Lecture Notes in Electrical Engineering 637

V. Bindhu
Joy Chen
João Manuel R. S. Tavares *Editors*

International Conference on Communication, Computing and Electronics Systems

Proceedings of ICCCES 2019

Springer

Lecture Notes in Electrical Engineering

Volume 637

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, Associate Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Executive Editor (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)

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, MetaPress, Web of Science and Springerlink ****

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

V. Bindhu · Joy Chen · João Manuel R. S. Tavares
Editors

International Conference on Communication, Computing and Electronics Systems

Proceedings of ICCCES 2019

Springer

Editors
V. Bindhu
Department of ECE
PPG Institute of Technology
Coimbatore, Tamil Nadu, India

Joy Chen
Department of Electrical Engineering
Dayeh University
Changhua City, Taiwan, Taiwan

João Manuel R. S. Tavares
Faculty of Engineering
University of Porto
Porto, Portugal

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-15-2611-4 ISBN 978-981-15-2612-1 (eBook)
https://doi.org/10.1007/978-981-15-2612-1

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

We are honored to dedicate the proceedings of ICCCES 2019 to all the participants, organizers and editors of ICCCES 2019.

Preface

This conference proceedings volume contains the written versions of most of the contributions presented during the ICCCES 2019. The conference has provided a platform to share and exchange the recent developments in a wide range of topics including computational intelligence, machine learning, signal and image processing, electronic devices and systems, antenna and wave propagation, wireless communication networks and so on. The conference has been a good opportunity for participants coming from various destinations to present and discuss the state-of-the-art topics in their respective research areas.

ICCCES 2019 tends to collect the latest research results and applications on computing, communication and electronics. It includes a selection of 71 papers from 215 papers submitted to the conference from various universities and industries present across the globe. All the accepted papers were subjected to a double-blinded peer-reviewing process by 2–4 expert referees. The papers are selected for its high quality and the relevance to the conference.

ICCCES 2019 would like to express our gratitude and appreciation to all the authors for their valuable research contributions to this book. We would like to extend our thanks to all the referees for expressing their constructive comments on all the research papers. In particular, we would like to thank the organizing committee for their tireless hard work. Finally, we would like to thank the Springer publications for producing this volume.

Dr. V. Bindhu
Conference Chair, ICCCES 2019

Head of the Department of ECE
PPG Institute of Technology
Coimbatore, India

Acknowledgements

We extend our gratitude to thank all who make this conference event successful. Moreover, we thank our institution, PPG Institute of Technology, Coimbatore, Tamil Nadu, India, for their immense support and a very efficient assistance during the conference event.

The conference organizers are particularly grateful to all the researchers, academicians and industrialists, who contributed their state-of-the-art research works with the conference. Furthermore, the organizers gladly acknowledge the efforts made by **Dr. V. Bindhu**, who helped out with the organization of conference event at all instances. We are very grateful to the conference keynote speaker **Dr. João Manuel R. S. Tavares**, who delivered valuable research insights and expertise, which may greatly assist the future research works of the conference attendees. We thank all our advisory and review committee members for their valuable reviews that greatly improved the manuscript.

Further, we wish to extend our thanks to all the potential authors, who contributed their research works to enhance the research quality of ICCCES 2019. Last but not least, we would like to thank all the session chairs and organizing committees for contributing their tireless efforts to this successful conference event.

About the Conference

It is with deep satisfaction that I write this Foreword to the Proceedings of the ICCCES 2019 held during November 15–16, 2019, at PPG Institute of Technology, Coimbatore, Tamil Nadu, India.

The conference was covering the latest research theories, methods and applications in all the areas of computing, communication and electronics.

This conference particularly encouraged the interaction of researchers, academicians and industrialists to share their research insights and results on the cutting-edge methods and applications. The research manuscripts contributed to the conference have the new innovative research works in the field of image and signal processing, machine learning, electronics devices and systems, Internet of things, wireless networks, antenna and wave propagation, intelligent systems and computational intelligence. The contribution of potential ICCCES 2019 authors has greatly helped to compile a high-quality ICCCES 2019 proceedings into existence. The local organizing committee members have put much effort in ensuring the success in each and every aspect of the conference meeting.

We hope that this conference program will strongly encourage the research efforts in the areas of design, analysis and implementation of computing, electronics and communication technologies. We feel honored and grateful to showcase the best recent research developments to you through this esteemed conference program.

We thank all the authors, participants and organizers for their contributions.

Dr. V. Bindhu
Conference Chair, ICCCES 2019

Head of the Department of ECE
PPG Institute of Technology
Coimbatore, India

Contents

About the Editors

Dr. V. Bindhu is a Professor at the Department of ECE, PPG Institute of Technology, Coimbatore, India. She completed her Ph.D. in Information and Communication Engineering at Anna University, M.E. at Maharaja Engineering College, and B.E. at Government College of Technology, Coimbatore.

Joy Chen received his B.Sc. degree in Electronics Engineering from the National Taiwan Technical University, Taipei, Taiwan, his M.Sc. degree in Electrical Engineering from Dayeh University, Changhua, Taiwan, in 1985 and 1995, respectively, and his Ph.D. degree in Electrical Engineering from the National Defense University, Taoyuan, Taiwan, in 2001. He is currently a Professor at the Department of Communication Engineering, Dayeh University at Changhua, Taiwan. Prior to joining Dayeh University, he worked as a technical manager at the Control Data Company (Taiwan). He has published about 40 international journal papers and acted as guest editor for several international journals. His research interests include AI, IoT development, wireless communications, spread spectrum technical, OFDM systems, and wireless sensor networks.

Dr. João Manuel R. S. Tavares graduated in Mechanical Engineering from the University of Porto - Portugal (1992). He holds an M.Sc. in Electrical and Computer Engineering, with a focus on industrial informatics, from the University of Porto (1995); and a Ph.D. in Electrical and Computer Engineering, from the University of Porto (2001). From 1995 to 2000, he was a researcher at the Institute of Biomedical Engineering (INEB). He is the co-author of more than 350 scientific papers published in national and international journals and conferences, co-editor of 18 international books, and guest editor of several special issues of international journals. In addition, he is editor-in-chief and associate editor of various journals and also a reviewer for several international scientific journals. He has also supervised and co-supervised several M.Sc. and Ph.D. books and been involved in numerous research projects, both as a researcher and as a scientific coordinator.

Additionally, he is co-author of 3 international patents and 2 national patents. His main research areas include computational vision, medical imaging, computational mechanics, scientific visualization, human–computer interaction, and new product development. He has been the co-chairman of various international conferences and numerous mini-symposia, workshops, and thematic sessions. In addition, he has been a member of scientific and organizing committees of several national and international conferences.

Enhancing the Performance of Software-Defined Wireless Mesh Network

Nithin Shastry and T. G. Keerthan Kumar

Abstract In a software-defined wireless mesh network, a centralized manner of managing and monitoring of the network occurs. The software-defined network (SDN) is the future of the upcoming generation network paradigm by separating control plane and data plane such that network management and optimization can be conducted in a centralized manner using global network information. In this paper, we proposed a novel architecture of software-defined wireless mesh networks to identify the issues of traffic balancing introduced due to node mobility. In order to reduce the overall response time of the SDN controller in the dynamic network topology, the new model predicts the probability of the link failure in the topology. Once the link failure is predicted, alternate selection of various routes proposed through the effective stability of traffic in the network is achieved and thereby overhead of the control plane is minimized. Utilizing ns-3 to efficiently address the above problem, we can enhance the network throughput and packet delivery fraction and minimize the delay in the network. Finally, performance is evaluated via extensive simulations.

Keywords Software-defined network (SDN) · Performance · Wireless mesh networks · Dynamic spectrum access · NS3 tools · OLSR daemon · Ad hoc network · Cognitive radio · Network coding · Radio spectrum management · Throughput

1 Introduction

The scope of study and use of wireless mesh network (WMN) has been increased in the last decade. In the WMN, the nodes of the WMN can connect automatically by searching the routers in order to maintain the connectivity needed in the network [1]. The nodes in the capacity of the network and the service coverage at the low cost can be accomplished using this kind of distributed network architecture [1]. But there are certain drawbacks with respect to the distributed network architecture [2].

