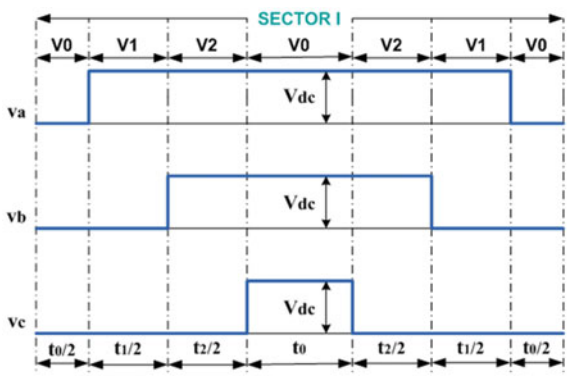


Fig. 2 Switching sequence for three-phase VSI using TESVPWM

(c) Decide Toffset; Toffset = $T_S/2$ - (T max +T min)/ V_d
(d) Inverter leg switching duration are achieved as
Ttx =Tp +Toffset; p = a, b and c.

3 MATLAB/SIMULINK Model

Convolutional neural networks (CNNs) are one of the most suitable neural networks for images. Our CNN model employs several components like kernel, activation function, pool, fully connected layers and softmax function.

Matlab/simulink model for three-phase system is built for simulation in following steps:

1. The three-phase sinusoidal input reference voltages are generated utilizing 'function' block at 120^0 phase shifted from each other.
2. The reference voltage is then transformed into two-phase equivalent utilizing 'function' blocks with the help of Clark's transformation equations. After this, two-phase equivalent is converted into polar form utilizing 'Cartesian to polar' block. Magnitude and corresponding angle are its first and second output of this function block.
3. With the help of two-phase equivalent dq coordinates, sector of the reference voltage vector is identified. Angle waveform is utilized for sector recognition.
4. Matlab function block is utilized for calculating the switching time and related switching state for every power semiconductor switch. Timer signal is compared with magnitude and angle of the reference with the help of the Matlab code. We get the gating signal for the all the semiconductor switches of three-phase two-level inverter. Figure 3 clearly indicates the time equivalent SVPWM switching generation for three-phase VSI at inductive load.
5. In this last step, this gating signal is applied to all the power switches of three-phase two-level inverter. The output of inverter is connected with three-phase inductive load as shown in Fig. 4. Desired output filtered phase voltages are received at the output.

4 Simulation Study

Figure 5 indicates the filtered output voltage having a RL load at the output terminals of VSI and Fig. 6 indicates the harmonic spectrum of output voltage for phase 'a'. Throughout the analysis, switching frequency of three-phase VSI, fundamental frequency and DC supply voltage are kept 5 kHz, 50 Hz and unity. The achieved reference voltage is 0.5103 peak (0.3608 rms). All the input voltages are 120° displaced from each other. The DC component in the output is 3.333e-05. The total harmonic distortion at load is 3.18%, which is very less as compared with other schemes.

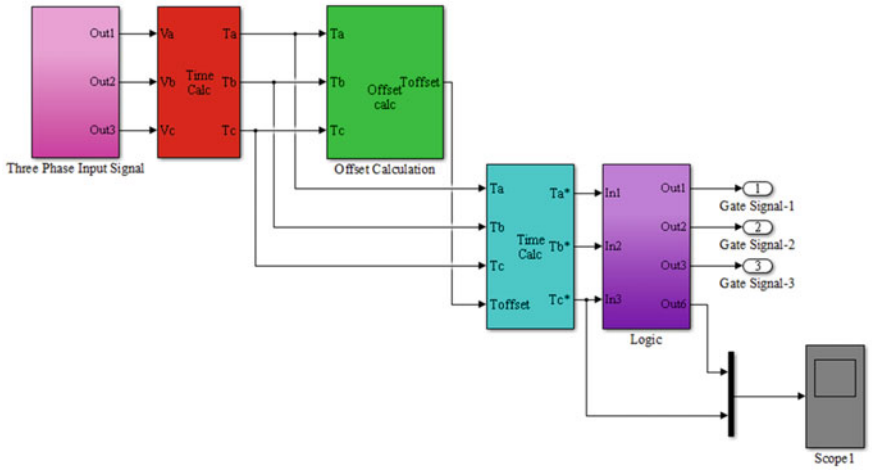


Fig. 3 MATLAB/SIMULINK Model for three-phase VSI

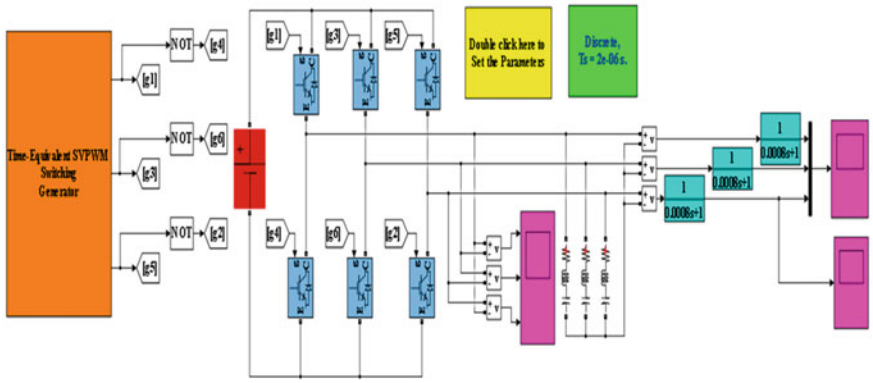


Fig. 4 Time equivalent SVPWM switching generation

5 Conclusion

To achieve the performance of three-phase voltage source inverter at non-linear inductive load, time equivalent space vector PWM scheme is presented in this article. It is concluded that the value of the fundamental component of voltage at inductive load is 0.5103 peak, which is very much higher as compared with other PWM schemes. The total harmonic distortion on load is also very less and equal to is 3, 18%. Hence it is found that this technique is suitable for three-phase medium voltage applications as compared with other schemes. The implementation of this scheme is also very simple, reliable and easygoing.

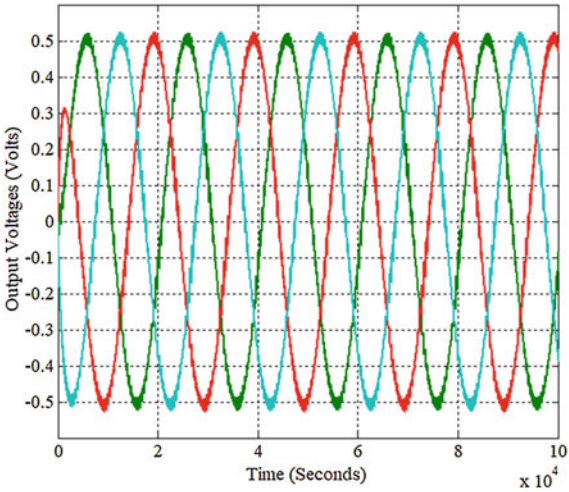


Fig. 5 Output voltage across load (Volts)

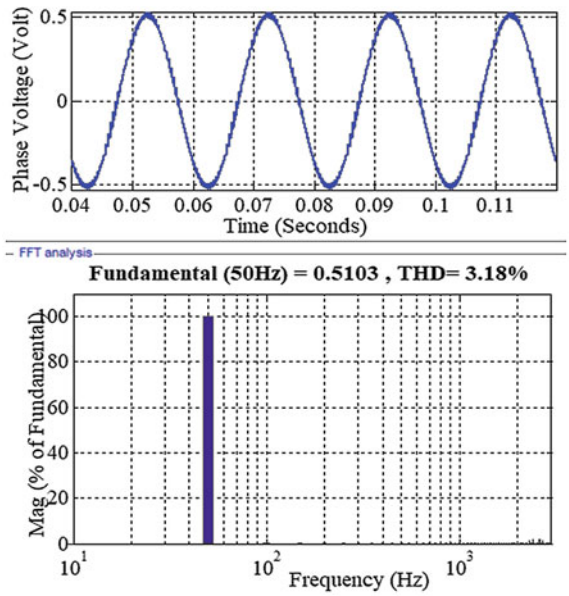


Fig. 6 THD for the output phase 'a' voltage

Modulation scheme of three-phase system for real-time power conversion has presented in [4], and PWM with harmonic injection for the multi-phase system has presented in [5]. Analysis of voltage space vectors [6] and PWM schemes are presented in [7]. Comparison of sinusoidal and space vector-based PWM has been given in [8]. Review on poly-phase machine development [9], multi-phase dual inverter system configuration [10], and on SVPWM [11] are in literature. PWM switching analysis for dual inverter using ANN has been presented in [12].

References

1. Kazmierkowski MP, Krishnan R, Blaabjerg F (2002) Control in power electronics- selected problems. Academic Press, California, USA
2. Arif Khan, Md., Iqbal A (2010) Comparative study between existing & proposed pwm techniques for three phase voltage source inverter. In: IEEE India international conference on power electronics (IEEE IICPE-2010)
3. Singh GK (2002) Multi-phase induction machine drive research—a survey. Electric Power Syst Res 61:139–147
4. Chung DW, Kim JS, Sul SK (1998) Unified voltage modulation technique for real time three-phase power conversion. IEEE Tram Ind Appl 34(2):374–380
5. Iqbal A, Levi E, Jones M, Vukosavic SN (2006) Generalised sinusoidal PWM with harmonic injection for multi-phase VSIs. In: IEEE 37th power electronics specialist conference (PESC) Jeju, Korea, 18–22 June 2006, CD_ROM paper no. ThB2–3, pp 2871–2877
6. vander Broek HW, Skudelny HC, Stanke GV (1988) Analysis and realization of PWM based on voltage space vectors. IEEE Trans Ind Appl 24(1):142–150
7. Holmes DG, Lipo TA (2000) Pulse width modulation for power converters—principles and practices. IEE Press, Wiley Publications
8. Wang FEE (2000) Sine-triangle versus space vector modul;ation for three-level PWM voltage source inverters. In: Proceedings of the IEEE-IAS annual meeting, Rome, pp 2482–24888
9. Gupta SK, Singh O, Khan MA, Kushwaha AK (2020) A review on developments of polyphase machines. J Inf Opt Sci 41(1):327–343. https://doi.org/10.1080/02522667.2020.1721628
10. Gupta SK, Singh O, Khan MA, Kushwaha AK (2019) A review on multiphase dual Inverter system configuration. In: 29th Australasian universities power engineering conference, Fiji
11. Gupta SK, Singh O (2018) A review on space vector pulse width modulation in different types of system. In: 42nd national systems conference (NSC-2018) ASEND, Trivandrum
12. Gupta SK, Singh O (2018) Pulse width modulation switching analysis for three phase dual inverter system using artificial neural network. In: 28th Australasian universities power engineering conference, Auckland

A Comprehensive Review on Active and Reactive Power Control of Grid Connected Converters

Arunesh Kumar Singh, Abhinav Saxena, Imran Ahmad, and Umakanta Choudhury

Abstract This paper presents the active and reactive power control of grid-connected converters. The converters are controlled in nature. The complete observation for controlling reactive and active power with different techniques is shown for consumable power of 1 kW. In most of the cases, solar PV of 2 kW array-based grid-connected inverter shows better performance in comparison to other systems. The performance of grid-connected converters with Z-source inverters gives better performance.

Keywords Converters · Active power · Reactive power · Grid · Solar · Z-source inverter

1 Introduction

There has been a rapid growth of integration of renewable energy based distribution generation (DG) in electrical power system grid recently. Because renewable energy sources (RES) aid mitigate emissions, reducing electricity transmission losses, and better utilize the locally available RES [1]. RES are intermittent in nature and change with variation in atmospheric conditions, which bring problems and challenges for the electrical grid. Some of the problems are voltage deviations, frequency fluctuations, and reverse power flow from grid to DG system [2]. Therefore, integration of distribution generation by exiting electrical grid can cause reverse power flow, fault ride through, voltage fluctuations, etc. So new generation inverters are designed to which can provide power system support functions such as compensation of reactive power, FRT, etc. [3]. There are many methods with a focus on reactive power

A. K. Singh · I. Ahmad (✉)
Jamia Millia Islamia, New Delhi, India

A. Saxena
JSS Academy of Technical Education, Noida, UP, India

U. Choudhury
Rajkumar Goel Institute of Technology, Aktu, Ghaziabad, U.P., India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_59

control grid connected converters. Grid-connected converters have mainly two important purposes, one extracting maximum power from PV array and two to providing reactive power to the grid.

Grid-connected converter may be two-stage topology systems in which boost converter and 1-phase inverter are separately preceded by filter and isolation transformer. Maximum power point tracking (MPPT) is simple to introduce by simply changing the boost converter switch to a two-stage topology, but it renders the overall network voluminous and contributes to unnecessary errors that decrease the overall device performance [4]. Therefore, single-stage power conversion unit (PCU) is practically more useful, single-stage power conversion unit is compact and less costly, however, it is challenging to implement MPPT and control algorithms in single-stage PCU. A lot of control methods for the control of reactive power in grid-connected converter are discussed in different research papers. Different types of controllers such as proportional-integral (PI), hysteresis band current (HBC), proportional-resonant (PR), and predictive controller are for use for finding the value of reference current [5–7]. The objective of this paper is to review the available technologies regarding reactive power control of grid-connected converter.

2 Different Methods of Reactive Power Control

PV system manually built to provide the gird's active control, however, due to widespread integration of the PV system into low voltage distribution networks, the feeders' tension profile can alter. When power generation from PV is greater than load and grid consumption a reverse power flow occurs which leads to overvoltage and the synchronization of defensive equipment and the protection and security of networks should not be disrupted. This is why a sensitive power management policy will be placed in order to maintain voltage within the appropriate range [8].

2.1 Unity Power Factor Method

This is the oldest method in which inverter runs without injection of reactive power into the grid. The disadvantages of this method is there might be over voltage in those place where high penetration of PV [9].

2.2 Active Power Dependent Method

This method adjusts the reactive power based on the active power out-put of the inverter, so it provides voltage regulation active power output variations [10].

2.3 Reactive Power(Q) Control

In this method, reactive power Q depends on local electrical power system voltage. The inverter either absorbs or generates reactive power as its terminal voltage deviates from specified critical upper or lower voltages VU and VL respectively [1, 11].

2.4 Two-Quadrant P-Q Control

In this way, the inverter works in a 2 quadrant P-Q plane, where the inverter produces inductive or capacitive reactive power at varying voltage levels [1, 12].

The inverter produces capacitive reactive power when the voltage falls below the lower voltage level, so when the voltage reaches the upper voltage point the inverter produces inductive reactive electricity.

2.5 Reactive Power Control at Night

At night, there is nonavailability of solar radiation, so PV system can provide active power but it can provide reactive power and can be used as static Var generator (SVG) [1, 3].

3 Voltage Efficiency Improvement in Low-Voltage Delivery Networks Use Single-Phase Inverter Reactive Control

A novel technique for successful voltage control in a three-phase four-wire low-voltage distribution network (LVDN) with strong PV device penetration is suggested. When huge number of rooftop PV systems are integrated with LVDNs, it disturbs the voltage profile of the feeders and deteriorates voltage regulation, particularly when the difference between the PV generation and the load usage profile occurs, severe voltage fluctuations can occur. When created during high PV [13, 14].

4 Using Lyapunov Function Based Current Controller

It proposed controlling the active and reactive power flow from sustainable sources of energy to a three-phase network microgrid. Not only would the proposed monitoring system follow the net. The current flow which also flows through gross harmonic system current distortion where traditional nonlinear loads are present. The Test Unit

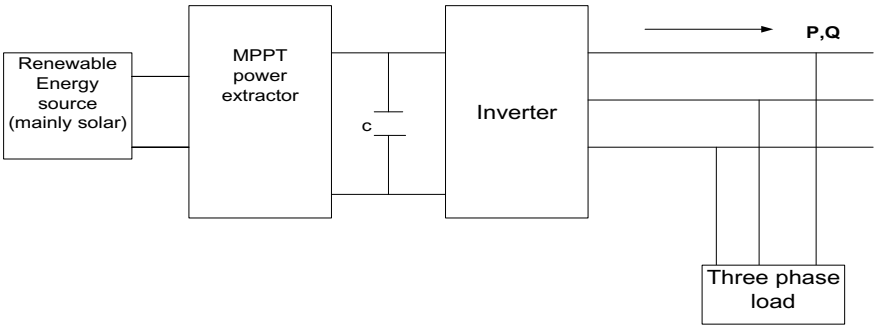


Fig. 1 Structure of system for control strategy system

forms system current keeping device voltage in mind unbalance, harmonics, and in line-side inductors unbalance. The efficiency of the control system is assured by the direct Lyapunov process. The proposed (digitally implemented) control framework provides a superior output over conventional multiples because of the absence of the PARK connection points, the control systems are proportional-integral and proportional-resonant, As the requirement for a phase lock loop in the control structure.

A New One Inverter simulation methodology to take account of imbalances in all system voltages. A new existing control strategy for a parallel renewable energy source-based three-phase inverter to link to the as shown in Fig. 1. The idea is for a simplified microgrid network. Strategy for Control is explicitly applied in the a—b—c picture, and can be taken. Take a note of unbalanced situations in both the system voltage as inductances and load side of panel. The proposed test system is used on a remote device and includes like the traditional route, no transformation block at Park, where several controls are applied synchronously in a dual frame of reference rotates and provides superior efficiency.

Hence the solution proposed functions without PLL is necessary. A three-step inverter simulation device linked to the grid is unequaled touch. Requisite for promoting the proposed regulation is also laid out operation. The proposed control strategy applies to the conventional three-phase inverter, six-switch topology (b–6 configuration), is shown in Fig. 1 in which MPPT extracts maximum power from solar and its output further fed to load via boost converter and inverter. The experimental findings reported illustrate the efficacy of the Supervisor [15, 16].

5 Efficient P–Q Theory

This paper provides an efficient power management strategy for single-process GCPVS by p–q theory applications clashed with the HBC system, where the reference current is obtained using updated p–q theory such that the working system is

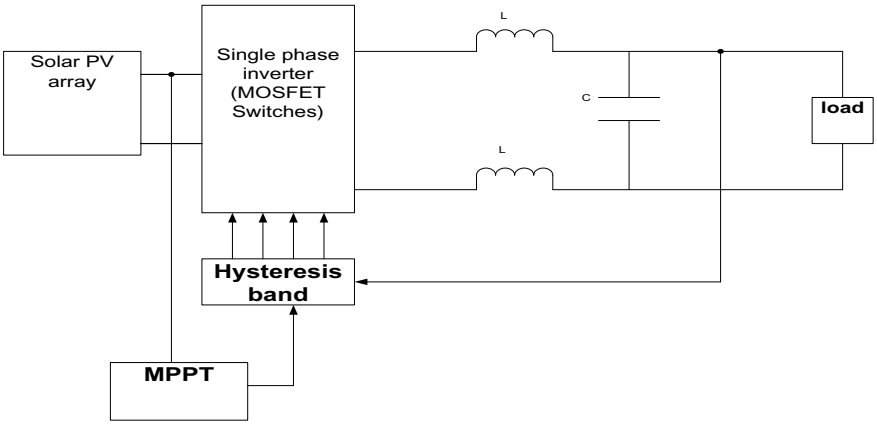


Fig. 2 Layout of p–q theory for one-phase inverter

successful and the reactive power flow is independently controllable. The present reference point is being determined by using modifications of p–q theory. It is shown in Fig. 2 in which switching of MOSFET-based inverter is done by using hysteresis band. In addition to the precision, the convergence rate of MPPT operation is enhanced by changing the perturbation stage size through relation to the maximum power point. Integrating power conversion and MPPT functions should render the device lightweight more functional and cost-effective [17, 18].

6 Z-Source Inverter

This article shows the regulation of the predictive capacity algorithm, which decouples responsive and reactive grid control photovoltaic (PV) structures are designed using a quasi-Z basis Turner (q ZSI). That is necessary in order to reach the evolving smart Inverter Network Interconnection specifications. The suggested controller features a simple, functional interface implementation, rapid dynamic response under changing transient state, and marginal steady-state monitoring error Active and reactive power control with decoupling in a typical Distributed Generation Systems (DGs). The grid-tied PEI for PV applications using qZSI is Built on an MPC System basis. The system will be analyzed for photovoltaic uses as distributed generation and auxiliary grid operation obligation at PCC. Suggested Precise regulation produces the comparisons to variables to take control. The program even tackles overcurrent issues during the grid.

Faults that have loop security overcurrent which is a restriction on cost function optimization. In fact, a total of power controllers is also used in the minimizing algorithm. The desire is to turn from one stress variable to another., Thus, it seeks to reduce risks from swapping.

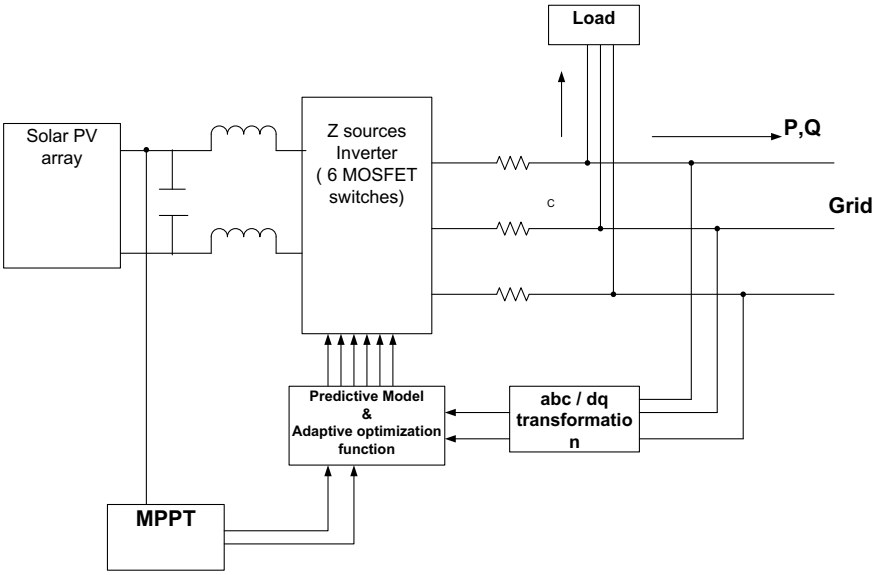


Fig. 3 Design of Z-source inverter for solar PV array

The machine was trackable and capable of delivering full PV power with fast monitoring dynamics with parallel individual accountability for reactive power for the load system. Can the program condition the grid therefore by providing the reactive power needed Grid service with device control factor [19, 20]. The complete structure is shown in Fig. 3 in which switching of the inverter is assessed by using predictive model and adaptive optimization based on abc-dq transformation.

The general equations for Z-source inverter are:

$$\begin{bmatrix} \dot{i}_{l1} \\ \dot{i}_{\alpha} \\ \dot{i}_{\beta} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{-R}{L} & 0 \\ 0 & 0 & \frac{-R}{L} \end{bmatrix} \begin{bmatrix} i_{l1} \\ i_{\alpha} \\ i_{\beta} \end{bmatrix} + \begin{bmatrix} \frac{-1}{L1} & 0 & 0 \\ 0 & \frac{-1}{L} & 0 \\ 0 & 0 & \frac{-1}{L} \end{bmatrix} \begin{bmatrix} v_{c1} \\ e_{\alpha} \\ e_{\beta} \end{bmatrix} + \begin{bmatrix} \frac{1}{L1} & 0 & 0 \\ 0 & \frac{1}{L} & 0 \\ 0 & 0 & \frac{1}{L} \end{bmatrix} \begin{bmatrix} v_{pv} \\ v^{\alpha}_{qzsi} \\ v^{\beta}_{qzsi} \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} \dot{i}_{l2} \\ \dot{i}_{\alpha} \\ \dot{i}_{\beta} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{-R}{L} & 0 \\ 0 & 0 & \frac{-R}{L} \end{bmatrix} \begin{bmatrix} i_{l2} \\ i_{\alpha} \\ i_{\beta} \end{bmatrix} + \begin{bmatrix} \frac{-1}{L1} & 0 & 0 \\ 0 & \frac{-1}{L} & 0 \\ 0 & 0 & \frac{-1}{L} \end{bmatrix} \begin{bmatrix} v_{c2} \\ e_{\alpha} \\ e_{\beta} \end{bmatrix} + \begin{bmatrix} \frac{1}{L1} & 0 & 0 \\ 0 & \frac{1}{L} & 0 \\ 0 & 0 & \frac{1}{L} \end{bmatrix} \begin{bmatrix} v_{pv} \\ v^{\alpha}_{qzsi} \\ v^{\beta}_{qzsi} \end{bmatrix} \tag{2}$$

Table 1 Comparative performance of all methods

Method	THD value (%)
Z-source inverter	7.65
Efficient p–q theory	9.98
Lyapunov function-based current controller	10.98
1-phase inverter reactive control	12.65

where

i_{l1} is terminal current 1,
I_{l2} is terminal current 2,
i_{α} current in stationary coordinates,
$i\beta$ current in stationary coordinates,
vc_1 is capacitor voltage at terminal 1,
vc_2 is capacitor voltage at terminal 2,
v_{pv} is voltage of solar PV array,
e_{α} is induced voltage in stationary coordinates of terminal 1, and
e_{β} is induced voltage in stationary coordinates of terminal 2.

The comparative performance for improving the harmonics or THD is shown in Table 1, which shows that Z-source inverter gives better results or lesser THD in comparison to all.

7 Conclusion

This paper discusses the active and reactive power management of grid-linked converters. The converters are nature-controlled. Full observation is seen for regulating reactive and responsive power with various techniques. In most instances, the grid-linked inverter centered on a solar PV array displays improved efficiency compared to other devices. It is observed that flow of reactive & active power is better analyzed in a Z-source inverter with minimum THD.

References

1. Zhao X, Chang L, Shao R, Spence K (2018) Power system support functions provided by smart inverters—A review. CPSS Trans Power Electron Appl 3(1):25–35
2. Datta A, Sarker R, Hazarika I (2019) An efficient technique using modified p–q theory for controlling power flow in a single-stage single-phase grid-connected PV system. IEEE Trans Industr Inf 15(8):4635–4645
3. Zeraati M, Golshan MEH, Guerrero JM (2019) Voltage quality improvement in low voltage distribution networks using reactive power capability of single-phase PV inverters. IEEE Trans Smart Grid 10(5):5057–5065

4. Jain S, Shadmand MB, Balog RS (2018) Decoupled active and reactive power predictive control for PV applications using a grid-tied quasi-z-source inverter. IEEE J Emerg Sel Top Power Electron 6(4):1769–1782
5. Lin F-J, Kuang-Chin Lu, Ke T-H, Yang B-H, Chang Y-R (2015) Reactive power control of three-phase grid-connected PV system during grid faults using Takagi–Sugeno–Kang probabilistic fuzzy neural network control. IEEE Trans Industr Electron 62(9):5516–5528
6. Dasgupta S, Mohan SN, Sahoo SK, Panda SK (2013) Lyapunov function-based current controller to control active and reactive power flow from a renewable energy source to a generalized three-phase microgrid system. IEEE Trans Ind Electron 60(2):799–813
7. Yang Y, Blaabjerg F, Wang H, Simões MG (2016) Power control flexibilities for grid-connected multi-functional photovoltaic inverters. IET Renew Power Gener 10(4):504–513
8. Han Y, Li H, Shen P, Coelho EAA, Guerrero JM (2017) Review of active and reactive power sharing strategies in hierarchical controlled microgrids. IEEE Trans Power Electron 32(3):2427–2451
9. Paál E, Tatai Z (2010) Grid connected inverters influence on power quality of smart grid. In: Power electron. Motion control, pp 35–39
10. Smith JW, Sunderman W, Dugan R, Seal B (2011) Smart inverter volt/var control functions for high penetration of PV on distribution systems. In: 2011 IEEE/PES power systems conference and exposition PSCE 2011, pp 1–6
11. Seuss J, Reno MJ, Broderick RJ, Grijalva S (2011) Improving distribution network PV hosting capacity via smart inverter reactive power support. In: IEEE/PES power system conference expo PSCE, pp 1–6
12. Olivier F, Aristidou P, Ernst D, Van Cutsem T (2016) Active management of low-voltage networks for mitigating overvoltages due to photovoltaic units. IEEE Trans Smart Grid 7(2):926–936
13. Abdelhakim A, Davari P, Blaabjerg F, Mattavelli P (2017) Switching loss reduction in the three-phase quasi-z-source inverters utilizing modified space vector modulation strategies. IEEE Trans Power Electron 33(5):4045–4060
14. Erginer V, Sarul MH (2013) A novel reduced leakage current modulation technique for Z-source inverter used in photovoltaic systems. IET Power Electron 7(3):496–502
15. Noroozi N, Zolghadri MR (2017) Three-phase quasi-z-source inverter with constant common-mode voltage for photovoltaic application. IEEE Trans Ind Electron 65(6):4790–4798
16. Zhou Y, Liu L, Li H (2013) A high-performance photovoltaic moduleintegrated converter (MIC) based on cascaded quasi-Z-source inverters (qZSI) using eGaN FETs. IEEE Trans Power Electron 28(6):2727–2738
17. Anderson J, Peng FZ (2008) Four quasi-z-source inverters. In: IEEE power electronics specialists conference, pp 2743–2749
18. Anurag A, Yang Y, Blaabjerg F (2015) Thermal performance and reliability analysis of single-phase PV inverters with reactive power injection outside feed-in operating hours. IEEE J Emerg Sel Top Power Electron 3(4):870–880
19. Hoke A, Shirazi M, Chakraborty S, Muljadi E, Maksimovic D (2017) Rapid active power control of photovoltaic systems for grid frequency support. IEEE J Emerg Sel Top Power Electron 5(3):1154–1163
20. Yang Y, Wang H, Blaabjerg F (2014) Reactive power injection strategies for single-phase photovoltaic systems considering grid requirements. IEEE Trans Ind Appl 50(6):4065–4076

An Energy Management System for Microgrid Resilience Improvement

Raj Kumar Yadav and Dipti Saxena

Abstract The introduction of renewable energy power such as solar energy, wind energy, diesel energy, etc., has led the system to have a backup system for various thermal and hydropower plants. During any problem in the primary system can lead to a complete blackout and hence the supply to the distribution system and the load gets affected. Secondary systems are used as a backup system for the primary system which provides continuous supply to load end when primary system fails to do so. In order to improve the power quality and environmental protection due to an improvement in the increased integration of the renewable energy power system decentralization is required. This algorithm provides efficient and improved operation as compared to other proposed techniques and algorithms. The decentralized microgrids can be connected to the large microgrids under special cases such as earthquakes, tsunami, and various natural and manmade events. An energy management system has been designed for these types of techniques for a microgrid, which is to be operated in islanded mode during various calamities. In this system, the power is stored in the battery system by importing from nearby grids. The proposed system is applied to microgrids consisting of a battery system.

Keywords Energy management system (EMS) · Microgrid · Grid-connected system · Resilient energy management system (REMS)

1 Introduction

In order to compensate or reduce the production of carbon and carbon emission in a power system, renewable energy sources are need to be increased in a main system as it reduces pollution. There is also a problem if too many renewable sources are used then the main power system may get affected and can lead to various changes

R. K. Yadav (✉)
Department of Electronics Instrumentation & Control Engineering, Government Engineering College Ajmer, Ajmer, Rajasthan, India

D. Saxena
Department of Electrical Engineering, MNIT Jaipur, Jaipur, Rajasthan, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_60

Table. 1 Blackout percentage of USA

Sr. no	Natural disaster/man made event	(%)
1	Natural disaster	44
2	Storms	43
3	Manmade events	34

in the power flow. Many countries are using this technique on a very large scale for their small plants and microgrids that operate in a stable manner [1–3].

They have the ability to utilize the power and maintain a balance between the systems. Whenever there is a microgrid operating in the grid-connected mode and a natural disaster occurs such as typhoons, earthquake, tsunamis, or some manmade events can lead to disconnection of the microgrid and islanded operation is achieved. The Natural and manmade disasters data of the United States of America (USA) data is shown in Table 1 [4–6].

Therefore, from the data provided above, the disruption can be predicted and the system can be prepared against them. If the microgrid has high resiliency, it can cope with these disturbances. The definition of resiliency is the ability to adapt or cope with the grid's disconnection and to maintain the power supply to the load after the disconnection of the main grid and recover the system as soon as possible. Various microgrids get damaged due to the disruption [7, 8].

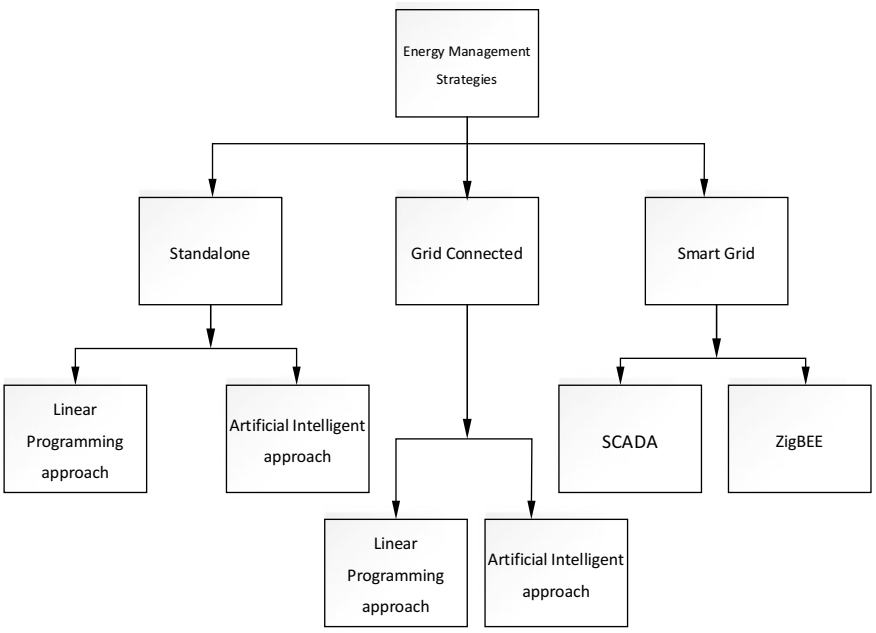


Fig. 1 Energy management strategies

The above Fig. 1 shows the different energy management systems. Standalone system operates individually as it is not connected to the grid whereas grid-connected system is a system whose output is connected to the grid. A smart grid is also a type of a grid that is used remotely with the help of various systems such as SCADA system. The occurrence of a natural disaster is low in some of the areas but it does not mean that the system has not to be prepared for the disruption. Even when the probability is low resilience must be high because it is a thing that cannot be predicted. When it comes to the highly dangerous areas or regions, the resiliency must be high so that they can work in an isolated mode without any disruption. Whenever there is a disruption with the low resiliency the microgrid gets damaged. Therefore, to prevent it the resiliency must be maintained [9–12].

Various algorithms have been introduced in order to prevent disruption but the operation can be improved by upgrading or changing various methods of operation and increase the efficiency of the operation. The microgrid must be able to work or operate under isolated mode and it must be capable of maintaining the power to the system for an amount of time till the connection is secured. It maintains the power that is required at the load end. Therefore, these systems must be implemented in various grids so that the smooth and continuous operation can be obtained and also the pollution is reduced due to the low carbon emission and hence it also protects the environment.

By summarizing all of this, the definition of a resilient microgrid is an ability to prepare, withstand, adapt to and recover from a disruption, while the ability to consistently deliver electric power to critical load is implemented. In this paper, an energy management system algorithm is proposed which is to operate in an accident in islanded mode. If the amount of procured reserve power is more than the one needed, the usage of generated power from renewable energy will be limited and if not enough reserve power is procured, the microgrid will not be able to operate in islanded mode until it recovers back to connected mode, due to which resiliency decreases.

Therefore, to improve the electric power efficiency given to microgrid an algorithm is formulated which calculates the reserve power. The algorithm is analyzed by comparing it with a normal energy management system.

The algorithm is used in order to improve resiliency. Due to various disasters which may lead to disconnection, it imports the power from the connected grids before the connection is break [12–15].

2 Algorithm

In this proposed paper, an energy management system having reserve power, for increasing the microgrid resiliency, is proposed. The algorithm of the method used is shown in the algorithm below. The microgrid data and status is obtained and the critical load and photovoltaic data is obtained. Whenever there is an accident in the system the system works in the islanded mode. Whenever there is a normal operation

the system works in a grid-connected system. References [15–18] the algorithm also detects the accident as shown in the flowchart. Initially, there is an assumption made that there is no accident occurred, but it is predicted. If the predicted accident has occurred the system operation is scheduled for islanded operation. Whenever in some cases there is a mismatch in the power generation and the power provided in the islanded mode. In order to prevent such conditions, the proposed algorithm is used, it provides the electric power until the production exceeds the critical load. The reserve power is procured and the expense is minimized [18–21].

Algorithm of REMS method

Step 1 Read data
Step 2 If fault occurs then go for standalone mode and else go for Grid Connected system
Step 3 Schedule Micro grid operation
Step 4 Continue till time is over
Step 5 Stop

Using the algorithm of reserve power procurement as shown below, the power can be stored or reserved by importing the power from various connected grids by purchasing or operating renewable energy source. The critical load and non-critical load for reserve and prediction data is shown in Tables 2 and 3 respectively.

Reserve Power procurement algorithm

Step 1 Read time when Photovoltaic is greater than load
Step 2 Calculate t - T(1) = a
Step 3 Calculate t - T(2) = b
Step 4 If b is greater a then sell through utility grid or store else sell to utility grid
Step 5 Stop

It is seen from Table 2 that reserve capacity of PV is generally more than load. Table 3 shows PV power is more compared to critical and non-critical load for some cases. The overall cost and the cost of renewable energy source like diesel generator should be optimized for low cost. But in most cases importing power from a connected grid is easy and the cost is also low as compared to other sources.

Table. 2 Reserve Procurement Data

S.No.	Time	Critical Load (Power in KW)	PV (Power in KW)
1	0	18	14
2	10	18	12
3	20	22	15
4	30	28	20
5	40	24	35
6	50	26	45
7	60	22	40
8	70	20	35
9	80	17	15
10	90	22	12
11	100	22	14
12	110	26	12
13	120	24	30
14	130	20	45
15	140	26	30

Table. 3 Prediction Data

S.No.	Time	Critical Load (Power in KW)	Non-Critical Load (Power in KW)	PV (Power in KW)
1	0	17	28	14
2	5	17	28	14
3	10	17	29	13
4	15	18	32	12
5	20	19	34	15
6	25	25	32	20
7	30	28	35	25
8	35	25	32	30
9	40	22	38	43
10	45	25	35	45
11	50	28	30	48
12	55	22	25	35
13	60	23	22	32
14	65	26	20	25
15	70	20	18	15

3 Conclusion

In this paper, a resilient energy management system (REMS) is proposed whose main objective is to increase the resiliency of a small-scale microgrid by reserve power procure effectively. In general, it is adapting with the stress which is analogous to the disturbances that occur in a system. During various natural disasters, the microgrid operation may become malfunctioned/disturbed. In such cases, microgrid must work in islanded mode. Whenever a microgrid operates in this mode, the grid must provide the electrical power to the loads until the microgrid gets back to work in the connected mode. It also causes less emission carbon and hence causes less pollution and is eco-friendly. The use of renewable energy sources is increasing which works as a secondary system and can provide electrical power for quite a duration of time. The hybrid systems provide high efficiency, good power output, the surge is reduced, and also they are highly reliable. The resiliency of the system is increased by the use of this energy management technique.