N. Shastry · T. G. Keerthan Kumar (✉)
Information Science and Engineering, Siddaganga Institute of Technology, Tumakuru, Karnataka 572103, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_1

The vital few among them are high communication latency and complicated network configurations. The global network management can be successfully accomplished through the usage of the software-defined networks. The characteristics of the SDNs are such that there exists a centralized control through its control plane which can facilitate the network management on large scale. These SDNs are widely applied where the nodes in the networks are often static and links have high-speed connectivity [1]. There are certain disadvantages and limitations in the existing conventional WMNs, which can be overcome through the application of the SDN architecture with the combination of the wireless mesh networks. In the data plane in order to provide effective communication between the conventional and SDN-based wireless devices, a routing scheme of hybrid approach is designed [2]. Initially, OLSR to open flow architecture was proposed. In the similar fashion, a hybrid architecture consists of software-defined network which is designed for ad hoc networks. The comparison of the performance was evaluated, and the result impacted that this was more effective when compared to OLSR protocol with respect to dynamic network topology [3]. Then, as the impact of the above result, many traffic scheduling algorithms were discussed with respect to the software-defined network-based wireless mesh network architecture. One of the most critical tasks that need to be addressed is the problem of traffic balancing which exists in the software-defined network control plane. The various link conditions and the allocation of the traffic based on various link conditions can be effectively implemented using controllers [4]. But the most challenging part with respect to functions of the traffic engineering of the control plane is the resolving of the link failure mechanism if exists in the control plane [5]. One of the solutions for the above problem with regard to link failure mechanism in the control plane is constant monitoring of the data plane and increasing the response time with respect to the uncertainty that exists in the link. However, the overhead that exists in the network and the degradation of the throughput can be introduced frequently through the up-gradation of the flow-tables [6]. An alternative solution that could be proposed is to first divide the whole network into many sub-divisions, thus having centralized regulator which is assigned to each and every region. The rest of this paper is organized as follows. Section 2 describes the literature survey. Section 3 formally defines the proposed system. Implementation and result analysis are presented in Sects. 4 and 5, respectively. Finally, we conclude the paper in Sect. 6.

2 Literature Survey

Cognitive radio (CR) is one of the best promising solutions in spectral underutilization. Cognitive radio network (CRN) has two types of users who are primary and secondary. The latter can use the band till former arrives. Here, for congestion avoidance and throughout increasing dynamic generation size adjustment algorithm as well as implementing interflow coding, we use ns-3 simulator [7]. Applying the utilization of the present spectral features of CRN and also not disturbing the PU access to the radio asset is an important fact here. When we consider the cooperative sending methods to increase the accuracy, we encounter the security threat when we collaborate with SUs since one node that is malicious can change the outcoming reports in accordance with personal benefit. Here, we are trying to explore how with the usage of many routes we can reduce the shortcomings of the node that is malicious with the application of multi-routes versus general uni-route results which are viewed through ns-3 [8]. The measurement of energy was generally for finding out signal energy and not for separating on the individual observation. Two-stage process is learning the components analogous to one in soft reports. In case of algorithm of hard reports, it is used in decreasing the difficulty of computing with the help of learning the 1-D mixtures in every CR and mixing the results using an algorithm that is acceptable as a novel. We detect the highest weight that is predominantly present in the graph that is directed. Using MATLAB we can simulate the evaluation of algorithms for the network performance with difference parameter. For cases where often collision occurs, the performance of soft reports kind of algorithm stands high. But when we take 802.1 simulation in ns-3 due to the capture effects of the channels, a deterioration can be observed [9]. Here, we design and implement and also check the protocol named dynamic spectrum access (DSA) in the CRNs. This is based on the assignment of the values of credit for proper and fair allocation of the resources among users. This is accomplished in the way of associating a value and subsequently finding out the level of trust, and later, resources and spectrum are duly allocated among the users in accordance with the values obtained. Carrier-sense multiple access (CSMA) [10] is relied upon Collision resolution and it performs well even under various scenarios like transmission over wireless access networks and providing accountability in sharing the files on cloud [11, 12].

3 Proposed System

The proposed system signifies a unique method of architecture that exists in a network, concept for designing the network for coupling the control plane and data transmitting plane, and network-oriented programmable media transfer between the nodes [13]. The main characteristic of SDN is shown as the following method of functions:

(A) Centralized control: Differentiation of the planes are done by the software-defined networking , in this the work of the forwarding plane is transmitting forward packets and simple process function in the network, workability, and alteration are effectively attained through centralized SDN control plane. The network management operation is simplified by the centralized network, hence increasing the speed of implementation which can be useful to the rapid development and invention of the network, and it is also beneficial to the global optimization of the network.

(B) Interface that consists of software-defined networking: In the present software-defined network, we need an open set of inputs that can be obtained in the network. In order to achieve the optimized congestion control in the network, we need to achieve the optimal traffic balancing and to reduce the packet loss in the network and to achieve the optimum traffic balancing that can be achieved in the network.

Advantages

- Delay reduction: In single-hop, there is a direct link between two devices in device-to-device communication which reduces the delay as compared to the minimum traditional network of two-hop link.
- Energy-efficient improvement: While communicating each device with BS in traditional cellular network, it requires high power; device-to-device communication can use lower energy due to smaller distance between each other in device-to-device communications which increases the energy efficiency of D2D devices.
- Coverage improvement: In device-to-device communication of multi-hop, D2D load balancing which is out of the range or in the closed spot regions could use other SDN load balancing networking devices as relays to communicate with the infrastructure network or to communicate with other receiver users, resulting in overall network development and improving coverage of network.

3.1 LEACH Algorithm

The above-mentioned protocol (as shown in Fig. 1) effectively works for the nodes considered as primary and secondary nodes.

In order to optimize the distribution of the power, it should be proportional such as sensory nodes and these nodes have the power to become cluster head using CH selector algorithm [14].

Step 1: There are various mechanisms that are associated with the above existing algorithm. The LEACH algorithm has two stages; namely, the first stage is setup stage and the last stage consists of the state of steady stage. For every nearest cluster head, the data that has been arrived is sent.
Step 2: In the second stage the cluster head is formed and setting up the cluster head selection and in order to effectively transmit and implementing effective slots for scheduling the time, phase consisting of the steady state will be implemented. The nodes present in the network are independent and they can determine if they have the possibility of choosing any nodes having sensing capability.

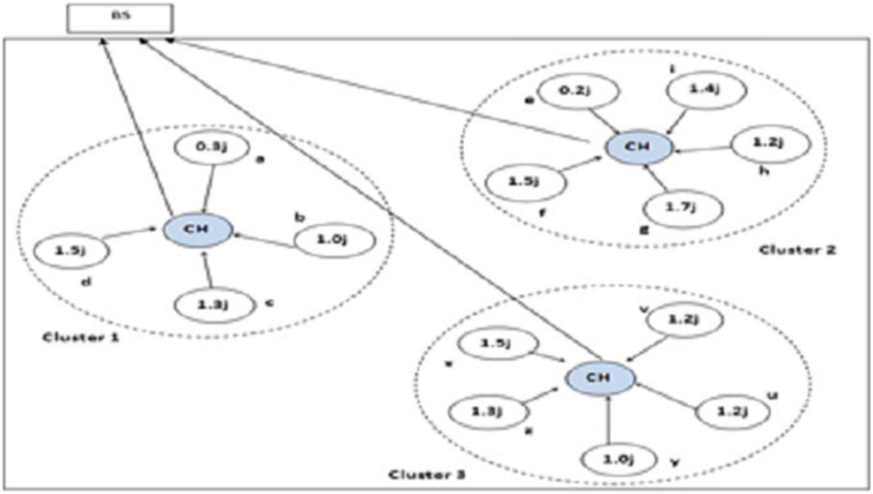


Fig. 1 Diagrammatic representation of LEACH algorithm

3.2 Mathematical Equation of SD-WMN

Let the above graph be denoted as $G = (V, E)$, here V denotes the vertices present in the graph and E denotes the edges present in the graph. The TD_v represents various demands associated with the traffic by considering the flow of data associated with one particular channel.

$$\mathrm{TD}_v = \left\{ f_p^{\mathrm{ID}:1}, f_p^{\mathrm{ID}:2}, \ldots, f_p^{\mathrm{ID}:n} \right\} \tag{1}$$

For any particular distribution associated with the various traffics, the below model denotes traffic for the model:

$$\left\{ f_{A1}^{\rho 1}, f_{A2}^{\rho 2}, f_{A3}^{\rho 3}, \ldots f_{An}^{\rho n} \right\} = \Delta_v \times \mathrm{TD}_v \tag{2}$$

where f_A^ρ indicates the flow rate associated with the particular node which sends through a node on the path. The below equation denotes the representation.

$$f_A^\rho = \left\{ \alpha_A^{f(\mathrm{ID}:1)} f_{P=\mathrm{in}}^{\mathrm{ID}:1}, \alpha_A^{f(\mathrm{ID}:2)} f_{P=\mathrm{in}}^{\mathrm{ID}:2}, \ldots, \alpha_A^{f(\mathrm{ID}:n)} f_{P=\mathrm{in}}^{\mathrm{ID}:n} \right\} \tag{3}$$

3.3 System Architecture

The topology consisting of the CRHAN [15] is formed on the basis of the network consisting of various cognitive radio devices. The various devices that are present in the CRHAN topology include HAN gateway, HAN controller with cognitive capability, and the number of cluster heads denoted by CHs. This type of network is associated with HAN [16] which is linked with the help of links consisting of multiple hops. The way in which communication takes place is in mesh way. The various functionalities are described briefly in the below paragraphs. In order to communicate with the Internet and other networks, we use HAN gateway and the connection to the Internet is provided with the means of various Internet service providers. In order to manage various usages related to HAN, we use a controller named cognitive HAN controller which needs to be constructed based on the environment associated with the HAN as shown in Fig. 2.

The construction consists of channels that are associated with the cluster heads and the information that is sent to the cluster heads is to be cast like the sensing that occurs in medium access control. Henceforth, the channels that are provided for the cognitive heads are to be effectively utilized for the clusters in the network. The channels associated with the cluster heads are called inbound channels. In order to manage the in-band channels access and other operations such as sensing, we need

Fig. 2 Diagrammatic representation of HAN controller

to use cluster heads. Cluster heads have the ability to request more than one channel for data transmission, and it has the capability for effective utilization of the channel present in the network. Apart from that, it also has the capability to perform sensing task and distributed the data across the networks in HAN devices which are based on the cognitive radio [17], and it also acts as a center of fusion for locating the devices that have sensing capabilities, and effective channel utilization can happen in order to have the data flow across the network.

The various applications that include safety and security are carried out with the help of devices associated with HAN. The function of these devices is to communicate and connect various clusters present in the network, and it needs to perform efficiently how the various data that needs to be transmitted needs to be carried out so that there should be less delay and effective utilization of the network. Figure 3 shows diagrammatic representation of cognitive HAN network.

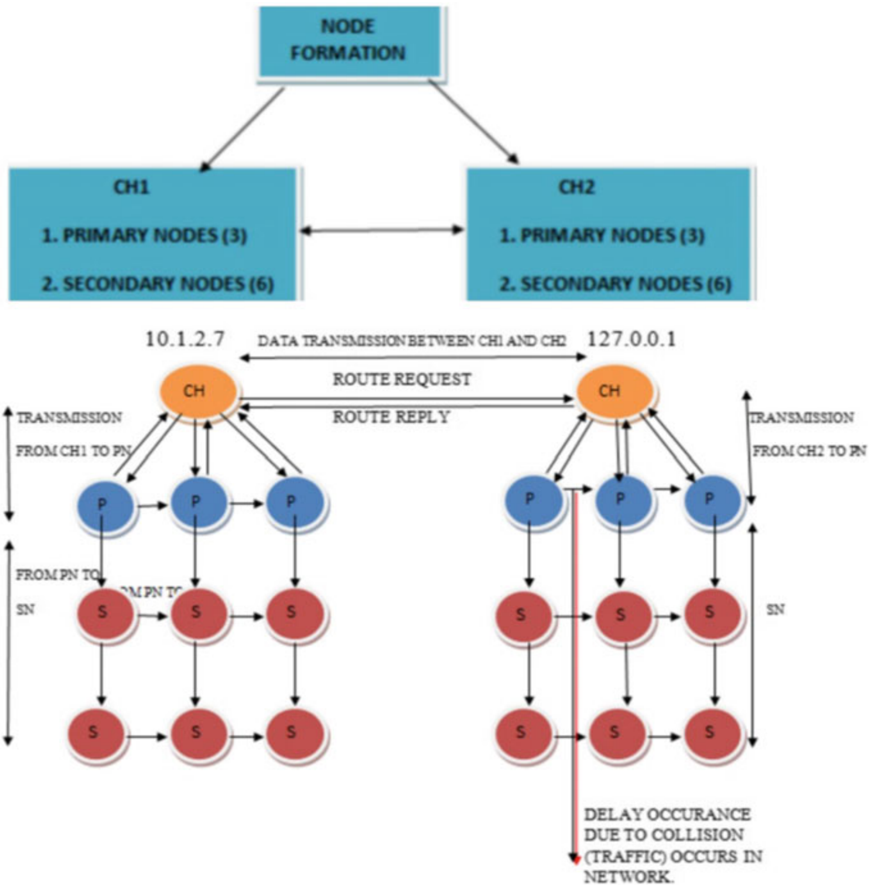


Fig. 3 Diagrammatic representation of cognitive HAN network

3.4 Performance Evaluation

1. Difference in various timings that are linked with Internet:

We can observe that the traffics present in the real world can change continuously within limited duration and occur in particular number of data that is transmitted [18]. We can also infer from the fact that for a brief time period, changes that occurred have been observed. Network users follow a same method because the fact occurs in Internet traffic. Thus, it causes difference in various timings that are linked with net.

2. Availability based on the slot time:

Probability in approximating the different path in low costs factor is difficult, and in particular, it cannot be solved. This proposed method has the lateral approach that is denoted as TABR. In this type of method, primarily there are two types of method, namely primary and secondary. Links can be shared among themselves in the existing paths which results in the reliability that exists between the connections.

Disadvantages

1. Throughput constraints: The throughput constraints of wired network are high when compared to wireless connection and it can be observed that wired network connection range is in the range of gigabytes per second, whereas wireless networks have the range in terms of around 2 Gbps.
2. High latency: When the node has no data to transmit, at that time energy conserving will be low or idle in the network. While transmitting the data between the different nodes when it goes through the idle, at that time we need to use the routing algorithm to wake up the node to reduce the delay.

4 Implementation

In this implementation section, we will briefly discuss various parameters that are associated with the successful implementation of the associated LEACH algorithm that is necessary for network implementation. The main concept is the formation of the cluster heads in the network which are denoted as cluster head 1 and cluster head 2 [19].

CLUSTER HEAD FORMATION:

PARAMETERS:

1. CH1 (Cluster Head 1).
2. CH2 (Cluster Head 2).
3. PRIMARY NODE (PN).
4. SECONDARY NODE (SN).

CH1 and CH2 Formation

Figure 4 consists of various nodes that are communicated wirelessly. In order to communicate with each and every node, we use various WiFi protocols such as yet another network simulator. It also consists of cluster head which controls and coordinates various other nodes that are connected wirelessly [20]. The user then can interact with the nodes from the intermediate base station so that the desired information for the user can be effectively sent from the nodes and the connection between the base station and the user is wired connection and the interaction can be effectively communicated.