References

1. Steward D, Zuboy J (2014) Community energy: analysis of hydrogen distributed energy systems with photovoltaics for load leveling and vehicle refueling. National Renewable Energy Lab.(NREL), Golden, CO (United States), Technical report NREL/TP-6A20–62781
2. Japanese smart energy products & technologies (2018) Japan bus alliance smart energy worldwide (JASE-W), Tokyo, Japan
3. 'Forging a new approach to using clean energy (2016) Southern California edison, Rosemead, CA, USA, Technical report
4. Nakanishi Y (2014) Kitakyushu smart community & technologies. In: Proceedings of the IEEE ISGT, pp 5–7
5. Woking borough council's (2005) Joint venture project, energy saving trust, London, UK
6. Stone M (2017) Sonnen tries different virtual power plant models in Germany, Australia and America. Greentech Media. Lst Accessed 30 Nov 2018
7. Hines P, Apt J, Talukdar S (2008) Trends in the history of large blackouts in the United States. In: Proceedings of the IEEE power and energy society general meeting-conversion and delivery of electrical energy in the 21st century, pp 1–8
8. Ahmad R, Asim MA, Khan SZ, Singh B (2019) Green IoT—issues and challenges. In: Proceedings of 2nd international conference on advanced computing and software engineering (ICACSE)
9. Asim M, Khan MS, Ahmad J, Umar T, Riyaz A (2020) Efficiency enhancement of solar panel using photodiode. In: Energy systems, drives and automations. Springer, Singapore, pp 215–223
10. Sinha S, Agarwal, P, Gupta NK, Asim M, Riyaz A (2020) Performance of solar cell under changing atmospheric condition. In: Energy ssystems, drives and automations. Springer, Singapore, pp 225–234
11. Balasubramaniam K, Saraf P, Hadidi R, Makram EB (2016) 'Energy management system for enhanced resiliency of microgrids during islanded operation.' Electr Power Syst Res 137:133–141
12. Garba M, Tankari MA, Lefebvre G (2017) Using of distributed energy ressources for microgrid resilience achieving. In: Proceedings of the IEEE 6th international conference renewable energy research and applications (ICRERA), pp 5–8

13. Hussain A, Bui V-H, Kim H-M (2019) Resilience-oriented optimal oper- ation of networked hybrid microgrids. IEEE Trans Smart Grid 10(1):204–215
14. Hussain A, Bui V-H, Kim H-M (2017) Optimal operation of hybrid microgrids for enhancing resiliency considering feasible islanding and survivability. IET Renew Power Gener 11(6):846–857
15. Khodaei A (2014) Resiliency-oriented microgrid optimal scheduling. IEEE Trans Smart Grid 5(4):1584–1591
16. Nagata T, Tao Y, Kimura K, Sasaki H, Fujita H (2004) A multi-agent approach to distribution system restoration. In: Proceedings of the 47th midwest symposium on circuits and systems, MWSCAS, p II
17. Toune S, Fudo H, Genji T, Fukuyama Y, Nakanishi Y (1998) A reactive tabu search for service restoration in electric power distribution systems. In: Proceedings of the IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence, pp 763–768
18. Luan WP, Irving MR, Daniel JS (2002) Genetic algorithm for supply restoration and optimal load shedding in power system distribution networks. IEE Proc Gener Transmiss Distrib 149(2):145–151
19. Eskandarpour R, Lotfi H, Khodaei A (2016) Optimal microgrid placement for enhancing power system resilience in response to weather events. In: Proceedings of the North American power symposium (NAPS), pp 1–6
20. Ren L, Qin Y, Wang B, Zhang P, Luh PB, Jin R (2017) Enabling resilient microgrid through programmable network. In: Proceedings of the IEEE power energy social general meeting, p 1
21. Chittum A (2017) What is the economic value of microgrid resilience?

A General Review of the Recently Proposed Asymmetrical Multilevel Inverter Topologies

Bilal Alam, Maaz Nusrat, Zeeshan Sarwer, Mohammad Zaid, and Adil Sarwar

Abstract This paper presents a comparative study of some of the topologies of asymmetrical multilevel inverters (MLI), which have been the focus of work of many researchers due to its many advantages over the conventional two-level inverter. Reduction in the total harmonic distortion (THD), lesser electromagnetic interference and increased fault-tolerant capability are some of the merits of MLIs. They can be classified into symmetrical and asymmetrical structures. The asymmetrical MLIs have a different magnitude of DC sources which results in obtaining more number of levels. The different modulation techniques used in MLIs are also discussed. The comparison parameters between the different topologies are the number of switches, number of driver circuits, number of DC sources, and number of levels. Comparison has been shown in the tabular form. Graphical analysis of the comparison has also been added in the paper for a better understanding.

Keywords Multilevel inverters (MLI) · Total standing voltage (TSV) · Asymmetrical MLIs · Pulse-width modulation (PWM)

1 Introduction

An inverter was introduced specifically to convert DC power to AC power at the required frequency and voltage. A conventional 2-level inverter is a square wave inverter that generates output with 2 voltage levels, i.e., +V and −V. Although it serves the purpose, it has several disadvantages such as electromagnetic interference (EMI), high total harmonic distortion (THD), high dV/dt stress, and also much less efficiency. Because of these limitations, these inverters are not used for power conversion in high power and high/medium voltage applications. Due to this, a new concept of multilevel inverters (MLI) was developed. The main difference between an MLI and a conventional two-level inverter was the presence of more number of voltage levels in the output waveform. This feature results in reducing the THD as

B. Alam · M. Nusrat · Z. Sarwer (✉) · M. Zaid · A. Sarwar
Electrical Engineering Department, Z.H. College of Engineering and Technology, Aligarh Muslim University, AMU, Aligarh 202001, India
e-mail: zeeshan.sarwer@zhcet.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_61

the waveform gets closer to a sinusoidal waveform. This was achieved by using modulation techniques also known as pulse-width modulation (PWM) techniques. In a PWM technique, the carrier signal is compared with respect to the reference or modulating signal and consequently, the output is obtained. The modulation techniques can be further classified into low-frequency and high-frequency modulation techniques. At high switching frequencies, a 3-level inverter produces significantly less harmonic distortion, low EMI, and reduced rating of dV/dt stress as compared to a 2-level conventional inverter. Moreover, over time new and advanced versions of multilevel inverters that are capable of producing almost sinusoidal output have been developed using the same and new modulation techniques with the help of different topologies of multilevel inverters.

The multilevel inverter has helped in mitigating many problems that were present in a 2-level inverter. The earlier conventional circuits of the multilevel inverter were flying capacitor (FC), neutral-point clamped (NPC), and cascaded H-bridge (CHB). Each of these structures has some merits and demerits over the other. Although, the FC structures don't require any current or voltage sensor which is an advantage but still, the voltage of the FC differs in each cell. Therefore, it requires bulky capacitors that are costly. NPC structures possess the problem of voltage balancing in capacitors. CHB type inverters do not require capacitors and hence the problem of capacitors voltage balancing is eliminated but these require many DC sources for producing more levels which becomes a disadvantage. Thus, researchers try to improve and modify the circuit of a multilevel inverter using different topologies. The aim is to reduce the count of components so that the loss per component may reduce as well as the cost. Since the beginning, advancement is being made using different topologies with better techniques of modulation.

In this paper, a review of the recently developed asymmetrically configured MLI topologies is presented. Each of the topologies has a different number of components which decide the reliability and cost of the inverter. For each of the component configurations, there are different numbers of voltage levels which a particular MLI topology can generate. The number of components in an MLI structure is directly related to the cost of the inverter therefore research work is going on to reduce the components. Researchers have proposed different MLI topologies that aim to reduce the total effective cost of the circuit without compromising the quality of the system. Moreover, the modulation techniques which are used in MLIs will also be discussed in the paper. The paper is organized into five sections: Sect. 1 contains the introduction, Sect. 2 describes the recent asymmetrical MLI topologies which contain the circuit description and the basic relations between their components. Section 3 discusses the various modulation techniques. Section 4 describes the comparative analysis between the various topologies. The comparison is presented in tabular form as well as in graphical form. Lastly, the conclusion is discussed in Sect. 5.

2 Recent MLI Topologies

This section discusses the structure of some of the recently proposed asymmetrical MLI topologies. Figure 1 represents the general stepped output waveform of a multilevel inverter. Continuous research is going in the direction to reduce the total number of components in an MLI structure and hence propose newer topologies. The different components include number of switches (N_{sw}), number of driver circuits (N_{dr}), number of DC sources (N_{DC}), etc. The switches used in MLIs need to block a certain voltage which determines its rating. The sum of blocking voltages of all the switches is termed as Total Standing Voltage (TSV). TSV is an important parameter in the topologies related to multilevel inverters. If the switches have to block lower voltage stress on the switches, it means the TSV of the topology is less which ultimately leads to the use of lower rating switches. The use of lower rating switches will reduce the cost of the inverter which is an advantage. The circuit diagrams are shown and the generalized relations N_{sw}, N_{dr}, N_{DC}, and TSV against the number of levels (N_L) are presented.

Figure 2a shows the new proposed multilevel inverter topology in [1]. It consists of a total of 10 switches, in which six switches (S_1 to S_6) are in the string and the other four switches are connected to the battery (S_7 to S_{10}). Moreover, the proposed structure has 4 DC sources. This topology produces 31 levels in output. The advantages of this topology are lower cost, lower losses, and low THD. The number of levels (N_L) can be related to the number of DC sources (N_{DC}) as shown in Eq. (1).

$$N_L = 2(N_{DC} + 1) \tag{1}$$

The MLI topology proposed in [2] is described in Fig. 2b. In this topology, 15 levels are achieved using 10 switches and 4 DC sources. There are only two magnitudes of DC sources that are required in the circuit topology. As one switch is bidirectional in

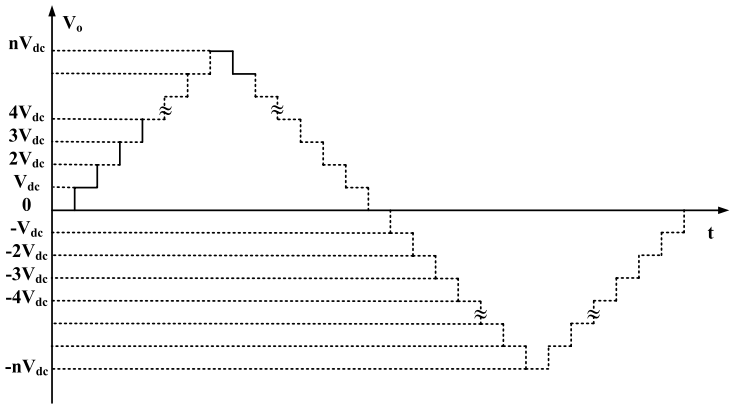


Fig. 1 Multilevel inverter waveform

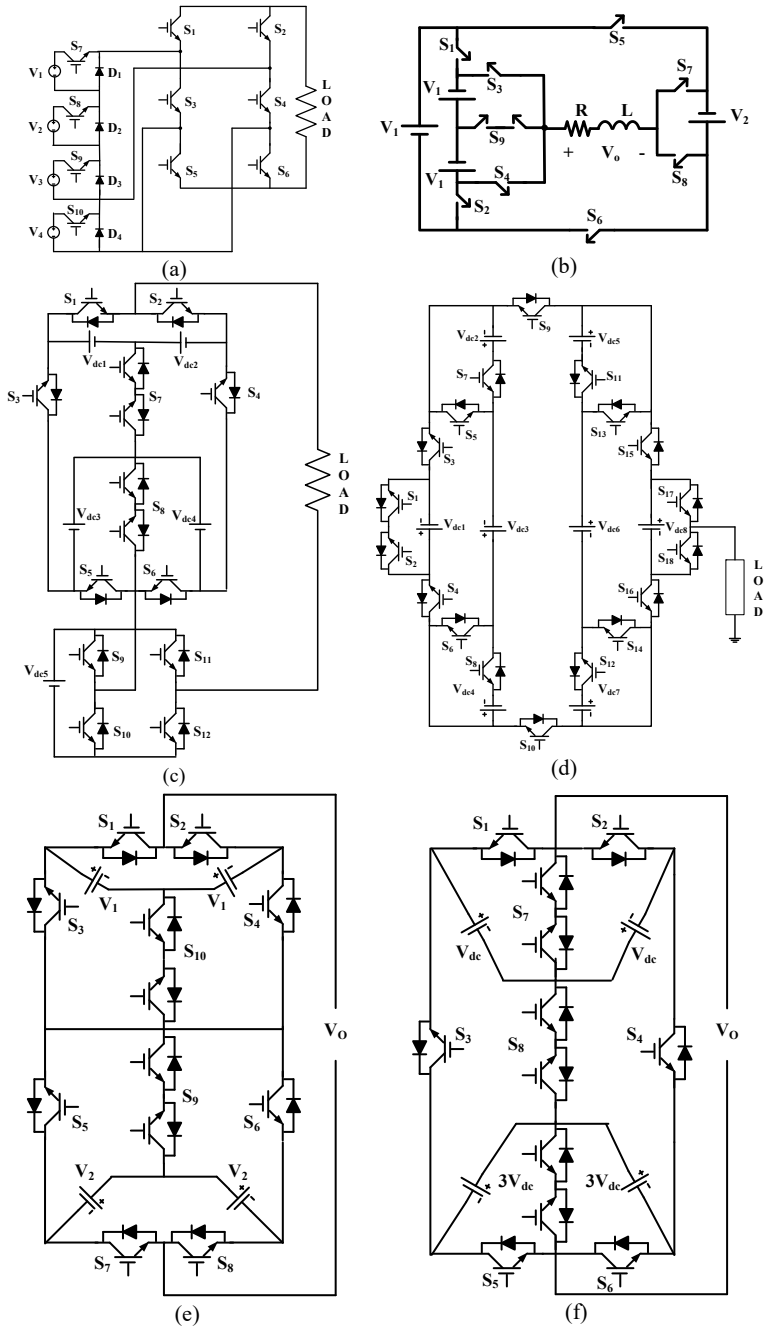


Fig. 2 MLIs proposed in **a** [1], **b** [2], **c** [3], **d** [4], **e** [5], **f** [6]

the proposed structure, 9 gate driver circuits are required. The THD for this MLI is 5.5%. The multilevel inverter proposed in [3] is depicted in Fig. 2c. As can be seen, it consists of 5 DC sources and 12 switches of which 10 are unidirectional switches and the remaining 2 are bidirectional switches. It can generate 27 levels in output level. The generalized relations for N_{DC}, $N_{sw,}$ and N_{dr} are shown in Eqs. (2)–(4), respectively.

$$N_{DC} = \left(\frac{N_L - 1}{4}\right) - 1 \tag{2}$$

$$N_{sw} = \left(\frac{3(N_L - 1)}{6}\right) + 1 \tag{3}$$

$$N_{dr} = \left(\frac{3(N_L - 1)}{6}\right) - 1 \tag{4}$$

A new topology of MLI with a lesser number of components is introduced in [4] which is shown in Fig. 2d. Here, 8 DC sources, 18 unidirectional switches, and 18 diodes are utilized to generate 33 levels of voltages. The different relations of various components are given by Eqs. 5–8.

$$N_{DC} = 8\frac{\ln(N_L)}{\ln(33)} \tag{5}$$

$$N_{sw} = 18\frac{\ln(N_L)}{\ln(33)} \tag{6}$$

$$N_{dr} = 18\frac{\ln(N_L)}{\ln(33)} \tag{7}$$

$$TSV = 5\frac{(N_L - 1)}{2} \tag{8}$$

A sub-module is proposed in [5] to generate 25 levels (12 positives, 12 negatives, zero levels) with 10 power switches and 4 DC links. Two of these DC links have the same magnitude of V_1 and the other two are V_2. Figure 2e shows the basic structure of the proposed topology. Basic relations are governed by Eqs. 9–12.

$$N_{DC} = \frac{N_L - 1}{2} \tag{9}$$

$$N_{sw} = 5\frac{(N_L - 1)}{12} \tag{10}$$

$$N_{dr} = \frac{N_L + 1}{2} \tag{11}$$

$$TSV = 18\frac{(N_L - 1)}{8} \tag{12}$$

The topology discussed in [6] consists of 4 DC sources, 2 bidirectional switches, and 6 unidirectional switches. Each unidirectional switch comprises an anti-parallel diode and IGBT. The structure is optimized to produce desired voltage levels with the least number of switches with low TSV. Figure 2f shows the circuit diagram of this topology and the basic relations are presented in Eqs. (13)–(16).

$$N_{DC} = \frac{N_L - 1}{4} \tag{13}$$

$$N_{sw} = \frac{3}{4}(N_L - 1) \tag{14}$$

$$N_{dr} = \frac{(N_L + 1)}{2} \tag{15}$$

$$TSV = 5\frac{(N_L - 1)}{2} \tag{16}$$

A typical asymmetrical MLI with 4 DC sources including 10 switches and 10 diodes is proposed in [7]. It will produce 13 levels (6 positives, 6 negative, zero levels) with reduced components. Asymmetric operation for high-voltage applications is also possible. The fundamental relations are given in Eqs. (17)–(20).

$$N_{DC} = \frac{N_L - 1}{3} \tag{17}$$

$$N_{sw} = \frac{5}{6}(N_L - 1) \tag{18}$$

$$N_{dr} = \frac{8(N_L - 1)}{6} \tag{19}$$

$$TSV = \frac{10(N_L - 1)}{6} \tag{20}$$

In [8] these are 2 DC voltages sources and 6 power electronic switches in which 2 switches are bidirectional and the remaining 4 are unidirectional switches. This proposed topology generates only 7 levels in their output voltages. In the proposed topology of [9], 21 levels in output voltages are achieved with 12 power electronic switches and 4 DC sources with different magnitude. In the novel proposed topology discussed in [10], there are 8 switches and 3 DC sources voltages with different magnitudes. It produced 15 levels in the output voltage with a reduced number of components. A novel cascade MLI topology [11] for the multilevel inverter to generate different levels of voltage comprises submodules, the basic part of

which generates 17 voltages levels with 14 components in total (8 unidirectional, 2 bidirectional switches, and 4 DC sources).

3 Modulation Techniques

Modulation techniques are the basic building block of the MLIs. In any modulation technique, a modulating or reference signal is compared with a carrier signal which in turn results in generating the switching signals. Figure 3 shows the classification of modulation techniques used in MLIs. These switching signals decide the operation of different switches, i.e., it governs the process of turning ON or OFF of the power electronic switches which help the power converter to make the transition from one state to another state. In MLI, the modulation technique generates a stepped waveform that resembles a sinusoidal signal. In many applications, the inverter output voltage should be controlled to overcome the change in input voltages and to meet the voltage and frequency. The harmonics of the output voltage depend on the modulation technique.

Based on switching frequency the modulation techniques can be classified as follows.

- Fundamental switching frequency techniques,
- High switching frequency techniques.

1. Fundamental switching frequency

In this switching strategy, the switching signals are generated using low or fundamental switching frequencies. This means that the number of ON/OFF cycles of a switch is less. At low switching frequency, the losses in the switches are less as compared to other methods. There can be different fundamental switching frequency

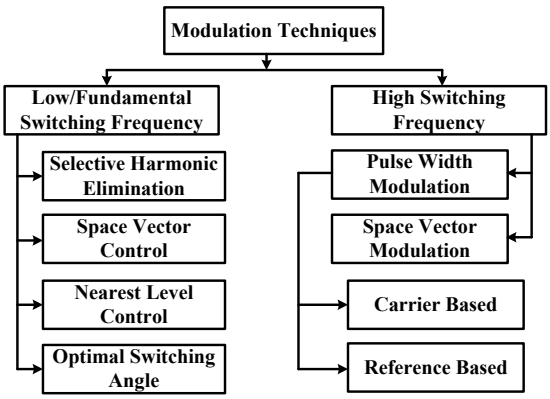


Fig. 3 Classification of modulation techniques

techniques. Some of them are Selective harmonic eliminations, Nearest level control, Optimal switching angle, etc.

A. Selective harmonic Eliminations (SHE)

This technique is also popular and is based on the fundamental switching frequency technique. This method was introduced in [12] to completely remove the defined harmonic contents. The main aim of selective harmonic is to prevent the lower order harmonic in a suitable range by using a switching angle. In this technique, the nonlinear transcendental equations have to be solved to get the values of switching angles. Different artificial intelligence techniques can be employed to solve these equations.

B. NLC (Nearest Level Control)

This modulation technique is basically a time-domain version of the space vector control PWM technique. In this, the output voltage is generated in the form of steps that helps to reduce error considerably. Generating the nearest level of voltage by comparing the output voltage level and the desired reference voltage is aimed in this technique.

C. NVC (Nearest Space Vector Control)

It is a type of space vector modulation technique that operates at low frequency. In this technique, the aim is to get the nearest vector with respect to the reference vector. A drawback of this type of modulation technique is that the low-order harmonics are not eliminated. This technique is more suitable where a higher number of levels has to be generated rather than in MLIs where lower numbers of levels are generated because of distortion problems.

2. High-frequency switching techniques

A PWM technique is a method in which the width of the modulation signal is adjusted with respect to the carrier signal.

A. Space Vector Modulation (SVM)

Space Vector Modulation is a high-frequency switching technique. It is a digital technique in which the PWM voltages are generated by using the reference signal. In this technique, many vector states are required. For a higher number of levels, this technique is not preferred because sector identification and switching sequence selection becomes a tedious task. Common mode voltage can be reduced along with switching losses by properly selecting a modulation vector.

B. Phase-Shift PWM technique

In this modulation technique, the number of carrier pulses is equal to (m−1) for 'm' modulating pulses. The amplitude and frequency of the carrier pulses are identical. In PS-PWM, the carrier pulses are phase-shifted by an angle Δ, where.

$$\Delta = 360^{\circ}/(m-1)$$

When the modulating signal is compared with the carrier signal, gating pulses are obtained.

C. Level Shift PWM technique

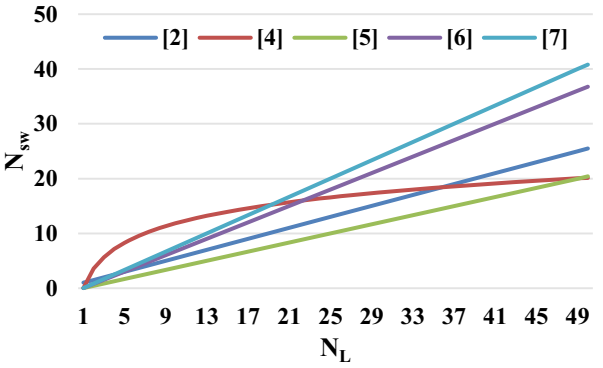
In the Level Shifted modulation technique, the carrier signals are shifted vertically and compared with the modulating signal. Here also, carrier pulses are identical and the total pulses required is (m−1) for 'm' modulating signals. The three schemes for this type of modulation technique are in phase disposition (IPD), phase opposition disposition (POD), and alternative phase opposition disposition (APOD).

4 Comparative Assessment

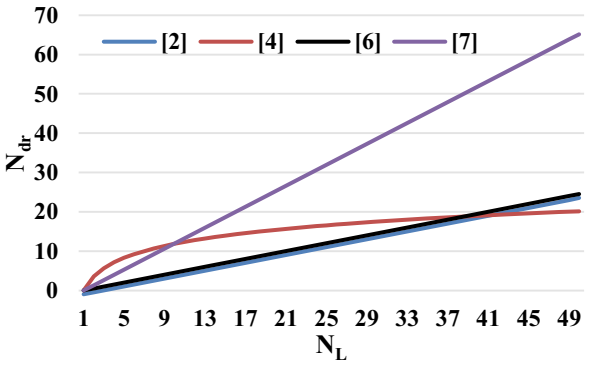
Table 1 compares the different reduced switch count MLI topologies. The quantitative features taken for the comparative analysis are N_{DC}, N_L, N_{sw}, and N_{dr}. By using 4 DC sources, the number of levels generated in [1, 2, 5–7, 9] and [11] are 31, 15, 25, 17, 13, 21, and 11, respectively. However, [1, 2, 5, 7, 11] uses 10 switches whereas [6] and [9] require 12 switches. 8 DC sources with 18 switches are used in [4] to generate 33 levels. To add some more information, the comparison is also shown in graphical form. Figure 4 shows the different plots used in the comparison. The variation of N_{sw}, $N_{dr,}$ and N_{DC} against NL is depicted in Fig. 4a, b, and c, respectively.

Table 1 Comparison of different asymmetrical MLIs

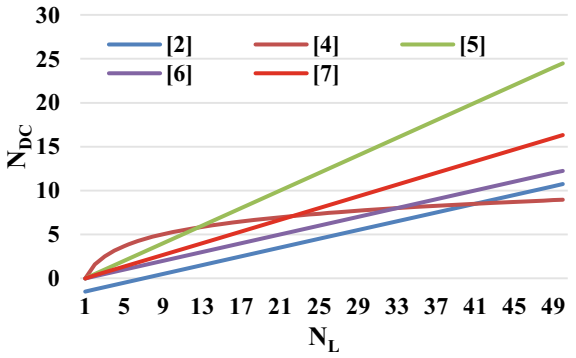
MLI	N_L	N_{DC}	N_{sw}	N_{dr}
[1]	31	4	10	10
[2]	15	4	10	9
[3]	27	5	14	12
[4]	33	8	18	18
[5]	25	4	10	12
[6]	17	4	12	9
[7]	13	4	10	16
[8]	7	2	6	8
[9]	21	4	12	12
[10]	15	3	8	8
[11]	17	4	10	10



(a)



(b)



(c)

Fig. 4 Variation of **a** N_{sw} versus N_L **b** N_{dr} versus N_L **c** N_{DC} versus N_L

5 Conclusion

This paper presents a comparative study of various asymmetrical multilevel inverter topologies. The comparative analysis is presented based on the total power electronic components. Reduction in the number of components reduces the overall cost of the structure and increases the reliability of the inverter. Conventional MLI topologies are now being replaced with other topologies requiring a smaller number of components. Further, new switched capacitor boost topologies have also become popular in which DC sources are replaced by capacitors. The focus of researchers is now to develop fault-tolerant MLI topologies with reduced switch count, low-cost high reliability. There are various modulation techniques that are used in MLIs. Further, the other work is to develop new modulation techniques to effectively control the working of MLI and to reduce the losses and THD in output voltage. Overall, MLI inverter remains an exciting area to work due to increased applications in renewable energy systems.

References

1. Raval KY, Ruvavara VJ (2018) Novel multilevel inverter design with reduced device count. In Proceedings of the 2018 International Conference on current trends towards converging technologies, ICCTCT
2. Sarwer Z, Siddique MD, Iqbal A, Sarwar A, Mekhilef S (2020) An improved asymmetrical multilevel inverter topology with reduced semiconductor device count. Int Trans Electr Energ Syst e12587
3. Prabaharan N, Palanisamy K (2016) A new hybrid asymmetric multilevel inverter with reduced number of switches. IEEE International Conference on Power Electronics, Drives and Energy Systems, PEDES
4. Vijeh M et al (2018) A new asymmetrical cascaded multilevel inverter with reduced number of components. In: IECON 44th Annual Conference of the IEEE industrial electronics society 1, pp 4429–4433
5. Hosseinzadeh MA, Sarbanzadeh M, Yazdani L, Sarbanzadeh E, Rivera M (2018) New sub-module inverter for cascaded multilevel inverter with reduced number of switch counts. In: Proceedings—IEEE Southern Power Electronics Conference, SPEC Janua, pp 1–6
6. Samadaei E, Sheikholeslami A, Gholamian SA, Adabi J (2018) A square T-type (ST-type) module for asymmetrical multilevel inverters. IEEE Trans Power Electron 33(2):987–996
7. Samadaei E, Gholamian SA, Sheikholeslami A, Adabi J (2016) An Envelope Type (E-Type) Module: Asymmetric Multilevel Inverters with Reduced Components. IEEE Trans Industr Electron 63(11):7148–7156
8. Kumar J (2016) A new 7-level asymmetrical multilevel inverter with reduced number of sources and switching components. In: India International Conference on Power Electronics, IICPE, pp 6–10
9. Agrawal N, Bansal P (2018) A new 21-level asymmetrical multilevel inverter topology with different PWM techniques. In: 2017 Recent Developments in Control, Automation and Power Engineering, RDCAPE 2017, 3, pp 224–229
10. Boora K, Kumar J (2017) A new general topology for asymmetrical multilevel inverter with reduced number of switching componenets. In: Recent Developments in Control, Automation and Power Engineering, RDCAPE, pp 66–71

11. Alishah RS, Hosseini SH, Babaei E, Sabahi M (2016) A new general multilevel converter topology based on cascaded connection of submultilevel units with reduced switching components, DC sources, and blocked voltage by switches. IEEE Trans Industr Electron 63(11):7157–7164
12. Patel HS, Hoft RG (1973) Generalized techniques of harmonic elimination and voltage control in thyristor inverters: Part I–harmonic elimination. IEEE Trans Ind Appl 3:310–317

Health Monitoring System Using IoT

Prateek Sharma, Prajjwal Soam, and Neeraj Joshi

Abstract In this twenty-first century of technology, humans have just become working machines who just work like robots, 24 h on a daily basis to get two-time meals. In this race, they just forgot to focus on a very important part, i.e., health. As initially there are numerous systems developed for monitoring in real-time tracking of the health of an individual. This type of monitoring systems is helping the government or an individual to focus upon the best gift of God, i.e., our body. This system is of higher accuracy and has precise criteria for judging the mentioned aspects. This health monitoring system has a great on-field outcome as a judge or as a moderator to directly ping measured details of the sensors attached to the user as well as to the admin with which any upcoming danger of the body can be avoided. In our system we are using a Heart Beat sensor, DHT-11 sensor, and Blood Pressure sensor along with Raspberry Pi, which takes sensor data and send it to the local server, i.e., Google Firebase cloud services (Google Firebase. https://en.wikipedia.org/wiki/Firebase. Accessed 02 Mar 2020) with real-time monitoring of a patient or working individual can be accessed and shared with Doctor as well as with the patient with the help of an Android application, which is developed using MIT App inventor. With future reference, the data generated from this system can help in the training of the machine, which can predict any upcoming diseases or emergency conditions for an individual.

Keywords IoT · Body sensors · Firebase · Health systems

P. Sharma · P. Soam (✉) · N. Joshi
Department of Electronics and Communication Engineering, Meerut Institute of Engineering and Technology, Meerut, U.P., India
e-mail: prajjwal.soam.ec.2016@miet.ac.in

P. Sharma
e-mail: prateek.sharma.ec.2016@miet.ac.in

N. Joshi
e-mail: neeraj.joshi@miet.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_62

1 Introduction

Nowadays, the population is increasing with a comparable speed of light so does their needs and health issues. So, to overcome this effect, the only thing possible is with the invention and use of advanced technology. As we see, there are lots of systems that are already available in the market. But they are limited to their own work application area. Our system is a step ahead in this fight. It is a single module that is using a well-known technology, i.e., IoT (Internet of things). IoT has become a very advanced field that has hundreds of techniques, which we can use to solve unsolved problems of our daily life. This module covers maximum issues related to health on-daily basis or we can say at each instant of time. It contains different sensors like—**DHT11, BMP180, Heart Beat Module, ECG, etc., a Raspberry Pi 3, Google Firebase, Android Mobile App using MIT App Inventor 2**. All the data is collected in numeric form and stored in Microprocessor, where it is converted into JSON variable, because only JSON form is required to transfer the data on Google Firebase whereas Google Firebase is an online local cloud, which provides us different functions like Real-Time Database, Firestore cloud, Storage, Hosting, Android app access, etc. An account is required to access all these facilities and this account is secured by Google Security. Our data is stored in a real-time database and it is accessed with the use of an Android App made by App Inventor tool provided by MIT previously Google. This app is provided to doctors, patients, and admin. Only admin has authority to Write and Read the data while doctor and patient can only read it.

2 Literature Survey

In [3], the authors highlighted this monitoring system which comprises heart pulse sensor, body temperature sensor, and Galvanic sensor, here both Arduino and Raspberry Pi are used which are sending the data of sensors to the storage cloud with the help of App developed using Android studio. This system is producing the reading in the form of graphs to produce a common consensus. As per the calibration of the sensors and the observed patterns, the use of this system seems successful, with some changes for future references.

In [7], the team authored a Patient monitoring system with the help of Raspberry Pi and Arduino along with sensors like ECG, Temperature sensor, Push switch, Camera module, LCD Display, etc., to extract different biomedical signals connected from different body parts of an individual. And the main feature of this paper is the auto-generated Video call for the emergency contacts mentioned initially. This is the real-time monitoring system that is helpful in critical monitoring of the patients. In [2], the author proposed an automated and smart system based on the Internet of things which senses the features like vital sign, pressure level, piezoelectric sensor, body temperature, etc., along with SIM900D (GSM Module) automatically. Here the

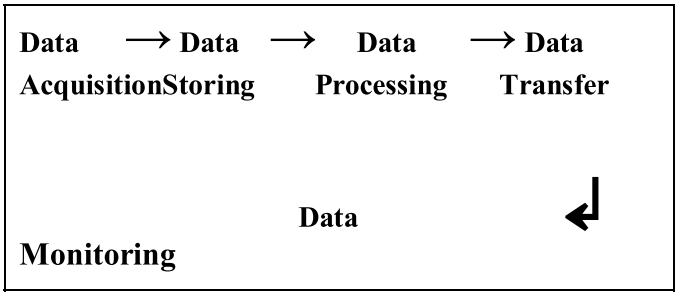
database helps the observer to have the whole record of an individual. But the main setback of this system is the unresponsiveness of their heart rate sensor which is not updating regularly. And the future of this system can be extended by adding more sensors like EEG, etc.

In [9], the authors focused on developing this system which has the capability to take multiple physical and internal readings through the sensors which can be worn on the body. Here we have used Arduino Fio Node and Base, which is working as the transmitter and receiver along with multiple sensors LM35 (Temperature sensor), Pulse rate, etc., along with LabVIEW software to build a Computer Graphics Interface (CGI). This system provides the opportunity to take real-time access.

3 Proposed Architecture

(a) Objective and System Approach

Our objective is to monitor the patient's health regularly with our sensors and to send the data to different users—Doctor, Patient, Admin for predetermining the health issues and also for better support at the time of any emergency. The combination of sensors produce the data which is collected and it is processed in the Raspberry Pi(storing and changing to specific form) and after that, it is transferred to a private cloud (real-time database), where it is available for access to a different authenticated user as shown in fig A. B



All the data flow is from Left to Right. The Data Transfer is to Mobile App via the cloud.

(b) Devices Used

1. Raspberry Pi 3

Raspberry Pi [5] It is a microprocessor of credit-card sized made by Raspberry Pi Foundations in 2012. It is a tiny single-board computer and Linux based is the operating system on which it worked and it is armed with Broadcom BCM2837B0 chip, 1.2 GHz 64-bit Quad-core processor with 1 Gb of RAM. It has an inbuilt Wi-Fi and Bluetooth module with 40 pins (in which 26 are GPIO's). This platform can be programmed with Python and scratch in which mouse, keyboard, monitor, etc., can

be connected. It is a System on a Chip (SoC) model; it can be used for different operations (Fig. 1).

2. Sensors

(a) DHT 11: Introduction to DHT-11 [4] It is a digital temperature and humidity sensor with a 3-pin circuit. It is integrated with a high performance 8-bit microcontroller with long-term stability. It works on the principle of NTC (Negative Temp Coefficients) for measuring the temperature. Each humidity sensor has accurate calibration of the humidity calibration chamber (Fig. 2).

(b) BMP180: Pressure Sensor-BMP 180 [8] It is a 5-pin sensor which works with the I2C('wire') protocol. The I2C uses only two pins to communicate with a lot of different devices. One carries a clock signal, and the other carries the data· It is the

Fig. 1 Raspberry Pi

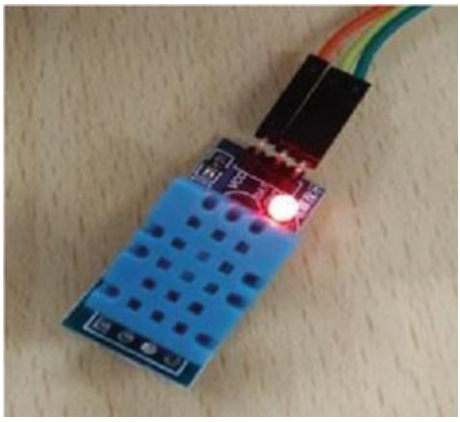


Fig. 2 Ultrasonic sensor

precision sensor developed by Bosch for measuring barometric pressure. It regulates a 3.3 V power supply (Fig. 3).

(c) HEARTBEAT SENSOR: Pulse sensor [6] It is a 3-pin sensor which is used to measure the Heartbeat or Pulse rate of a person on the principle of photoplethysmography (Fig. 4). This sensor comprises VCC (5 V) and ground as well as with Analog Pin. This is a highly precise low-cost heartbeat analog sensor which can be used with platforms like Arduino, Raspberry Pi, etc.

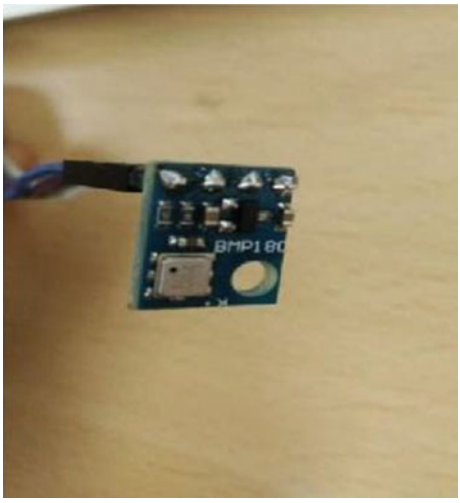


Fig. 3 BMP180 Sensor

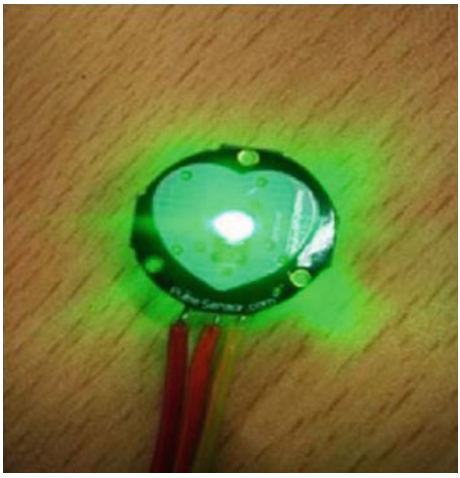


Fig. 4 Heartbeat sensor

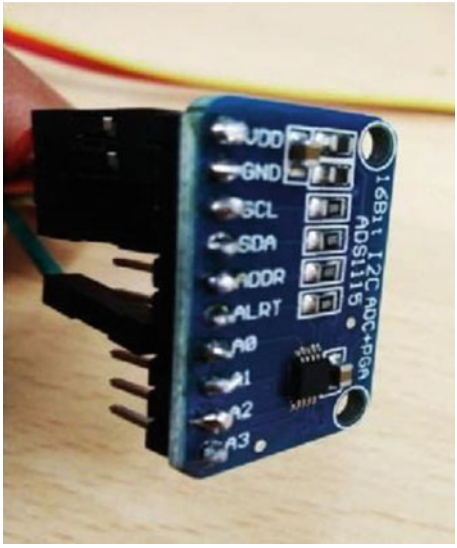


Fig. 5 ADS1115 A/D converter

3. ADS1115: ADS1115 16-Bit ADC–4 Channel with Programmable Gain Amplifier [1] It is a 16-bit 4-channel I2C compatible Analog to Digital converter. It is a low-power and high-precision A/D converter, which can perform conversion at data rates of up to 860 samples per second. The ADS1115 operates either in continuous mode or in single-shot mode the devices are automatically power down after one conversion in single-shot mode. Therefore, power consumption is very low as compared to A/D converters (Fig. 5).

4 Methodology

The system is divided into two parts—Hardware and Software.

The Hardware part contains Sensors and Raspberry Pi, while the Software part contains Cloud processing and Android mobile app. Here, Raspberry Pi is the brain or the controlling unit while sensors are the basic units that provide different parameters to the processing unit (Fig. 6).

DHT 11 is a sensor that works on different principles for measuring temperature and humidity. For measurement of temp, it works on the principle of Negative Temp Coefficients (NTC). A Negative Temperature Coefficient thermistor is a temperature sensing sensor made of semiconductor material that has a mixture of several metal oxides. The materials it possesses are the charge carriers that allow current to flow through the thermistor showing a large change with respect to a small change in temp. It measuresRelative Humidity for humidity. The formula for calculating is as follows.

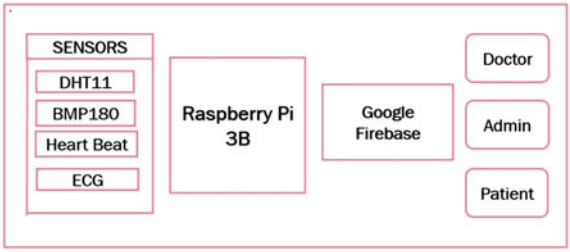


Fig. 6 Proposed system architecture

Relative humidity is

$$\mathrm{Rh} = (\rho_w/\rho_s)^*100$$

where

Rh - Relative Humidity.

ρ_w - Density of water vapor.

ρ_s - Density of water vapor at saturation.

Higher Rh decreases the resistance between the electrodes, while lower Rh increases the resistance between the electrodes. This change is used to calculate the humidity. [8] BMP180 is a barometric sensor used for the calculation of different parameters like pressure, altitude, vertical velocity. Barometric pressure is pressure applied by the weight of air down on earth. The BMP180 is a piezoresistive sensor that detects pressure and Temperature both. Piezoresistive sensors are made up of semiconducting material that changes resistance when atmospheric pressure is applied.