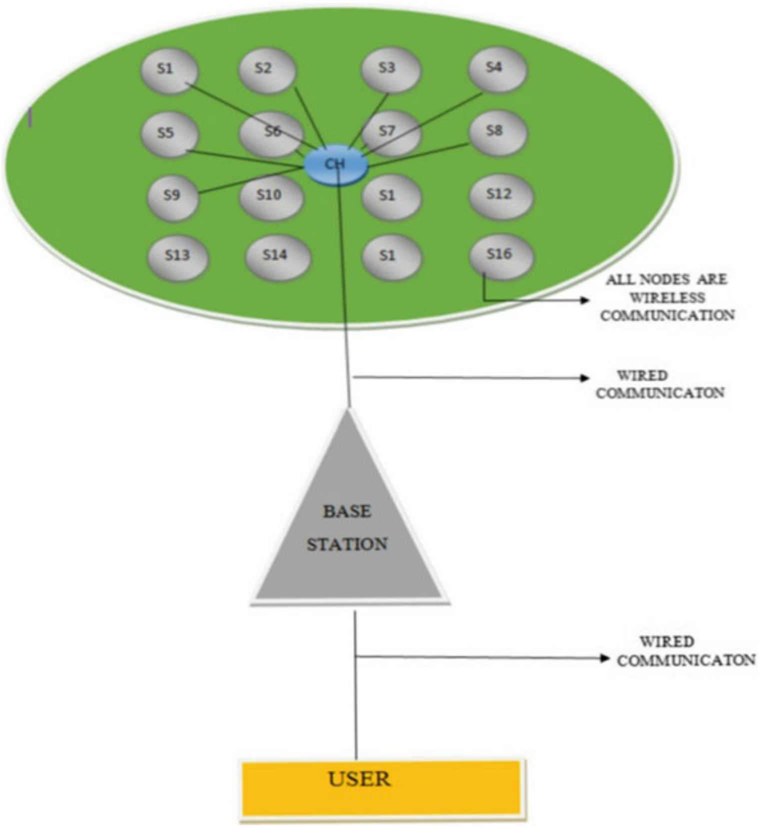


Fig. 4 Group of nodes with cluster head

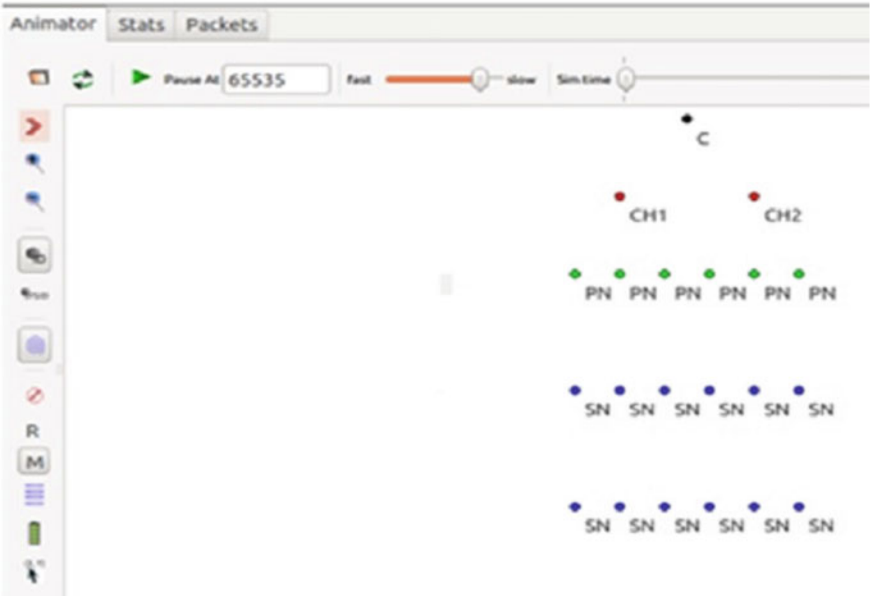


Fig. 5 Representation of node creation in net animator

5 Results and Analysis

5.1 Node Creation in Net Animator

Figure 5 consists of a single gateway with the two cluster nodes. It consists of primary nodes and the secondary nodes. Primary nodes are represented by PN and the secondary nodes are represented by SN. It also contains the stationary nodes and the mesh nodes. Mesh nodes are used for data transmission, while the stationary nodes are used for data receiving.

5.2 Data Transmission Between the Nodes in Net Animator

Figure 6 represents data transmission between nodes in net animator. To see the transmission of data through those nodes, we use net animator. It clearly denotes various nodes as well as various WiFi protocols in order to transfer data between various nodes in the network.

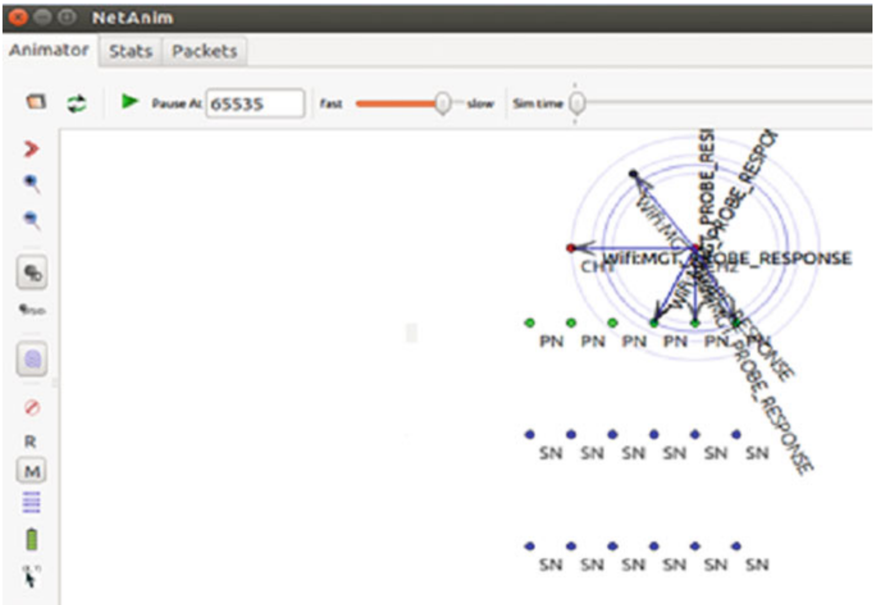


Fig. 6 Representation of data transmission between the nodes in net animator

5.3 *Graphical Approach*

5.3.1 Bit Error Rate (BER)

Figure 7 represents the bit error rate. It tells about the number of packets dropped from the whole packets. So, our simulation result shows graph of packet drop versus simulation time. The red color line indicates the results of the proposed method.

6 Conclusion and Future Enhancement

In this paper, we elaborately dealt with the various issues that are associated with the wireless mesh network and the various cumbersome that exists in maintaining the issue that exists in mobility associated with the various nodes. We have also used support vector machine schema that successfully predicts the failure of the link associated with the network, and we can easily choose an alternate path for the message to be transmitted in the network. We have efficiently utilized the various features and the application of the network simulator-3 to implement and enhance the working of the network in the specified scenarios where congestion occurs and the choosing of the alternate path in order to transmit data so that we can avoid the packet loss and effectively increase the throughput of the network. We have successfully

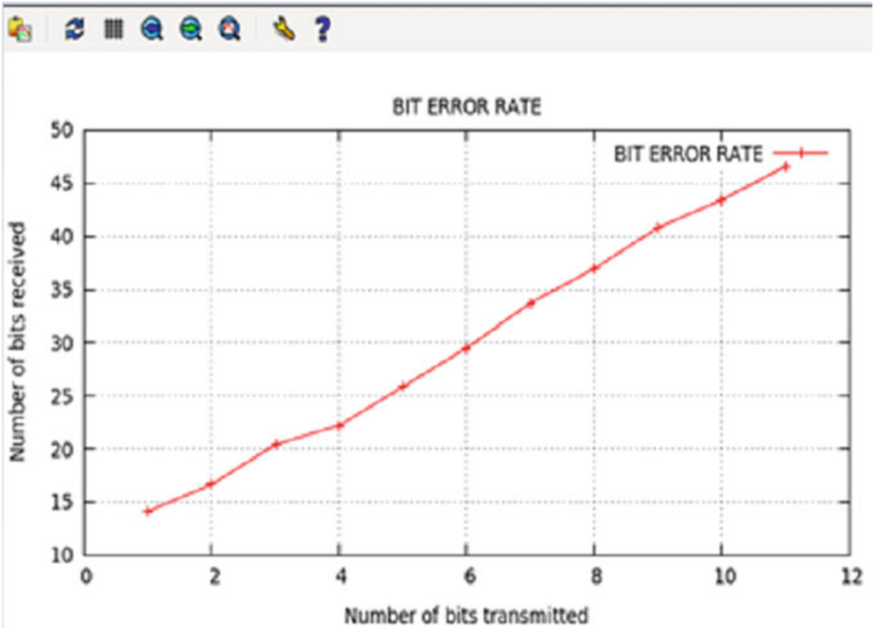


Fig. 7 Representation of BER

reduced the overall possibility that can occur to lose packet in the network, thus avoiding the bit error rate to minimal and increasing the overall performance of the system. Thus, the implementation of the corresponding system in the real-world scenario will have effective impact on overall efficiency of the packet delivery and stability of the network. In future efficiency can be enhanced using various other simulators like OMNeT++ and enhanced version of prevention of congestion in the network can be done so that the efficiency of software-defined wireless mesh network can be increased.

References

1. Peng, Y., Guo, L., Deng, Q., Ning, Z., Zhang, L.: A novel hybrid routing forwarding algorithm in SDN enabled wireless mesh networks. In: 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems, New York, NY, pp. 1806–1811 (2015).https://doi.org/10.1109/hpcc-css-icess.2015.271
2. Detti, A., Pisa, C., Salsano, S., Blefari-Melazzi, N.: Wireless mesh software defined networks (wmSDN). In: 2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Lyon, pp. 89–95 (2013).https://doi.org/10.1109/wimob.2013.6673345

3. Abolhasan, M., Lipman, J., Ni, W., Hagelstein, B.: Software-defined wireless networking: centralized, distributed, or hybrid? Netw. IEEE **29**, 32–38 (2015). J. Netw. Comput. Appl. **61** n.C, 199–221 (2016). doihttps://doi.org/10.1016/j.jnca.2015.11.012
4. Yu, H.C., Quer, G., Rao, R.R.: Wireless SDN mobile ad hoc network: from theory to practice. In: 2017 IEEE International Conference on Communications (ICC), Paris, pp. 1–7 (2017). https://doi.org/10.1109/icc.2017.7996340
5. Labraoui, M., Boc, M., Fladenmuller, A.: Self-configuration mechanisms for SDN deployment in wireless mesh networks. In: 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), Macau, pp. 1–4 (2017).doi: 10.1109/WoWMoM.2017.7974352
6. Magdalene, W., Let, G.S.: Implementation of dynamic generation size adjustment algorithm for cognitive radio ad-hoc network. In: 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, pp. 119–122 (2016). https://doi.org/10.1109/wispnet.2016.7566103
7. Fathy, M., Tammam, A., Saafan, A.: Mitigating the impact of malicious behavior via utilizing multiple routes in a cooperative sensing cognitive radio network. In: 2017 IEEE 15th Student Conference on Research and Development (SCOReD), Putrajaya, pp. 247–252 (2017). https://doi.org/10.1109/scored.2017.8305384
8. Laghate, M., Cabric, D.: Cooperatively learning footprints of multiple incumbent transmitters by using cognitive radio networks. IEEE Trans. Cogn. Commun. Netw. **3**(3), 282–297 (2017). https://doi.org/10.1109/TCCN.2017.2710309
9. AlShammari, T., Hamdaoui, B., Guizani, M., Rayes, A.: Overcoming user selfishness in DSA systems through credit-based resource allocation. In: 2014 IEEE International Conference on Communications (ICC), Sydney, NSW, pp. 318–323 (2014). https://doi.org/10.1109/icc.2014.6883338
10. Choi, H., Lee, I., Lee, H.: Delay analysis of carrier sense multiple access with collision resolution. J. Commun. Netw. **17**(3), 275–285 (2015). https://doi.org/10.1109/JCN.2015.000050
11. Fu, C.P., Liew, S.C.: TCP Veno: TCP enhancement for transmission over wireless access networks. IEEE J. Sel. Areas Commun. **21**(2), 216–228 (2003). https://doi.org/10.1109/jsac.2002.807336
12. Keerthan Kumar, T.G., Virupakshaiah, H.K., Nanda K.V.: Ensuring an online chat mechanism with accountability to sharing the non-downloadable file from the cloud. In: 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Bangalore, pp. 718–721 (2016). https://doi.org/10.1109/icatcct.2016.7912093
13. Chen, Z., et al.: A novel bandwidth estimation algorithm of TCP westwood in typical LTE scenarios. In: 2015 IEEE/CIC International Conference on Communications in China (ICCC), Shenzhen, pp. 1–5 (2015).https://doi.org/10.1109/iccchina.2015.7448600
14. Alrshah, M.A., Othman, M.: Performance evaluation of parallel TCP, and its impact on bandwidth utilization and fairness in high-BDP networks based on test-bed. In: 2013 IEEE 11th Malaysia International Conference on Communications (MICC), Kuala Lumpur, pp. 23–28 (2013). https://doi.org/10.1109/micc.2013.6805793
15. Jiang, X., Jin, G.: CLTCP: an adaptive TCP congestion control algorithm based on congestion level. IEEE Commun. Lett. **19**(8), 1307–1310 (2015). https://doi.org/10.1109/LCOMM.2015.2447541
16. Wang, J., Wen, J., Zhang, J., Xiong, Z., Han, Y.: TCP-FIT: an improved TCP algorithm for heterogeneous networks. J. Netw. Comput. Appl. **71**, pp. 167–180 (2016). ISSN 1084-8045. https://doi.org/10.1016/j.jnca.2016.03.020
17. Le, T.A., Hong, C.S., Razzaque, M.A., Lee, S., Jung, H.: ecMTCP: an energy-aware congestion control algorithm for multipath TCP. IEEE Commun. Lett. **16**(2), 275–277 (2012). https://doi.org/10.1109/LCOMM.2011.120211.111818
18. Lee, H.-J., Lim, J.-T.: Congestion control for streaming service in IEEE 802.11 multihop networks. Commun. IET. **4**, 1415–1422 (2010). https://doi.org/10.1049/iet-com.2009.0376

19. O'malley, S.W., Brakmo, L.S., Peterson, L.L.: TCP Vegas: New techniques for congestion detection and avoidance. ACM Comput. Commun. Rev. (CCR) **24** (1994). https://doi.org/10.1145/190809.190317
20. Wei, D.X., Jin, C., Low, S.H., Hegde, S.: FAST TCP: motivation, architecture, algorithms, performance. IEEE/ACM Trans. Netw. **14**(6), 1246–1259 (2006). https://doi.org/10.1109/TNET.2006.886335

Performance Comparison of Machine Learning-Based Classification of Skin Diseases from Skin Lesion Images

Shetu Rani Guha and S. M. Rafizul Haque

Abstract Skin is one of the main parts of the human body. At the same time, skin will be easily infected and damaged by various kinds of skin diseases. Skin disease is a major health hazard across the globe. Nowadays, many people are suffering from skin diseases. It is tedious and time consuming for doctors to manually diagnose them. Recently, machine learning techniques have been successful in the detection and recognition of different types of objects in the images which have been applied to recognize various types of diseases from the medical images. Various machine learning techniques have been used to recognize and classify skin diseases from the images. Here, three machine learning techniques support vector machine (SVM), VGGNet and Inception-ResNet-v2 have been implemented to classify seven types of skin diseases from skin lesion images. Performance of these models has been evaluated and compared by using precision and recall values. Inception-ResNet-v2 has been found to be superior based on the classification performance among these three models.

Keywords Skin lesion · VGG16 · SVM · Inception-ResNet-v2

1 Introduction

Skin is the primary part of a human body which helps to cover the muscles and bones in addition to the entire body. Skin is open to the external environment; hence, the sickness and disease happen more to the skin. Nowadays, many people are undergoing skin diseases. So, appropriate care to skin disease is vital. It is one of the most well-known diseases in the human population, and its frequency is expanding noticeably. Melanoma is the deadliest type of skin malignant growth. In spite of the fact that melanoma counts just 4% of all skin malignancies, it is responsible for 75% of all skin disease death tolls [1]. One out of five Americans will be suffering from skin

S. R. Guha (✉) · S. M. Rafizul Haque
Khulna, Bangladesh

S. M. Rafizul Haque
e-mail: rafijul@cse.ku.ac.bd

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_2

cancer by the age of 70. If the symptoms are identified, this disease can be treated in its early stage and can be cured; but if it is identified too late, it can spread deeper into the skin, propagate to different parts of the body and can be risky, as it becomes hard to treat [2]. In our dataset, we have found seven types of skin diseases; name of these seven types of diseases is mentioned in Sect. 4.1. Skin malignancies fall into commonly two classes: melanoma skin disease and non-melanoma skin disease. Correct identification of skin disease is a challenging task. Conventional diagnosis of skin disease is not only tedious and time consuming but also prone to error. Human experts are not available everywhere, and disease recognition is performed by unaided eye which is not always accurate. Analysis of the disease also becomes hard because of its different shapes, sizes and location of a skin lesion in the body. This research applies the ideas of machine learning and computer vision for classification of skin diseases.