Pulse sensor [6] Heartbeat sensor works on the principle of plethysmography. As we know Heartbeat rate is defined as the ratio of time between two successive heartbeats. In every heartbeat, our blood is circulated in our body then it gets squeezed into capillary tissues which resulted in an increase and decrease in the size of capillaries. This little change in the size of capillary can be checked when light is passed through it. Based on that principle, our sensor works. In the process, sensor is placed in front of the human finger and then due to regular decrease and increase in the size of capillaries light intake can be checked. And by the process heartbeat can be recorded when this sensor is interfaced with Raspberry Pi along with an A/D converter, i.e., ADS1115 because Raspberry Pi works only on the digital input.

The Data from DHT11 is in digital form, so it is directly read by Raspberry Pi and then it is stored in Memory. This data provides us the value of temp and humidity of the Body. The unit of temperature is degrees Celsius but it is converted into Fahrenheit by programming the logic. The humidity is given in percentage. Now we have 2 parameters now for the next one, we have used the BMP180 sensor, which works on I2C protocol or 2-wire protocol in which one line is there as a serial clock and the other line is there to read/write serial data to the slaves or modules attached

to it. It provides the pressure value in mbar and temperature in degree Celsius. Due to I2C protocol for data communication, speed is high as compared to DHT11. The accuracy is very high as compared to temp values given by both sensors. For the Heartbeat part, we have used the Heartbeat sensor which provides analog readings on connecting to the power supply. This sensor is accurate and very small so it has a lot of application areas. Its value is directly fed to ADS1115. ADS1115 module is a 4-channel Analog to Digital convertor, it contains a convertor chip over it. It is capable of converting 4 analog values simultaneously with very high accuracy and speed. It takes continuous values from the heartbeat sensor and provides digital values to the Raspberry Pi. We have also used an ECG sensor to take different readings from the heart. The study of electrical signals of the heart is done by ECG hence it provides the most significant and accurate results with any small change in heartbeats. These different values are collected by Raspberry Pi and stored in the memory.

Now in the data processing part, all the data is collected and processed by changing its form to JSON variables by programming in Python in Raspberry Pi. JSON variables are forms of string in which all the data is modified to string form. Only JSON form is allowed to transfer to Google Firebase. Here, Google Firebase acts as a cloud that stores the data and provides it to the authenticated or authorized users. All the data is stored in a real-time database, it is easy, safe, and free of cost up to some limits. When the data is sent from Raspberry pi to Firebase, a new random value node is generated, which contains the value of different sensors inside it. This is further provided to 3 different users via Android Mobile App. One user is our Doctor, second is client/ Patient, and third is Admin. All the access is given through Gmail ID; hence it is secured with Google Security and there is no chance of data leaking or any cyber hacking. The Android mobile app is generated with App Inventor 2 and it contains a virtual desktop with all the health information of any patient as an output from different sensors.

5 Result

The data which is received is pure and accurate that is available for the doctor all the time. Various processes go parallel on a single system and data is collected on one end and transferred on the other end. The initial state with no reading has particular functionality on the start with a background connection with a cloud that refreshes every two seconds (Figs. 7 and 8).

The result is very helpful in creating a parallel database of different patients which is in the real-time state of data-transferring. Doctors have the access to check the database or data of any patient with just a click. After getting the details, a report will be made by the doctor itself, with the medication of any seriousness, and it will be transferred to the patient will the same portal. This process will continue as the reports get back to normal. After the treatment, the system will store the results and will remove that patient from mainframe so to create the space for the next user (Figs. 9 and 10).

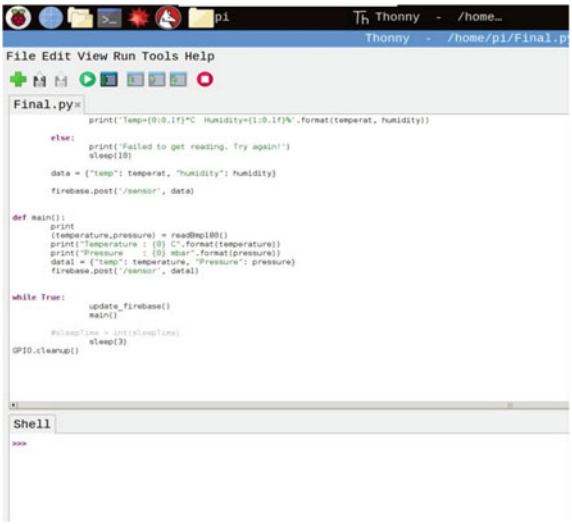


Fig. 7 Python code on Raspberry Pie

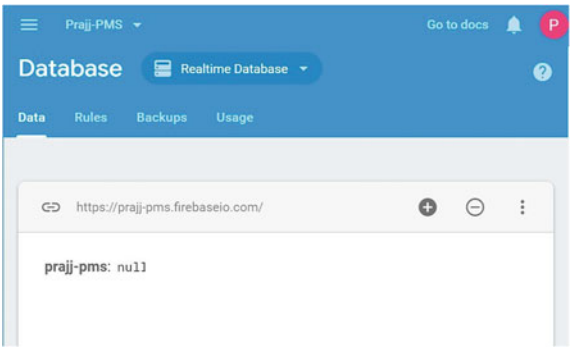


Fig. 8 Reading on firebase

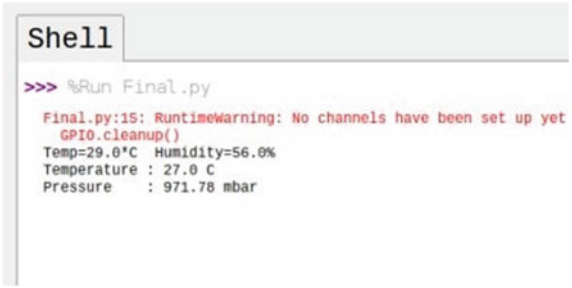


Fig. 9 Readings on firebase shell

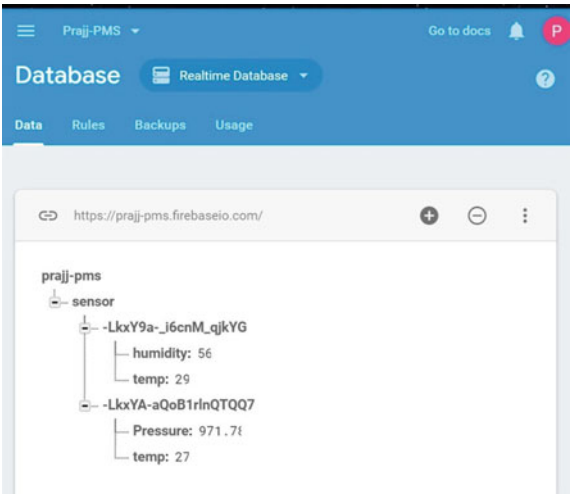


Fig. 10 Dual readings on firebase shell

The following Fig. 11 shows the multiple reading from sensors on Shell and the Firebase. As we can see on Firebase, every time a false function is created which is having the value from sensors. This false function is unique for each time the reading appears on the Shell.

6 Conclusion

IoT has opened a number of methods to solve the problems in the field of health care. The data generated by sensors will provide a deep understanding of patient's health conditions at a regular interval of time. It will help the doctor with proper medications and will lead them away from any future emergencies that may cause due to improper treatment and also, this system will provide all the data of any patient about its previous diseases. This information will be very useful for the present treatment of any client/patient. This solution will change the way of treatment by neglecting the distance parameter, which sometimes plays a crucial role in the living or death of any patient. It will also reduce the cost of treatment and increases the speed so the doctor may visit more numbers of patients throughout the day.

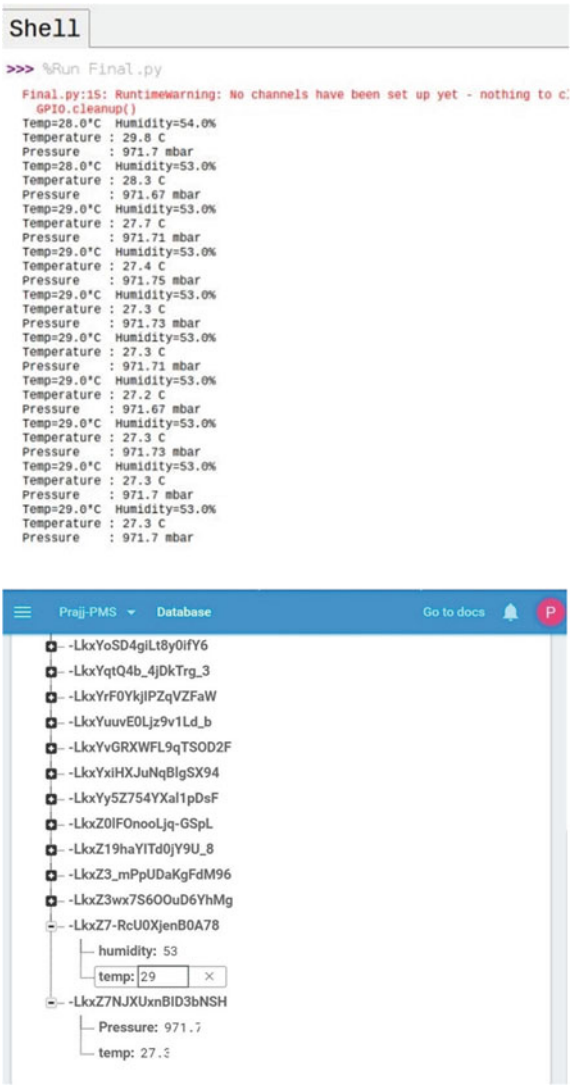


Fig. 11 **a** Readings on shell **b** Same readings on firebase

References

1. ADS1115 16-Bit ADC-4 Channel with programmable gain amplifier (2020) "ADS1115 16-Bit ADC-4 Channel with Programmable Gain Amplifier" https://www.adafruit.com/product/1085. Accessed 2 Mar 2020
2. Ahmed ZU, Mortuza MG, Uddin MJ, Kabir MH, Mahiuddin M, Hoque MJ (2018) Internet of things based patient health monitoring system using wearable biomedical device. In: 2018 International Conference on Innovation in Engineering and Technology (ICIET), Dhaka, Bangladesh, pp 1–5

3. Hamim M, Paul S, Hoque SI, Rahman MN, Baqee I (2019) IoT based remote health monitoring system for patients and elderly people. In: 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, pp 533–538
4. Introduction to DHT-11 (2020) "Introduction to DHT-11" https://www.theengineeringprojects.com/2019/03/introduction-to-dht11.html. Accessed 2 Mar 2020
5. Raspberry Pi (2020) "Raspberry Pi" https://en.wikipedia.org/wiki/Raspberry_Pi Accessed on 02nd March 2020
6. Pulse sensor (2020) "Pulse sensor" https://components101.com/sensors/pulse-sensor. Accessed 2 Mar 2020
7. Rahman A, Rahman T, Ghani NH, Hossain S, Uddin J (2019) IoT based patient monitoring system using ECG sensor. In: 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, pp 378–382
8. Pressure Sensor-BMP 180 (2020) "Pressure Sensor- BMP 180" https://wiki.eprolabs.com/index.php?title=Pressure_Sensor-BMP_180. Accessed 2 Mar 2020
9. Vippalapalli V, Ananthula S (2016) Internet of things (IoT) based smart health care system. In: 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, pp 1229–1233

A Compact Dual Rectangular Slot Monopole Antenna for WLAN/WiMAX Applications

Chandan, R. K. Ratnesh, and Amit Kumar

Abstract A compact dual rectangular slot monopole antenna with microstrip fed is discussed in this article. It consists of dual rectangular slots with the ground, where it generates a dual band for WLAN and WiMAX applications. The low profile with an overall size of $20 \times 20 \times 0.8$ mm^3 achieves good impedance matching. The simulated the result is presented in this article. The successful implementation of impedance bandwidth is 23.10% from 2.3 to 2.9 GHz at 2.6 GHz and impedance bandwidth in higher band is 13.20% from 4.9 to 5.6 GHz at 5.3 GHz.

Keywords Dual rectangular slot · Ground plane · Microstrip line feed

1 Introduction

In recent time, microstrip patch antennas have more popular in the 1970s due to their low profile, easy fabrication, lightweight, and their ability to integrate other devices. But some disadvantages of conventional microstrip patch antenna like lower gain, narrow bandwidth, low profile, and lower efficiency. Multiband antenna better performance because this antenna is used for multiple applications. It increases the technology, which means antenna size is reduced day by day. In this article, dual band with a compact size of the antenna and good impedance matching is achieved. This article is generating the impedance bandwidth with a better gain and efficiency is more than 90%.

In [10], circularly polarized hexagonal ring microstrip patch antenna with asymmetrical feed and DGS was used, in this reference antenna size is larger. And similar to dual-polarized, dual band patch antenna loaded with modified mushroom unit cell [11] is used, and in this article antenna dimension has more area occupied than

Chandan (✉)
Department of Electronics and Communication Engineering, Institute of Engineering & Technology, Dr. RML Avadh University Ayodhya, Ayodhya, U.P, India

R. K. Ratnesh · A. Kumar
Department of Electronics and Communication Engineering, Meerut Institute of Engineering and Technology, Meerut, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_63

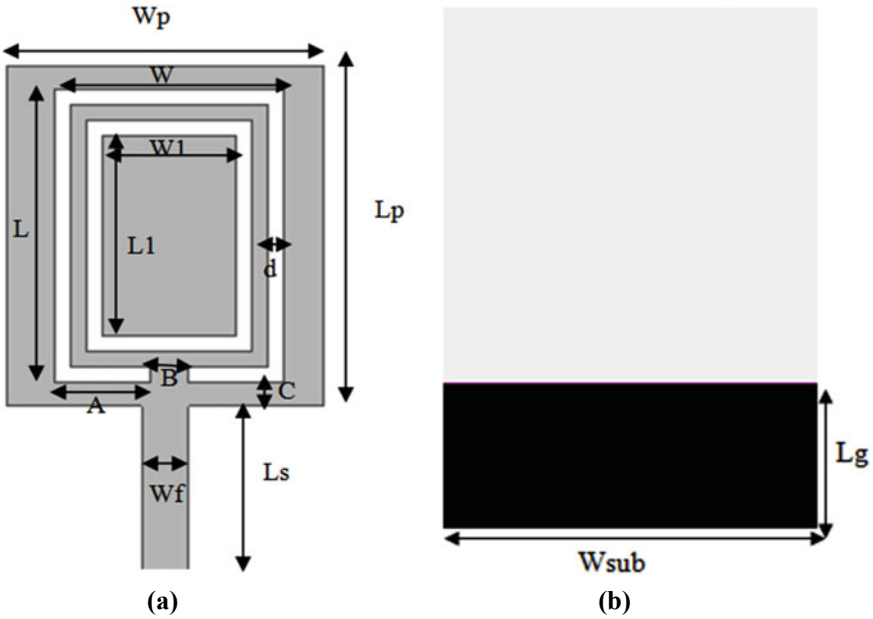


Fig. 1 Geometry of the antenna, **a** top view and **b** bottom view

the proposed antenna. A compact triple band monopole printed monopole antenna for WLAN/WiMAX application, in this paper antenna size is more than proposed antenna [9] and more papers have been studied [1–6, 12, 13]. The proposed antenna is better in performance than the existing antennas.

2 Proposed Antenna Design

The dual band rectangular slot antenna is shown in Fig. 1 and the antenna is designed on FR4 substrate with dielectric constant and height of 4.4 and 0.8 mm, respectively.

The antenna is designed by a rectangular patch with two rectangular slot cut and partial ground plane as shown in Fig. 1. Two bands generate after this procedure follows, single rectangular cut in the patch, lower band achieved but higher resonance not achieved. After that one other rectangular slot cut in the patch to generate a higher band with good impedance bandwidth. Further optimized geometry parameters are given in Table 1.

3 Configuration of the Proposed Antenna

Table 1 Dimensions of the antenna

Parameters	Wp	Lp	L	W	L_1	W_1
Values (mm)	20	22	19	13.5	15	8.5
Parameters	A	d	C	Wf	Ls	Lg
Values (mm)	6	0.5	1.5	2.5	11	10

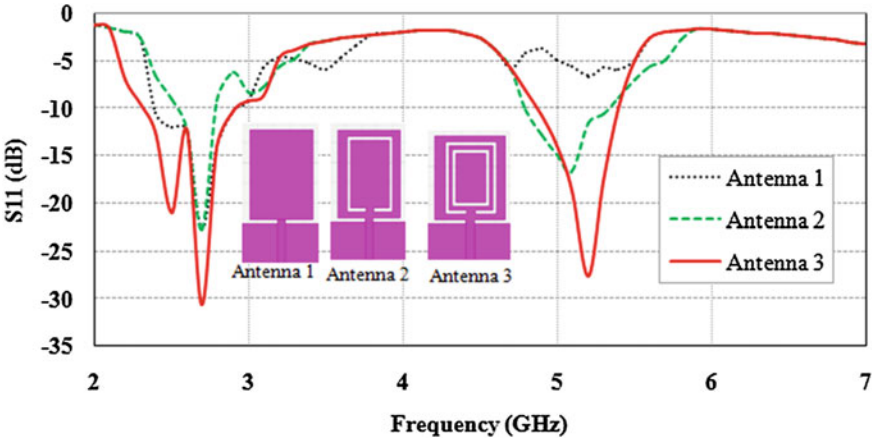


Fig. 2 Simulated return loss S11 of configuration antennas

In the configuration of the different antennas processes, rectangular monopole antenna implements (Fig. 2) which resonates at 2.4 GHz. Antenna 1 is the simple rectangular monopole antenna, this antenna have generated only a single band, has not achieved desired application. Further improved the proposed antenna in antenna 2, this antenna follows step 1 with one rectangular slot cut in the patch, which is called the antenna 2. Antenna 2 generates dual bands, but not achieved desired application and further in step 2 is called antenna 3. Antennas 3 are generated dual band, this proposed antenna is suitable for desired application such as WLAN/ WiMAX application. The successful implementation of impedance bandwidth is 23.10% from 2.3 to 2.9 GHz at 2.6 GHz and impedance bandwidth is a higher band is 13.20% from 4.9 to 5.6 GHz at 5.3 GHz.

4 Simulated Results

The return loss of the antenna is presented in Fig. 3.

The antenna covered $S_{11} \leq 10$ return loss and each band satisfies of application for WLAN/WiMAX. The impedance bandwidth in lower band is 23.10% from 2.3

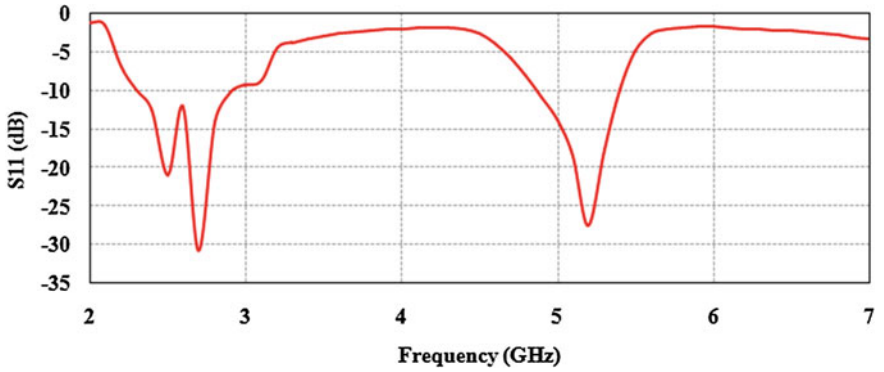


Fig. 3 Return loss S11 of the proposed antenna

to 2.9 GHz at 2.6 GHz and impedance bandwidth in the higher band is 13.20% from 4.9 to 5.6 GHz at 5.3 GHz.

In this article, a dual band with dual rectangular slot cut, rectangular monopole antenna for WLAN/WiMAX application has been implemented. The proposed antenna has been achieved to gain and efficiency of Figs. 4 and 5. Fig. 4 indicates the gain of the antenna which is 2.8 dBi for lower band at 2.3 GHz resonance and 5.2 dBi for higher band at 5.6 GHz resonance. Fig. 5 is simulated efficiency of the proposed antenna are 89% for a lower band at resonance 2.3 GHz and 94% efficiency for a higher band at 5.6 GHz resonance frequency. The simulated radiation pattern has at 2.3 and 5.6 GHz, it's seen Fig. 6. In this Fig. 6a, E plane and H-plane are shown in Fig. 6b. In Fig. 7, it is shown the surface current distribution, this indicates the behavior of the antenna. It can be observed from Fig. 7a patch and ground current, highly red in color which indicates more impedance matching in this area and similar to Fig. 7b.

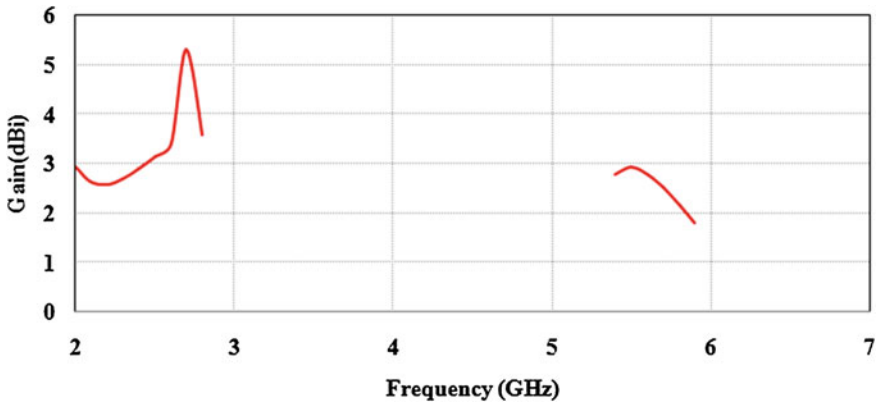


Fig. 4 Gain of the proposed antenna

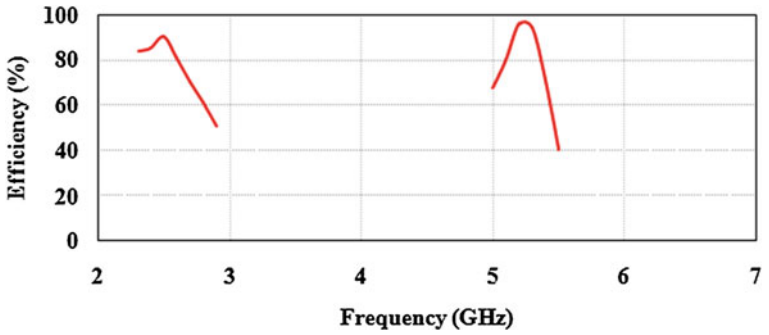


Fig. 5 Efficiency of the proposed antenna

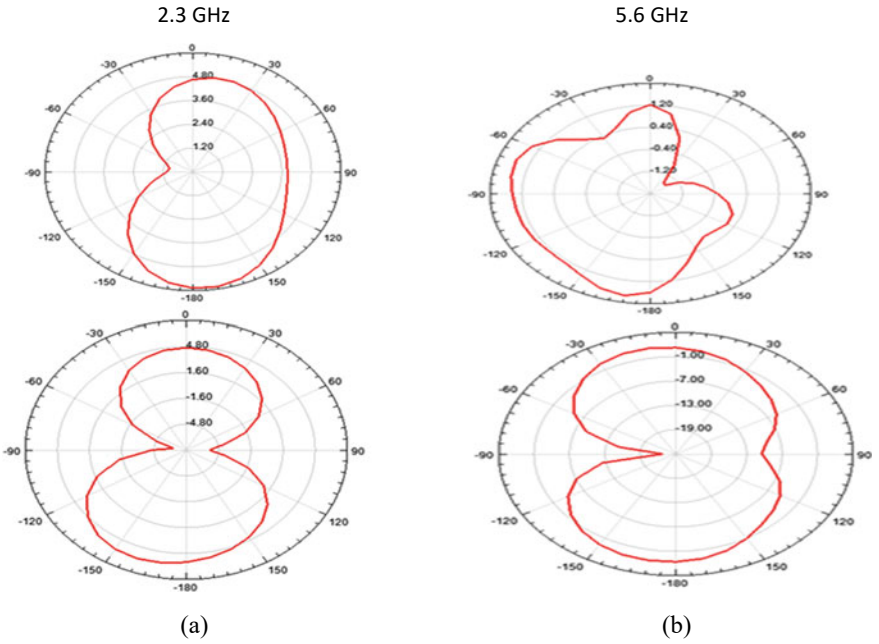
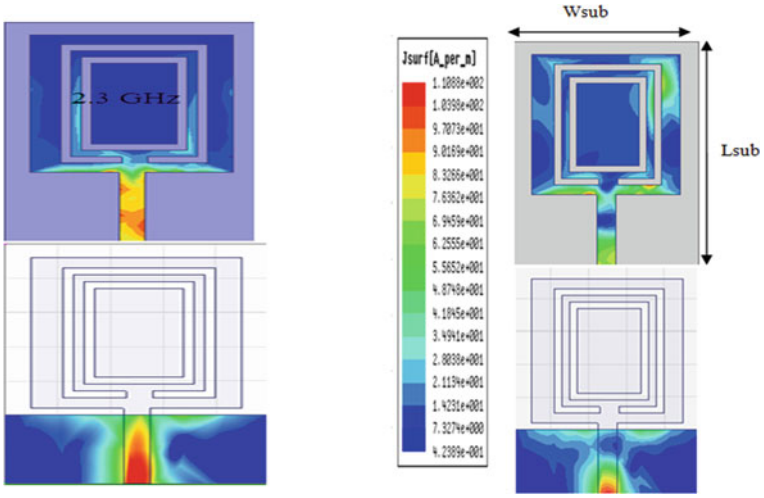


Fig. 6 **a** and **b** Radiation pattern of the proposed antenna

5 Conclusion

In this article, a dual band rectangular slot cut monopole antenna for WLAN/WiMAX application with ground plane also microstrip fed is done. The proposed antenna has a compact size than the reference antennas, it's seen in Table 2. The successful implementation of impedance bandwidth is 23.10% from 2.3 to 2.9 GHz at 2.6 GHz



(a) (b)

Fig. 7 **a** and **b** Surface current distribution of the antenna

Table 2 Comparison of the proposed antenna with the reference antenna

References	Compact size	Impedance bandwidth %	Gain	No. of band
[10]	50 × 50 × 1.6	13.4	8.14	3
[11]	60 × 60 × 1.6	3.5 and 1.97	2.44	3
[9]	34 × 18 × 1.6	3.00	0.28	
[13]	71 × 52 × 1	28.0, 26.0 14.0 and 12.5	2.5	4
Proposed antenna	20 × 20 × 0.8	23.10 and 13.20	2.8 dBi and 5.2 dBi	2

and impedance bandwidth in higher band is 13.20% from 4.9 to 5.6 GHz at 5.3 GHz. And the gain of dual band are 2.8 and 5.3 dBi.

References

1. Chandan, Rai BS (2014) Bandwidth enhancement of Wang shape microstrip patch antenna for wireless system. In: IEEE Fourth International Conference on communication systems and network technologies (NITTR BHOPAL). pp 11–15
2. Chandan, Rai BS (2014) Dual-band Wang shaped microstrip patch antenna for GPS and Bluetooth application. In: IEEE Sixth International Conference on computational intelligence and communication networks (Udaipur). pp 69–73

3. Chandan, Rai BS (2016) Dual-band monopole patch antenna using microstrip fed for WiMAX and WLAN applications. In: Springer Proceedings Information Systems Design and Intelligent Applications, vol 2, pp 533–539
4. Chandan (2020) Truncated ground plane multiband monopole antenna for WLAN and WiMAX applications. IETE J Res 1–6
5. Chandan, Bharti GD, Bharti PK, Rai BS (2018) Miniaturized Pi (π)—Slit monopole antenna for 2.4/5.2.8 applications. In: AIP Conference Proceedings, American Institute of Physics. pp 200351–200356
6. Chandan, GD Bharti, Srivastava T, Rai BS (2018) Dual band monopole antenna for WLAN 2.4/5.2/5.8 with truncated ground. In:AIP Conference Proceedings, American Institute of Physics. pp 200361–200366
7. Gangwar A, Alam MS (2017) A high FoM monopole antenna with asymmetrical L-slot for WiMAX and WLAN application. Microwave Opt Technol Lett 60(1):196–202
8. Garg R, Bhartia P, Bahl IJ (2001) A microstrip antenna design handbook, Boston, MA, Artech House
9. Li L, Zhang X, Yin X, Zhou L (2016) A compact triple band printed monopole antenna for WLAN/WiMAX application. IEEE Antennas Wirel Propag Lett 16(1):1853–1855
10. Maurya RK, Kanaujia BK et al (2019) Circularly polarized hexagonal ring microstrip patch antenna with asymmetrical feed and DGS. Microwave Opt Technol Lett 62(4):1–7
11. Saurabh K, Sarkar D, Srivastavva KV (2014) Dual-polarized dual-band patch antenna loaded with modified mushroom unit cell. IEEE Antennas Wirel Propag Lett 13(1):1357–1360
12. Srivastava T, Rai BS (2017) L-slotted microstrip fed monopole antenna for triple band WLAN and WiMAX applications. In: Springer in Proceedings Theory and Applications, vol 516, pp 351–359
13. Srivastava T, Rai BS (2016) Multiband monopole U-slot patch antenna with truncated ground plane. Microwave Opt Technol Lett 58(8):1949–1952

A Low-Profile Three-Stub Multiband Antenna for 5.2/6/8.2 GHz Applications

Sparsh Singhal, Pranjal Sharma, and Chandan

Abstract A low-profile multiband antenna achieved with three stub is proposed in this paper. The proposed antenna achieved three bands of frequencies range: 4.8 GHz to 5.4 GHz; 5.8 GHz to 6.4 GHz; 8 GHz to 8.4 GHz their resonating frequencies are 5.2 GHz; 6 GHz; 8.2 GHz. These frequencies are suitable for WiMAX application. Miniaturization is obtained by taking a patch of and area 14×18 mm^2. The proposed structure has s11 < -10 dB for bandwidth of about 11.76%(4.8−5.4 GHz); 9.83%(5.8–6.4 GHz); 4.8%(8–8.4 GHz). The proposed antenna is simulated.

Keywords Monopole antenna · Multiband antenna · FR-4 substrate and probe feed technique

1 Introduction

This antenna is having features like low profile, low cost, compact size and one of the most important applications is obtaining multiband which is used for different applications. The demand for seeking works like WiMAX, WiFi, and so forth which has been possible through wireless system. If the user wants to have their important information at any time, the wireless networks will be suitable. Nowadays, the need for communication is required more. WiFi has been played an important role in the communication world and will remain forever. Growing office community led to growing wireless transmission significantly. Wireless technology is used for a long time. The WiMAX application is used for communicating, from different places and the number of its users increases every day. The hardware devices connect with each

S. Singhal (✉) · P. Sharma · Chandan
Department of Electronics and Communication Engineering, Meerut Institute of Engineering and Technology, Meerut, India
e-mail: sparsh.singhal.ec.2017@miet.ac.in

P. Sharma
e-mail: pranjal.sharma.ec.2017@miet.ac.in

Chandan
e-mail: chandan.kumar@miet.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_64

other without using a wire or cable. There are many processes for designing antennas. Here, the use of monopole microstrip antenna is used due to easy construction, low cost, and required features of the antenna with low profile devices. The multiband antennas are used in different bands, due to which these antennas can be achieved in terms of size. There are several ways to design multiband antennas. This antenna uses three stubs and a rectangle ground plane for WLAN/WiMAX are used for multiband and broadband communication. The frequency we obtain is valid for different applications such as WiMAX(wireless broadband access to buildings), WiFi 6E, radar applications. The frequency band required for this application is all we obtained and also advantage is easily simulation on which we work.

Also, we get knowledge about disadvantages so that we can work on it and overcome it and these are narrow bandwidth, low efficiency, lower gain. Here we used an antenna of area $14 \times 18\text{mm}^2$and also it consists of cut in ground part and height of the antenna is 1.6 mm. After simulation, we used it in commercial and vast applications. We get multiband of different frequency range at bands (i.e., (4.8 GHz−5.4 GHz-WiMAX application); (6.6 GHz−7 GHz-WiFi 6E); (8 GHz−8.4 GHz-radar)). As we know, our size is small which helps in balancing bandwidth, gain, and efficiency.

We take a result at a frequency of 2.4 GHz. Chandan and Rai [17]—there is an antenna created which is used for WiMAX application, WiFi in this research paper. This antenna consists of a rectangular patch with 2 L-shaped stubs and 3 slots over the ground which are separated by a substrate. As a result, the simple design enables the antenna to operate in such frequencies of 1.9, 2.4, 3.5, 5.2, and 7.6 GHz which are across the WWAN/WiMAX/WLAN bands. Chandan and Rai [16] Moreover, researches are done and in one research There is an antenna which composes of a square patch etched with an X-shaped slit, and with an improved ground plane the proposed antenna provides a relative wide 3-dB AR bandwidth of 24.2% (2.22–2.83 GHz), 25.9% (2.92–3.79 GHz), and 9.3% (5.45–5.98 GHz), which successfully covers the bands of WLAN (2.4–2.484 GHz and 5.725–5.850 GHz) and WiMAX (3.3–3.6 GHz). On further studying about WiMAX and WLAN application [1–8]. The proposed antenna is of different types like monopole, Wang-shaped patch antenna these all used microstrip feeding due to which dual band, enhanced bandwidth features were achieved. Its applications are Bluetooth and GPS, wireless. One antenna which is proposed is the pi-shaped antenna for a particular range of frequencies, i.e., 2.4\2.8 GHz. The features were miniaturization of several antennas that are U-slotted patch antenna, L-slotted monopole antenna which were used in these truncated ground were used.

2 Antenna Design

The proposed microstrip antenna is shown in Fig. 1, which is comprised of a substrate with an area of 40×30 mm^2. Which is made up of FR-4 material which consists of a rectangular patch of area 14×18 mm^2. There is a strip of area 3×11.6 mm^2. Whose one end is connected to the substrate and other to path this strip is

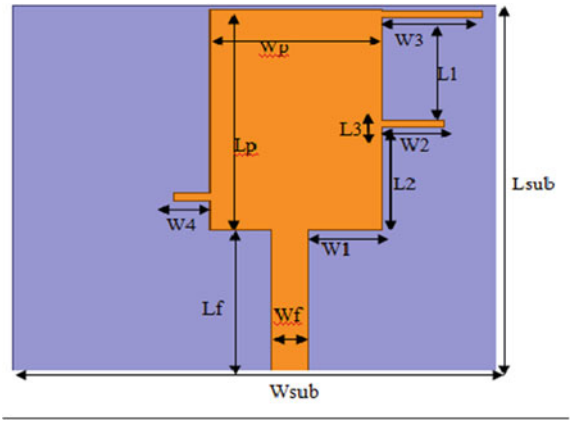


Fig. 1 Proposed antenna

united with patch and in down of strip input is applied by the help of lumped port so that we received the multiband which is suitable for WiMAX and other applications. According to simulated results, as shown in the figure above, we obtained a multiband of frequencies having range 4.8 GHz–5.2 GHz; 5.8 GHz−6.1 GHz; 8 GHz–8.4 GHz which have the application like WiMAX, WiFi 6E, radar system. The figure shows the ideal prototype of required multiband microstrip antenna. Here Fig. 1 shows the footprint of the proposed microstrip patch antenna and the overall footprint of the proposed design is 40 × 30 mm^2. Fig. 2 shows the simulated results of return loss which should be less than −10 (means at least 90% input power is given to device and reflected power is less than 10%. This value is sufficient for many applications).

The proposed parameters are optimized using HFSS software. Also, the designed guidelines of the proposed structure are shown in the figure which clearly shows that how multiband behavior is obtained by a simple printed antenna (Table 1).

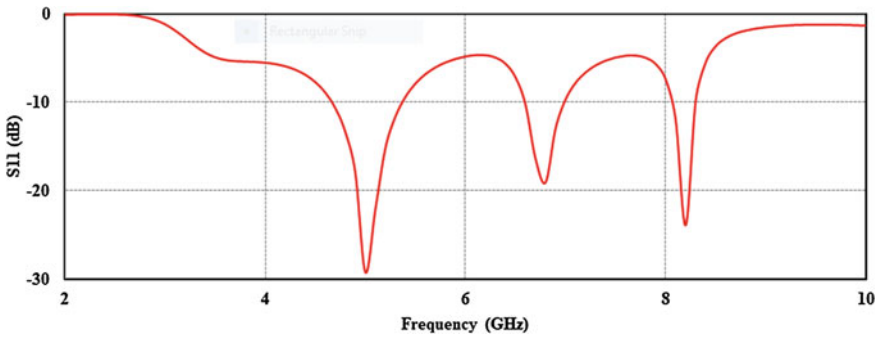


Fig. 2 Return loss of the proposed antenna

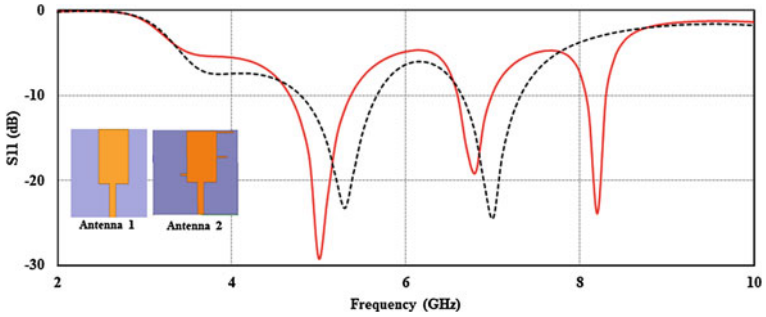


Fig. 3 Configuration of the proposed antenna

3 Result Discussion

On the substrate we take one more rectangle of length 40 × 11.6 mm^2 and name it as ground. Firstly, we take only a patch of 14 × 18 mm^2 area but the result obtained is only up to single band which doesn't meet the requirement we want for a multiband due to the application of WiMAX . Secondly, we further attach the strip of an area 8 × 0.5 mm^{2}and unite it with the patch to achieve the multiband but this also doesn't give multiband and also doesn't fulfill WiMAX application so we further process. Thirdly, we simultaneously did two modifications; one, we attached one more strip of an area 5 × 0.8 mm^2 and unite it with a patch and the other we take an E-shape cut in the ground to achieve the target but this time also we achieve the result as dual band which is not suitable for WiMAX. Further fourthly, we take one more strip, i.e., (p3) of an area 4 × 0.5 mm^{2}and unite it with the patch this time our result start coming but not that much required so in the last we take one more strip, i.e., (p4) of an area 8 × 3.85 mm^2 inside the patch and unite it with a patch and this time ideal result is achieved and we get multiband required fir WiMAX application. This much adjustment is only done to achieve multiband because we work on WiMAX application (Figs. 3 and 4).

4 Result Discussion

Figure 1 shows the simulated design of the proposed antenna which comprises of size 40 mm, 30 mm, 1.6 mm. The antenna shows return loss at 5 GHz; at 6 GHz; at 8.2 GHz resonating frequencies, at this frequency antenna proposed maximum power (Fig. 0.2). The antenna is used for wireless transmission.it is a monopole antenna whose characteristic is its radiation which is maximum in horizontal directions. Its types are whip, helical, umbrella, etc. Each antenna has its own design, characteristics, etc. In today's scenario communication is not possible without an antenna specially a wireless application. It plays an important role nowadays in terms of

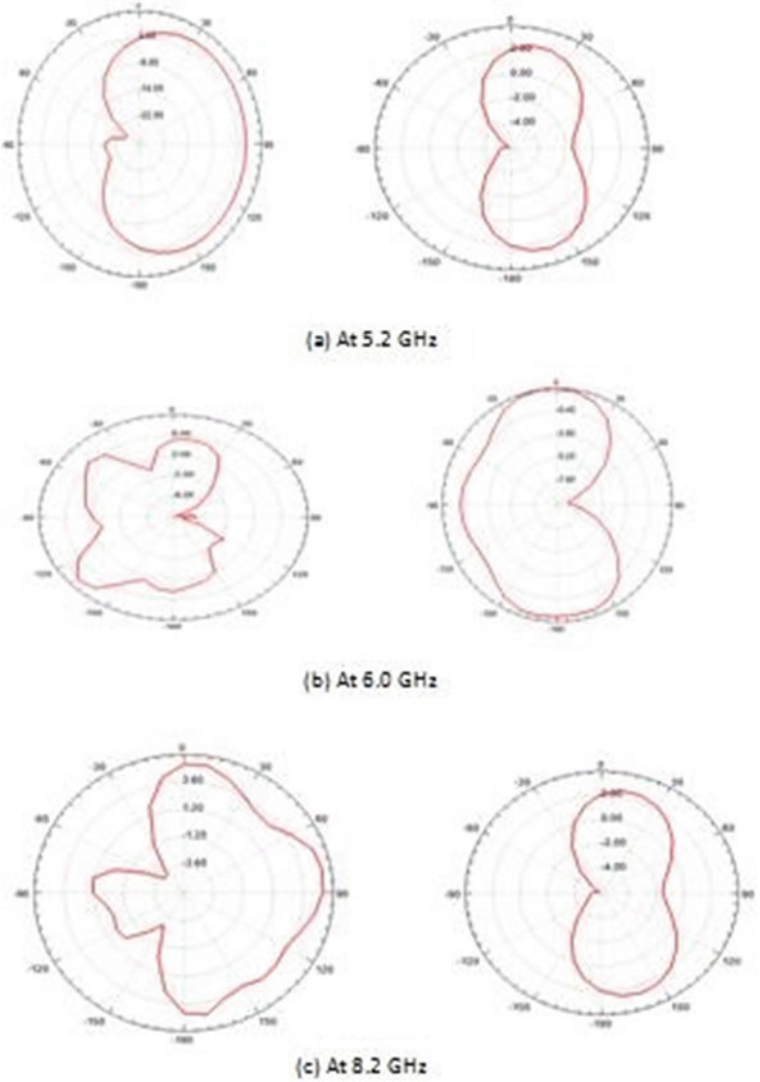


Fig. 4 Radiation pattern of the proposed antenna

Table 1 Dimension of the proposed antenna (mm)

Lsub	30	L2	6
Wsub	40	L3	8
Lf	11.6	Lp	18
Wf	3	Wp	8
L1	5	W1	9.5
W2	5	W4	4

modern communication and one of its types is microstrip patch antenna which has low size, low cost, simple design. The patch antenna takes any shape but in this paper, we use three stubs with it for configurations. This patch is of low profile and is used as long-range communication also for WiMAX applications. The proposed antenna of multiband is used for various applications and one antenna met the different requirement and fulfill it. So, the main purpose is to obtain the multiband due to which configuration will have to be done.

5 Conclusion

The proposed patch antenna array of three stub shapes was simulated and implemented using FR-4 material substrate. As the analysis of HFSS software, we observe that the antenna resonated at 2.4 GHz. In this paper, a low profile antenna with three stubs that operate at WiMAX and WiFiWiFi applications was successfully simulated. The result after implementation shows that the antenna produces return loss less than 10 dB at frequency ranges 4.8 GHz–5.4 GHz; 5.8 GHz–6.2 GHz; 8 GHz–8.4 GHz and their resonating frequencies are 5.2 GHz; 6 GHz; 8.2 GHz, which lies in the range required for wireless transmission. The given antenna is easy in simulation and has simple configuration. It also gives a fine radiation pattern. Hence, this antenna is appropriate for wireless applications such as WiMAX, Wi-Fi 6E, etc.

References

1. Chandan, Rai BS (2014) Bandwidth enhancement of Wang shape microstrip patch antenna for wireless system. In: IEEE Fourth International Conference on communication systems and network technologies (NITTR BHOPAL), 11–15
2. Chandan, Rai BS (2014) Dual-band Wang shaped microstrip patch antenna for GPS and Bluetooth application in: IEEE Sixth International Conference on computational intelligence and communication networks (Udaipur), pp 69–73
3. Chandan, Rai BS (2016) Dual-band monopole patch antenna using microstrip Fed for WiMAX and WLAN applications. In: Springer Proceedings information systems design and intelligent applications, vol 2, pp 533–539
4. Chandan (2020) Truncated ground plane multiband monopole antenna for WLAN and WiMAX applications. IETE J Res 1–6
5. Chandan, Srivastava T, Rai BS (2017) L-Slotted microstrip fed monopole antenna for triple band WLAN and WiMAX applications. In: Springer in Proceedings theory and applications, vol 516, pp 351–359
6. Chandan, Srivastava T, Rai BS (2016) Multiband monopole U-slot patch antenna with truncated ground plane. Microwave Opt Technol Lett 58(8):1949–1952
7. Chandan, Bharti GD, Bharti PK, Rai BS (2018) Miniaturized Pi (π)—slit monopole antenna for 2.4/5.2.8 applications. In: AIP Conference Proceedings, American Institute of Physics, pp 200351- 200356
8. Chandan, Bharti GD, Srivastava T, Rai BS (2018) Dual band monopole antenna for WLAN 2.4/5.2/5.8 with truncated ground. In: AIP Conference Proceedings, American Institute of Physics, pp 200361–200366

9. CuiYang Y, Liu LB, Li R (2016) Multiband planar antenna for LTE/ GSM/UMTS and WLAN/Wi-MAX handsets. IET Microwave Antennas Propag 10:502
10. El Yassini A, Ibnyaich S, Chabaa S, Zeroual A (2020) Miniaturized broadbandmultiband planar antenna with a symmetric quartercircular ground plane for WLAN/WiMAX standards. Microwave Opt Technol Lett 1–12
11. Ganesh MK, Rekha PL, Ramesh SP (2015) Tuneable and dual band rectangular microstrip antenna for bluetooth and WiMAX applications. Microwave Opt Technol Lett 57:1986–1991
12. Gangwar SP, Gangwar K, Kumar A (2018) Dual-band modified circular slot antenna for WLAN and WiMAX applications. Progress Electromagn Res C 85:247–25D
13. Gangwar SP, Gangwar K, Kumar A (2019) Dual band modified circular ring shaped slot antenna for GSM and WiMAX applications. Microwave Opt Technol Lett
14. Gautam AK, Bisht A, Kanaujia BK (2016) A wideband antenna with defected ground plane for WLAN/Wi-MAX applications. AEU Int J Electron Community 70:354–358
15. Hanae E, Amar Touhami N, Mohamed A, El Amrani S, Tazon A, Boussouis M (2014) Miniaturized microstrip patch antenna with defected ground structure. Prog Electromagn Res C
16. Hua MJ, Wang P, Zheng Y, Yuan SL (2013) Compact tri-band CPWfed antenna for WLAN/WIMAX applications. Electronic Letters 49(18):1118–1119
17. Wei CY, Liu JC, Bor SS, Hung TF, Chen CC (2013) Compact singlefeed circular slot antenna with asymmetrical C-shaped strips for WLAN/Wi-MAX triband and circular/elliptical polarizations. Microwave Opt Technol Lett 55:272–278
18. Wu T, Shi XW, Li P, Bai H (2013) Tri-band microstrip fed monopole antenna with dual polarisation characteristics for WLAN and Wi-MAX applications. Electron Lett 49(25)1597–1598

A Review on Renewable Energy System

Shraddha Sharma

Abstract Various organizations all across the globe adopted a mechanics called 'Clean Development Mechanism' (CDM). Clean and renewable energy resources are developed due to the shortage of naturally available energy sources like fossil fuels and natural gases. Research-based on renewable energy has become an essential topic in the twenty-first century due to the higher energy crisis. On the other hand, the energy extraction process, like the combustion of fossil fuels, gives rise to the enormous pollution level in the world; thus, the utilization of these energy sources leads to the rapidly decreasing their reverse. Natural resources like sunlight, wind, tidal, geothermal, etc., are available all across the world, and free of cost are popularly known as renewable energy resources or non-conventional energy resources. Among all non-conventional energy resources, solar energy or sunlight energy is most popular as well as widely used for the extraction of useful energy and power. Solar energy has become one of the most promising alternatives for conventional energy resources. In contrast to conventional energy resources, renewable energy resources are much eco-friendly and much cleaner and produce energy without the harmful effect of pollution.

Keywords Solar photovoltaic · Solar thermal · Wind energy · Renewable energy · Photovoltaic cells

1 Introduction

There are particular types of cells and equipment used for the extraction of useful electrical energy from the solar energy of sunlight or solar irradiance these cells are known as solar cells or photovoltaic cells or PV solar cells. The name "photovoltaic" has been derived from the two words "photo" meaning light and "volt" meaning the measurement of electricity which directly suggests the generation of voltage (electrical energy) from the photon (incident sunlight energy). The transformation of the

S. Sharma (✉)
Delhi Technological University, Delhi 110042, India
e-mail: shraddha.sharma@iilm.edu

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_65

light energy in electrical energy is based upon the fact that in a semiconductor fixed electron can be converted into a freely moving conduction electron. Ashwani et al. [5, 6, 21] using the PV cell to harness the sun's energy is a rapidly expanding science. Heat energy can also be obtained from solar thermal power plants, where the light energy is obtained from the sun with the help of solar collectors. Sunlight directly falls on the solar collectors with high intensity, and these collectors can concentrate the full sunlight, and heat energy can be obtained at the focus of the collectors, which can be collected for heat utilization. Konstantinos [11, 14, 16, 22] wind energy can be used for the generation of electricity, and it is also known as a renewable source of energy. The wind is available all over the world free of cost, and it is one of the eco-friendly and readily available sources of energy. Wind energy converts the form of useful energy such as electricity with the help of a windmill, which can be installed in the open hilly areas where the wind would be available without any interruption and at a very high intensity so that this can be able to move the blades of windmills and the motor connected to the axes of the windmill can move quickly to produce the electrical energy. Apart from solar and wind energy Hydropower, Biomass, Geothermal, and Ocean energy are also considered as renewable energy sources [17]. India faces on an average 300 sunny days per year and due to its ideal geographical position India receives 1600–2200 kilowatt-hours per meter 2. According to NISE (National Institute of Solar Energy) in every state, nearly 3 percent of the land is available for solar power plants and has the potential of approximately 750 GW solar energy production. States with high solar potential are Rajasthan, Maharashtra, Jammu and Kashmir, and Madhya Pradesh [23]. India's wind energy program started in the 1990's to provide pollution-free energy in Asia [4]. In 2014 India became the fifth largest wind energy-producing country. According to MNRE, India is on the way to produce 60000 MW of electricity by 2022 using wind power. In 2010, the potential of onshore wind energy was 49.1 GW while in 2012 it was updated to 102.1 GW at a hub height of 80 m. According to an estimate presented by Lawrence Berkeley National Laboratory, India has a wind energy potential of 2000–3000 GW [15]. Hydroenergy is the most exploited energy in India. Turbine machines are used to convert the kinetic energy of flowing water into electricity and India is among the top 10 hydro energy power generators worldwide [26]. Nearly 20,000 MW of energy is generated via hydropower.

2 Principle of Photovoltaic Cell

Bell Telephone researchers developed the first photovoltaic cell. The diagram showing the solar radiation and its types is given in Fig. 1. Initially, at first, PV cells were used primarily in space to power the U.S. space satellite. But current solar photovoltaic cells are being standard in many different applications. Solar cells or photovoltaic cells are semiconductor material, a thin wafer of silicon, which constitutes some properties of both metals and insulators, making it uniquely capable of converting solar light into electricity. The top of the wafer has a tiny amount of phosphorus added to it. This makes the top portion of a wafer having a sufficient amount of

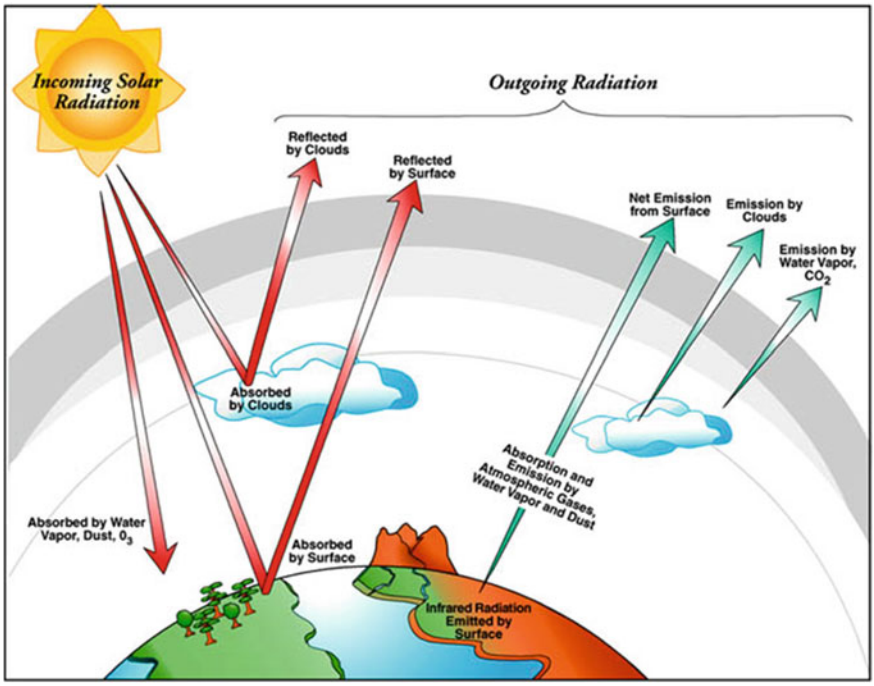


Fig. 1 Types of radiation from the Sun (Credit—Steve Ackerman and Tom Whittaker)

free, negatively charged electrons. The photovoltaic system directly converts sunlight into energy [10, 12, 13, 20]. Figure 2 shows the cross-section of real solar cells. The starting material of solar cells popularly known as the base is a p-doped or p-type semiconductor. It means it is a semiconductor crystal having an excess of acceptor impurities. The emitter of the photovoltaic cell is an n-doped or n-type semiconductor, i.e., excess of donor impurities, and it is highly doped as compared to the base by some magnitude. Thus p–n junction formed is unsymmetrically doped. The boundary between the two regions is called the p–n junction.

When the light of appropriate frequency falls on the junction, additional electron–hole pairs are generated in both regions (due to the breaking of covalent bonds). As light is absorbed by the p–n junction of the semiconductor diode, photons of light can transfer their energy to the electron, allowing the electron to flow through the material as an electrical current. In the absence of illumination, the reverse saturation (dark) current in the p–n junction diode is proportional to the concentration of thermally generated minority charge carriers and id denoted by I_d. When the PV cell is placed in the sun, the radiant energy excites the free electrons [3, 7, 18, 19]. If the complete circuit is made by connecting the side of the wafer, electrons transfer their energy from atom to atom from the n-type through the wire to the p-type. The photovoltaic cell now starts producing energy, i.e., transfer of energy from moving of electrons. Nowadays, in various power generation industries, the solar photovoltaic array is

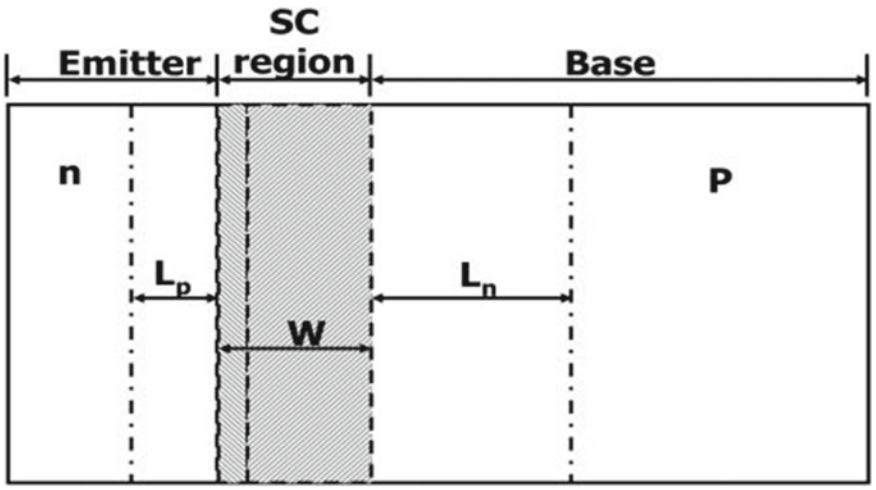


Fig. 2 Cross-section of real solar cell [1]

installed on the roof of the building itself, thus in such a case it will become more difficult to ignore the effects of shades as the shade of building will definitely fall onto the solar panel throughout the day and in every season. Thus, these shadows decrease the level of overall power generated by the solar cells. Now to solve this problem and to obtain the desired amount of electrical power a large number of solar modules will be required, but it leads to the higher cost of the installation, and because of the higher installation cost solar photovoltaic module will become useless as compared to the initial situation. Arora et al. [4, 7, 17, 23, 25] this is the reason why the impact of partial shading has been discussed in recent years. The study of partial shading involves various time-consuming and cost-effective procedures and became very complex such as it required appropriate weather condition, field testing is very costly as well as time-consuming. On the other hand, the numbers of shaded and illuminated cells are keeping on changing during the whole experiments. Thus to obtain the accurate and appropriate results, it is better and more convenient that the study would be carried out on some simulation model with the help of computer.

If a load connected along the wire such as a light bulb forming the circuit, the bulb will glow as the electricity will pass through it. Here the conversion of sunlight energy into electrical energy is an entirely silent and instant process. When the reverse-biased junction is illuminated, the number of newly created electron–hole pairs is proportional to the number of the incident photon [15, 26].

Each absorbed photon creates an electron and holes (an electron in the valence band moves up to the conduction band, leaving a hole in the valence band). The built-in electric field forces the electron toward the n-region and the hole toward the p-region. The two sides of the junction thus become charge oppositely, so a voltage is generated across the junction. This voltage drives a current in an external circuit, i.e., an electrical current can be produced. Solar cells are used in space vehicles

for charging batteries in the daytime and then taking electrical power from them at night. Energy harvesting is one of the most recent applications of solar cell. New technologies are continually being developed to make PV cells thinner and more flexible. There are now PV roofing shingles. Rather than putting panels on your roof, solar shingles can be used that match the conventional shingles for a more pleasing look. Scientists are developing photovoltaic cells that can be fit into the home window and on thin, flexible film that can be attached to the outside of the home.

3 Solar Thermal Power Plant

The sun can provide a vast amount of energy, and this energy has the potential to fulfill all the needs of the entire population existing on the earth. According to the International Energy Agency, the worldwide installed capacity of solar thermal power plants is triple from 4 to 12 GW by 2020. The challenge is to utilize this climate-neutral energy inefficient and cost-effective manner. The Sahara Desert has an edge nearly Southern Moroccan town, and it is planning to build its own thermal power plant.

Including the very first solar thermal parabolic trough power plant, and it will have to be proved as the largest among other such kinds of power plants. One plant having the intention to provide power up to 500 MW is proposed by Ouarzazate after the completion of the second construction phase. Proving climate-friendly electricity supply is the first step going to be taken by the State Moroccan Energy Agency. In big countries like the U.S. and Spain, the technology of parabolic trough power plants has already been adopted with the solar thermal power plant.

Using giant parabolic mirrors, sun rays can be made concentrated on a kind of tube known as an absorber tube through which the medium for transferring heat flows. That medium most of the time is thermal oil. The main aim is to feed it onto the power plant block, where water can be evaporated with the help of a steam turbine which is capable of generation of electrical energy. *Although this technology has been proven itself in commercial areas.* There are four main research areas presented by the BINE Themen info brochure. They are:

- Fresnel system,
- Parabolic trough collector technology,
- Tower power system, and
- Tests and quality assurance.

Approximately, 95% of the solar commercially operated thermal power plants are based on a parabolic trough system. Its cost can further be reduced by enhancing its collector's designs and by making its operation and maintenance to be automatic. There is problematic thermal oil that has to be used now as the medium for heat transferring because of its aqua toxicants and its property to leak in the pipeline system, when dissolved in groundwater it can cause serious ecological damage. A diagram showing the various types of solar concentrators is given in Fig. 3.

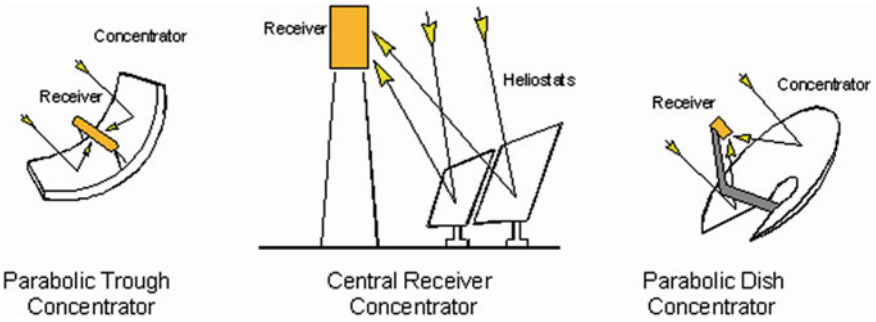


Fig. 3 Types of solar collectors. (https://www.powerfromthesun.net/Book/chapter01/chapter01.html)

4 Working of Parabolic Trough Solar Thermal Power System

Only vertically incident radiations on the optical axis are to be concentrated on the focal point. Thus the concentrator can continually track the path followed by the sun. For this purpose, hydraulic drive system would be used. Electric motors are used by the smaller collectors also. The sensor, which is used for the determination of the position of the collectors, which is relative to the elevation of the sun, is controlled by the drive system. The sun's elevation is also calculated numerically and the position of the sensor for collectors or by the combination of these two. Concentrated solar radiation is transformed into heat and allows to transfer to the heat transfer medium, i.e., oil flowing inside the absorber tube. This is a steel tube having a coating of optical material for maintaining high absorbance in the range of solar spectrum wavelength. Parabolic trough collector has a structure that is modular in shape. All the collectors involved in the process are connected to one another and mounted in rows on supporting pylons so that they would be torsional stiff.

A diagram is given that shows the various losses in sun energy while converting into the electrical energy when incident on the surface of the concentrator mirror and due to this reason the intensity of sunlight falling on to the concentrator surface is higher than the intensity of the sunlight reflecting from the concentrator surface. The losses associated with it are known as the End Losses. Figure 4 shows the end losses in the solar thermal power plant.

5 Wind Energy

A converted form of solar energy is the wind energy; it can be produced by nuclear fusion of two gases such as Hydrogen (H) and Helium (He) in its core. Heat and electromagnetic radiation are streamed out from the sun into all over the space when

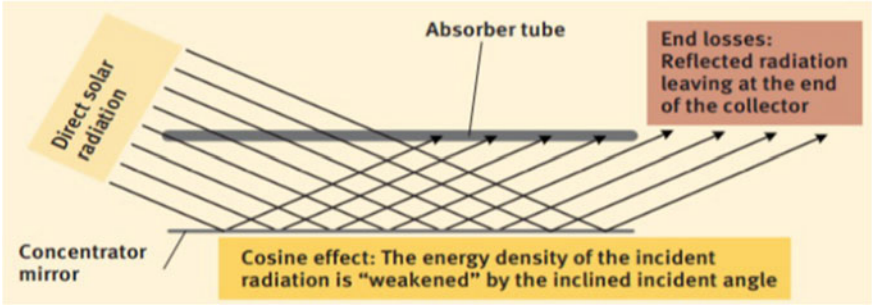


Fig. 4 Cosine effects and end losses

these two gases H → He fusion process takes place. For new power generation, wind energy is considered as the primary energy source in the world energy market. Because atmospheric pressure gradient movement of air takes place, it results in the generation of wind energy. Wind flows from the region of the lower concentration of wind to the region of higher concentration of wind. Wind speed and wind power depend on the atmospheric pressure gradient. Windmills are the most important and successfully utilized wind energy generation system. China is using windmills for several previous years.

When the object having a given mass moves with a translational or rotational speed exists kinetic energy moving air's kinetic energy when the object moves are determined by

$$E_k = \frac{1}{2}mv^2 \tag{1}$$

where 'm' is the mass of air, and 'v' is the speed of the wind over a specified time. When the kinetic energy is differentiated with respect to time, wind power can be obtained

$$P_w = \frac{dE_k}{dt} = \frac{1}{2}mv^2 \tag{2}$$

Only a small fraction of wind energy is converted into electrical energy. Blades of wind turbines rotate when the wind passes through the wind turbine. The flow rate of the corresponding mass of wind is

$$m = \rho Av \tag{3}$$

where ρ is the density of air, and 'A' is the swept area of the blades. Figure 5 shows the swept area of the blades. Thus the power available in the wind can be obtained as

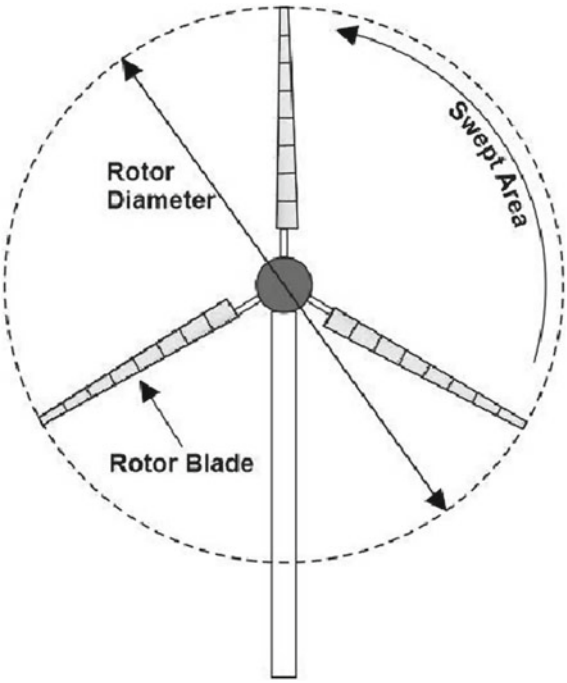


Fig. 5 Swept area of wind turbine blade

$$P_w = \frac{1}{2}\rho A v^2 \tag{4}$$

The swept area of blades can be calculated by the given below equation:

$$A = \pi\left[(l+r)^2 - r^2\right] = \pi(l+2r) \tag{5}$$

6 Windpower Density

At any particular site, the wind power density is used as the comprehensive index for the evaluation of wind resources. It is defined as the available wind power through the particular cross-sectional unit area in a unit time in airflow.

Table 1 indicates the wind power density at two standard wind measurement heights, i.e., 10 m and 50 m.

Table 1 Wind power class w.r.t height

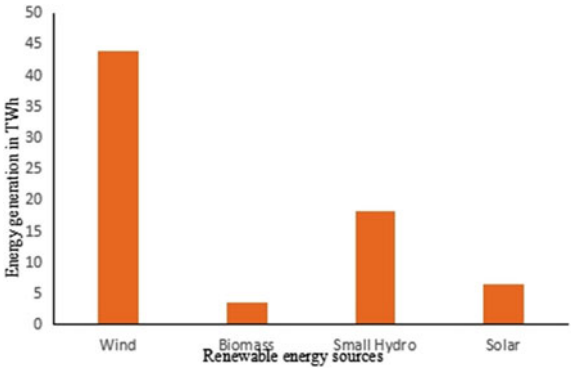
Wind power class	10 m height		50 m height	
	Wind power density (W/m^2)	Mean wind speed (m/s)	Wind power density (W/m^2)	Mean wind speed (m/s)
1	<100	<4.4	<200	<5.6
2	100–500	4.4–5.1	200–300	5.6–6.4
3	150–200	5.1–5.6	300–400	5.6–6.4
4	200–250	5.6–6.0	400–500	6.4–7.0
5	250–300	6.0–6.4	500–600	7.0–7.5
6	300–350	6.4–7.0	600–700	7.5–8.0
7	>400	>7.0	>800	>8.0

7 Analysis

China, India, and Pakistan collectively have nearly 40% of the world population and due to growing economy the energy need is also increasing [2]. This can be understood by per capita energy consumption in various countries as per the following table:

Country	Per capita energy consumption (kWh)
USA	13,361
France	7756
Germany	7217
China	2942
Turkey	2474
India	644
Sri Lanka	636.3
Pakistan	457
Afghanistan	119.8
Bangladesh	278.1
Nepal	454.1

In India, renewable energy sources constitute wind, biomass, small hydro, and solar energy. The World Institute of Sustainable Energy predicted [9] that the energy renewable energy production would rise to 220 TWh by 2032. In 2015, the total renewable energy production was 72.1 TWh [8] out of which the contribution of wind, biomass, small hydro, and solar were 43.9 TWh, 3.5 TWh, 18.2 TWh, and 6.5 TWh, respectively, this can be understood by the following bar graph.



According to Tiewsoh et al. [24], electricity model based on demands from 2015 to 2030 predicts 2 to 3 folds of enhanced consumption. The main problem in India is the dependence on coal energy production. Emphasis is on coal energy production which should be shifted towards renewable energy sources. Future options to fulfill the need for energy may be solved by social acceptance of nuclear power, evaluating the environmental impact of renewable energy sources, and gradually replacing high carbon emission sources for electricity generation with low emission sources.

8 Conclusion

Global warming, environmental pollution, and energy security are some rising concerns in the field of energy generation for the fulfillment of the present and future needs. Thus the need for renewable energy sources for useful energy generation is necessary. Here the author tried to introduce some of the natural resources for the generation of electrical energy such as solar energy solar thermal energy, wind energy etc. Eco-friendly energy resources like wind, solar, hydropower, geothermal, hydrogen, and biomass as the required replacement for fossil fuels. A short description of a solar photovoltaic cell is given by the author. The help of a suitable diagram has described the basic working procedure of a solar photovoltaic cell. The concept of solar thermal power plants is given along with the figure showing the solar collectors and their types. Eventually, the short description of wind energy leading to the wind power density at the two heights has been shown with the help of a proper table.

References

1. Abunahla HN (2013) Thesis, "Modelling and Simulation of Photovoltaic Modules", University of Sharjah
2. Ahmed S, Mahmood A, Hasan A, Sidhu GAS, Butt MFU (2016) A comparitive review of China, India and Pakistan renewable energy sector and sharing opportunities. Renew Sustain Energy Rev 57:216–225
3. Alola AA (2019) The trilemma of trade, monetary and immigration policies in the United States: accounting for environmental sustainability. Sci Total Environ 658:260–267
4. Arora DS et al (October 2010) Indian renewable energy status report. Background report for DIREC
5. Ashwani K, Kapil K, Naresh S, Satyawati S (2010) Renewable energy in India: current status and future potentials. Renew Sustain Energy Rev 2434–2442
6. Bakshi R (2002) "Wind energy", the Indian scenario. IEEE 1:344–345
7. Bekun FV, Alola AA, Sarkodie SA (2019) Toward a sustainable environment: nexus between CO2 emissions, resource rent, renewable and nonrenewable energy in 16-EU countries. Sci Total Environ 657:1023–1029
8. Central electricity authority (2015) Annual report 2014–15, Government of India, Ministry of Power, Central Electricity Authority: New Delhi, India
9. Planning Commission (2006) Integrated energy policy: report of the expert Committee, Government of India: New Delhi, India
10. Ghosh S, Sengupta PP (2011) Energy management in the perspective of global environment crisis: an evidence from India. IEEE, pp 1–5
11. Konstantinos D (2011) Solar energy in Egypt: significant business opportunities. Renewable Energy 36:2305–2311
12. Kothari DP (2000) Renewable energy scenario. IEEE 1:634–636
13. Li G, Hwang Y, Radermacher R (2012) Review of cold storage materials for air conditioning application. Int J Refrig 35:2053–2077
14. Load Generation and Balance Report, Central Electricity Authority, Ministry of Power, Government of India, 2011–12
15. Mapping India's Renewable Energy Growth Potential, Status and Outlook (2013) Renewable energy India expo
16. Moosavian SM, Rahim NA, Selvaraj J (2013) Energy policy to promote photovoltaic generation. Renew Sustain Energy Rev 25:44–48
17. Pal K, Yadav P, Tyagi SK (2017) Renewable sources in India and their applications. In: Chandel A, Sukumaran R (eds) Sustainable biofuels development in India. Springer, Cham
18. Sarkodie SA, Strezov V (2018) Assessment of contribution of Australia's energy production to CO2 emissions and environmental degradation using statistical dynamic approach. Sci Total Environ 639:888–899
19. Shahbaz M, Sinha A (2019) Environmental Kuznets curve for CO2 emissions: a literature survey. J Econ Stud 46(1):106–168
20. Singh R, Sood YR (2010) Analysis of renewable promotional policies and their current status in Indian restricted power sector. IEEE, pp 1–8
21. Sisodia G, Singh P (2016) The status of renewable energy research on Indi issues and challenges. Energy Procedia 95:416–423
22. Solangi KH, Islam MR, Saidur R (2011) A review on global solar energy policy. Renew Sustain Energy Rev 15:2149–2163
23. State wise (2015) Estimated solar power potential in the Country, (Solar R&D Division) ministry of new and renewable energy, government of India
24. Tiewsoh LS, Jirasek J, Sivek M (2019) Electricity generation in India: Present state, future outlook and policy implications, Energies 12:1361(1–14)
25. Wang J, Dong K (2019) What drives environmental degradation? Evidence from 14 Sub-Saharan African countries. Sci Total Environ 656:165–173
26. World Energy Council (2001) Survey of energy resource 2001 (Chapter on hydropower). www.worldenergy.org/wec-geis/publicatiopns/reports/ser/hydro.asp

Low-Profile H Slot Multiband Antenna for WLAN/Wi-MAX Application

Siddharth Vashisth, Sparsh Singhal, and Chandan

Abstract This paper presents a microstrip antenna that gives multiband when we use a microstrip feeding at the input of the antenna. The proposed antenna is made up of FR-4 substrate. The multiband that we obtained is used for different modern applications and also we get a different range of frequencies at each band. Here, the bandwidth is also satisfactory as we required and also return loss obtained is less than 10 dB as per the requirement. In this paper, various rectangular patch antenna arrays are designed for the application of WLAN (wireless local area network) at 2.4 GHz. The single patch antenna is designed using probe feeding technique. Arrays of 2 × 1 are also designed using the edge feeding technique.

Keywords Low profile antenna · FR-4 Substrate and probe feed technique

1 Introduction

Today's world is characterised by wireless communication and it is being served as one of the biggest contributions of technology to mankind. A wireless communication system uses the free space as the communication channel instead of wires and cables. Here the electrical signals from the transmitter need to be converted into the EM signals for propagation through free space and then the EM signals received by the receiver need to be converted back into electrical signals. The transducer which converts electrical signals to EM signals and vice versa at the transmitter and receiver side, respectively, is called antenna. Hence antenna is playing a vital component in a wireless communication system. Since the wireless devices are becoming more and more compact on one side and increase in the number of applications in a single device such as Wi-Fi, WiMax, Bluetooth, GPS, 5G, etc. [5]. There are some designs

S. Vashisth (✉) · S. Singhal · Chandan
Department of Electronics and Communication Engineering, Meerut Institute of Engineering and Technology, Meerut, India
e-mail: siddharth.vashisth.ec.2017@miet.ac.in

S. Singhal
e-mail: sparsh.singhal.ec.2017@miet.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_67

containing different shapes like U shaped, T shaped etc. [6, 7]. Parasatic radiation elements of different sizes and stacked patch antenna result in thicker antenna [4]. On the other side, there is a need for antenna design, which is compact in size, light in weight, lower in cost and supports multiple frequency band operation [8]. All the above-mentioned needs are fulfilled by the microstrip patch antenna.

2 Antenna Design

The diagram of the antenna structure is shown in Fig. 1. The values of the parameters are also enlisted in Table 1. The antenna is made on a cost-effective FR-4 epoxy substrate of thickness 1.6 mm. There are three monopoles having different lengths, microstrip feed line. In the typical design procedure of the microstrip antenna, the desired resonant frequency, thickness and dielectric constant of the substrate are known or selected initially. In this design of rectangular microstrip antenna, FR4

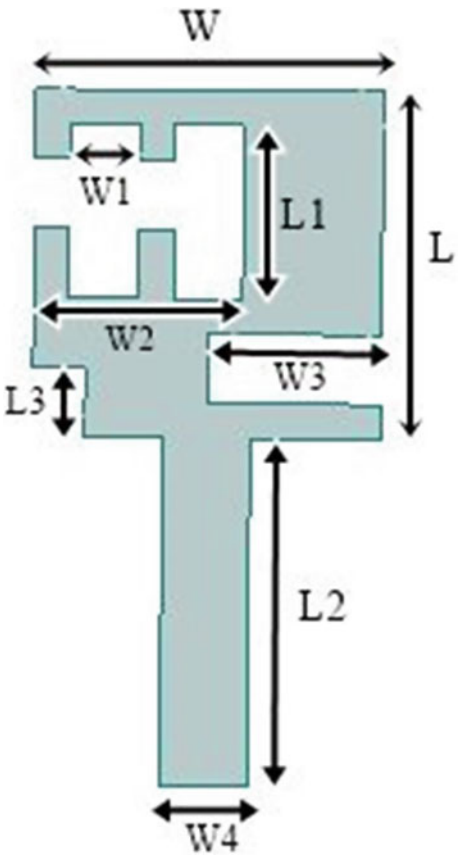


Fig. 1 Geometry of the proposed antenna

Table 1 Dimensions of the proposed antenna (in mm)

All Dimensions in mm	
L	10
L1	5
L2	10
L3	2
W	10
W1	2
W2	6
W3	5
W4	2.5

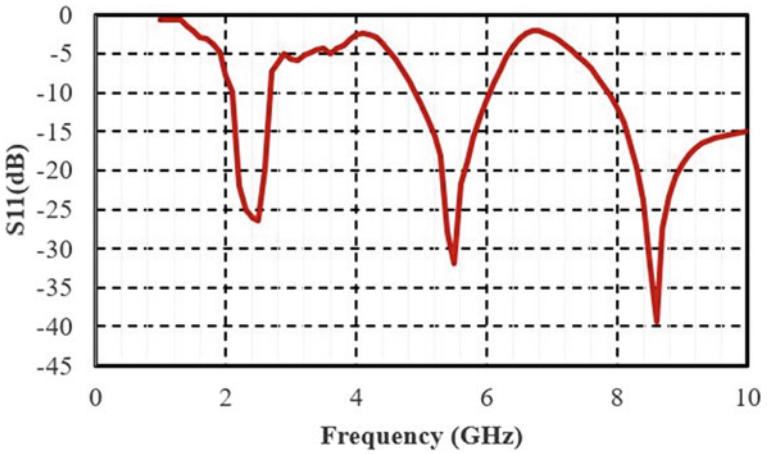


Fig. 2 Return loss of the proposed antenna

dielectric material ($\varepsilon r = 4.4$) with dielectric loss tangent of 0.02 is selected at the substrate with 1.6 mm height.

The antenna has shown −26.44 dB return loss at 2.4 GHz, −31.93 dB return loss at 5.5 GHz and −39.41 dB return loss at 8.6 GHz resonant frequency and obtained bandwidth is 2.4 GHz, 5.45 GHz and 8.56 GHz, which is shown in Fig. 2. At the resonant frequency, antenna radiates maximum power.

3 Results and Discussion

Figure 1 shows the simulated prototype of our proposed antenna, which consists of a compact size of 50 mm, 115.7 mm and 1.6 mm. The antenna has shown −26.20dB return loss at 2.4GHz, −31.97dB return loss at 5.5GHz and −39.61dB return loss

at 8.6GHz resonant frequency and obtained bandwidth is 2.45GHz, 5.85GHz and 8.55GHz, which is shown in Fig. 2. At the resonant frequency, antenna radiates maximum power. Antennas play a vital role in the field of wireless communications. Some of them are parabolic reflectors, slot antennas, patch antennas and folded dipole antennas. Each one of the antennas is superior in its own characteristics, design and applications. We can state antennas are one of the most essential things in wireless communication without which the world could not have imagined the present day of technology.

In Fig. 3a, we have designed a patch in which three slots are cut and we have obtained the results as shown in the figure. In the above step, we get a single band only between 8 and 9 and we are working for triple band so it is of no use for us now so we have to do further slot additions in our design.

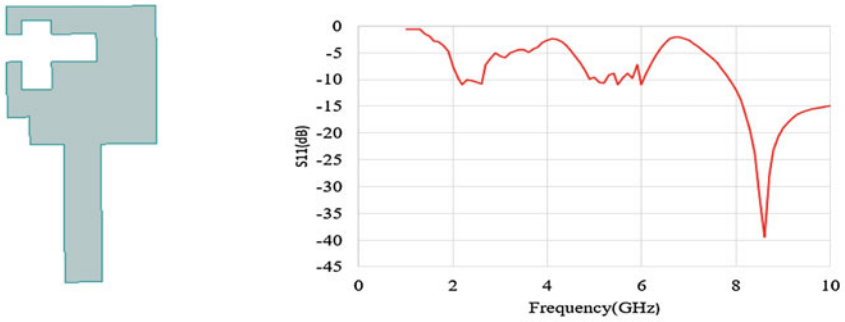
In Fig. 3b, we have done some modifications to improve our result so we have added one more slot to achieve our result so that we can get triple band. After adding a slot, we have received another band, which is not for our final result so we have to do further modifications.

In Fig. 3c, when we cut one more slot then we have achieved our result, which is triple band. This time we get usable frequency bands with usable application at different frequencies. Return loss should be less than −10 dB for acceptable operation. The antenna has −12.55 dB return loss at 2.4 GHz, −38.17 dB return loss at 3.9 GHz and −28.61 dB return loss at 5.6 GHz (Table 2).