Automatic diagnosis of diseases is now a very important application of digital image processing. Different methods of image processing can be successfully applied to detect different features from the images of the disease-affected organs or body parts. Various machine learning algorithms, such as k-nearest neighbor, artificial neural network and its different variants, unsupervised learning methods like K-means clustering and support vector machine (SVM) can use these features of the images for training and can generate promising results for classification of the diseases. Among different variants of artificial neural networks and deep neural network, more specifically, convolutional neural network (CNN) has been proved to be very successful for classification of images. Here, K-means clustering method has been used to segment the skin lesion in the images of International Skin Imaging Collaboration 2018 (ISIC-2018) dataset. Then, two models of CNN, VGG16, Inception-ResNet-v2 and SVM have been used separately for classification of seven types of skin diseases, in these images.

This paper is composed as follows. The past works identified with this are described in the second part. The third part represents the diagram of CNN, and the fourth part manages the methodology. The fifth part describes the experimental results of the mentioned work, and the last part closes the work.

2 Related Work

Skin detecting different kinds of skin diseases from a lesion image is kind of difficult. Recently, a number of works have been done on this topic because it is quite hard to enhance the performance. For the segmentation of the lesion, authors used a mix of K-means grouping, wavelet investigation, and morphological tasks. In this section, many important techniques are described which have been used in some existing works for analysis of skin diseases.

In a study, Alessandro et al. [3] proposed a programmed segmentation instrument for epi-illumination (ELM) and transillumination (TLM) pictures. For segmentation of the lesion, authors used a mix of K-means grouping, wavelet investigation and morphological tasks. They used SVM classifier for finding a mean exactness of 81.0% and mean training exactness of 90.4%, and the arrangement was returned multiple times.

In an investigation, Litjens et al. [4] applied deep learning in an interesting method to enhance the efficiency of medical image examination. In their work, they used deep learning technique for overall image processing tasks, for example, image segmentation, extraction and classification and so on. In addition, a clear concept is discussed for neuron, retinal, aspiratory, computerized pathology, cardiac and abdominal.

In a study, Codella et al. [1] introduced a hybrid process to detect melanoma utilizing a convolutional neural system and SVM. In this examination, Codella and his partners built a framework uniting current improvements in deep learning and AI methods for skin sore division and arrangement [5].

3 Overview of Convolutional Neural Network

Deep learning has been proved to be very efficient for image classification problem. CNN is widely used for image classification, recognition, object detection [6], etc. General structure of CNN is shown in Fig. 1. Different procedures such as batch normalization, dropout and shortcut connections can also be used to enhance classification accuracy.

3.1 Convolutional Layer

In CNN architecture, the first layer is the convolution layer. It removes the highlights of an image applying easy filtering. The nonlinearity layer is observed to be the second layer in CNN architecture. Tanh, sigmoid or rectified linear unit (ReLU) is the popular functions used as nonlinear activation function. In most of the recent applications, ReLU is popularly used as the activation function. The parameters of

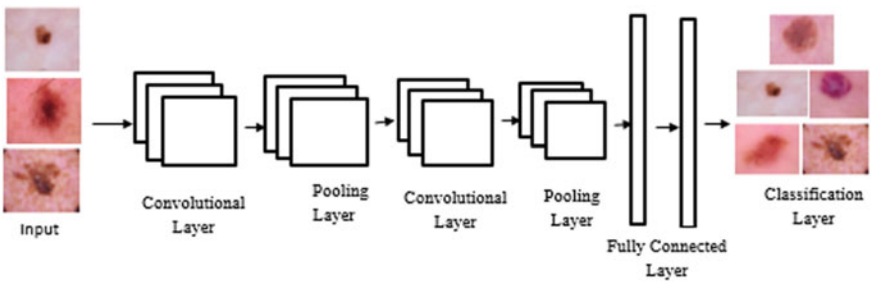


Fig. 1 Overview of CNN architecture

convolutional layer are: number of kernels, kernel size, padding, stride, activation function, regularization value and so on.

3.2 Pooling Layer

The pooling layer follows the nonlinear layer. It works with the width and height of the image and reduces the number of examples important by keeping the quantity of vital data so as to accelerate the process. Three kinds of pooling layers are accessible, for example, max, normal and aggregate pooling. This pooling layer can work autonomously and monitor overfitting.

3.3 Fully Connected Layer

After the end of sequence of convolutional and pooling layers, it is essential to assign a fully connected layer. This layer will have a complete association with all the initiation capacities present in the past layers and retrieve the important information present in the image. Normalization is an important task for finding better results. The normalization is required on the grounds that the parameters chosen for the initiation capacity may have various units. So as to prepare the information superbly, normalization must be done. To break down successive information, a recurrent neural network is broadly applied.

4 Proposed Methodology

This work applies machine learning classifiers to use a portion of the images from the dataset for training and the rest of the images, which were not used in training, to classify the skin diseases. Depending on the features, the classification is performed using SVM, VGG16 and Inception-ResNet-v2 classifiers. The proposed method has focused on recognizing the skin lesion which can assist the doctors to take a proper step immediately to take care of their patients.

4.1 Dataset

One of the problems in conducting research on this topic is the unavailability of freely accessible dataset. Professional dermatologists have access to large number of datasets, but most of them are not openly accessible to the researchers of this area. The procedure of manual classification of the dataset is monotonous and tedious. A

Table 1 Skin lesion images on dataset are distributed in seven classes

Disease name	Number of images
AKIEC	327
BCC	514
BKL	1099
DF	115
MEL	1113
NV	6705
VASC	142

massive categorized dataset is essential for training CNN. The dataset 'ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection' has been used for this work, and in this dataset, images vary in their source, area, size and highlights of the skin lesions [7, 8]. In this dataset, we have found seven types of skin diseases; these are: actinic keratoses (AKIEC), basal cell carcinoma (BCC), benign keratose lesions (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevi (NV) and vascular lesions (VASC). Skin lesion images in this dataset are distributed in seven classes as shown in Table 1.

4.2 Proposed Methodology

The proposed methodology of our framework has two fundamental stages: training stage and testing stage. The first step of our methodology is preprocessing, and it is performed mostly to eliminate the regions in the images that are not affected by skin lesions. Then, the features from the lesion area of the image are extracted and trained in our classifier model. The proposed framework has focused on recognizing the skin infections which helps the specialists to take proper attention of their patient. The steps present in our methodology are shown in Fig. 2.

4.2.1 Image Preprocessing

In the image preprocessing phase, we preprocessed our image step by step. We take an image as an input, and the images are resized to 224×224 pixels. Then, we used a Gaussian filter to cast off the noise and make the image noise-free. The filtering technique also enhances the image quality by replacing the intensity value of the adjacent pixel that may have noises. At that point, the contrast of the image is improved for better visualization of various regions of the image, for example, ordinary skin and lesion existent in the skin part.

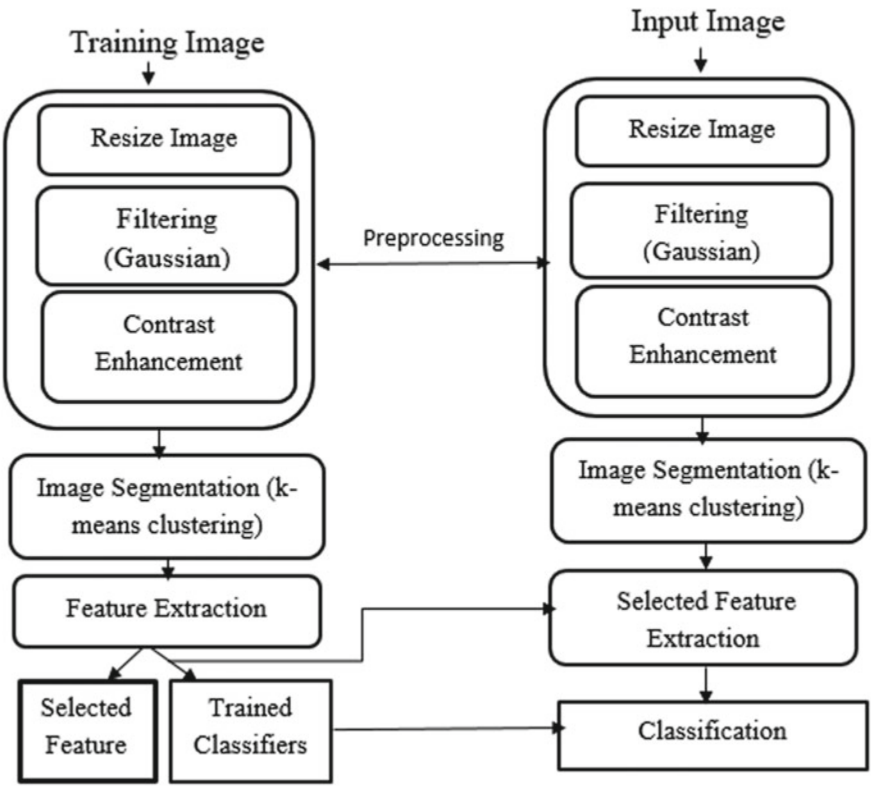


Fig. 2 System architecture of the proposed system

4.2.2 Image Segmentation

The segmentation is useful for analyzing the significant regions of an image. After segmentation, we can get the lesion portion of the images. For segmentation, we used K-means clustering. At first, we take the preprocessed images, then convert the image from RGB color space to Lab color space and categorize the colors into number of groups in a^*b^* space using K-means clustering. Here, the number of iterations was 10. Then, the original image is segmented using color and isolates the infected lesion portion.