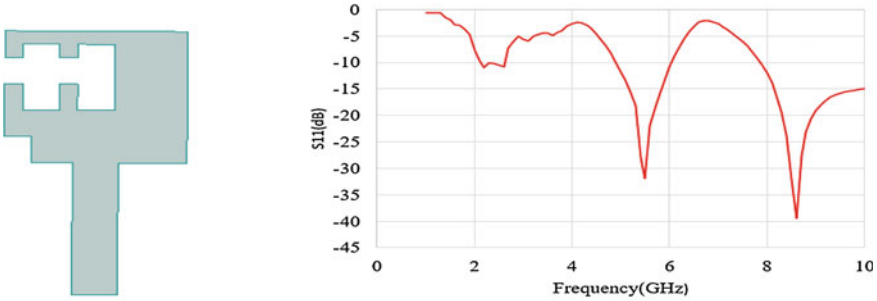
We have a frequency versus gain graph, which shows the gain for the proposed antenna design at two different frequencies (Figs. 4, 5, 6, 7 and 8).

4 Conclusion

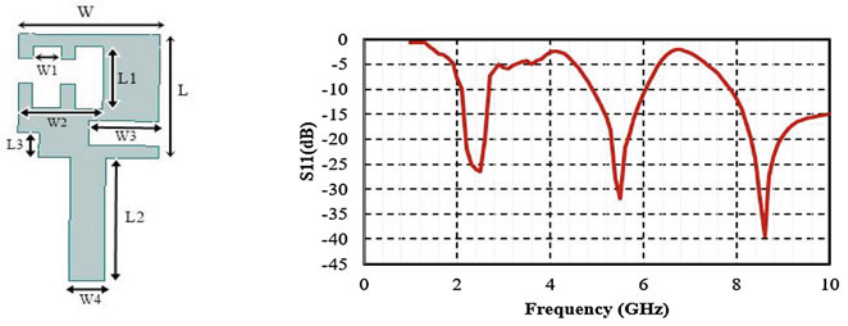
Low-profile H slotted antenna for different wireless applications is investigated and successfully simulated in this paper. The antenna has shown −26.20 dB return loss at 2.4 GHz, −31.97 dB return loss at 5.5 GHz and −39.61 dB return loss at 8.6 GHz resonant frequency and obtained bandwidth is 2.45 GHz, 5.85 GHz and 8.55 GHz at the maximum radiated power.



(a)



(b)



(c)

Fig. 3 Step modification of antenna

Table 2 Comparison of the proposed antenna with reference antenna

Ref	Number of band	Compact size of patch
[3]	2	70.5 mm × 87 mm
[13]	2	8.9 cm × 6.4 cm
Proposed antenna	3	20 mm × 10 mm

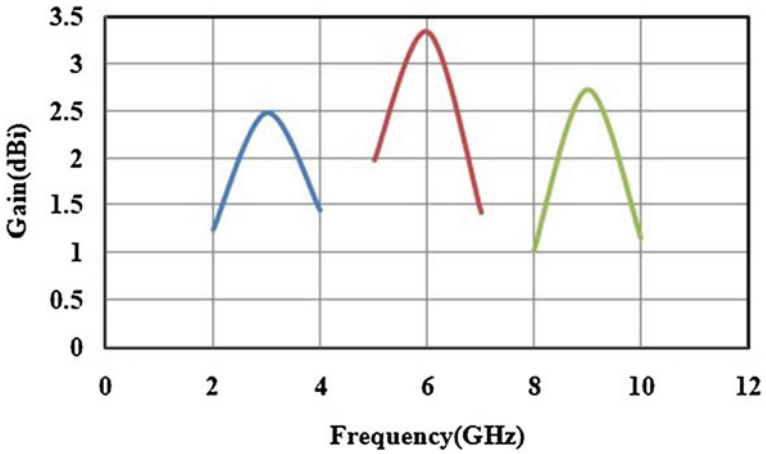


Fig. 4 Gain curve of the proposed antenna

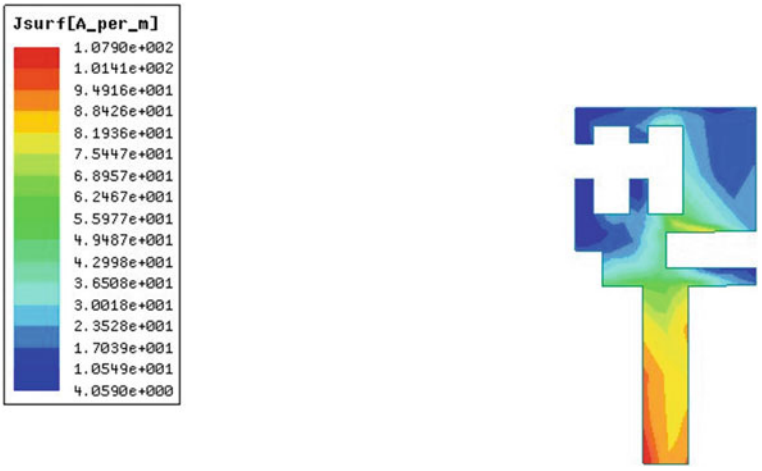
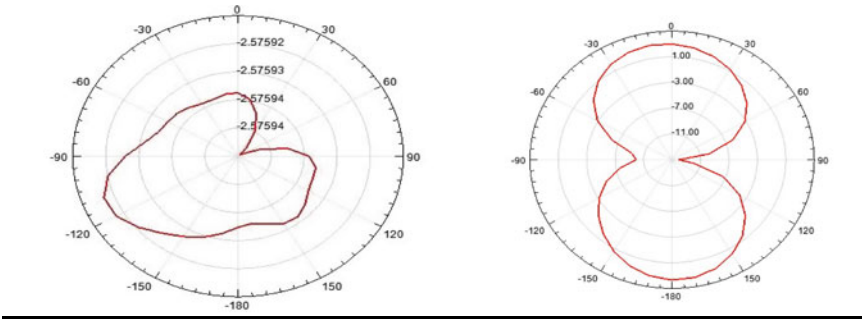


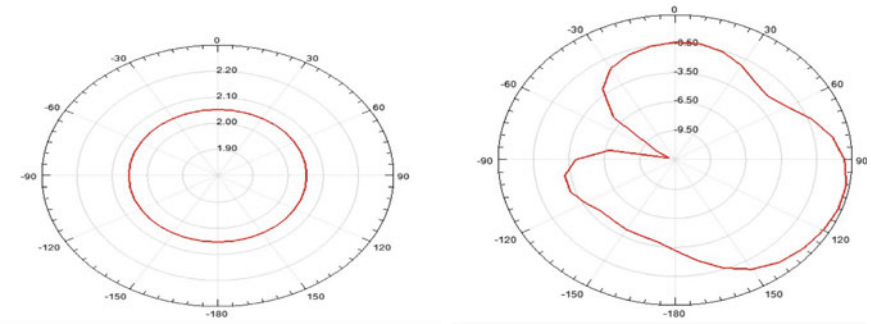
Fig. 5 Current distribution pattern at 2.4 GHz

Fig. 6 Current distribution pattern at 5.5 GHz

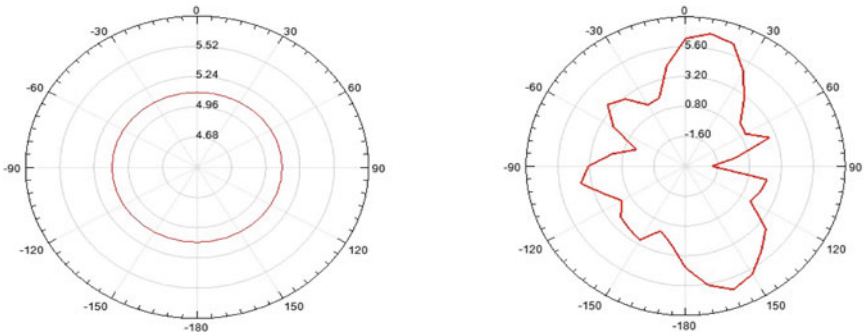
Fig. 7 Current distribution pattern at 8.6 GHz



(a) At 2.4GHz



(b) At 5.5GHz



(c) At 8.6GHz

Fig. 8 Radiation pattern at different frequencies

References

1. Ashrf A, Ufuk Türeli MS (2018) Design, simulation, and fabrication of a small size of a new spiral shaped of circular microstrip patch antenna. Microwave Opt Technol Lett 60(12):2912–2918
2. Chandan, Rai BS (2014) Bandwidth enhancement of Wang shape microstrip patch antenna for wireless system. In: IEEE Fourth International Conference on communication systems and network technologies (NITTR BHOPAL), pp 11–15
3. Chandan, Rai BS (2014) Dual-band Wang shaped microstrip patch antenna for GPS and Bluetooth application. In: IEEE Sixth International Conference on computational intelligence and communication networks (Udaipur), pp 69–73
4. Chandan, Rai BS (2016) Dual-band monopole patch antenna using microstrip Fed for WiMAX and WLAN applications. In: Springer Proceedings information systems design and intelligent applications, vol 2, pp 533–539
5. Chandan (2020) Truncated ground plane multiband monopole antenna for WLAN and WiMAX applications. IETE J Res 1–6
6. Chandan, Srivastava T, Rai BS (2017) L-slotted microstrip Fed monopole antenna for triple nand WLAN and WiMAX applications. In: Springer in Proceedings theory and applications, vol 516, pp 351–359
7. Chandan, Srivastava T, Rai BS (2016) Multiband monopole U-slot patch antenna with truncated ground plane. Microwave Opt Technol Lett 58(8):1949–1952
8. Chandan, Bharti GD, Bharti PK, Rai BS (2018) Miniaturized Pi (π)—slit monopole antenna for 2.4/5.2.8 applications. In: AIP Conference Proceedings, American Institute of Physics, pp 200351–200356
9. Chandan, Bharti GD, Srivastava T, Rai BS (2018) Dual band monopole antenna for WLAN 2.4/5.2/5.8 with truncated ground. In: AIP Conference Proceedings, American Institute of Physics, pp 200361–200366
10. Ghosh B, Ghosh S, Kakade AB (2008) Investigation of gain enhancementof electrically small antennas using double-negative, single-negative,and double-positive materials. Phys Rev E 78(2):026611
11. Kumar G, Ray KP (2003) Broadband microstrip antenna, Artech House Inc
12. Kushwaha N, Kumar R (2013) An UWB fractal antenna with defectedground structure and swastika shape electromagnetic band gap. Prog Electromag Res B 52:383–403
13. Salonen P(1997) Dual-band E-shaped patch wearable textile antenna. Antennas and Propagation Society International Symposium C.A. Balanis, Antenna Theory, 2nd edn New York: John Wiley & Sons, Inc
14. Singh D (2000) Small H-shaped antennas for MMIC applications. IEEE Trans Antennas Propag 48(7):1134–1141

Dual Circular-Inverted L Planar Patch Antenna for Different Wireless Applications

Shivam Choudhary, Yash Sharma, Shubham Kumar, Shivam, and Chandan

Abstract This paper includes a dual operation and dual circular L planar patch antenna, which is fed by the transmission line. Since the circular L planar patch antenna is formed at C band, which is used in many satellite communication works. It is also used in probe feeding. The proposed antenna is designed FR4 & Air, i.e. we can say on a 2 layer substrate which have an area of 30 mm by10 mm. It has two frequencies, which are 2.4 GHz and 5.5 GHz. For the dual bandA (−10 dB) bandwidths having return loss of 4.3% and 4.6% are S11 characteristic respectively.

Keywords L-shape · Dual-band · Patch antenna · Antenna for wireless applications

1 Introduction

The microstrip patch antenna has a dielectric substrate, also ground plane on the different sides. We know that antennas are the soul and whole wireless communication cannot possible without the world would not have imagined the present day of technology. This portrayed an unreplacable role in today's world of wireless communication systems. It (patch) could be any convenient shape but circular and rectangular design frequently used configurations. It has a advantage of, low profile planar configuration, capability to portable with microwavei.c. technology and low fabrication costs, it is good for applications such as cell phones, radar systems, pagers, wireless communication systems and satellite communication systems [3, 17]. Parasitic radiation elements of different sizes and stacked patched antenna result in a thicker antenna [18]. There are some designs containing different shapes like

S. Choudhary (✉) · Y. Sharma · S. Kumar · Shivam · Chandan
Department of Electronics and Communication Engineering, Meerut Institute of Engineering and Technology, Meerut, India

Y. Sharma
e-mail: yash.sharma.ec.2017@miet.ac.in

Shivam
e-mail: shivam.r.ec.2017@miet.ac.in

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_68

Table 1 Dimensions of the proposed antenna (*Unit* mm) are shown above

All dimensions in mm			
L	30	W	10
L1	4	W1	2
L2	2	W2	4
L3	6	W3	8
L4	2	W4	9
L5	6	W5	3
L6	4	W6	3
L7	8	W7	8
R1	2	R2	2

rectangle, circle, triangle and L shape have been reported [12, 14, 15]. The most imp. advantages of using transmission line feeding is to simple to matched and easy to fabricate. This paper contains a dual circular-inverted L planar patch antenna for different wireless applications having size (30 * 10) mm and h = 1.6 mm, respectively.

2 Antenna Design and Structure

It consists of many parameters, which are investigated using Ansoft HFSS software version 11 (Table 1).

In the figure, we have two return losses for two different frequencies; one is 2.4 GHz having a return loss of −25 dB and 5.2 GHz having a return loss of −19 dB.

3 Results and Discussion

Figure 1 shows the fabricated prototype of our proposed antenna, which constitutes a compact size of 30 mm, 10 mm and 1.6 mm. The antenna has shown −20.55 dB return loss at 2.4 GHz and −19.17 dB return loss at 5.2 GHz resonant frequency and obtain bandwidth is 2.45 GHz and 5.2 GHz, which is shown in Fig. 2. At the resonant frequency, the antenna radiates maximum power. Antennas play a vital role in the field of wireless communications (Fig. 3).

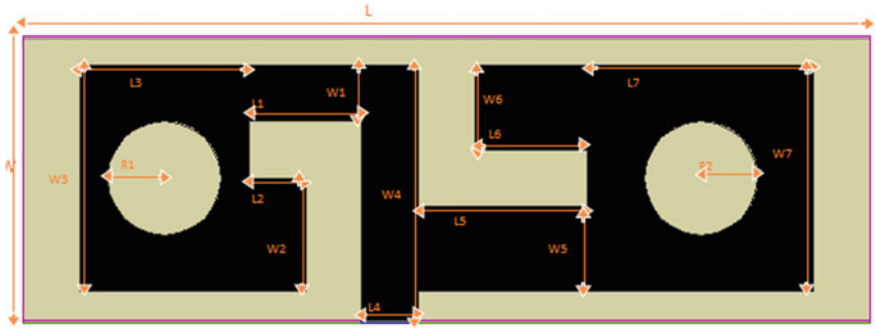


Fig. 1 Proposed antenna design

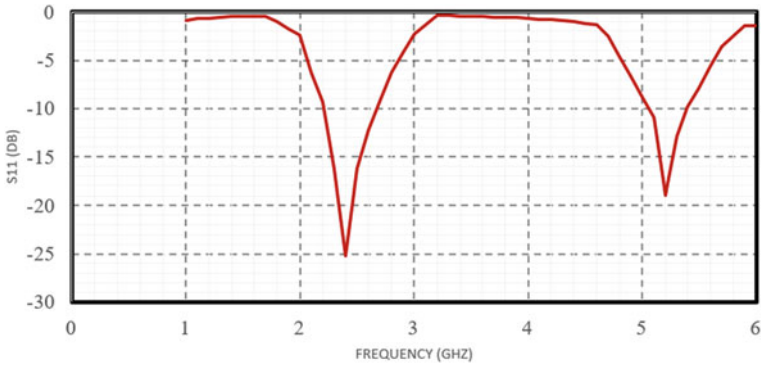


Fig. 2 Frequency versus gain curve

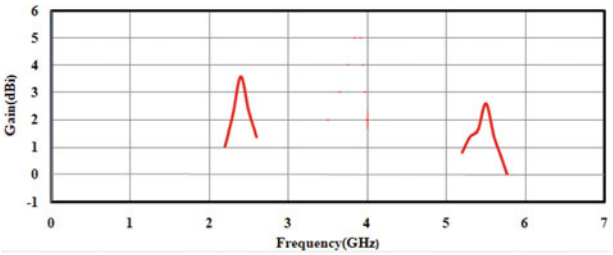
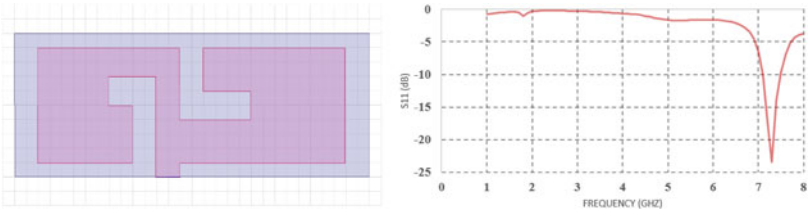
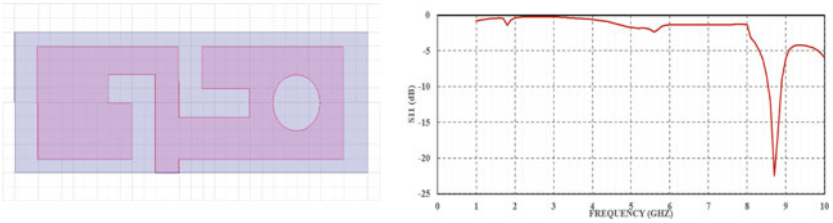


Fig. 3 Gain curve of the proposed antenna



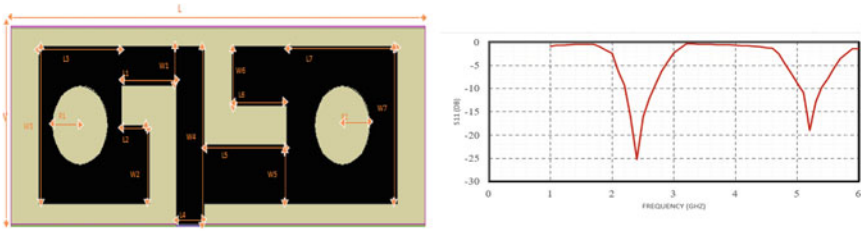
(a)

In (a), we have a design in the inverted L shape on which we get the result as shown in the figure. In the above step, we get a band only between 7 and 8 as we are working on dual band then it is useless for us to make an antenna for single band so we make some more changes to antenna design to obtain more band, which is useful to some applications.



(b)

In (b), we modify our design in order to get better bands so we just cut a circle from the above design. Now we get a single band but this time at 8–9 which is also not too good for application, so we have to work on our design further.



(c)

In (c), we cut one more circle and analyses our result. This time we get dual band at two frequencies with usable applications, one at 2.4 GHz and other at 5.2 GHz. Return loss should be less than −10 dB for an acceptable operation. The antenna has shown −20.55 dB return loss at 2.4 GHz and −19.17 dB return loss at 5.2 GHz resonant frequency and obtained bandwidth is 2.45 GHz and 5.2 GHz.

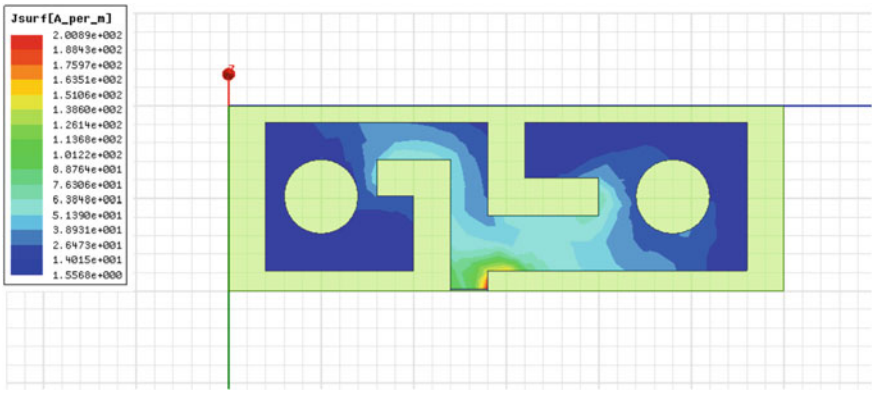


Fig. 4 Current distribution pattern at 2.4 GHz

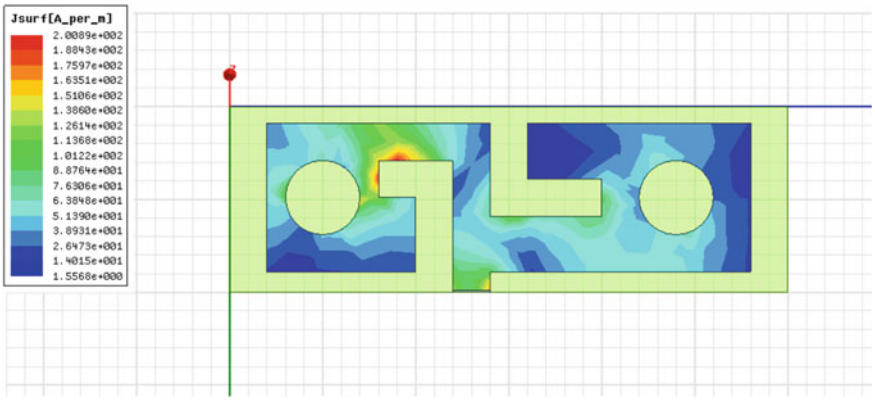


Fig. 5 Current distribution pattern at 5.2 GHz

Current Distribution Pattern at Different Frequencies

See Figs. 4, 5 and 6.

4 Conclusion

In this paper, circular inverted L planar patch antenna for different wireless applications is investigated and successfully simulated. The antenna has shown −20.55 dB return loss at 2.4 GHz and −19.17 dB return loss at 5.2 GHz resonant frequency and obtained bandwidth is 2.45 GHz and 5.2 GHz. For dual-band operation, our design shows perfect characteristics.

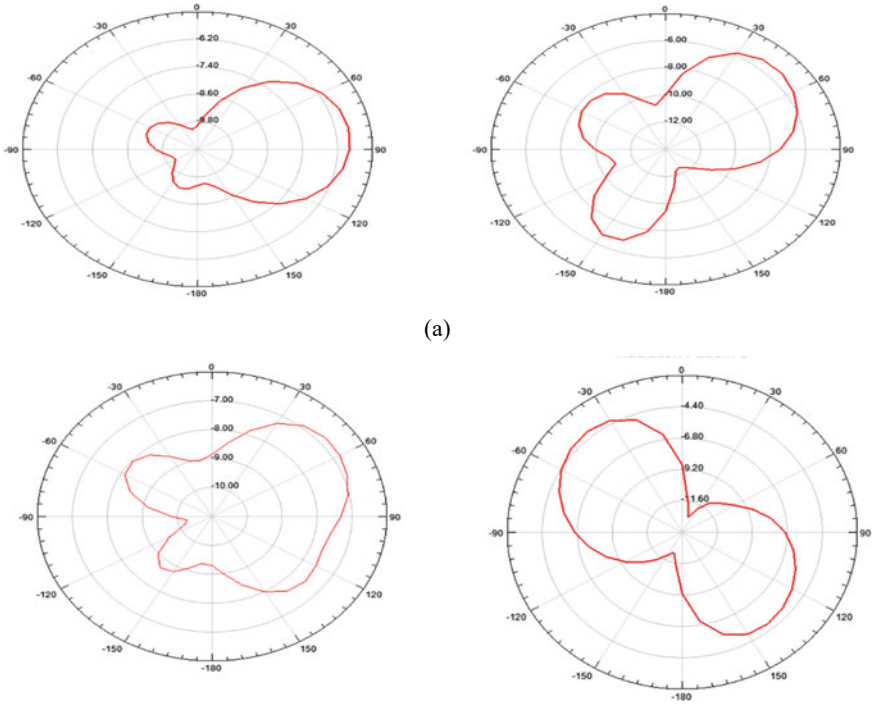


Fig. 6 **a**, **b** Radiation pattern at different frequencies

References

1. Al T, Biradar RC (2017) A triple-band highly miniaturized antenna for WiMAX/WLAN applications. Microwave Opt Technol Lett 60(2):466–471
2. Aoad A, Ufuk Türeli MS (2018) Design, simulation, and fabrication of a small size of a new spiral shaped of circular microstrip patch antenna. Microwave Opt Technol Lett 60(12):2912–2918
3. Brar RS, Saurav K et al (2018) A quad-band dual-polarized monopole antenna for GNSS/UMTS/WLAN/WiMax. Microwave Opt Technol Lett 60(3):538–545
4. Chandan, Rai BS (2014) Bandwidth enhancement of Wang shape microstrip patch antenna for wireless system. In: IEEE Fourth International Conference on communication systems and network technologies (NITTR BHOPAL), pp 11–15
5. Chandan, Rai BS (2014) Dual-band Wang shaped microstrip patch antenna for GPS and Bluetooth application. In: IEEE Sixth International Conference on computational intelligence and communication networks (Udaipur), pp 69–73
6. Chandan, Rai BS (2016) Dual-band monopole patch antenna using microstrip Fed for WiMAX and WLAN applications. In: Springer Proceedings information systems design and intelligent applications, vol 2, pp 533–539
7. Chandan (2020) Truncated ground plane multiband monopole antenna for WLAN and WiMAX applications. IETE J Res 1–6

8. Chandan, Srivastava T, Rai BS (2017) L-slotted microstrip Fed monopole antenna for triple band WLAN and WiMAX applications. In: Springer in Proceedings theory and applications, vol 516, pp 351–359
9. Chandan, Srivastava T, Rai BS (2016) Multiband monopole U-slot patch antenna with truncated ground plane. Microwave Opt Technol Lett 58(8):1949–1952
10. Chandan, Bharti GD, Bharti PK, Rai BS (2018) Miniaturized Pi (π)—slit monopole antenna for 2.4/5.2.8 applications. In: AIP Conference Proceedings, American Institute of Physics, pp 200351–200356
11. Chandan, Bharti GD, Srivastava T, Rai S (2018) Dual band monopole antenna for WLAN 2.4/5.2/5.8 with truncated ground. In: AIP Conference Proceedings, American Institute of Physics, pp 200361–200366
12. Deng C, Lv X, Feng Z (2018) Wideband dual-mode patch antenna with compact CPW feeding network for pattern diversity application. IEEE Trans Antennas Propag 2628–2633
13. Di Carlo CA, Di Donato L et al (2017) A circularly polarized wideband high gain patch antenna for wireless power transfer. Microwave Opt Technol Lett 60(3):620–625
14. Feng S, Zhang Li et al (2019) A compact broadband differential-fed microstrip patch antenna with 5.8 GHz WLAN band-notched under quad-mode resonance. Microwave Opt Technol Lett 62(4):1716–1723
15. Maurya RK, Kanaujia BK et al (2019) Circularly polarized hexagonal ring microstrip patch antenna with asymmetrical feed and DGS. Microwave Opt Technol Lett 62(4):1702–1708
16. Mbinack C (2018) Dual-band microstrip-fed SQUARE ring antenna input and output performances analysis for Wi-Fi application. Microwave Opt Technol Lett 61(4):1011–1015
17. Shen Voon C, Yeap KH et al (2017) A compact double-Psi-shaped dual band patch antenna For WLAN/LTE applications. Microwave Opt Technol Lett 60(5):1271–1275
18. Tu ZH, Wang YX, Nie N (2019) Low-profile circularly polarized microstrip patch antenna with pattern diversity. Microwave Opt Technol Lett 62(4):1570–1576

Effective Application and Sustainability of Lean Six Sigma to Improve Process Performance

Md. Ehsan Asgar, Rajeev Ranjan, Shrey Atre, Rahul Bhandari, and Saba Zaidi

Abstract Six Sigma is a technique devised by Bill Smith of Motorola in 1985, operates with the help of numerous statistical, data-based tools and techniques in order to yield an appropriate result that substantially increases profits and ensures delivery of value by deteriorating the variation and eradication of defects. With global economy facing simultaneous slowdowns in the past few years, organizations are keen on finding an opportunity to curb the costs and lead the competition. Thus, a procedure known as Lean Six Sigma is used, that can aid in achieving such economic desires. Lean Six Sigma has been the one to show significant cost reduction by increasing the rate of material flow and information, making the process more efficient, and improves the grade of the product that is composed by the process. Lean Six Sigma has successfully been implemented by organizations to significantly improve a variety of processes in many different domains. The intent of this review paper is to link Lean Six Sigma with certain sustainability factors and sustainability itself.

Keywords Process improvement techniques · Process performance · Six sigma · Sustainability · Lean six sigma · LSS · CI

1 Introduction

Manufacturing companies always seek ways and means to cut on wasteful expenditure right from the Resourcing to finishing -goods stage and for this endeavor is maximum profit with greater satisfaction in the excellence of the product. Innovative technology, competitiveness helps in managing operations as per social and environmental norms [1]. Therefore, when manufacturing the management to find true outcomes must balance different elements in a combination so that the outcome

Md. E. Asgar · R. Ranjan · S. Atre (✉) · R. Bhandari
HMR Institute of Technology and Management, New Delhi-110036, India

S. Zaidi
Learnet Skills Limited, New Delhi-110020, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_69

emerges knowledgeably [2]. Therefore, manufacturing companies require the capability of adjusting to the demand and supply of the market. For this, a keen eye is to be kept on the supply line and backup plan to regulate the same both ways [3]. It has been found that those companies who implemented lean manufacturing or the lean production process have experienced greater benefits and results. Many use the technique of Six Sigma method aiming to remove defects thereby not only compromising outcomes but also helps in three different ways the applications are helpful in cutting costs. Increasing competition in the company that how they can increase their profits at the same time not compromising on the quality control, companies that implemented mainly manufacturing were to be seen having gained a little less, which is a combination of Lean manufacturing along with Six Sigma. In the final stage, the implemented tools of Lean Six Sigma are effective in helping manufacturing and the production sector. The supervisors and technicians work out schedules and analyze what problems they face, therefore next, they start troubleshooting, which helps in complex work environments to get the true output, which is both effective and value-added [3].

2 Terminology

One must lay the groundwork to understand practical applications and the terms for understanding the terminology Lean Six Sigma are important. Lean applications involve maximum people so that the goals are achieved, and they can be implemented in many different contexts. This Six Sigma improves productivity through process design and fusion occurs, which achieves a statistical control of the process improving the speed at which the work is done [4]. But as per Lindeman, Six Sigma lacks theoretical underpinning. To improve the process and achieve high levels, we must go into the root cause of defects in order to achieve product availability [5].

One can say that this algorithm of Six Sigma has five phases [6]:

1. Defining
2. Measuring
3. Analyzing
4. Improving
5. Controlling.

One problem is to define a particular tool that is to be employed to measure the scale of the problem in front; what causes it and how it can be removed that is all the possible remedies and measures. As from research point of view, it is believed that this approach should be expanded and studied to include environmental safety, integrity and social responsibility of the job in the organization toward the country, society and toward their own conscious [7]. Concluding with the above mentioned, stages for foolproof drive and change in IT form; it is advised to engage all stakeholders into the system so that everyone is involved at a constant pace for effective change and smooth operation of the work [8]. Therefore, using this methodology and business

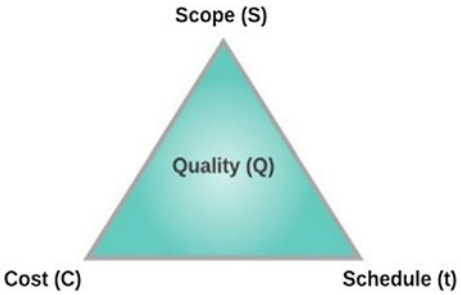


Fig. 1 Pillars of Six Sigma [10]

strategy to head toward a competitive global market, the tools and techniques need to be efficiently applied [9]. Therefore, to maximize the results and minimize the delays, we should see the scope, check on the cost and work out the schedule to achieve desired quality (Fig. 1).

2.1 Defining

In the earliest stages of the process, defining is a crucial factor. Whether it is the defining of the problem, defining of the goals, defining is the foundation of the process.

Develop a Problem Statement

- **Determine the process that is causing problems**
  This stage requires the team to have access to the existing data regarding the ongoing problem. They should be able to determine and make a logical decision about the process that might be the cause of the problem.
- **Confirmation of the Priority and Impact of the problem**
  Now the team is required to create a problem statement, which includes:
  - **Severity**: Severity of the problem is to be observed and noted. Specific data are required but in case it is not available right away, the team might fill the blanks later, perhaps during the phase.
- **Confirm resources are available**
  Are people receptive to the issue and can dedicate time towards solving the issue? It is critical to involve a Black Belt or a Team leader in such crucial situations.

2.2 Measure

Measure means to weigh the expectations of the clients, the shareholders and the most importantly the customers. This can be done by studying the competitor analysis.

2.3 Analyze

The assumption about the cause of the problem is one of the first steps taken while initiating the project. If these notions have remained unreviewed, a "jump to conclusion" situation might occur among the team, even without realizing the principal issue underlying. Unregulated root causes prompt various issues including poor time management, wastage of resources and bewilderment. The ideal way is to study the entire project and data to get to the root cause with ease. Data analysis, process observations and relevant tests are some of the methods of verification. It is crucial to be very coherent about the sources of problems in the process of the project before considering the solutions. The Analyze phase is dependent on such processes and verification.

- **Graphical Representation of the Data for in-depth analysis**
  The team is required to present the data collected using graphs and charts, so that the problem as well as the overall improvement after the implementation of the solution is evident. The graphical representation eases out the conversations and makes way for a clear understanding between different departments as well. It is pertinent to select the right charts and graphs that provide everyone with valuable inputs about the causes of process issues and the clear picture of any kind of improvement made after the implementation of the solutions. Below are the graphs showing the before and after scenarios of implementation of LSS by a company producing "Bolts" in accordance with the order they received. An evident improvement can be observed with these graphs. These charts are used by various industries be that Education, Food, Aerospace or Automobile, around the world to successfully implement the LSS and demonstrate the results (Figs. 2 and 3).
- **Verify the Cause(s) of the Problem**
  Before the Improve phase, the data are used by the team to analyze, process and compare in order to sniff out the root cause of any problem or situation. This can be done merely by watching or conducting hypothesis tests.

2.4 Improve

Improve phase consists of brainstorming for the right solution, regulate and initiate changes if they are required in the process, implementation of solutions and collection of data to affirm measurable improvement. A logical and continuous improvement is the basis of the innovation and provision of elegant solutions, thus, enhancing the customer experience. Following are the steps to be considered in the improve phase:

- Brainstorm solutions that might fix the problem
- Select the best and practical solution(s)
- Implement the solution(s).

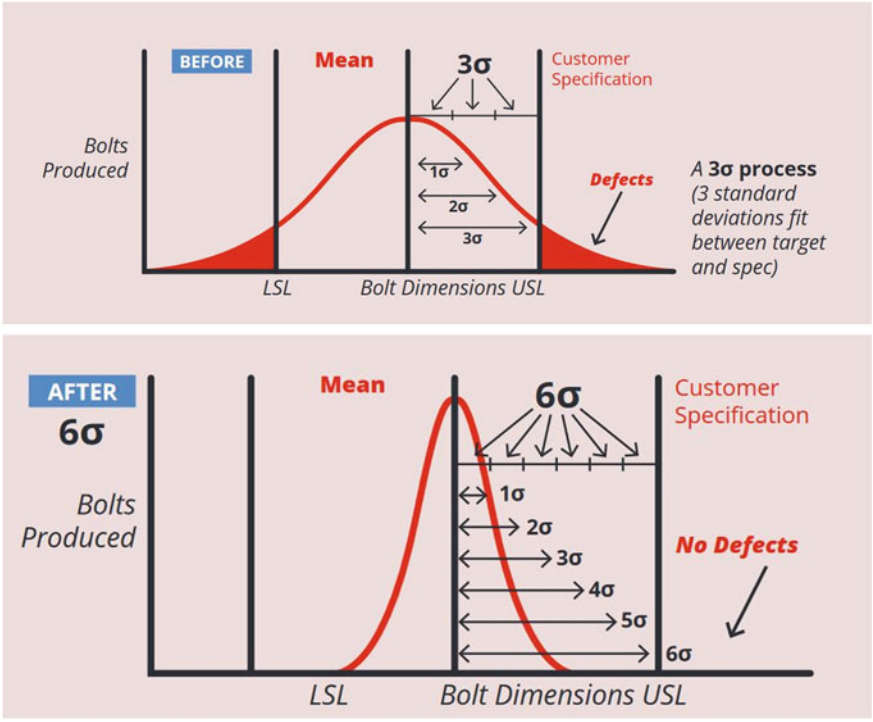


Fig. 2 Graphical representation of before and after scenarios of implementing lean Six Sigma with in-depth analysis [33]

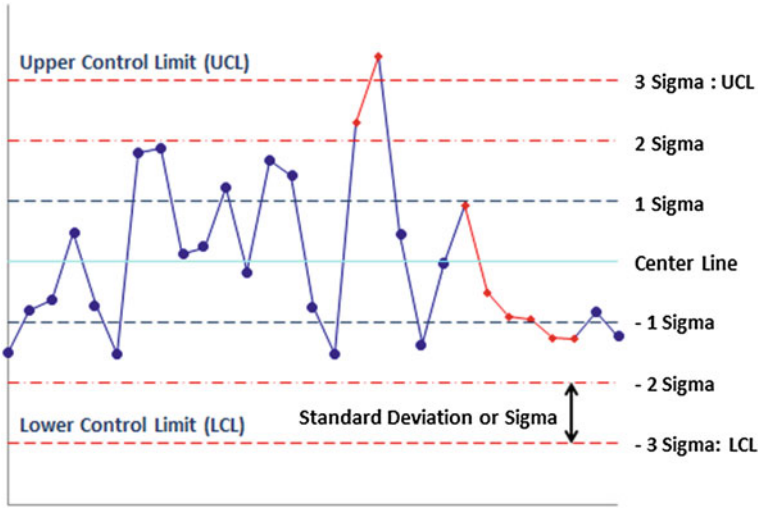


Fig. 3 Lean Six Sigma control chart showing limits and stability analysis [34]

2.5 Control

During the Control phase of the analysis, the whole team is focused on devising a plan to estimate the accomplishments of the reorganized procedure and building up a resource plan in the event that there is a plummet in execution. Once completed, these plans are handed over to the process owner to maintain continuous upkeep. The process owner then makes sure that the process is continuously monitored and maintained. This results in control on flow of information and the accuracy of the process, hence, allowing accurate repeatability.

3 The Methodology of the Research

The primary intent of this research is to recognize and link the "Lean Six Sigma" with that of sustainability so that the issues and the opportunities could be very well-typed for productive outcomes [11]. The best way is to identify content and all crucial steps, which should be taken to discover and to conceptualize the detailed study. This involves four stages, which are defining units, classifying them, the material evaluation along quality publication followed by delimiting in a way in order to have a structured and effective literature review [12]. The next criteria are all about customer satisfaction, improved quality, improved productivity, increased profit and employee involvement along with reduced turnover rate [13]. Now the priority rates that are set after the decision-making technique are subjected to alternative reasons for the achievement of goals [14].

4 Augmentation of Lean Six Sigma

Augmentation of Lean Six Sigma can be identified by the following identifiers [15]:

(1) Signals and Alerts in order to divert attention from task.
(2) Improve the decision-making through experience and learning.
(3) Feedback forgets cities development.
(4) Communication for efficient work.

The methods that provide significant importance to the dynamics and activities of cognitive work are to be considered. The sustainability and the urgent need of Six Sigma in the need of research field, more and more organizations are required to outsource things to make operations sustainable [16]. It can be observed that on the research of integration of Lean Six Sigma, it is presumably to be found, that management focuses on Lean Six Sigma and sustainability to have a better place in all geographical areas of applications concerned [17].

5 Sustainability of Lean Six Sigma

Sustainable in general business terms might cite to the safeguard of the attributes and all the resources that make an organization capable of standing toe to toe and even surpassing their rival organizations in the same industry. Thus, it can be said that sustainability might assume the number of definitions depending upon the perspective of the viewer. It can be observed as to mean endurable, tenable, tolerable, acceptable, plausible, justifiable and debatable. Bhasin [18]. Initially, it was only a representation of the environmental concerns. Now, we can see that the word sustainable has taken a broader shape and its incorporation might not only be limited to a specific zone. Thus, designating sustainability as an incorporation of social and economic issues; making it a composite of social, economic and environmental factors [19]. An organization is ready to commence changes toward economic surge, social advancement and environmental protection continuously in order to grasp a better future and entwine such changes eventually as the part of the organizational policy is worthy of being titled sustainable. Burke and Gaughran [20] When observed through the lens of economics, sustainability is the continuous addition of value and creation, which falls in perfect alignment with the principles of Six Sigma [21]. It involves reduced ramification, improved veracity and better effectiveness throughout the business process [22].

5.1 Sustainability Factors for Lean Six Sigma

The application of a sustainable strategy is dependent upon the number of critical factors [23, 24]. Both the internal and external factors make up for this implementation. The internal determiners are corporate governance and stakeholder engagement. The social and cultural factors alongside the legal system in the country constitute the external factor [25]. A few of them are mentioned below [26]:

- **Continuous Improvement (CI) culture**: Deming defined the continuous improvement philosophy as a "Consistent improvements that increase success and reduce failures [27]". In this, Bessant added "a company-wide process of focused and continuous incremental innovation [28]." When observed from the perspective of Six Sigma, this is a culture, which is the reflection of sustained improvement with the primary objective of waste reduction in every possible business process within the organization. The continuous effort is all about searching root problems, sources of variation and waste alongside ways to eliminate them in the long run. It has to be something done by each employee day in day out.
- **Innovative culture**: While CI is the source for sustainability in LSS, innovation makes up for the competitiveness in LSS. Juergensen [29] Various studies conducted in this direction have shown that new improved and innovative culture strengthens the competitiveness of the firm and the employees. It is only through creativity and innovativeness that managers sometimes find success with the most

unorthodox and out-of-the-book methods in the field of company management, which lays the foundation for future success [30].

- **LSS knowledge of the Employees**: Employees' knowledge about LSS and sustainability is definitely a crucial factor. This naturally involves the knowledge of DMAIC. As the firm progress, most of the members should be inclined or rather made to achieve a Green belt in Six Sigma at least. The number of Black Belts and other lower belts should constantly be on the increasing side of the scale. This will better the overall working and manufacturing structure of the firm. Hence, continuous employee training in order to develop a better understanding of them is quite essential [31].
- **Communication**: Naturally, communication plays an important role for a better performance altogether in the firm and naturally in the active implementation of Lean Six Sigma. In several publications, it is emphasized that efficient and frequent communication provides the momentum in LSS continuation [26]. The LLS sustainability requires effective top-down communication in order to provide employees with clear objectives and consistent mission statements [32]. Just as discussed in the first phase of Six Sigma brainstorming is a part of communication. Therefore, an effective communication is essential when it comes to correct Lean Six Sigma implementation.

6 Conclusion

Growing organizational interest in Six Sigma has seen quite a tremendous growth in recent years of the Industrial Calendar. Project-driven organizations are leaning more and more toward Lean Six Sigma day by day. Multiple factors are responsible for influencing six sigma projects; few of them being greater involvement of management and commitment toward the organization, control skills, effective cultural changes and perpetual trainings. The core of the better and successful implementation of this method lies in understanding the topic first hand. Numerous approaches have been available at hand to increase the overall performance of divergent business sectors and strategies. Lean Six Sigma is something, which relies heavily on rigorous, data-driven, result-based approach in order to reach business process improvement. It uses strategies evolved in the lap of giants of the market such as Motorola, GE, Toyota and Xerox to name a few. Integration of such strategies to make tools a powerful engine is created to improve quality, efficiency and speed. It also is a crucial factor to ensure minimum wastage and thus creating the overall sustainability factor of the process pretty high. Lean Six Sigma has proven to be one of the most potent tools to improve overall proficiency and quality. One of the major drawbacks of this method is to maintain its sustainability factor. If the project is only dependent upon the support of management to move forward, then staff lower in the hierarchy may not be able to propagate it further, or at least they cannot be trusted with it. Focus and support for LSS are the essential factors that can very well define its sustainability.

References

1. Wong WP, Wong KY (2014) Synergizing an ecosphere of lean for sustainable operations. J Clean Prod 85(15):51–66
2. Klein GA (1999) Sources of power: how people make decisions. MIT
3. Barley SR, Orr JE (1997) Between craft and science: technical work in US settings, Cornell University Press
4. Hines P, Holweg M, Rich N (2004) Learning to evolve: a review of contemporary lean thinking. Int J Oper Prod Manage 1;24 (10):994–1011 and George ML. Lean Six Sigma for service. Mc Graw-Hill; 2003
5. Linderman K, Schroeder RG, Zaheer S, Choo AS (2003) Six Sigma: a goal theoretic perspective. J Oper Manag 21:193–203
6. Bergman B, Klefsjö B (2003) Quality from customer needs to customer satisfaction. Student litteratur, Lund, Sweden
7. Kuei CH, Madu CN (2003) Customer-centric six sigma quality and reliability management. Int J Qual Reliab Manag 20(8):954–964
8. Woods DD, Hollnagel E (2006) Joint cognitive systems: patterns in cognitive systems engineering. CRC Press
9. Snee RD (2010) Lean Six Sigmae getting better all the time. Int J Lean Six Sigma 1(1):9–29
10. https://www.projectsmart.co.uk/understanding-the-project-management-triple-constraint.php
11. Easterby-Smith M, Thorpe R, Lowe A (2002) Management research: an introduction. Sage Publications, London
12. Mayring P (2003) Qualitative inhaltanalyse—Grundlagen und Techniken. Qualitative content analysis, (8th edn) Beltz Verlag, Weinheim, Germany
13. Alhuraish I, Robledo C, Kobhi A (2016) Assessment of Lean manufacturing and Six Sigma operation with decision making based on the analytic hierachy
14. Dezert J, Tacnet JM, Batton-Hubert M, Smarandache F (2010) Multi-criteria decision making based on DSmT-AHP. Workshop on the theory of belief functions. Belief Functions and Applications Society (BFAS). France
15. Sterman JD (2000) Business dynamics: systems thinking and modelling for a complex world
16. Garza-Reyes JA (2015) Lean and green e a systematic review of the state of the art literature. J Clean Prod 102:18–29
17. Garza-Reyes JA (2015) Green lean and the need for Six Sigma. Int J Lean Six Sigma 6(3):226–248
18. Bhasin S (2011) Measuring the leanness of an organisation. Int J Lean Six Sigma 2:55–74
19. Brundtland GH (1985) World commission on environment and development. Environ Policy Law 14:26–30
20. Burke S, Gaughran WF (2007) Developing a framework for sustainability management in engineering SMEs. Rob Comp-Integr Manuf 23:696–703
21. Carroll AB (1999) Corporate social responsibility evolution of a definitional construct. Bus Soc 38:268–295
22. Chaharbaghi K, Lynch R (1999) Sustainable competitive advantage: towards a dynamic resource-based strategy. Manage Decision 37:45–50
23. Gelei A, Losonci D, Matyusz Z (2015) Lean production and leadership attributes-the case of Hungarian production managers. J Manuf Technol Manage 26:477–500
24. Hult GTM, Hurley RF, Knight GA (2004) Innovativeness: its antecedents and impact on business performance. Ind Market Manage 33:429–438
25. Jackson A, Boswell K, Davis D (2011) Sustainability and triple bottom line reporting-What is it all about. Int J Bus Humanities Technol 1:55–59
26. Kowang TO, Yong TS, Rasli A, Long CS (2016) Lean Six Sigma sustainability framework: a case study on an automotive company, 9(5):279–283
27. Jeyaraman K, Teo LK (2010) A conceptual framework for critical success factors of lean six sigma: implementation on the performance of electronic manufacturing service industry. Int J Lean Six Sigma 1:191–215

28. Bessant J (2005) Enabling continuous and discontinuous innovation: learning from the private sector. Public Money Manage 25:35–42
29. Juergensen T (2000) Continuous improvement: mindsets, capability, process, tools and results. Juergensen Consulting Group Inc, Indianapolis Kimber D, Lipton P (2005) Corporate governance and business ethics in the Asia-Pacific region. Business Soc 44:178–210
30. Robert BP (2008) Management systems: the bedrock for building a sustainable business. GreenBiz Group
31. Welford R (2005) Corporate social responsibility in Europe, North America and Asia. J. Corporate Citizenship 33–52
32. Wu CW, Chen CL (2006) An integrated structural model toward successful continuous improvement activity. Technovation 26:697–770
33. https://goodelearning.com/subject-areas/lean-six-sigma
34. https://www.qimacros.com/control-chart/stability-analysis-control-chart-rules/

Determination of Apple, Lemon, and Banana Ripening Stages Using Electronic Nose and Image Processing

S. Suthagar, K. S. Tamilselvan, M. Priyadharshini, and B. Nihila

Abstract The objective of this paper is to determine the ripening stages of fruits like the unripe stage, ripeness stage, and overripe stage by using electronic nose and image processing. The electronic nose consists of an array of sensors such as MQ-3, MQ-6, MQ-8, MQ-135 was placed inside the fruit chamber to detect the various odors emitted from the fruits. Each sensor is capable of detecting various gases that were emitted from apple, lemon, and banana. Based upon the value detected by these sensors, ripening stages of these fruit samples were done. Parallel to this image processing technique was done to determine the ripening stages of these fruit samples by their color and shape. Finally, the results obtained by electronic nose and image processing techniques were compared along with manual verification by the human being to improve the accuracy of the ripening stages of apple, lemon, and banana.

Keywords Ripening stages · Electronic nose · Image processing · Sensors · Fruits

1 Introduction

Industrial and agricultural field plays an important role in the economic development of any country. Fruit producers should know the right time for harvesting fruit and at the same time, it should reach the customer before it gets overripe. Fruit's maturity stage is the main factor that affects the fruit quality during ripening and marketability after ripening. Identifying the fruit maturity phase will help the farmers to avoid early harvesting, which means harvesting unripe fruit. It will avoid unwanted transportation of unripe and overripe fruit to save money and time. By proper determination of the fruits, the ripening stage will not allow any wastage of fruit but at the same time, it

S. Suthagar (✉) · M. Priyadharshini · B. Nihila
Department of Electronics and Communication Engineering, Kongu Engineering College, Erode, India

K. S. Tamilselvan
Department of Electronics and Communication Engineering, KPR Institute of Engineering and Technology, Coimbatore, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_70

ensures good quality of fruits reaches the market for customers. This might surely help farmers to prevent the unwanted pre-harvest and post-harvest fruits.

2 Literature Review

Prabha [8] Image method technique is to search out the ripening stage like an unripe stage, ripe stage, overripe stage of banana fruit by its color, and size price of their footage precisely. Throughout this paper, the mean color intensity from the bar chart, area, perimeter, major axis length, and minor axis length from the size values was extracted from the activity image. Analysis of variance between each maturity stage on these choices indicated that the mean color intensity and house feature were tons of important in predicting the ripening stage of banana fruit. Their unit of measurement two classifier rules notably mean color intensity recursive rule and house formula were developed and their accuracy on maturity detection was assessed. Hence, this method helps the farmers at the right time of harvest by avoiding pre-harvest of fruit.

Devalatkar and Koli [4] It develops a technique not only to speed up the processing time but also minimizes error in the finding of the age factor of fruits. This proposes an effective fruit grading system based on computer vision techniques. It also determines the defective or non-defective and unripe or ripe or overripe stage of the fruit.

Arakeri [1] It consists of a group of sensors to recognize the odor emitted from fruits. By this, they can eliminate or reduce the lack of quality by using an ethylene gas sensor that can detect fruit maturation. It is also used for continuous monitoring of the freshness of any fruit at the time of sale.

Raut and Bora [9] An electronic nose that is wireless for identifying the ripening stages of fruits. Zigbee, an electronic nose, which is a group of sensors, LabVIEW GUI, and fruits are the sensing system for this project. LabVIEW stand-alone application is for data analysis and principal component analysis-based data classification technique is used to determine the performance of an electronic nose.

Kaur and Gupta [7] It develops a non-destructive instrument that is used for the classification of various fruits. This consists of an array of metal oxide semiconductor gas sensors. The project is based on the response of a sensor array to different VOC emitted from the fruit during its ripening stages. These are analyzed by using PCA and this method is also used for grading and classification of different climacteric fruits, which depends on the ripening stages of fruits.

3 Methodology

The ripening stages of banana, lemon, and apple were determined and classified by using image processing techniques and electronic nose. The result obtained by these two processes was compared to determine the ripening stages. This comparison will

give us a better result. The process to determine the ripening stages is explained below.

3.1 Determining the Ripening Stages Using Image Processing

The objective of this work is to determine the ripening stages (unripe, ripe, over-ripe) of fruits samples of banana, lemon, and apple. For this, the image processing technique is used to classify the numerous ripening stages of fruits samples by mean color intensity. There are several stages involved in the determining of fruit ripeness stages. They are image acquisition, edge detection, RGB value, and classification supported maximum color value. A block diagram of classification for fruit ripening stages is shown in Fig. 1.

3.1.1 Image Acquisition

The acquisition is that the initiative involved the method of classifying the fruits ripening stages. The fruit image is captured by using a camera. This captured fruit image is given as input for this classification process.

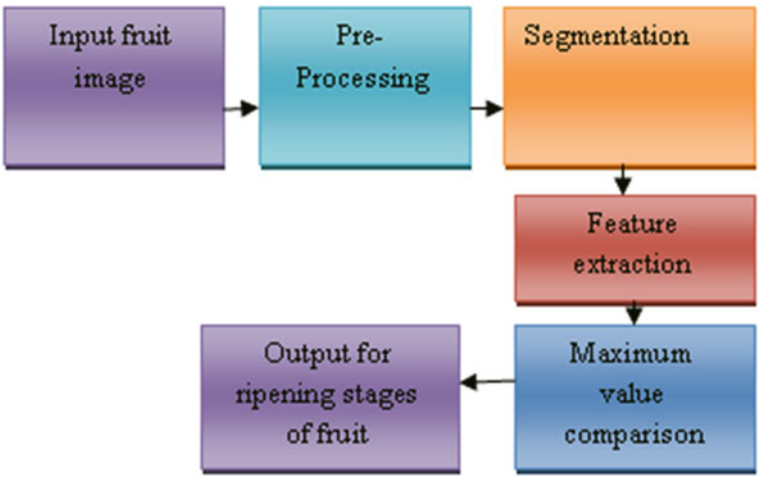


Fig. 1 Block diagram of classification for fruit ripening stages

3.1.2 Edge Detection

Edge detection is one of the image segmentation techniques. This may determine the lines or edges present in an input fruit image. Suthagar et al. [11] The important use of edge detection is to simplify the image data to attenuate the data that are to be processed. There are alternative ways to try edge detection. Most methods are Laplacian edge detection and gradient edge detection. The acceptable sort of edge detection is to be chosen on the idea of requirements for vertical, horizontal tangential edges present within the input image. Sobel filter is employed for this classification work.

3.1.3 RGB Value and Classification

RGB is the format of any color image. The input fruit image is represented with three matrices of sizes matching the image format. Suthagar et al. [10] These matrices correspond to one of the red, green, blue. This RGB value is obtained by using an image processing tool. Then, the maximum color value in an input image can also be obtained. Based upon the maximum color intensity value of an input fruit image, the ripening stages of fruits can be classified.

3.2 Classification Using Electronic Nose

This process comprises of electronic nose, fruit chamber, Arduino board, and LCD. After all technical initiation of the electronic nose, the fruit is placed inside the fruit chamber. Geethapriya and Marypraveena [5] Fruit chamber is tightly closed with an array of sensors inside. The sensor array consists of MQ 3 sensor, MQ 6 sensor, MQ 8 sensor, and MQ 135 sensor, which senses the gases emitted from the fruits. Each sensor has the capability of sensing gases in particular ranges. The processor takes the input from the sensors and it displays the value in LCD. The block diagram of classification for fruit ripening stages is shown in Fig. 2.

3.2.1 Array of Sensors

MQ3 sensor is suited for sensing or detecting the gases of ethanol with a standard detection range is 0.1–100 ppm and it has a fast response and high sensitivity. It also has a stable and long life. Kanade and Shaligram [6] MQ6 sensor detects the gases of alkanes (propane, butane), and olefins with a standard detection range are 0.3–50 ppm. It also has a fast response, a stable and long life. MQ8 sensor is suited for detecting the gases of hydrogen with a standard detection range is 0–3 ppm. Chen et al. [3] It also has a stable and long life at low cost. MQ135 sensor detects

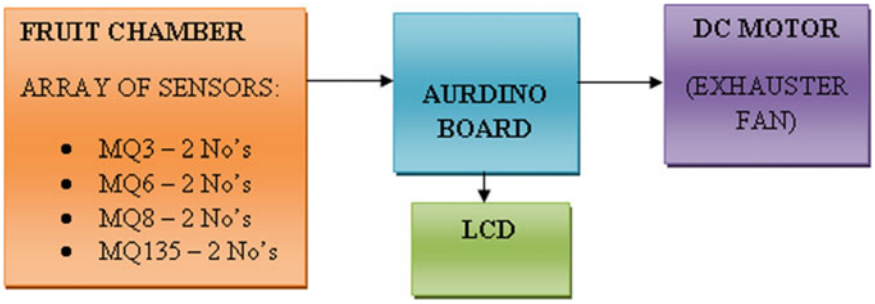


Fig. 2 Electronic nose for classification of fruit ripening stages

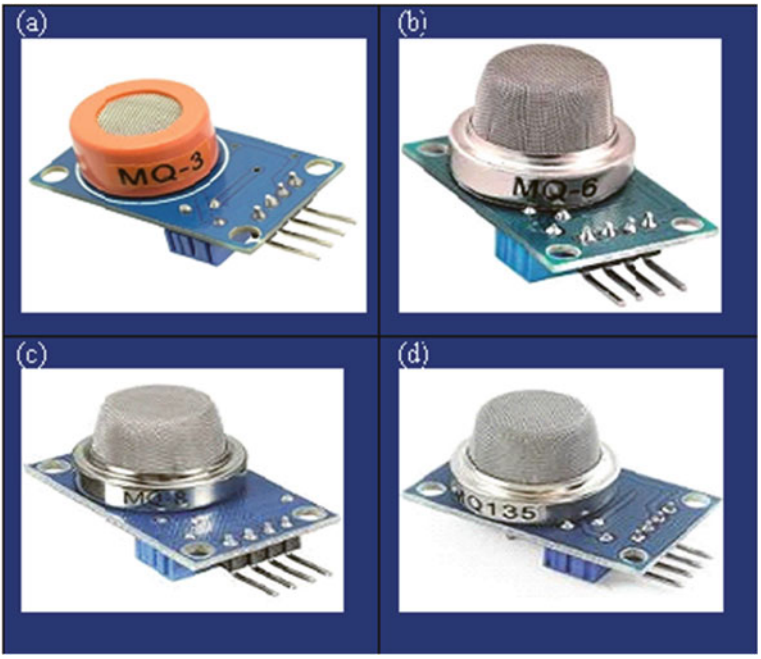


Fig. 3 Shows **a** MQ3, **b** MQ6, **c** MQ8, and **d** MQ135 sensors

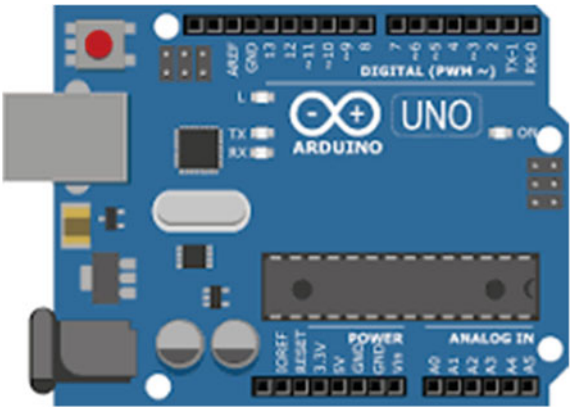
the gases of ammonia, sulfides with standard detection is 0.01–100 ppm. It also has a fast response and high sensitivity stable (Fig. 3).

The target gas, concentration, and operating temperature of MQ3, MQ6, MQ8, and MQ135 sensors are shown in Table 1.

Table 1 Specifications of MQ3, MQ6, MQ8, and MQ135 Sensors

Sensors	Target gases	Concentration	Operating temperature
MQ-3	Alcohol, ethanol, and smoke	(0.1–100) ppm	10 °C–0 °C
MQ-6	LPG, butane, propane, LNG	(0.3–50) ppm	−10 °C–50 °C
MQ-8	Hydrogen	(0–3)ppm	−10 °C–50 °C
MQ-135	NH_3, NOx, alcohol, benzene, smoke, CO2	(0.01–100) ppm	−10 °C–45 °C

Fig. 4 Arduino board



3.2.2 Arduino Board

An Arduino board acts as a microcontroller kit. The Arduino pin diagram is shown in Fig. 4. The operation voltage of the Arduino board is 5 V with 14 digital input/output pins and 6 analog input pins. Each digital input pin takes 40 mA and the output pins provide 40 mA DC.

Each analog input/output pin provides 10-bit resolutions. The electronic nose has been interfaced with this Arduino board. Arduino board takes the analog output from the odor sensor and provides the display in LCD.

3.2.3 Operating Procedure of Electronic Nose

The proposed electronic nose system was tested with the odors of three fruits, namely, lemon, banana, and apple. The odors were prepared by placing samples of fruit in a fruit chamber sealed with a membrane. The operational procedures were as follows:

- The fruit chamber was opened and the exhauster fan was ON to prepare the chamber to place fruits. The exhauster fan is ON state to pump the odor present inside the fruit chamber.

- Once the fruit is placed inside the chamber it was closed, and the sensor's resistance was given 60 s to reach a steady state. The classification result obtained by sensors appeared on the LCD.
- The fruit that is paced inside the chamber was removed after taking the reading. After this, empty chamber is kept opened and the exhauster fan is made to ON state for 60 s.
- After 60 s, another fruit is placed inside the chamber and closed tightly. This process continued for various fruits and reading was noted. The sensors and the exhauster fan made by DC motor both fit into the fruit chamber and the transparent box, which contains the interface circuitry and the Arduino board.

4 Results and Discussion

4.1 Classification Using Image Processing

The results obtained by the image processing technique which determines the fruit and classifies the ripening stages of various fruits by their shape and color are shown in Fig. 5. Figure 5a, c, and e represents the different ripening stages of banana fruit on various days, and Fig. 5b, d, and f shows the ripening stages of an input banana on various days.

In this system, the ripening stages of various fruits can be classified. Figure 6a,

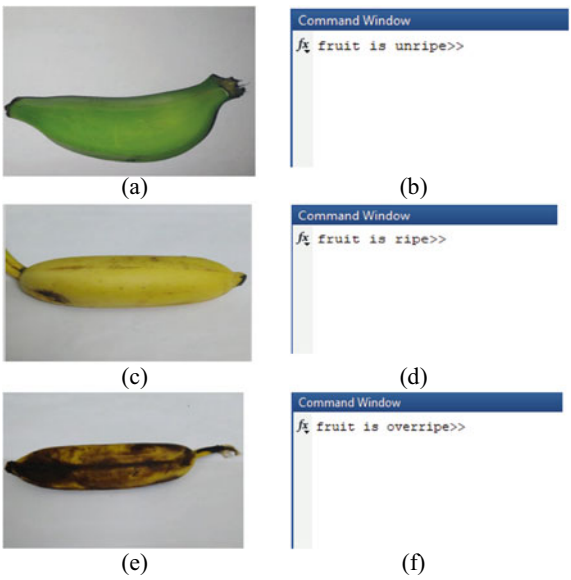


Fig. 5 **a**, **c**, **e** are the input images and **b** **d**, **f** are the corresponding results

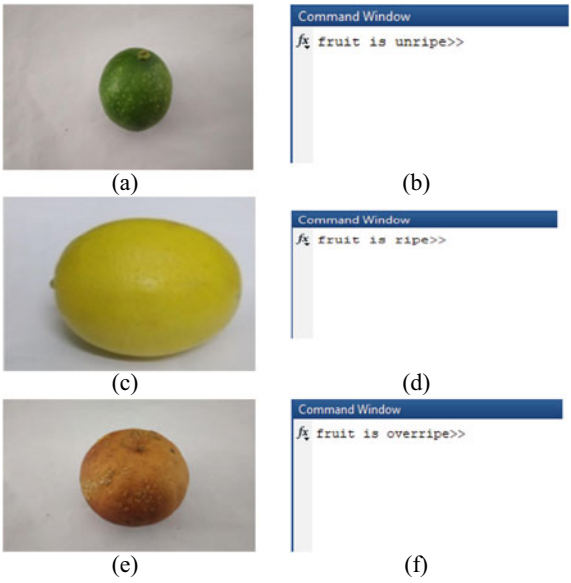


Fig. 6 **a**, **c**, **e** are the input images and **b d**, **f** are the corresponding results

c, and e represents the different stages of lemon fruit images, and Fig. 6b, d, and f shows the corresponding ripeness stages of an input lemon fruit, which are given below.

Figure 7a, c, and e represents the different stages of apple fruit images, and Fig. 6b, d, and f shows the corresponding ripeness stages of an input apple fruit, which are given below

4.2 Classification Using Electronic Nose

The ripening stages of various fruits were determined and classified by using an electronic nose. Figure 8 represents the experimental setup of an electronic nose system. Figure 8a, b, and c represents the experimental setup of an electronic nose system for apple lemon and banana fruit.

Unripe banana, lemon, and apple are taken for testing by using an electronic nose, which consists of MQ3, MQ6, MQ8, and MQ135 sensors. Everyday sensor reading is monitored at 6.00 am, 12.00 pm, and 10.00 pm for all three fruits. The average value of each sensor is treated as the day reading of each sensor for that respective fruit. The day-wise and time-wise readings taken from different sensors for three fruits are given in Table 1. It was observed that banana fruit is unripe for the value of the MQ3 sensor lies between 0.0116 ppm and 0.0139 ppm, the MQ6 sensor value lies between 0.6281 ppm and 0.4954 ppm, MQ8 sensor value lies between 0.0060 ppm and 0.0129 and the MQ135 sensor value lies between 0.3288 ppm and 0.3372 ppm. It was also observed that lemon fruit is unripe for the value of MQ3 sensor lies

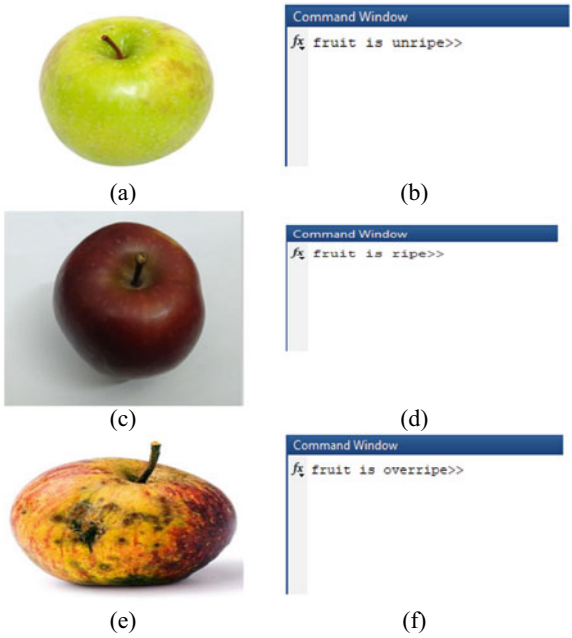


Fig. 7 **a**, **c**, **e** are the input images and **b d**, **f** are the corresponding results

between 0.0282 ppm and 0.0307 ppm, MQ6 sensor value lies between 0.5484 ppm and 0.6120 ppm, MQ8 sensor value lies between 0.0060 ppm and 0.0370 and MQ135 sensor value lies between 0.3976 ppm and 0.4550 ppm. Similarly, it was observed that for apple fruit is unripe for the value of MQ3 sensor lies between 0.0044 ppm and 0.0103 ppm, MQ6 sensor value lies between 0.5568 ppm and 0.6205 ppm, MQ8 sensor value lies between 0.0309 ppm and 0.0509 ppm and MQ135 sensor value lies between 0.3962 ppm and 0.4258 ppm.

From Table 2, it was observed that banana fruit is ripe for the value of MQ3 sensor lies between 0.0175 ppm and 0.0253 ppm, MQ6 sensor value lies between 0.0706 ppm and 0.0708 ppm, MQ8 sensor value lies between 0.1599 ppm and 0.1633ppm, and MQ135 sensor value lies between 0.3673 ppm and 0.4037 ppm. It was also observed that lemon fruit is unripe for the value of MQ3 sensor lies between 0.0340 ppm and 0.0415 ppm, MQ6 sensor value lies between 0.7725 ppm and 0.8341 ppm, MQ8 sensor value lies between 0.1641 ppm and 0.1739ppm, and MQ135 sensor value lies between 0.5805 ppm and 0.6565 ppm. Similarly, it was observed that for apple fruit is unripe for the value of MQ3 sensor lies between 0.0338 ppm and 0.0630 ppm, MQ6 sensor value lies between 0.7767 ppm and 0.8016 ppm, MQ8 sensor value lies between 0.1779 ppm and 0.2034 ppm and MQ135 sensor value lies between 0.5825 ppm and 0.6380 ppm.

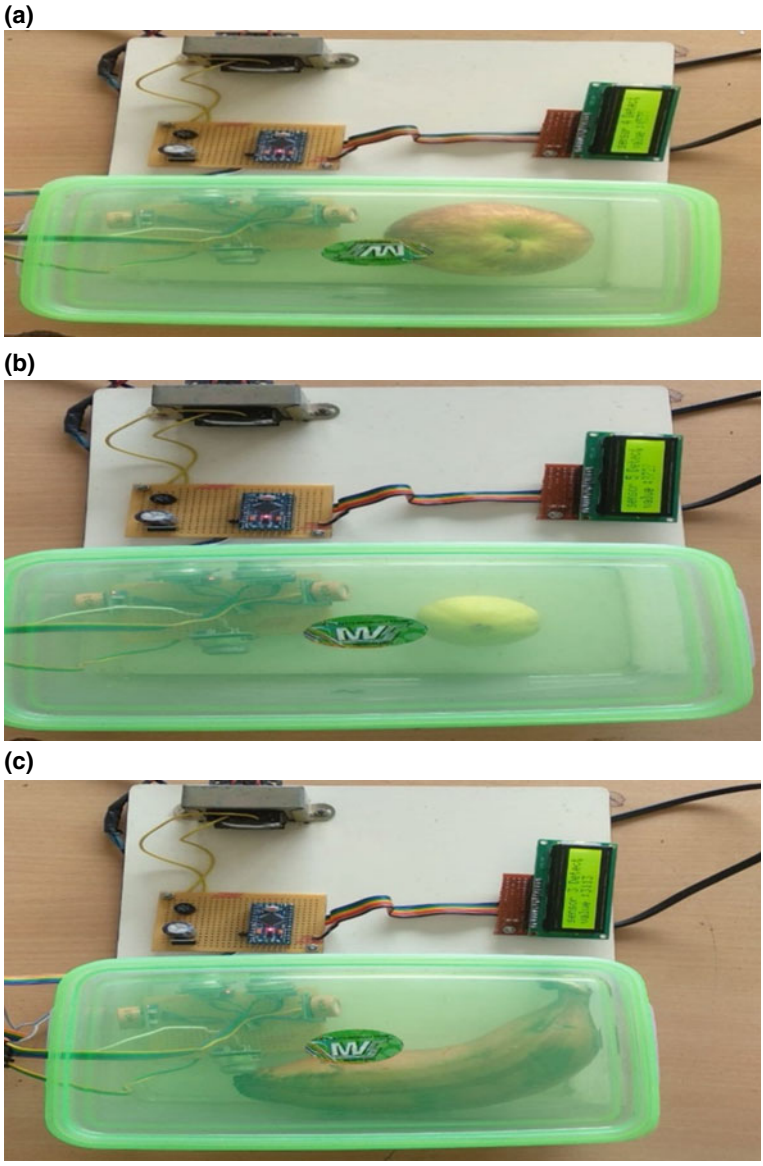


Fig. 8 **a** Electronic nose system for determining the ripening stages of apple fruit **b** Electronic nose system for determining the ripening stages of lemon fruit **c** Electronic nose system for determining the ripening stages of banana fruit

Table 2 Shows the obtained results for different ripening stages of various fruits

Fruits		Banana				Lemon				Apple			
*Sensors		MQ3	MQ6	MQ8	MQ135	MQ3	MQ6	MQ8	MQ135	MQ3	MQ6	MQ8	MQ135
Day 1	6.00am	0.0129	0.6496	0.0034	0.3266	0.0271	0.5466	0.0106	0.3900	0.0012	0.5466	0.0106	0.3900
	12.00 pm	0.0119	0.6345	0.0049	0.3297	0.0277	0.5487	0.0117	0.3922	0.0043	0.5518	0.0223	0.3987
	10.00 pm	0.0101	0.6002	0.0097	0.3301	0.0298	0.5499	0.0229	0.4105	0.0076	0.5721	0.0598	0.3999
	Average	0.0116	0.6281	0.0060	0.3288	0.0282	0.5484	0.0151	0.3976	0.0044	0.5568	0.0309	0.3962
Day2	6.00am	0.0136	0.5991	0.0101	0.3354	0.0301	0.6102	0.0334	0.4376	0.0091	0.5945	0.0421	0.4087
	12.00 pm	0.0137	0.5550	0.0127	0.3367	0.0309	0.6122	0.0376	0.4398	0.0101	0.6212	0.0519	0.4288
	10.00 pm	0.0145	0.3321	0.0159	0.3394	0.0312	0.6137	0.0399	0.4877	0.0117	0.6457	0.0588	0.4399
	Average	0.0139	0.4954	0.0129	0.3372	0.0307	0.6120	0.0370	0.4550	0.0103	0.6205	0.0509	0.4258
Day3	6.00am	0.0154	0.0706	0.1555	0.3652	0.0325	0.7700	0.1603	0.5739	0.0271	0.7654	0.1716	0.5739
	12.00 pm	0.0175	0.0706	0.1568	0.3676	0.0337	0.7721	0.1619	0.5776	0.0287	0.7751	0.1801	0.5834
	10.00 pm	0.0197	0.0707	0.1675	0.3691	0.0358	0.7754	0.1701	0.5899	0.0457	0.7896	0.1819	0.5901
	Average	0.0175	0.0706	0.1599	0.3673	0.0340	0.7725	0.1641	0.5805	0.0338	0.7767	0.1779	0.5825
Day4	6.00am	0.0226	0.0706	0.1616	0.3991	0.0399	0.8101	0.1717	0.6287	0.0567	0.7968	0.1987	0.6197
	12.00 pm	0.0245	0.0706	0.1637	0.4021	0.0407	0.8266	0.1734	0.6534	0.0612	0.8004	0.1998	0.6356
	10.00 pm	0.0287	0.0707	0.1645	0.4099	0.0438	0.8655	0.1765	0.6873	0.0712	0.8076	0.2117	0.6587
	Average	0.0253	0.0708	0.1633	0.4037	0.0415	0.8341	0.1739	0.6565	0.0630	0.8016	0.2034	0.6380
Day5	6.00am	0.0602	0.0127	0.1812	0.4640	0.0588	0.9077	0.1819	0.8131	0.0854	0.9027	0.2603	0.8131
	12.00 pm	0.0634	0.0128	0.1835	0.4677	0.0591	0.9176	0.1843	0.8223	0.0889	0.9077	0.2734	0.8255
	10.00 pm	0.0687	0.0129	0.1899	0.4697	0.0601	0.9366	0.1888	0.8338	0.0912	0.9456	0.2834	0.8459
	Average	0.0641	0.0128	0.1849	0.4671	0.0593	0.9206	0.1850	0.8231	0.0885	0.9187	0.2724	0.8282

(continued)

Table 2 (continued)

Fruits		Banana				Lemon				Apple			
*Sensors		MQ3	MQ6	MQ8	MQ135	MQ3	MQ6	MQ8	MQ135	MQ3	MQ6	MQ8	MQ135
Day6	6.00am	0.0717	0.0129	0.1902	0.4802	0.0723	0.9877	0.1904	0.8999	0.0997	0.8967	0.2998	0.8911
	12.00 pm	0.0724	0.0129	0.1937	0.4831	0.0754	0.9901	0.1944	0.9272	0.1003	0.9021	0.3103	0.9110
	10.00 pm	0.0751	0.0129	0.1951	0.4859	0.0793	0.9967	0.1978	0.9577	0.1103	0.9117	0.3288	0.9227
	Average	0.0731	0.0129	0.1930	0.4831	0.0757	0.9915	0.1942	0.9283	0.1034	0.9035	0.3130	0.9083

*All values are in ppm.

As mentioned in Table 2, banana fruit is overripe for the value of MQ3 sensor lies between 0.0641 ppm and 0.0731 ppm, MQ6 sensor value lies between 0.0128 ppm and 0.0129 ppm, MQ8 sensor value lies between 0.1849 ppm and 0.1930 ppm and MQ135 sensor value lies between 0.4671 ppm and 0.4831 ppm. It was also observed that lemon fruit is overripe for the value of MQ3 sensor lies between 0.0593 ppm and 0.0757 ppm, MQ6 sensor value lies between 0.9206 ppm and 0.9915 ppm, MQ8 sensor value lies between 0.1850 ppm and 0.1942 ppm and MQ135 sensor value lies between 0.8231 ppm and 0.9283 ppm. Similarly, it was observed that for apple fruit is overripe for the value of MQ3 sensor lies between 0.0885 ppm and 0.1034, MQ6 sensor value lies between 0.9187 ppm and 0.9035 ppm, MQ8 sensor value lies between 0.2724 ppm and 0.3130 ppm and MQ135 sensor value lies between 0.8282 ppm and 0.9083 ppm.

After measuring the values by various sensors, the ripening stages of all the tree fruits were determined. To improve the efficiency of the system, data collected from the electronic nose are compared with results obtained by image processing technique and manual verification of fruits on the respective days are shown in Table 3. Accuracy of the ripening stages was calculated for this system to show that the ripening stage determination of banana is 94.44%, lemon has 83.33%, and apple has 100%. This method of verifying the fruit ripening stages produces better results when compared with other methods, which were done earlier.

5 Conclusion

This proposed system has better efficiency when compared with other existing systems. The accuracy of ripening stages was well determined by comparing the result produced by electronic nose and image processing techniques. Also to improve the accuracy of the system, it was compared with manual verification by the expert in that field. This system will help farmers to prevent unwanted pre-harvest and post-harvest fruits. This will also reduce the labor work in the packaging phase of the industry. Thus, the determination of apple, lemon, and banana ripening stages using electronic nose and image processing was done with good accuracy.

Table 3 Shows the comparison between the result obtained by the electronic nose, image processing techniques, and manual verification

Fruits	Banana				Lemon				Apple			
Methods	E	IP	M	Accuracy %	E	IP	M	Accuracy %	E	IP	M	Accuracy %
Day 1 *Stages	UR	UR	UR	100	UR	UR	UR	100	UR	UR	UR	100
Day 2 *Stages	UR	UR	UR	100	UR	UR	UR	100	UR	UR	UR	100
Day 3 *Stages	R&UR	R&UR	R&UR	100	R&UR	UR	UR	66.66	R&UR	R&UR	R&UR	100
Day 4 *Stages	R&UR	R	R	66.66	R&UR	R&UR	R&UR	100	R&UR	R&UR	R&UR	100
Day 5 *Stages	R	R	R	100	R	R&UR	R&UR	66.66	R	R	R	100
Day 6 *Stages	R	R	R	100	R	R	R	66.66	R	R	R	100

#EN-Electronic Nose; #IP-Image Processing technique; #M-Manual; *UR-Unripe stage; *R&UR-Neither Ripe nor Unripe stage; *R-Ripen stage.

References

1. Arakeri MP (2016) Computer vision-based fruit grading system for quality evaluation of Tomato in agricultural industry. Procedia Comput Sci 79:426–433
2. Baietto M, Wilson AD (2015) Electronic–nose application for fruit identification, ripeness and quality grading. Sensors 15(1):899–931
3. Chen LY, Wong DM, Fung CY, Chiu CI, Chou TI, Wu CC, Chiu SW, Tang KT (2018) Development of an electronic-nose system for fruit maturity and quality monitoring. IEEE ICASI
4. Devalatkar P, Koli S (2016) Identification of age factor of fruit (Tomato) using Matlab-image processing. Int J Recent Trends Eng Res 02(07)
5. Geethapriya N, Marypraveena S (2017) Evaluation of fruit ripeness using electronic nose. Int J Adv Inf Sci Technol (IJAIST) 6(5) ISSN: 2319:2682
6. Kanade A, Shaligram AD (2014) Development of an E-nose using metal oxide semiconductor sensors for the classification of climacteric fruits. Int J Sci Eng Res 5(2)
7. Kaur K, Gupta OP (2017) A machine learning approach to determine maturity stages of Tomatoes. Orient J Comput Sci Technol 10(3):683–690
8. Prabha J (2015) Satheesh Kumar, "Assessment of banana fruit maturity by image method technique." J Food Sci Technol 52(3):1316–1327
9. Raut K, Bora V (2016) Assessment of fruit maturity using digital image processing. Int J Sci Technol Eng 3(01)
10. Suthagar S, Meeravali S, Sudhakar K (2012) Performance comparison of various levels of fusion of Multi-focus image using wavelet. Int J Eng Res Technol 1(6)
11. Suthagar S, Tamilselvan KS, Balakumar P, Rajalakshmi B, Roshini C (2020) Translation of sign language for Deaf and Dumb people. Int J Recent Technol Eng 8(5):4369–4372 ISSN: 2277–3878

Material Selection and Analysis of Bicycle Frame

Sourav, Ankit Sharma, and Mrinal Pandey

Abstract In this paper, different materials have been selected using Ashby's charts of young's modulus with density (E vs. ρ) and strength with density (σ vs. ρ). The selected materials were analyzed for static forces on a diamond-typed bicycle frame in SOLIDWORKS, a computer-aided design software. The primary focus is to select a material of high strength, high stiffness, and lower mass as possible. The different materials selected for this purpose from Ashby's charts are CFRP, aluminum alloys, titanium alloys, and steel alloys. For analysis, the rider's weight distribution is considered as 30%, 45%, and 25% on the handlebar, bottom bracket, and rider's seat, respectively. Based on various parameters such as maximum displacement, maximum stress-induced, and maximum stress to density ratio, it has been concluded that CFRP is the best-suited material for high strength, stiffness with lower mass for a bicycle frame.