4.2.3 Feature Extraction

Feature extraction is the initial stage in image classification. Sometimes, the input data size is too huge, which is extremely difficult to process in its raw form. For resolving this, the input data can be changed into a number of features [8]. After removing all the noises from the image, we apply thresholding technique and extract the features of the image.

4.2.4 Classification

Machine learning technique is used to classify data into a specified number of classes. In this work, SVM, VGG16 and Inception-ResNet-v2 classifiers are used to train the skin lesion images. In this work, we have classified seven types of skin diseases using these classifiers.

Support Vector Machine (SVM)

SVM is a supervised machine learning technique. For classification and regression problems, SVM is used. SVM performs classification by making hyperplane that splits instances of various classes. The data that is nearby to the discriminating function is called a support vector. It utilizes a technique called the kernel trick to change the information, and after that depending on these changes, it searches an ideal limit within the probable outputs. In our implementation, parameters used are kernel, gamma and regularization. Here, the used kernel type is 'rbf', regularization parameter is C and gamma is 'auto'.

VGG16

The CNN has displayed excellent performance in solving numerous computer vision and machine learning problems. VGGNet is a very simple CNN, and its architecture is easy. In our work, we used VGG16 and it consists of 16 convolutional layers. A large number of applications use VGG16 because it has weight design configuration. It is also used as a feature extractor; that is why, we have used it in our work. Some of the parameters and their values used in VGG16 are: The kernel size is 10, stride is 2, the batch size is 128, and we used a same padding.

Inception-ResNet-v2

Inception-ResNet-v2 is a CNN model. Again, it is a deep classical neural network as it has 164 layers [9]. For this classical model, the input size is 229×229 pixels. Here, we resized all the images to 224×224 pixels. This model was retrained on our dataset by fine-tuning overall layers, one global average pooling, it was applied to the output of the last convolutional layer, and finally it was propagated to the softmax layer to classify seven disease classes. Likewise, it is demonstrated that this model can accomplish higher accuracy at a lower epoch and furthermore can learn rich feature for a wide range of images.

5 Experimental Results

In our work, the used platform was Python, Keras and TensorFlow. In deep learning system, Keras was a structure of Python and used to execute the neural system design. Additionally, Keras provides an opportunity for the clients to utilize extra Python conditions, with SciPy and PIL. In artificial intelligence, TensorFlow is an open-source library, utilizing information flow outlines for structure models. It engages specialists to make enormous scale neural systems with different layers. TensorFlow is essentially utilized for classification, recognition and prediction. In this work, we have performed our experiment into two stages: training stage and testing stage. We have analyzed the performance of the three models comparing their classification accuracies for each of the diseases. After training with a portion of the images of the dataset using SVM, VGG16 and Inception-ResNet-v2, and then testing with rest of the images in the dataset which were not used for training, we obtain the accuracy of our model. The complete accuracy of the result of each individual model is shown in Table 2.

Best performance, in terms of accuracy, among these algorithms is obtained by using Inception-ResNet-v2 model. The number of epochs for VGG16 and Inception-ResNet-v2 model was 40. The overall result is calculated using 351 images for testing, and 3212 images are used for training purpose. In our work, we use same training and validation dataset. In our dataset, we find that the images are distributed in seven types of skin disease classes.

Here, Fig. 3 represents the confusion matrix for all the models. Based on the confusion matrix, it is observed that all three models performed poorly on the classification of vascular lesions and dermatofibroma. Better performance was achieved for classification of melanoma and benign keratose lesions for these three models. Here, we also see that using these three models, we get moderate performance for classification of melanocytic nevi, basal cell carcinoma and actinic keratoses. Inception-ResNet-v2 325 and VGG16 305 have achieved best performance overall.

Here, Fig. 4 represents the result analysis of VGG16 and Inception-ResNet-v2 models. For evaluation of the performance, we used two commonly used metrics: recall or sensitivity and precision. The portion of genuine positives that are correctly distinguished is called a recall. Specificity is the reverse of recall or sensitivity. The portion of genuine positives from the cases that are predicted as positives is called precision.

Table 2 Result summary

Model name	Accuracy
SVM	0.9037089
VGG16	0.9153 ± 0.0343
Inception-ResNet-v2	0.9623

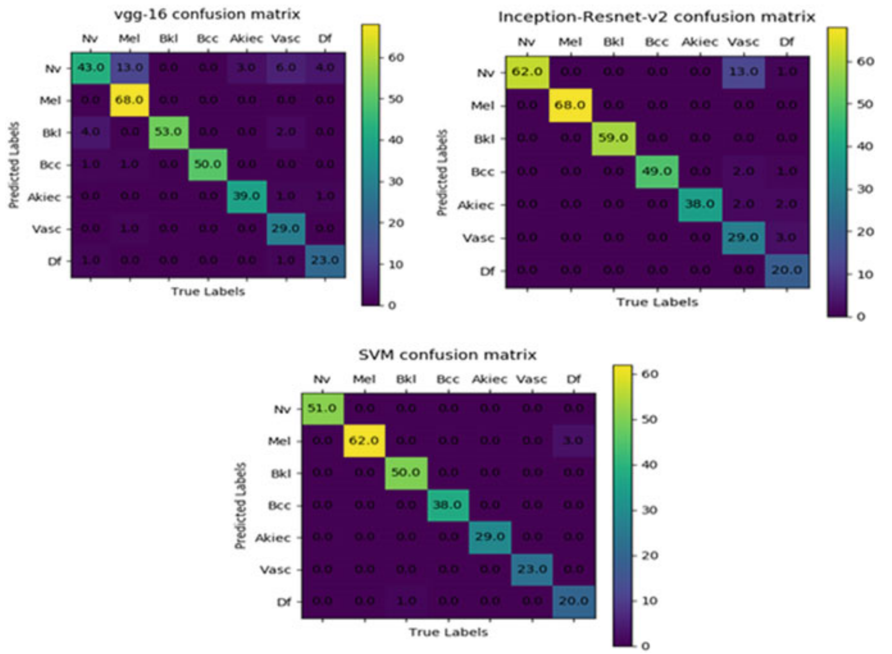


Fig. 3 Confusion matrix of every model

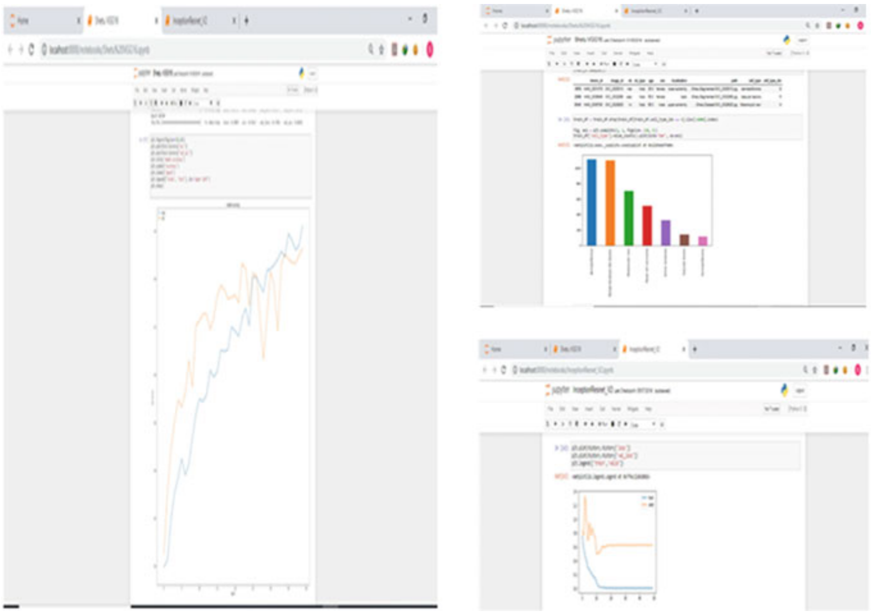


Fig. 4 Result analysis of VGG16 and Inception-ResNet-v2 models

We show the performance analysis of our classifiers based on the precision and recall in Figs. 5 and 6, respectively. The precision and recall of the result for each specific disease are presented as a bar in the chart in Figs. 5 and 6, respectively. It can be observed from Fig. 5 that the best precision values are achieved for melanoma (MEL), benign keratose lesions (BKL) and basal cell carcinoma (BCC) diseases for each of the classifiers. As a whole, application of machine learning on the classification of skin disease has been found satisfactory in terms of classification accuracy which has demonstrated excellent efficiency of deep CNN models Inception-ResNet-v2, VGG16, etc.