Keywords Material selection · Bicycle frame · Analysis · Strength · Ashby's charts · CAD · Simulation · CFRP · Alloys

1 Introduction

Material is one of the most critical considerations for any mechanical design. It should be readily available, cost less, and provide excellent strength and rigidity for members to function correctly [1]. Material selection for bikes has been evolved from

Sourav
Design, Analysis and Study of Ashby's Charts, HMR Institute of Technology and Management, New Delhi, India
e-mail: souravmae@gmail.com

A. Sharma
Selection of Material using Ashby's Chart and Analysis, Think & Learn Pvt. Ltd., BYJU, Bangalore, India
e-mail: ankit00726@gmail.com

M. Pandey (✉)
Validation of Material Selection and Analysis, HMR Institute of Technology and Management, New Delhi, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_71

wood, steel tubes, aluminum tubes, carbon fiber, etc. from the 1800s to the present; a wide range of research has done on these materials for their best application in bike frames. A basic bike frame is generally quadrilateral consisting of two triangulations for a rigid structure. The length of the frame tube from the bottom bracket to the upper side of the seat tube represents the size of the frame [2], and it also depends on the category of bike and the biker. The bike frame must be capable of absorbing the forces act on it and must support all the components [3, 4]. Generally, stress analysis on frames is done using simulation software like Ansys, Solidworks, Pro-E, etc.

2 Methodology

2.1 Ashby Charts for Selecting Materials

The two most common Ashby chats are E versus ρ, i.e. young's modulus versus density and σ versus ρ, i.e. strength versus density. The charts are provided with guidelines used for selecting appropriate material as per requirements. As the frame's members act as a tie having constant cross-section area, and members contain only axial loads; therefore, a line with slope equals one has chosen as a reference line for both the graphs. Hence, the guideline having linear relation, i.e. E/ρ and σ/ρ, is selected for selecting a range of materials for high stiffness to mass ratio and high strength to mass ratio, respectively. In common practice, ceramics are avoided as they can fail catastrophically without any indication; therefore, one can select materials from polymers and metals. Guidelines that are selected can be shifted parallel to the general range of material, the higher the position of these guidelines in the general range means the better the strength and stiffness material we get for bike frame as shown in Figs. 1 and 2. From Fig. 1, it can be observed that the guideline shifted is touching CFRP and the nearby materials are Al alloys, Ti alloys, steel alloys having the similar elastic modulus compared with CFRP, in fact, steel alloys have more strength compared with CFRP, but meanwhile, they also have more density, which implies more mass as the volume of members are constant due to constant cross-section area. So, we need to take the ratio of E/ρ of the selected materials, and the one who is having the highest ratio is the best material for high stiffness and low mass. Table 1 shows the selected materials and their ratio. It can be observed from Table 1 that CFRP has a high E/ρ ratio. Hence, it is selected. Similarly, to get high strength with low mass material, the highest ratio of σ/ρ is selected, i.e. CFRP is shown in Table 2.

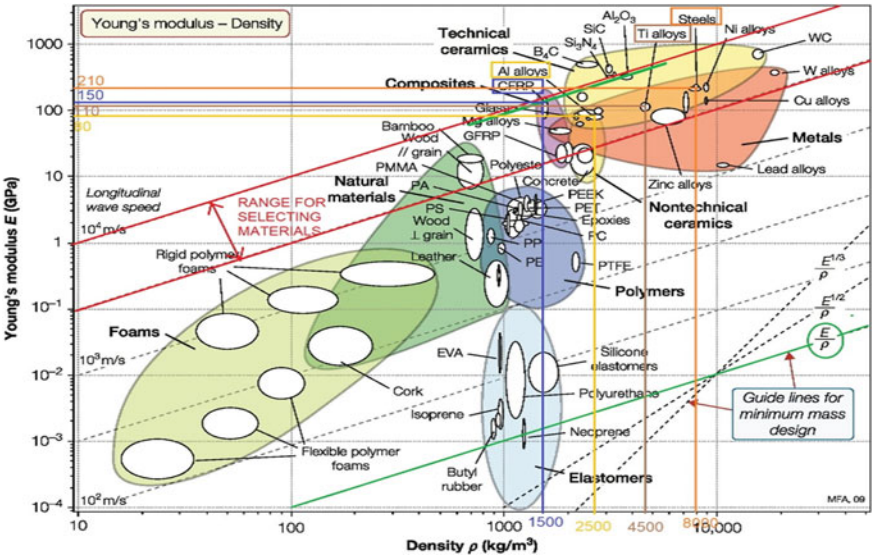


Fig. 1 Ashby chart of E versus ρ

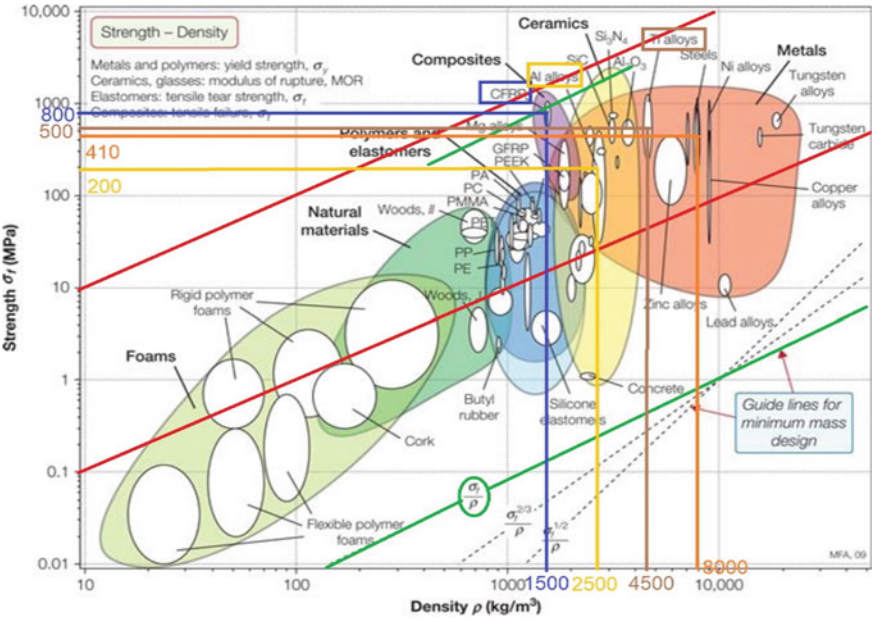


Fig. 2 Ashby chart of σ versus ρ

Table 1 Materials and their E/ρ ratio

S. no.	Material	E (GPa)	ρ (Kg/m^3)	E/ρ
1	CFRP	150	1500	0.1
2	Al alloys	80	2500	0.032
3	Ti alloys	110	4500	0.0244
4	Steel alloys	210	8000	0.02625

Table 2 Materials and their σ/ρ ratio

S. no.	Material	σ (MPa)	ρ (Kg/m^3)	σ/ρ
1	CFRP	800	1500	0.533
2	Al alloys	200	2500	0.08
3	Ti alloys	500	4500	0.111
4	Steel alloys	410	8000	0.05125

2.2 *Design and Analysis of Bicycle Frame*

A basic structure of the bicycle frame is of Diamond-type with proper triangulations to make the frame rigid so that it can transfer the force to each member effectively. A similar design has made using Solidworks, which is a CAD and simulation software. Figure 3 shows the basic frame structure with nomenclature.

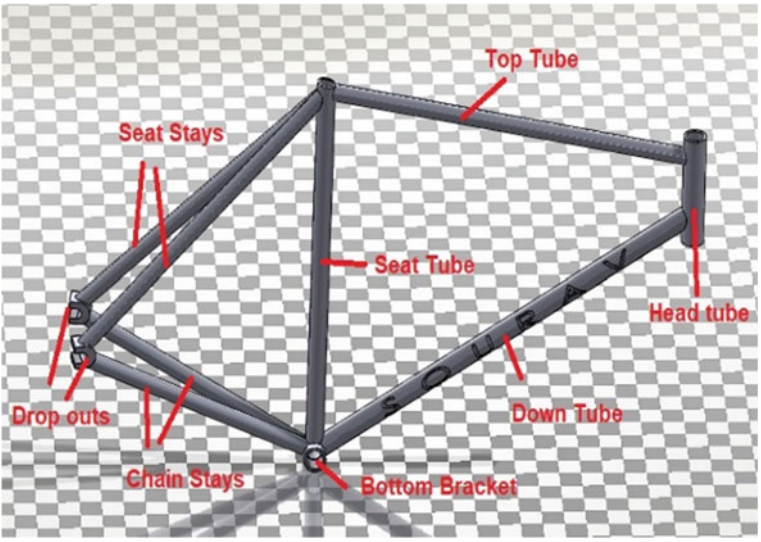


Fig. 3 Bicycle frame nomenclature

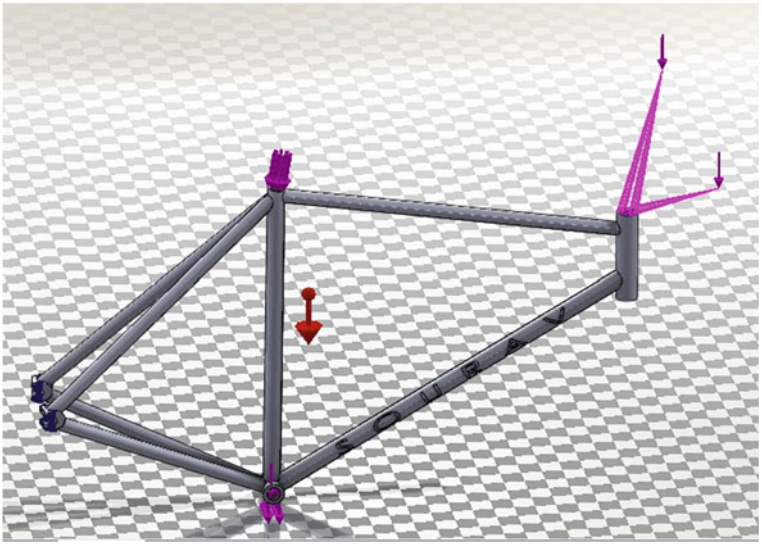


Fig. 4 Forces acting on the bicycle frame

In the static analysis, it is assumed that the weight of rider is distributed as 30% on the handlebar, 45% to the bottom bracket through which paddles are attached, and 25% on the seat on which the rider is sitting. For the static analysis of the frame, the rider mass has been considered 90 kg, which means 90 * 9.81 = 882.9 N in weight, and hence the weight distribution is 264.87 N on the handlebar, 397.305 N on the bottom bracket and 220.75 N on the bike seat. Figure 4 shows the different forces acting on the bicycle frame.

3 Results

The static simulation has performed on a bicycle frame with four different materials under consideration, and the results of stress and displacements are shown in Figs. 5, 6, 7, 8, 9, 10, 11 and 12.

The results of the maximum stress induced (MPa) and maximum displacement (mm) analysis of bicycle frames subjected to different materials are listed in Table 3.

4 Discussions

Table 3 shows that the maximum displacement of CFRP is the lowest among all, i.e. 0.275 mm; hence, the CFRP's frame will be more rigid. Further, to find out the material that can provide sufficient strength to bicycle frame with the lowest mass

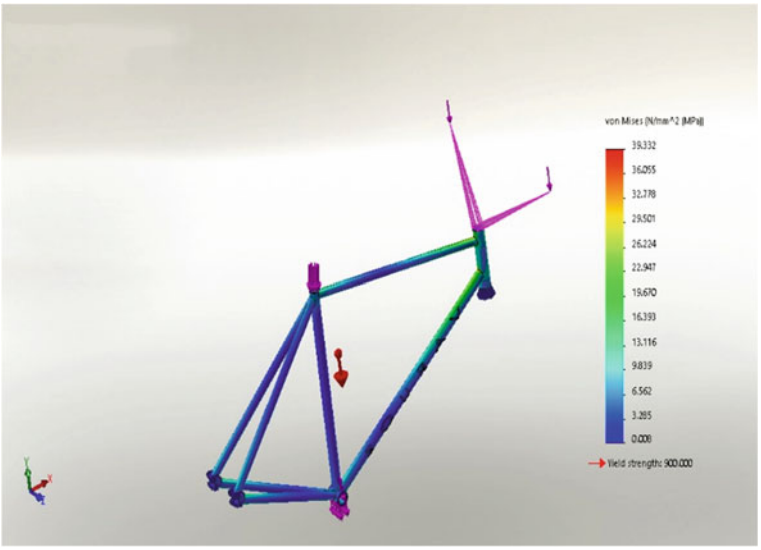


Fig. 5 Stress analysis of CFRP bicycle frame

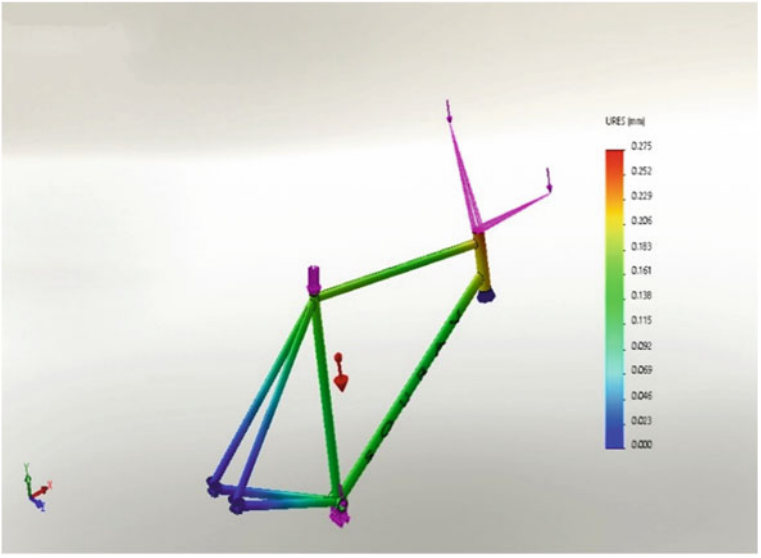


Fig. 6 Displacement analysis of CFRP bicycle frame

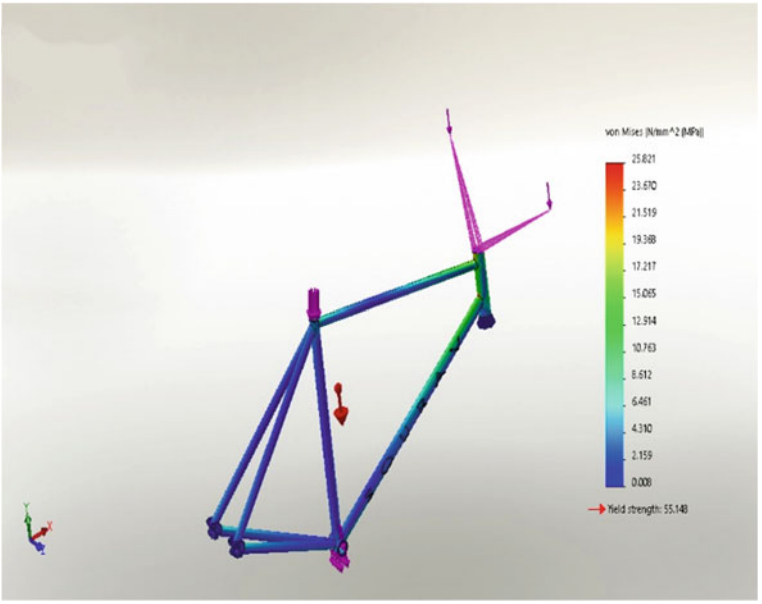


Fig. 7 Stress analysis of Al alloy bicycle frame

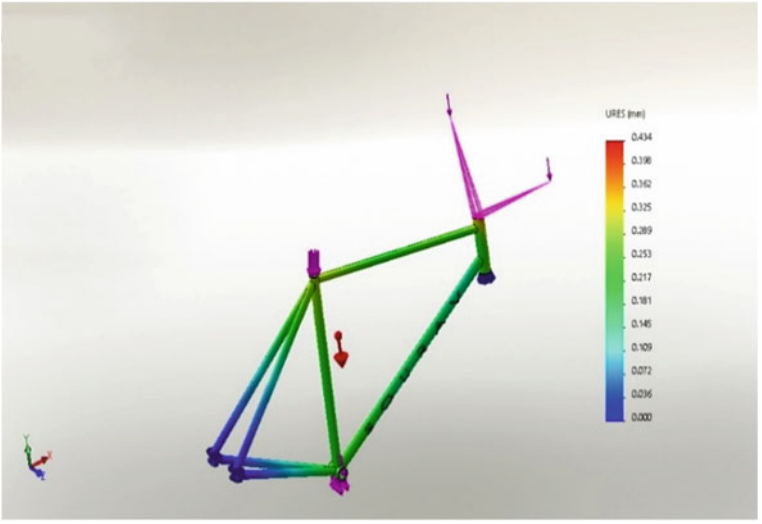


Fig. 8 Displacement analysis of Al alloy bicycle frame

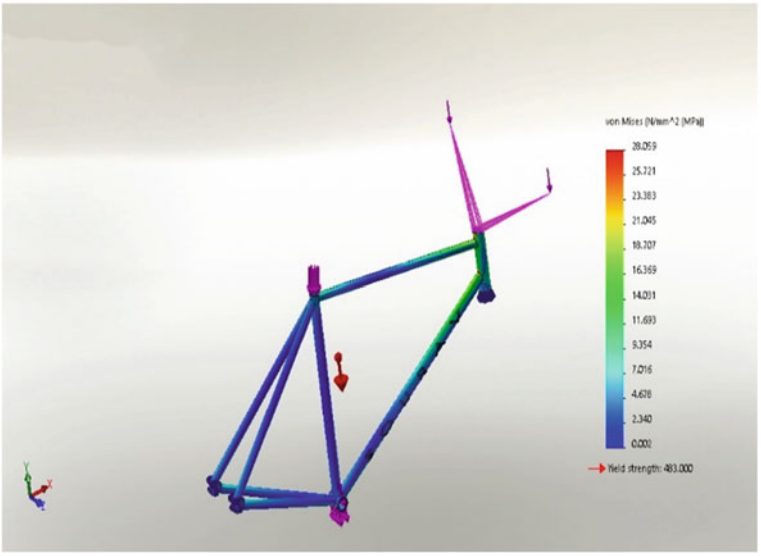


Fig. 9 Stress analysis of Ti alloy bicycle frame

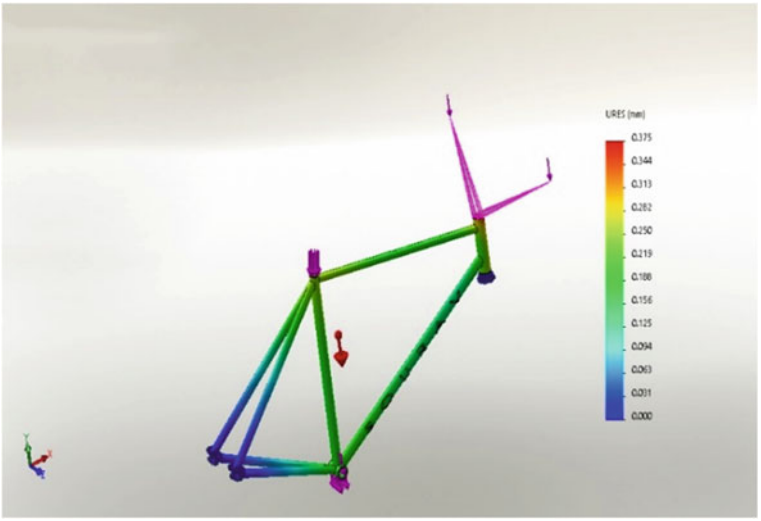


Fig. 10 Displacement analysis of Ti alloy bicycle frame

possible, we need to select the material with the highest strength to mass ratio. Table 4 shows the stress to density ratio for all the chosen materials, and it is highest for CFRP, i.e., 0.02622; hence CFRP's frame will have the highest strength providing the lowest mass.

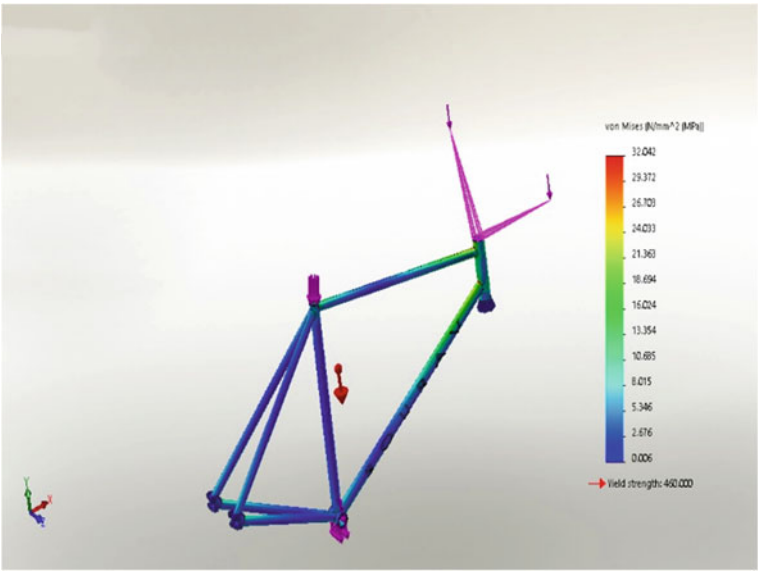


Fig. 11 Stress analysis of steel alloy bicycle frame

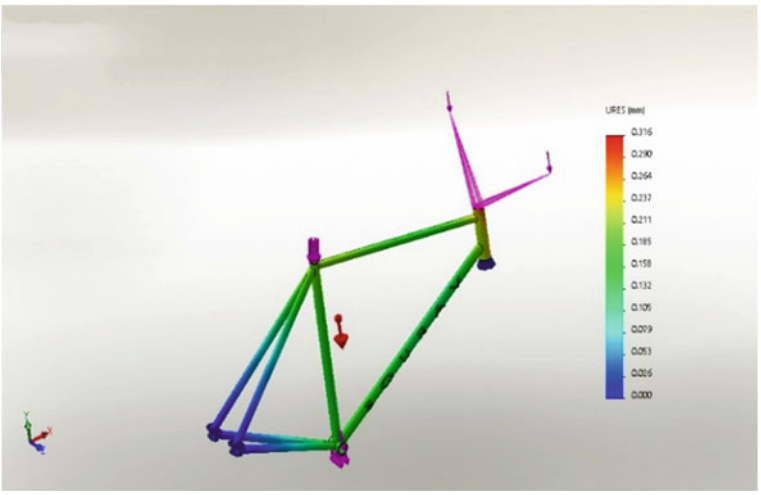


Fig. 12 Displacement analysis of Steel alloy bicycle frame

5 Conclusions

By analyzing the above static analysis' results and observations, it is manifest that from Ashby's chart, one can select a wide range of materials as per the requirement,

Table 3 Displacement and stress analysis results

S. no.	Material	Maximum displacement (mm)	Maximum stress (MPa)
1	CFRP	0.275	39.332
2	Al alloy	0.434	25.821
3	Ti alloy	0.375	28.059
4	Steel alloy	0.316	32.042

Table 4 Maximum stress to density ratio

S. no.	Material	Max. stress/density
1	CFRP	39.332/1500 = 0.02622
2	Al alloy	25.821/2500 = 0.01032
3	Ti alloy	28.059/4500 = 0.006235
4	Steel alloy	32.042/8000 = 0.004

and the scrutinization of chosen materials can be done through simulation software. Conclusions that are drawn from the present work are as follows:-

1. CFRP, aluminum alloys, titanium alloys, and steel alloys have been selected using Ashby's charts of strength with density (σ vs. ρ), and young's modulus with density (E vs. ρ), considering high strength, high stiffness, providing low mass to a bicycle frame.
2. CFRP is selected as the best-suited material for the objective by analyzing maximum displacement results, maximum stress induced in the frame, and calculating the strength to density ratio.

References

1. Chien-Cheng L, Song-Jeng H, Chi-Chia L (2017) Structural analysis and optimization of bicycle frame designs. Adv Mech Eng 09(10)
2. Rontescu C, Cicic D, Amza CG, Chivu OR, Dobrota D (2015) Choosing the optimum material for making a bicycle frame. Metalurgija-Sisak then Zagreb 54:679–682
3. Bhale Swapnil M (2017) Design and structural analysis of a frame of a quad bike (ATV). IRJET 04(1):943–947
4. Soden PD, Adefeya BA (1979) Forces applied to a bicycle during normal cycling. J Biomech 12:527–541

Wheel Robot Review

Vibha Burman and Ravinder Kumar

Abstract Wheeled robots are winning popularity in the area of robotics applications because of their properties invading movement in different orientations. The wheeled robots have the property to operate in complex environments with the abilities of specialized mobility and flexibility. The fields of applications where wheeled robots are applicable are medical waste collections, black/light vision area, point-to-point dedicated path movement, assistance in carrying bulky objects, and many more. This report will provide the wider aspects of wheeled configurations used in assorted applications of robotics in industries as well as different control systems. In this paper, two main areas are discussed. The first section consists of different wheeled robot navigation systems using a kind of path planning algorithms for the safest path. The second section consists of obstacle avoidance techniques.

Keywords Omnidirectional wheels · Mecanum wheels · Point-to-point dedicated path · Safest path · Obstacle avoidance

1 Introduction

Wheeled robots are extensively gaining popularity in mechanical and scientific arenas. To crisscross in compact areas and also for escaping obstacles, wheeled robots should have good mobility and flexibility. This effectiveness depends on the design of wheels.

A sort of omnidirectional wheeled platform is classified into special and conventional wheels. Special wheels possess an active navigation technique and non-resistive motion guidance. Standard wheels are classified into two classes, small wheels, and driving wheels. In 1973, the Mecanum wheel was invented by Mr. Ilon, a Swedish engineer (Mecanum Company). Thus, it is known as Mecanum (Swedish) wheel. The first constructed mobile robot with the help of Mecanum wheels was named "Uranus" at Carnegie Mellon University [1].

V. Burman (✉) · R. Kumar
Department of Mechanical and Automation, Indira Gandhi Delhi Technical University of Women, Kashmere Gate, New Delhi 110006, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_72

Wheeled robots are best suited for structured areas like flat platforms or regular terrain areas, whereas, walking machines are best suited for unstructured areas because they offer stability in a large range of situations, merely they are mechanically complex and need more control resources [2].

In [3], the kinematic model of "Uranus", WMR (wheeled mobile robot) was prepared in the Robotics research foundation of Carnegie Mellon University for performing studies in supervision, computer visualization, route arrangement, and sonar sense. "Uranus" provides omnidirectional movement with a balance, rectangular wheel based. Each omnidirectional wheel comprises the hub of a wheel and consists of three degrees of freedom (DOFs). The first DOF provides rotary motion about the point in the middle of the wheel and base; the second DOF provides the direction of the wheel route and the third DOF lies on the route of the roller rotary motion.

Slippage is categorized into two broad classifications based on vision cameras and proprioceptive sensors. Vision cameras have limitations on computations and less vision in black areas, whereas, proprioceptive sensors have their limitations in error and drift in the estimation of body velocity. A variety of machine learning algorithms (i.e., Support Vector Machines (SVM), Multi-layer perceptron, K-Means, and Self-Organizing Maps) are applied in estimating slippage would be decision trees [4].

For wheeled robot path navigation, fuzzy logic control is applied smartly. The revolutionary robots have a strong ability to portray strong concepts using genetic algorithms. Neural Network (NN) is suitable for uncertain and non-linear situations and has strong adaptive capabilities, therefore, it is popular for robot controller designing [5].

Finding out the robot's current location is known as localization. Localization methods can be characterized either in the category of comparative or absolute. While the previous class includes measuring wheel revolving and rate of revolution and acceleration, and the final class includes dynamic beacons, artificial and natural sight detection and map similar [6].

The accessible off-road movement arrangement for movable robots is dependent on two arrangements, first, the straight path or wheels with an alike car-like architecture and second, discrepancy constrain routing arrangement. Recently, off-road movement systems, steering and movable in contracted in addition to confidential places would be complicated to which an omnidirectional wheel robot might fulfill those abilities very well [7].

The demand for simulation arises because they are easier to execute, cheaper and computationally expensive algorithms can also be sunk up. The solutions are transferable to the real robots later [8]. Simulation software provides an ideal situation for simulating any mobile, legged and flying robots. This software lets programming interfaces of C, C++, and Java, or from other types of software throughout TCP/IP. It should provide a complex environment creation for robot simulations.

Simulations can be categorized into the following:

- Multi-agent simulations: some robots cooperate to achieve a global aim.

- Artificial Intelligence: validate psychology assumption, mostly learning, by simulating smart mobile robot behaviors.
- Control explore: develops professional control algorithms to carry out the difficult mobile robot activity.
- Robot architecture: shapes mobile robots presenting the position and attributes of the external element of a device such as sensors and actuators.

The Robotic Operating System [9] is a communication situated system invented to combine different alternate systems that are applied in a robotic. It is collected by C++ and Python libraries. The Robotic Operating System can be completed as a circulated arrangement, in which nodes can natter by subscribing to topics to convey or receive messages [10].

At present, this is one of the highest rising methods in the robot machinery with the improvement of technology, by which we can solve problems. Applications of robots in thousands of industries are prepared to offer various applications for developing the efficiency of the innovation of the ware. By applying robots can promote and make simpler the manufacturing work and can facilitate in the world all over the place [11].

In this paper, we use and enlarge the idea of the safest path planning and obstacle avoidance methods. Our target is to find only the safest path among the real and prefixed positions of the automaton. The format of the report is in the subsequent order; the literature survey discusses the related works, focal point on motion scheduling algorithms for stationary and moving obstacles. In the literature review, the first part discusses the various wheel configurations (Table 1) and software algorithms (Table 2) and later discusses the diverse hardware elements (Table 3). The conclusions discuss the new method and plans.

Table 1 Depiction of the different categories of wheels

Universal wheel (simple)	Mecanum wheel	Powered steered wheel	Castor wheel
+ Lowest load, compacted drawing + Easy mechanical drawing + Commercially existing	+ Compacted drawing + Elevated load capability	+ Nonstop wheel contact + Elevated load capability + Strong to ground surroundings	+ Nonstop wheel contact + Elevated load capability + Little scrub force throughout the routing + Strong to ground surroundings
– Irregular wheel makes contact with/variable drive radius – Responsive to ground indiscretion	– Irregular wheel contact – Elevated responsiveness to ground indiscretion – Composite wheel drawing	– Irregular wheel contact – Elevated responsiveness to ground indiscretion – Composite wheel drawing	– Large drawing – Broadcast signal and power from corner to corner revolving joints – Composite mechanics

Table 2 Depiction of the concepts and algorithms used by different authors

Concepts used	Authors
Fuzzy control strategy	Jincong et al. [21]
Artificial immune system (AIS) algorithm	Huang [22]
Ant colony optimization	Song et al. [23]
Neural network	Meyer [24]
Artificial neural network	Pol and Murugan [25]
Spiking neural network (SNN), genetic algorithm	Wang et al. [5], Greenwald and Artz [6], Batlloria et al. [26], Hagras et al. [27]

Table 3 Depiction of the components used by different authors

Components used	Authors
Ultrasonic sensor	Wang et al. [5], Jincong et al. [21]
PI, PD and PID controller	Sikander and Prasad [30], Belzunce et al. [28], Algoz et al. [31]
IMU controller	Algoz et al. [31]
Motor driver L298N	Barros [10]

2 Literature Review

Omnidirectional automobiles with Mecanum wheels include several restrictions. Consequently, an automobile with Mecanum wheels as illustrated in Fig. 1 is liable to slippage, and as a consequence, utilizing the equivalent amount of wheel rotary motion, the agile roaming gap is different as of longitudinal roaming distance. The second disadvantage is with the purpose of the contact end connecting the wheel along with ground progress all along a line similar to the wheel axis, even so, the wheel is constantly in touch through the earth. The lateral motion displays horizontal sensations. The very last disadvantage is its capacity to conquer obstacles is dependent on the roaming route [12].

The Swedish (Mecanum) wheel is designed apart from its rollers to be fixed on top of the angle. A Mecannum omnidirectional wheel has three DOF collected of wheel

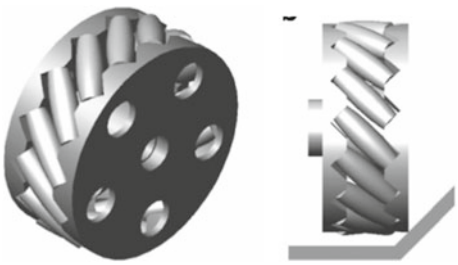


Fig. 1 Mecannum wheels utilized in [12] prototype

rotary motion, roller, rotary motion, and rotational slide concerning the perpendicular axis transient throughout the place of the middleman.

Within the omnidirectional wheel, the wheel velocity is capable of sorting out the apparatus within the dynamic route and the passive direction. The active element is pointed at the axis of the roller within contact to amid the ground, whereas the passive individual is vertical to the roll axis. The method includes miniature rollers positioned in the orderliness of the external diameter of a rack to permit common wheel rotary motion. Until nowadays, to revolve in the route comparable in the instruction to the wheel axis, the cycle is carried through with this motion for the cause that the rollers are mounted perpendicular to the bloc of the circular motion of the bike. When two or additional of these wheels will mount on top of an automobile display place their mutual controlled and unrestrained movement allows intended for omnidirectional mobility [12].

The term omnidirectional is used in the way of explaining the capability of an arrangement to move instantly to some course. Robotic automobiles are habitually planned in favor of the planar motion; they get together on a storehouse ground, track, desk, etc. Within 2D space, a body has three DOF. It is good at interpreting commands in addition to revolving concerning its center of gravity [12].

Legged robots have exceptional land flexibility. They offer alluring potential in the areas of sharpness and obstacle avoidance. On the soft ground or in unstructured environments, legged ones are preferred for locomotion rather than wheeled ones. Whereas, wheeled robots provide ample durability, mechanical clarity, and active behavior. They are fast, dominant concerning load to weight ratio, balanced, and easy controlling operations [13].

Legged and wheeled robots altogether rely on the mechanical communications between the wheels (or feet) and the primer for achieving and supervising the displacements. The restriction is that the contact pairs are defined; thus, they provide very little ability to control the shift between the above two [14]. The wheeled robot grabs maximum speed, high energy efficiency, and technology preparedness and it provides lower mechanical and control complexity. Wheeled robots provide low obstacle intersection, step/stair climbing, and slow climbing capabilities as well as low walking capabilities on soft and uneven terrains as explained in [15].

The Swedish wheel is those omnidirectional wheels that are well constructed. In whatsoever event, three Swedish wheels are recommended to construct a holonomic omnidirectional robot. As omnidirectional robots are created without the pursuit of dynamic steering of the wheel part; the mechanical items of actuating parts consist of simple structures. Whereas, designing the wheel is quite difficult. The major drawback of the Swedish wheel is steep vibration occurred majorly of discontinuous contact which use to occur during motion. To overcome this, various mechanical designs are proposed [16].

In autonomous mobile robots, two factors of smart mobility and management are considered. Mobility potential aims with real attributes of the automaton to help build it possible on the way to be in question. Mobility manages with the assistance of “intelligence” in governing the robot to essentially accomplish its full mobility facilities [17].

For locomotion of robot mimicking humans, a single helix contained by the proximity of every joint is attained with inflexible elements similar to humans with respect to the closeness of each human link [18].

Wheeled Mobile Robot (WMR) is a set of mechanical structures. These structures illustrate during constraints of kinematic as they are not fused and cannot be omitted. The consequence of the above scenario is that the standard routing and control algorithms are no more applicable which were amended in the robotic axis where constraints are not recognized. Four varieties of land space for the understanding architecture of WMR, namely, posture kinematic, configuration kinematic, configuration dynamical, and posture dynamical model have been developed [19]. Wheeled robots are an enticing model of non-holonomic systems. The contact points in the center of the wheel besides en route to the floor amuse the demands of clean, gently sloping as well as non-tripping [20].

Neural Network (NN): It is comprised of exceedingly synchronized elementary blocks that demonstrate neural network design philosophy. These are containers that can be used for bringing together an additional multipart method. These frameworks do not accomplish through the means of conventional methods of executing correct mathematical procedures.

This algorithm allows a great amount of speeding up with the statistics processing capacity of knowledge. In robotics, artificial intelligence is required to do modeling in kinematics and dynamics. This algorithm also provides pathway planning and trend direction in mobile robots.

Advantages of NN

- It is the potential to instruct a NN by regulating the principles of weights linking elements to do a fuzzy task.
- The target entity generally attracts the robot where whole location space, whereas, hindrance worldwide drives away from the golem.

Artificial Neural Network technique (ANN): ANN algorithm works for route-discovery management of movable robots. This technique uses multiple-layer feed-forward ANN of building pathway scheduling controller with the help of evaluating the non-linear function. ANN works best for routine management of movable robots. The motion controlling ANN is skilled with a comprehensive broadcast during a time algorithm, which offers a potential arena in favor of prevention from obstruction.

Advantages of ANN

- ANN is used for exploration, optimization, pattern recognition, and learning problems to build an obtainable finest solution.
- It is sufficient in creating an intelligent judgment that can be incorporated into the car.
- The pathway scheduling, task is shortened to a confidential problem, to wit, five states–action mapping associations.

Spiking Neural Network technique (SNN): SNN is well suited by pattern recognition and classification. It translates the linear and non-linear high-dimensional cluster problems successfully. Manner control of movable robots is supported on SNN by applying ultrasonic sensory data. SNN is adjusted by an unsupervised Spike-based Hebbian learning algorithm.

Advantages of SNN

- It offers aid in deriving the design of the robot controller.
- SNN can be suitable in wheeled robots while tracking input signals coming from ultrasonic sensors.
- The SNN with training rules is effective and can be carried out easily.
- The ultrasonic sensor calculates distance using the method of the fourth dimension of flight (TOF), for detecting obstacle avoidance, sonar sensors are used, and in the controller, discrete-time is used.

Fuzzy Logic: The obstacle avoidance technique uses a fuzzy control scheme. The basic fuzzy controller is utilized to develop a connection between reaching the target and obstacle avoidance strategy. The grouped ultrasonic sensors are three in number as left-front, right-front and front sensors. Whenever a target is placed in whichever direction, the controller will move in the opposite respective direction due to if–then rules in fuzzy set theory [21].

Artificial Immune System (AIS): Artificial immune system supports smart movement organizes a problem of movable robots. Workspace support bears the high cost and not a feasible option for appropriate simulation in real time. This fault can be overcome by using the Personal Computer hardware/software codesign tools in FPGA (Field Programmable Gate Array) chip. FPGA technologies have been a new archetype for artificial intelligent computing and proceed as professional and influential funds to comprehend complicated algorithms in various regulations [22].

Ant Colony Optimization (ACO) algorithm: As a global optimization algorithm, the ACO algorithm is an arbitrary exploration optimization technique that holds a broad relevance area. The primary difficulty that desires to be resolved is en route for formulating the search space for ACO as bulky as achievable to discover a global minimize path. The preceding knowledge is applied to executing the arrangement possibility to worldwide excellently [23].

Initially, the torque controlling technique was done manually while estimation was dependent on robot heaviness and total load. To overcome this, a movement investigation was applied to Gazebo Simulator. A gazebo is a simulation tool for robotic software that contains the ability of precise and efficient dynamical simulation of robots, sensors, and objects in a three-dimensional world. The CAD model was prepared by Solidworks [7].

Previously, neural network consequences have been frequently restricted to the development of simple sensory-motor mechanisms. But recently, some achievements are developing more cognitive architectures have been reported. Uncertain situations arise whether automatic approaches will level up with increasing difficulty and

whether they will finally compete with human abilities for designing resourceful robots [24].

The robotic operating system has enabled a control system that typically avoided low-level programming throughout code recycle and encloses the construction. The robotics operating system could manage the two key components of the robot: manipulator and driving system [28].

After researchers [29] applied the facts of the qualified location of additional statistics like obstruction and landmark nodes, the procedure of progression otherwise within integration through additional search method has urbanized especially. It considerably enhanced the efficiency of using genetic algorithms toward resolving the pathway scheduling problem, except pathway discontinuity persists. Other authors try to apply arc linking using two straight line segments, leave out an increase in difficulty in providing answers. Currently, the broad relevance of movable robot and vehicle class, model, smooth pathway scheduling problems contain steadily appreciated. Therefore, the numerous curves in personal computer graphics commonly function, such as the Hermite curve, Parametric Polynomial Curve, Bezier curve with B-Spline. They are given for obstruction escaping pathway scheduling.

Other most widely used controllers in industries are PI, PD, and PID. The PID controllers are appropriate for two-wheeled mobile robots due to their degree of freedom in which one for angle and another one for speed and depending upon the applicability and another one for the yaw [30].

Energy utilization of movable robots is reduced by raising the energy competence of robot composition, as well as by adding more motors and their drivers. Minimization of robot relevance also throws in energy conservation. Motion scheduling has newly appeared as a solution for cutting down energy consumption for the period of robot motion. The least measure of energy, velocity scheduling was projected where battery saving energy up by 10%. Energy-proficient action arrangement for robot discovery was saved up by 42% [32].

For developing efficiency in output, various energy-efficient movement scheduling is a movement toward industrial robots. Energy-efficient action arrangements were also running for different robot configurations like space manipulator, humanoid, rover, soccer, and bio-robots for performing manipulations in the cubicle. Energy-efficient action scheduling for movable robots has lately fascinated substantial consideration to assemble necessities for movable robots at numerous relevance. Successful pathway scheduling was considered to allow the robot to complete a sequence of work by using less energy. Energy limitations were measured in exploring the problem of numerous sensor exposures [33].

3 Conclusion

The artificial neural network technique is possible to arrange a fight-free way and speed meant for the golem at a standardized time. Using the greatest goal-achieving technique, an obstruction occurs in the road of the pathway of the robot. Hence, a different approach is needed to the obstruction during the activity. Taking into account, the ambiguity of sensors for obstruction and robot, this departure set off instantly in conflict.

We can conclude that the ANN algorithm working in comparison with the genetic algorithm, the latter one is safer in terms of the puzzle-like position. Also, computation time and the requirement of retention of the genetic algorithm are very heavy. On the reverse, the ANN has its limitations in situations where the overall path finding is very large and experiences a disorganized environment. The ANN algorithm in comparison with A* algorithm operates under low-size inputs. If the input size exceeds the memory limit, the computation size increases. Artificial neural networks are, in general, utilized for search optimization, pattern identification, and learning problems to provide natural and most favorable solutions.

Acknowledgements I would like to thank Mr. Bramha Swaroop Tripathi, Senior Project Associate, Department of Computer Science, IGDTUW for guidance in the software and hardware implementation of this project.

References

1. Doroftei I, Grosu V, Spinu V (2007) Omnidirectional mobile robot-design and implementation. Bioinspiration and robotics walking and climbing robots, pp 511–528. IntechOpen
2. Siegwart R, Lamon P, Estier T, Lauria M, Piguet R (2002). Innovative design for wheeled locomotion in rough terrain, robotics and autonomous system, pp 151–162
3. Muir PF, Neuman CP (1987) Kinematic modeling for feedback control of an Omni-directional wheeled mobile robot. Autonomous robot vehicles, IEEE, pp 1772–1778
4. Gonzalez R, Chandler S, Apostolopoulos D (2019) Characterization of machine learning algorithms for slippage estimation in planetary explorations rovers. J Terrramech 82:23–34
5. Wang X, Hou Z-G, Zou A, Tan M, Cheng L (2008) A behavior controller based on spiking neural networks for mobile robots. Neurocomputing 71: 655–666. Elsevier
6. Greenwald L, Artz D (2004) Teaching artificial intelligence with low-cost robots. Accessible hands-on artificial intelligence and robotics education. American Association for Artificial Intelligence, pp 35–41
7. Sharifi M, Young MS, Chen XQ, Clucas D, Pretty C (2016) Mechatronic design and development of a non-holonomic Omni-directional mobile robot for automation of primary production. Cogent Eng 3:1–13
8. Michel O (2004) Cyberbotics Ltd. WebotsTM: Professional mobile robot simulation. Int J Adv Robot Syst 1:40–43
9. Koenig N, Howard A (2004) Design and use paradigms for gazebo, an open-source multi-robot simulator. In: Proceedings of the 2004 IEEE/RSJ international conference on intelligent robots and systems (IROS 2004), vol 3, pp 2149–2154
10. Barros TTT (2014) A mobile manipulator controller implemented in the robot operating system. In: Conference ISR ROBOTIK, pp 121–128

11. Gyenes Z, Szádeczky-Kardoss EG (2018) Motion planning for mobile robots using the safety velocity obstacles method, pp 389–394. IEEE
12. Doroftei I (2010) Conceptual design of an omni-directional mobile robot. In: Proceedings of SYROM 2009, Springer, pp 115–127
13. Conduraru (Slătineanu) A, Doroftei I, Conduraru I (2013) An overview on the design of mobile robots with hybrid locomotion. Adv Mater Res 837:555–560
14. Wieber P-B (2008) Viability and predictive control for safe locomotion. In: RSJ International conference on intelligent robots and systems, pp 1103–1108. IEEE
15. Chung W, Iagnemma K (2016) Wheeled robots. Mobility of wheeled robots, pp 575–593
16. Bruzzone L, Quaglia G (2012) Review article: locomotion systems for ground mobile robots in unstructured environments. Mech Sci 49–62
17. Moore KL, Flann NS (2000) A six-wheeled omni-directional autonomous mobile robot. In: Proceedings of the 1999 IEEE international symposium on intelligent control/intelligent systems and semiotics, IEEE Control Systems Magazine, pp 53–66
18. Muscolo GG, Caldwell D, Cannella F (2017) Biomechanics of human locomotion with constraints to design flexible-wheeled biped robots. In: IEEE International conference on advanced intelligent mechatronics (AIM), pp 1273–1278. IEEE
19. Campion G, Bastin G, D' AndrCa-Novel B (1996) Structural properties and classification of kinematic and dynamic models of wheeled mobile robots. In: IEEE transactions on robotics and automation, vol 12, pp 47–62
20. d'Andrka-Novel B, Bastin G, Compion G (1992) Dynamic feedback linearization of nonholonomic wheeled mobile robots. In: International conference on robotics and automation, pp 2527–2532. IEEE
21. Jincong Y, Xiuping Z, Zhengyuan N, Quanzhen H (2009) Intelligent robot obstacle avoidance system based on fuzzy control. In: The 1st international conference on information science and engineering (ICISE2009), pp 3812–3815. IEEE Computer Society
22. Huang H-C (2013) Intelligent motion control for four-wheeled holonomic mobile robots using FPGA-based artificial immune system algorithm advances in mechanical engineering, pp 1–11. Hindawi Publishing Corporation
23. Song G, Sun L, Yang X (2019) Design and implementation of self-balancing and navigation robot based on ROS system. In: The 31st Chinese control and decision conference, pp 5597–5602. IEEE
24. Meyer J-A (1998) Evolutionary approaches to neural control in mobile robots. In: IEEE International conference on systems, man, and cybernetics, SMC'98 Conference proceedings, vol 3, pp 2418–2423
25. Pol RS, Murugan M (2018) A review on indoor human aware autonomous mobile robot navigation through a dynamic environment. In: International conference on industrial instrumentation and control (ICIC), pp 1339–1344. IEEE
26. Batlloria R, Laramee CB, Land W, Schafferb JD (2011) Evolving spiking neural networks for robot control. Proc Comput Sci 6:329–334
27. Hagras H, Pounds-Cornish A, Colley M, Callaghan V, Clarke G (2004) Evolving spiking neural network controllers for autonomous robots. In: International conference on robotics and automation, pp 4620–4626. IEEE
28. Belzunce A, Li M, Handroos H (2016) Control system design of a teleoperated omnidirectional mobile robot using ROS. In: IEEE 11th Conference on industrial electronics and applications (ICIEA), pp 1283–1287
29. Sudhakara P, Ganapathy V, Sundaran K (2018) Genetic algorithm based optimization technique for route planning of wheeled mobile robot. In: 4th International conference on advances in electrical, electronics, information, communication, and bio-informatics (AEEICB-18). IEEE
30. Sikander A, Prasad R (2017) Reduced-order modeling based control of two-wheeled mobile robots. J Intell Manuf. Springer
31. Algoz A, Bakhtiyar AH (2018) A control system for a 3-axis camera stabilizer

32. Liu S, Sun D (2014) Minimizing energy consumption of wheeled mobile robots via optimal motion planning. IEEE/ASME Trans Mechatron 19:401–411
33. Pol RS, Murugan M (2015) A review on indoor human aware autonomous mobile robot navigation through a dynamic environment, survey of different path planning algorithms and methods. In: 2015 International conference on industrial instrumentation and control (ICIC), pp. 1339–1344. IEEE

Aperture-Coupled Rectangular Dielectric Resonator Antenna for Multiband Applications

Sachin Kumar Yadav, Amanpreet Kaur, and Rajesh Khanna

Abstract This paper introduces the simulation of a rectangular dielectric resonator (RDR) antenna for utilizing in Ku-band for tri-band. This antenna can be used for 5G cellular, satellite communication as broadcast satellite services (BSS) and fixed satellite services (FSS) applications. These terms, FSS and BSS, are used for video data transmission. The band 11.7–12.5 GHz segment is allocated to BSS and 12.5–12.7 GHz band is allocated to the FSS. The proposed RDR antenna is excited by aperture-coupled fed. RDR has cut a corner, set a metallic strip for controlling the multiband purpose and also control the resonance of the antenna. The antenna is reported peak gain 9.01 dBi and high radiation efficiency is above 90%. It is a high gain antenna so can be easily used in 5G applications.

Keywords RDR antenna · Multiband antennas · Ku-band

1 Introduction

The dielectric resonator (DR) antenna has been developed with major advantages in the wireless communication system. In 1983, Long et al. [1, 2] were published an article in the research field of DR. It has several advantages like as cheapo, low loss, ease of excitation, compact, etc. [3, 4]. Radiator element DR has different shapes like cylindrical, rectangular [1, 2], annular, spherical DR and hemispherical, triangular, and conical [5–10]. DR antenna has several advantages over microstrip patch antenna (MSA) like high gain, super gain [11] and higher radiation efficiency. DR antennas are provided 10–20% impedance bandwidth as compared with MSA. Different types of fed mechanism are used in DR antennas like coaxial, MS, aperture coupled MS, conformal strip and coplanar waveguide fed [12–21]. Waveguide slot

S. K. Yadav (✉) · A. Kaur · R. Khanna
Thapar Institute of Engineering and Technology, Patiala, Punjab, India

A. Kaur
e-mail: amanpreet.kaur@thapar.edu

R. Khanna
e-mail: rkhanna@thapar.edu

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_73

has attractive features [22, 23] to reduce the feed line losses for higher frequency applications. The resonant frequency of DR antenna depends on these factors, the ratio of height/length or width/length of DR [24]. The RDR antenna easily operated widen impedance bandwidth due to near field, represented by the electric field. These are represented by the TE, TM, HE, HEM mode theory like fundamental and higher order modes TE_{111}, $TE_{113,}$ and TE_{123} [25–27].

In this paper, the RDR antenna has $\varepsilon_{DR} = 10.2$ (permittivity) simple structure and slot is using for fed connectivity between feedline and RDR. The antenna is implemented to truncate nature of tri-band and resonance by metallic strip and DR height. The peak radiation efficiency is 97.2% and gain is 9.01 dBi. So antenna reported high gain for satellite applications.

2 Antenna Design and Configuration

The proposed RDR antenna is compact and multiband for satellite communication, which using RDR cut one side corner to achieve different bands. The RDR relative permittivity (ε_{DR}) 10.2 and tangent loss 0.018 with size $10 \times 4 \times 12$ mm^3 etched on top of ground plane ($14 \times 14 \times 0.035$) mm^3 have a slot and ground plane is fixed on top of substrate ($14 \times 14 \times 0.75$ mm^3) permittivity 4.4 (FR_4) with loss tangent 0.018 as shown in Fig. 1a–d. The RDR antenna has fed through a slot (0.5×4 mm^2) and aperture provides coupling between fed line and DR. Antenna variables and their values are shown in Table 1. A metallic strip with length, width, height ($0.7 \times 0.35 \times 5.5$ (Hp)) is used in the cut section of DR for controlling the bands and resonances of the antenna.

3 Simulation Results and Discussion

The multiband RDR antenna was simulated in CST 2016 and obtained tri-multiband with an MS fed line of 50 Ω is being insured. The simulated reflection coefficient S_{11} (below –10 dB) is presented in Fig. 2 with tri-band characteristics, and Table 2 shows the bandwidth and gain of the proposed antenna. It is observed that tribands utilize, in Ku band, first two for satellite communication because uplink requires high gain as compared with downlink that antenna fulfilled the criteria and third band for transmission of video data. The simulated antenna is achieved three bands (band 1, band 2 and band 3) in the S_{11} graph as shown in Fig. 2 with the resonance of 12.1, 14.3 and 16.3 GHz and also shows the perfect real impedance (50 Ω) matching in Fig. 3.

Figure 3 shows the surface current distribution at the three resonant frequencies. Current distribution on metallic part of antenna at 12.1 GHz shows in Fig. 3a single resonance path and going in higher side frequency increase the resonance paths that are shown in Fig. 3b and c. So antenna existed three bands. The radiation efficiency

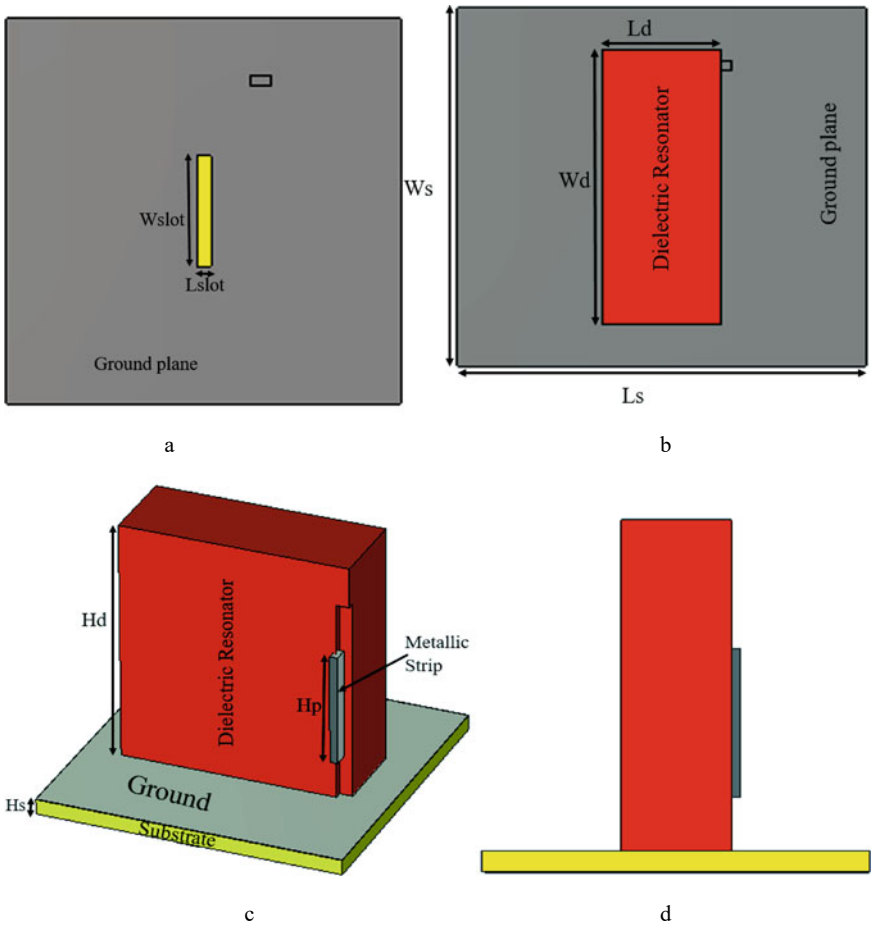


Fig. 1 RDR antenna views: **a**, **b** Front. **c** Perspective. **d** Side

Table 1 Dimensions of RDR antenna

Parameters	Dimensions (mm)	Description
L_s	14	Length of substrate
W_s	14	Width of substrate
H_s	0.8	Height of substrate
W_d	10	Width of RDR
H_d	12	Height of RDR
L_d	4	Length of RDR
L_{slot}	0.5	Length of slot
W_{slot}	4	Width of slot
l_f	10.5	Length of feedline
w_f	1	Width of feedline

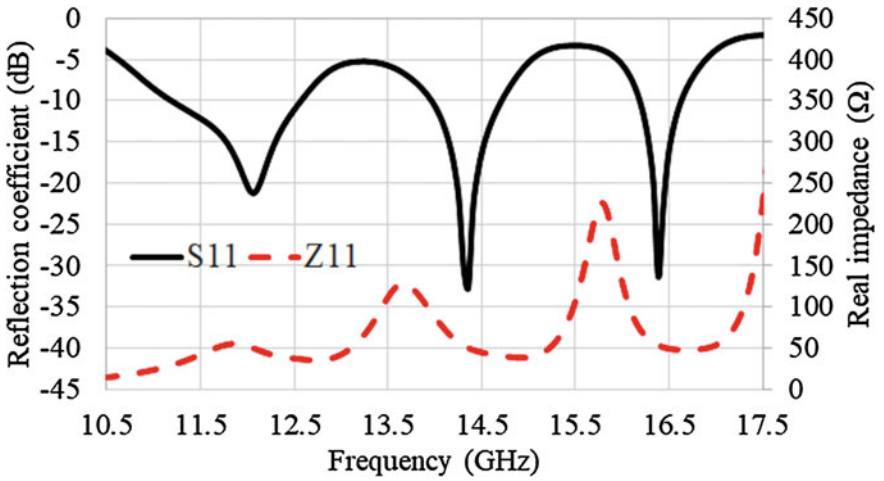


Fig. 2 The return loss of RDR antenna

Table 2 Frequency range, resonant frequency, BW of RDR antenna

Band	Frequency range (GHz)	Resonant frequency (GHz)	Bandwidth (GHz) %	Peak gain (dBi)	
1	11.3–12.6	12.1	10.92	9.01	
2	13.9–14.7	14.3	5.59	7.84	
3	16.3–16.6	16.3	1.82	6.98	

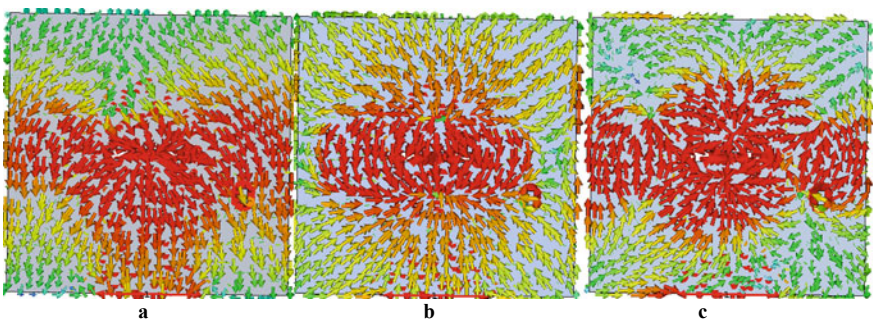


Fig. 3 Surface current distribution at **a** 12.1 GHz, **b** 14.3 GHz, **c** 16.3 GHz

and the RDR antenna are shown in Fig. 4. The peak radiation efficiency of the antenna is 97.2% and peak gain is 9.01 dBi. The 3D radiation pattern of RDR antenna has shown in Fig. 5a–c at 12.1, 14.3 and 16.3 GHz. It is shown that the directional pattern and antenna reported high gain according to Fig. 5.

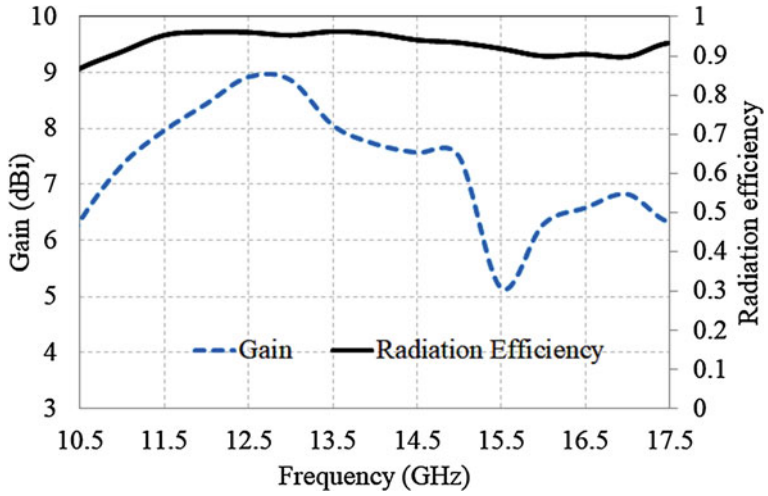


Fig. 4 Radiation efficiency and gain of RDR antenna

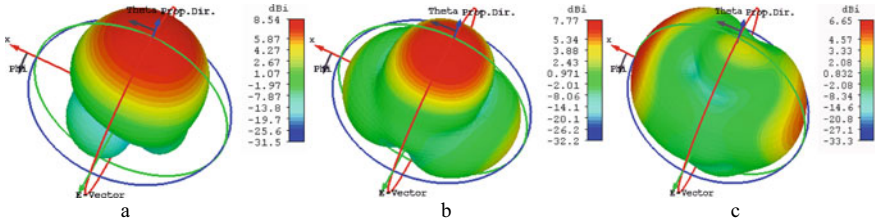


Fig. 5 3D radiation pattern at **a** 12.1 GHz, **b** 14.3 GHz, **c** 16.3 GHz

This proposed RDR antenna is compared with [28]. In reference, antenna haves range from 6.5 to 12.18 GHz with 4.8 dBic gain and circularly polarized. But the proposed compact antenna is linearly polarized for 5G application with peak gain 9.01 dBi as shown in Figs. 4 and 5.

3.1 Parametric Analysis

The performance of RDR antenna depends on the height, width and relative permittivity of DR. Antenna bounded all these parameters, hence radiation efficiency, directivity, gain and reflection coefficient influenced and apply parametric analysis on the height of RDR, the best value of S_{11} at 12 mm also to achieve good gain as shown in Fig. 6a and b. Another parameter, metallic strip height is set by the parametric sweep and to choose the best value at 5.5 mm for S11 and gain, shown in Fig. 6c and d.

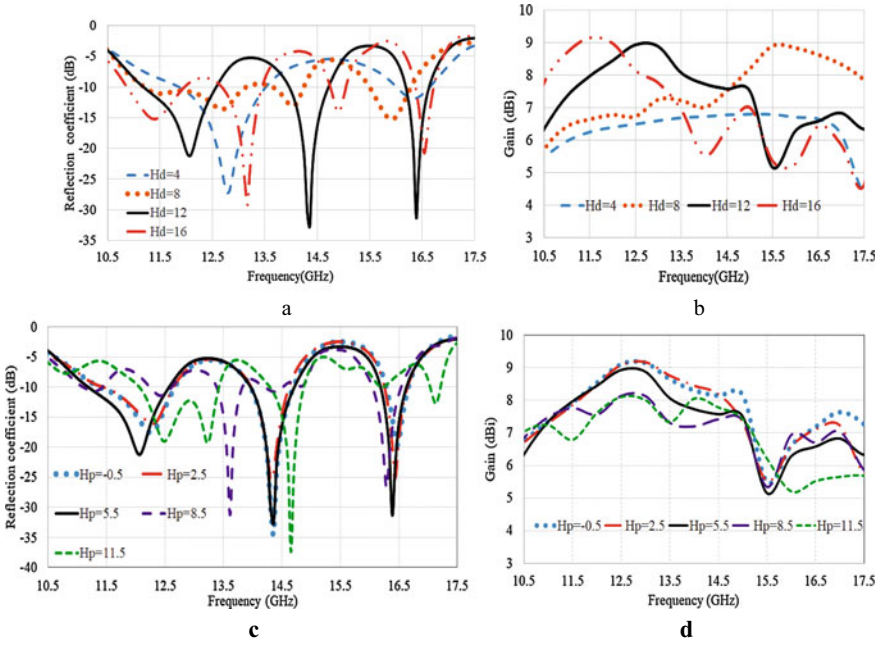


Fig. 6 Proposed RDR height of DR varies **a** S11. **b** Gain. Proposed RDR height of metallic strip varies **c** S11. **d** Gain

4 Conclusion

The proposed RDR antenna has cover Ku band applications in microwave radio of EM (electromagnetic) spectrum, particularly application in radar, satellite communication and wireless computer network. Ku satellite in DBSS (direct broadcast satellite services) and FSS (fixed services satellite) provides more flexibility to television transmission data. The first two bands are using for satellite uplink and downlink applications (12 GHz uplink and 14 GHz downlink) with a gain of 9.01 and 7.52 dBi. The simulated peak gains are 9.01, 7.8, 6.98 dB at 12.1, 14.3 and 16.3 GHz, and maximum radiation efficiency is reported 97.2%.

References

1. Long SA, Allister MMc, Shen L (1983) The resonant cylindrical dielectric resonator antenna. IEEE Trans Antenna Propag 406—412
2. Long SA, Allister MMc, Conway GL (1983) Rectangular dielectric resonator antenna. Electron Lett 19:218—219
3. Petosa A (2007) Dielectric resonator antenna handbook. Artech House, Norwood, MA, USA
4. Luk KM, Leung KW (2003) Dielectric resonator antennas. Research studies, Baldock, UK

5. Leung KW, Wong WC, Luk KM (1998) Annular slot coupled dielectric resonator antenna. Electron Lett 34:1275–1277
6. Mongia RK, Ittippiboon A, Bhatia P (1993) Electric monopole antenna using a dielectric ring resonator. Electron Lett 29:1530–1531
7. Leung KW, Luk KM, Yung EKN (1994) spherical cap dielectric resonator antenna using aperture coupled. Electronic Letters 30:1366–1367
8. Leung KW, Luk KM, Lai KYA (1993) theory and experiment of a coaxial probe fed hemisphere DRA. IEEE Trans Antennas Propag 41:1390–1398
9. Ittipiboon A, Mongia RK, Antar YMM (1993) Aperture fed rectangular and triangular dielectric resonators for use a magnetic dipole antennas. Electronics Letter 29:2001–2002
10. Kishk AA, Yin Y, Glisson AW (2002) conical dielectric resonator antenna for wideband application. IEEE Trans Antennas Propag 50:469–474
11. Fakhte S, Draizi H, Vadjad MH (2012) A high gain dielectric resonator loaded patch antenna. Prog Electromag Res 30:147–158
12. Luk KM, Leung KW, Chow KY (1997) Bandwidth and gain enhancement of dielectric resonator antenna with the use of stacking element. Microw Opt Technol Lett 14:215–217
13. Petosa A, Thirakoune S (2011) Rectangular dielectric resonator antennas with enhanced gain. IEEE Trans Antennas Propag 59:1385–1389
14. Yarga S, Sertel K, Volakis JL (2005) A dielectric resonator antenna using degenerate band edge crystals. IEEE Trans Antennas Propag 57:799–803
15. Junker GP, Kishk AA, Glission AW (1994) Input impedance of dielectric resonator antennas excited by a coaxial probe. IEEE Trans Antennas Propag 42:960–966
16. Kranenburg RA, Long SA (1988) Microstrip transmission line excitation of dielectric resonator antenna. Electron Lett 28:1156–1158
17. St. Martin JTH, Antar YMM, Kishk AA (1990) Dielectric resonator antenna using aperture coupling. Electron Lett 24:2015–2016
18. Leung KW, Luk KM, Lai KYA (1995) Theory and experiment of an aperture coupled hemispherical dielectric resonator antenna. IEEE Trans Antennas Propag 43:1192–1198
19. Leung KW (2000) Conformal strip excitation of dielectric resonator antenna. IEEE Trans Antennas Propag 48:961–967
20. Kranenburg RA, Long SA, Williams JT (1991) Coplannar waveguide excitation of dilectric resonator antennas. IEEE Trans Antennas Propag 39:119–122
21. Al Salmaeh MS, Antar YMM (2002) Coplanar waveguide fed slot coupled rectangular dielectric resonator antenna. IEEE Trans Antennas Propag 50:1415–1419
22. Eshrah IA, Kishk AA, Yakovlev AB (2005) Excitation of dielectric resonator antennas by a waveguide probe: modelling techniques and wideband design. IEEE Trans Antennas Propag 53:1028–1037
23. Eshrah IA, Kishk AA, Yakovlev AB (2005) Theory and implementation of dielectric resonator antenna excited by a waveguide slot. IEEE Trans Antennas Propag 53:483–494
24. Mongia RK, Bhatia P (1994) Dielectric resonator antennas—a review and general design relations for resonant frequency and bandwidth. Int J Microw Millimeter-Wave Comput Aided Eng 4:230–247
25. Rashidian A, Klymyshyn DM (2009) On the two segmented and high aspect ratio rectangular dielectric resonator antennas for bandwidth enhancement and miniaturization. IEEE Trans Antennas Propag 57:2775–2780
26. Fang XS, Leung KW (2011) Designs of single-, dual-, wide-band rectangular dielectric resonator antennas. IEEE Trans Antennas Propag 59:2409–2414
27. Petosa A, Thirakoune S, Ittipiboon A (2009) Higher-order modes in rectangular DRAs for gain enhancement. In: 13th International symposium on antenna technology and applied electromagnetics and the Canadian radio sciences meeting, vol 6, pp 1–4
28. Abedian M, Rahim SK, Danesh S, Jamaluddin MH, Islam MT (2016) Compact wideband circularly polarised dielectric resonator antenna. Electron Lett 53(1):5–6

Design and Fabrication of Electric Unicycle

Mohan Kumar, Yogesh Kumar, Thiruveedi Rahul, Ramit Dhingra, and Parul Sharma

Abstract This abstract presents a unique idea of electric unicycles in this field of automation industry. The common problem faced by society today is time management, reaching to places on time in a certain workspace. The main reasons are fatigue, lethargy, and exhaustion. This decreases the standard of one's lifestyle. This vehicle helps in short-distance travel purposes, also this vehicle targets reducing the spacing problems caused by conventional two wheelers in parking areas. Being an electric vehicle, it cannot produce any pollution. This electric unicycle is fully electric, lighter in weight, easy to operate after knowing its skills to ride, cheaper in cost, and totally pollution free. This electric unicycle operates on an electric motor and vehicle itself stabilizes by rider's skills. This electric unicycle is considered to be swift and mobile mode of transport. Moreover, within a practice period of 30 min, one can learn how to balance it. The design includes 215 W brushless direct current motors. The fabrication of the frame will be done by using circular pipes. The vehicle at maximum can attain a speed of 15–20 km/hr and travel up to 10 km in a single charge. It weighs approximately 40 kg and its load carrying capacity is around 100 kg. The electric unicycle can be used in areas like universities, airports, industries, etc.

Keywords One wheeled electric vehicle · Brushless direct current (BLDC) hub motor · Self-balance · Pollution free · Noise free · Centre of gravity · Stabilization

1 Introduction

The increase in pollution at the time is favoring toward electric vehicles. Consumers are being cautious and are considering electric vehicles more advanced, reliable compared with a typical petrol or diesel engine, as they are environment friendly. It goes without saying that, electric vehicles aid in reducing pollution. As of now, the necessity of electric vehicles in India is rising [2]. In this review paper, we have targeted to design and analyze a new concept of electric vehicles. The electric

M. Kumar · Y. Kumar · T. Rahul · R. Dhingra (✉) · P. Sharma
HMR Institute of Technology and Management Hamidpur, Delhi 110036, India

Guru Gobind Singh Indraprastha University, Sector—16C, Dwarka, Delhi 110075, India

J. Singh et al. (eds.), *Innovations in Cyber Physical Systems*, Lecture Notes in Electrical Engineering 788, https://doi.org/10.1007/978-981-16-4149-7_74

unicycle is a downsized version of a conventional motorcycle that is driven by a hub motor [5]. A decent speed can be achieved in just seconds and is environmentally friendly. This compact vehicle is just as fun to ride as it is sensible for the small commute. Subsequent paragraphs, however, are indented.

1.1 Components of Electric Unicycle

The design of electric unicycle is totally based on its components, the major components used in the unicycle are as follows.

1.1.1 Brushless Direct Current (BLDC) Hub Motor

A conventional DC motor is equipped with a commutator and brushes in order to reverse the current through the rotating armature coil. The brushes come in sliding contact with commutator, which may cause sparks that in long run result in wearing of commutators and brushes. To overcome this drawback, brushless motor was designed. Essentially, they consist of a sequence of stator coils and a permanent magnet rotor. A current-carrying conductor in a magnetic field experiences a force; likewise as a consequence of Newton's third law of motion, the magnet will also experience an opposite and equal force. With the brushless DC motor, the magnet is free and the current-carrying conductor is made to move. The wheel hub motor is minimalistic design, where the motor is housed inside of the hub of the wheel itself. This design is extremely efficient as it deducts the transmission setup as of conventional vehicles [6]. Hub motor wheel is compact and can produce decent rpm (Figs. 1 and 2).

1.1.2 Accelerometer

An accelerometer is a device that measures the acceleration of the vehicle. It has a broad application in electrical industries. An accelerometer is also capable of dispensing commands to a stepper motor. Accelerometers used in aircraft industries are highly sensitive. Each one of learning of the accelerometer will transmit a sign or signal to the stepper motor and further these signs direct the throttle for acceleration [8].

1.1.3 Handlebar and Steering Mechanism

More specifically, the handle bar is linked to the body, body is linked to the wheel forks, and further wheel, forks are further linked to the wheel. It is well known that when a rider rotates the handle of the unicycle, the center of gravity of unicycle/rider

Fig. 1 Hub wheel assembly

Fig. 2 Hub motor

combo changes and that influences the direction of the writer. Rotating the unicycle's handlebar to the left or right discretely push and pull the wheel fork thereby applying force to alter the direction of the rider.

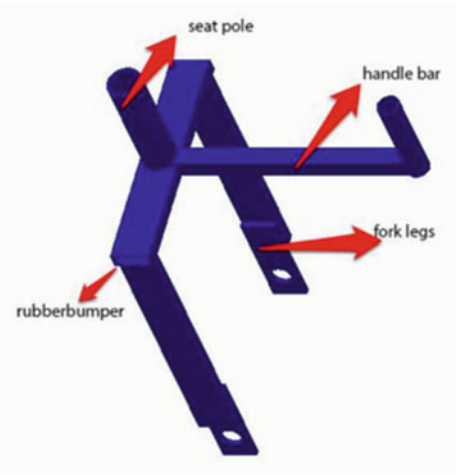


Fig. 3 Fork

2 Design and Implementation

The process began with acquiring data, information, and knowledge of the concept and studying already published research papers. With this gained knowledge, developing a design and analyzing our unique model of electric unicycle [7]. The standard components that had to be kept similar for the functioning of the electric unicycle are the fork assembly, the main chassis. In designing, we focused on a few key aspects that are; ease of manufacturing, optimal center of gravity, durability, and design flexibility. To achieve the optimum design, we performed many iterations.

2.1 Fork

Fork design mechanism is used to deal with alignment issues in motor rim combination. In order to maintain symmetry between wheels and spindle, center plane of wheel is set to coincide with center plane of spindle. If they are not aligned properly, it results to failure of tyre not being in plane central to the rider. The fork structure includes a rubber bump at its horizontal end to prevent the electric unicycle from getting damage in case of collision (Fig. 3).

2.2 Mechanical System Implementation

The design process was completed with one key aspect that the unicycle should be fully rideable, with some practice beforehand to polish the skills required for

Fig. 4 Frame assembly

riding the unicycle. The main focus was on the weight balancing of the batteries, and the placement of the seat on the unicycle [3,4]. The mechanical design included the chassis, axle of the wheel, center plate, and a fork assembly to hold the wheels together. These can be seen in the rendered solid model. The seat is placed on the top of the plate. We have also joined a cage to the fork, to prevent the rider from accidentally falling forward or backward. The steering mechanism is simply a steering rod with the help of which the unicycle can be turned. The footpegs are placed at a decent height on the chassis. The brakes are mechanical and are placed on the side of the wheel (Fig. 4).

3 Inverted Pendulum Principle

The present issue with the balancing of the unicycle corresponds to the issue of the 'inverted pendulum' or we can say 'reverse pendulum' problem. A reverse pendulum is specifically a pendulum in which its concentration is above its pivot. Whereas a conventional pendulum is steady and balanced, the reverse pendulum/inverted pendulum is congenitally unsteady and might be actively stabilized to past upstanding [1] (Fig. 5).

3.1 Design of Unicycle on Solid Works and Ansys

(See Figs. 6 and 7).

Fig. 5 Inverted pendulum

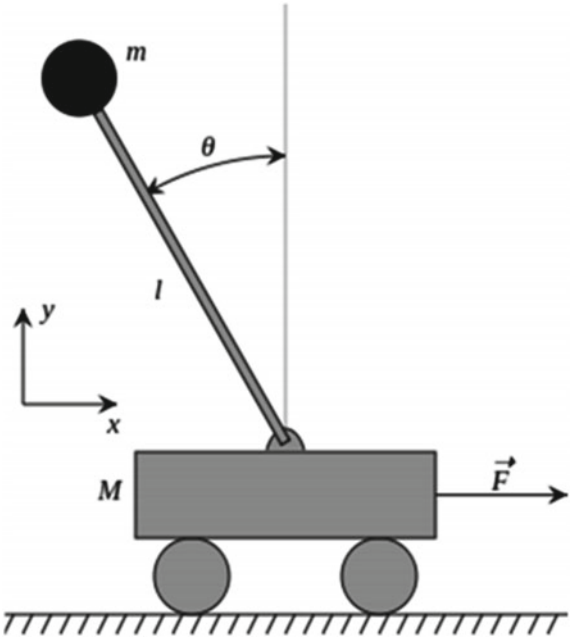


Fig. 6 Software design

4 Conclusion

This research paper gives a brief demonstration of the design and fabrication of an electric unicycle, which is electrically powered by a battery and runs on a brushless DC motor. Whereas, in the composition part, some contrasting approaches for the evolution of this electric unicycle have been revised right after its origination or invention.

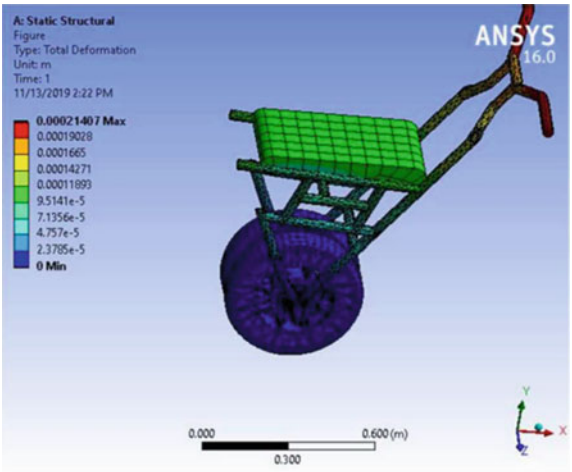


Fig. 7 Ansys model

References

1. Chouhan M, Yusufi A (2016) One wheeled self-stable electric bike. Department of Automobile Engineering, MIST, vol 3, issue 03
2. Pavan Kumar SV, Maddukuri BS (2015) Design and fabrication of ICYCLE, vol 3, no 5
3. Gadge H, Kadam N, Dhatavkar S (2018) Design, development and analysis of self-balancing electric bike, vol 9, issue 5
4. Cazzolato B (2010) Modelling, simulation and control of an electric unicycle
5. Arioui H (2011) Review on single track vehicle and motorcycle simulators
6. Agnihotri A, Dey P, Kalita D, Dutta SS (2015) An investigation study based on emerging demand of electric unicycle vehicles, vol 5, issue 5
7. https://einsty.com/how-selfbalancing-bike-works/
8. https://en.wikipedia.org/wiki/Electric_unicycle