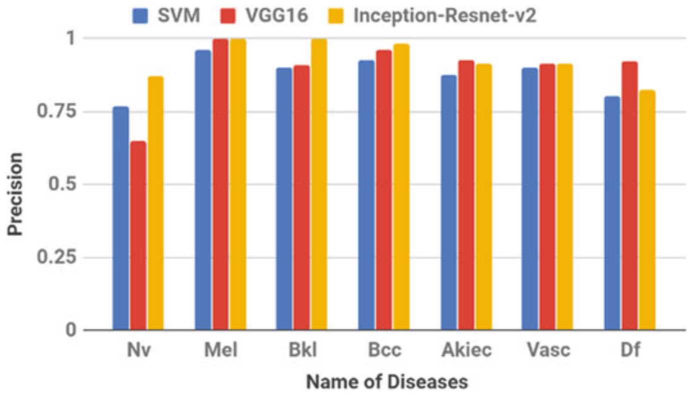


Fig. 5 Precision values obtained by the classifier models for all seven diseases

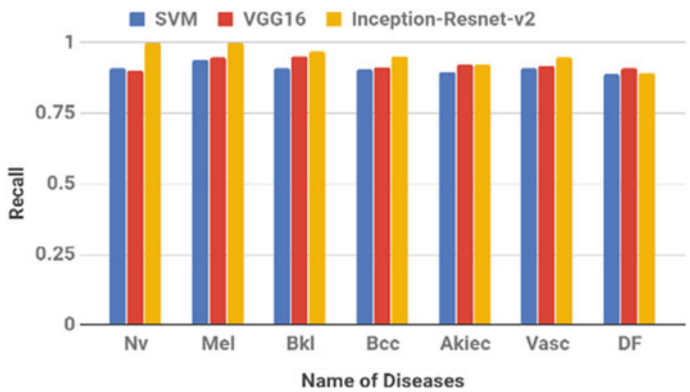


Fig. 6 Recall values obtained by the classifier models for all seven diseases

6 Conclusion

Skin is a very significant part of human body and is thus very important for the well-being of a person. In this work, we have implemented three machine learning models to enhance the exactness of skin disease classification. This investigation expects to build a strong and easy framework for skin disease classification of human body. Among the three models used, Inception-ResNet-v2 produces better results than the other models. In this work, we classify seven major skin diseases and it is observed that we can get better precision for classifying melanoma, basal cell carcinoma and vascular diseases using all of these models. Increasingly assorted datasets (changed classifications and various ages) with substantially more dermoscopy images and adjusted examples per class are required for further improvement. In the future, we have the plan to analyze more skin diseases or other diseases using several machine learning techniques and it will give a viable computerized and quicker conclusion framework.

Acknowledgements This research is funded by the Information and Communication Technology (ICT), Division of Ministry of Posts, Telecommunications and Information Technology, Government of People's Republic of Bangladesh.

References

1. Codella, N., Cai, J., Abedini, M., Garnavi, A., Halpern, A., Smith, J.R.: Deep learning, sparse coding and SVM for melanoma recognition in dermoscopy images. In: International Workshop on Machine Learning in Medical Imaging, pp. 118–126 (2015)
2. Azmi, N.F.M., Sarkan, F.M., Haslina, M.S., Yazriwati, Y., Suriayati, C.: ABCD rules segmentation on malignant tumor and benign skin lesion images. In: 3rd International Conference on Computer and Information Sciences (ICCOINS), pp. 66–70. IEEE (2016)
3. Alessandro, B.D., Atam, P.D., Nizar, M.: Computer-aided analysis of epi-illumination and transillumination images of skin lesions for diagnosis of skin cancers. In: Annual International Conference on Engineering in Medicine and Biology Society (EMBC), pp. 3434–3438. IEEE (2011)
4. Litjens, G., Kooi, T., Bejnordi, B.E., et al.: A survey on deep learning in medical image analysis. Elsevier Med. Image Anal. **42**, 60–88 (2017)
5. Codella, N., Nguyen, Q.B., Pankanti, S., Gutman, D., Helba, B., lpern, A., Smith, J.R.: Deep learning ensembles for Melanoma recognition in dermoscopy images. IBM J. Res. Dev. **61**(4) (2016)
6. Tajbakhsh, N.: Convolutional neural network for medical image analysis: full training (or) fine tuning. IEEE Trans. Med. Image **35**(5), 1299–1312 (2016)
7. https://challenge2018.isic-archive.com/task3/
8. Maria, M.V.J., Rosado, L., Ferreira, M.: A new risk assessment methodology for dermoscopic skin lesion images. In: International Symposium on Medical Measurements and Applications (MeMeA), pp. 570–575. IEEE (2015)
9. Szegedy, C., Loffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In AAAI, 4, p. 12 (2017)

FastICA Algorithm Applied to Scattered Electromagnetic Signals

M. Pushyami Rao, R. Sunitha and Dhanesh G. Kurup

Abstract In this article, a method based on blind source separation (BSS) is applied for separating multiple scattered electromagnetic signals. For BSS, we used fast independent component analysis (FastICA) algorithm. It is shown that individual echoes from the targets can be separated from multiple electromagnetic echoes collected at different spatially separated antennas. For generating the test data, we used a numerical electromagnetic tool. It is concluded that FastICA has the potential for separating echoes from multiple targets in the area of radar systems.

1 Introduction

In radar systems, the received signal has echoes from multiple targets such as aircrafts and guided missiles. This mixture of signals is due to scattering of radar pulse after hitting the airborne targets. Many times we need the capability to differentiate varied targets as well as obtain intelligence of a particular target from the mixture of signals.

Separation of individual signals from mixture of multiple echoes is a highly underdetermined problem, as there is no prior knowledge about the scattered signals or their methodology of mixing. This poses severe engineering challenges in the area of radar signal processing.

Blind source separation (BSS), which involves separating independent components from a mixture of signals has the potential to be applied in the area of radar systems [1]. The prefix "blind" in BSS indicates that a priori information about the source signals is not available. This technique has been undergoing intense study for the last few decades in many domains such as signal processing and telecommunication [2, 3]. The very essence of BSS refers to the question of finding unknown mixing

M. P. Rao · R. Sunitha · D. G. Kurup (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India
e-mail: dg_kurup@blr.amrita.edu

R. Sunitha
e-mail: r_sunitha@blr.amrita.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_3

methodology and information about the sources of the signals from mixed signals which are available for the separation process [4–9]. This technique encompasses statistical signal processing, information theory and artificial neural networks.

One of the most well-recognized statistical methods used for blind source separation is independent component analysis (ICA). ICA is widely used in various fields such as telecommunication, signal processing, and image processing [10–13]. FastICA is a powerful and well-known algorithm for ICA, which was proposed by Hyvarinen and Oja in the late 1980s [3]. FastICA is an algorithm that works on fixed-point iterative theory, which optimizes a nonlinear contrast function. The main goal of FastICA is to maximize the non-Gaussian nature of the source signals. The algorithm uses kurtosis to resolve all the independent components [14]. Advantages of FastICA algorithm over other methods include low memory requirement and less computational time, which are attractive for developing many practical applications [15, 16].

To the best of the author's knowledge, there have been no studies so far for evaluating the performance of FastICA algorithm for separating individual signals from a mixture of scattered electromagnetic signals. In this article, we study the performance of FastICA in a radar application [17] and compare the agreement between unmixed source signals and signals separated from scattered echoes.

2 Theory

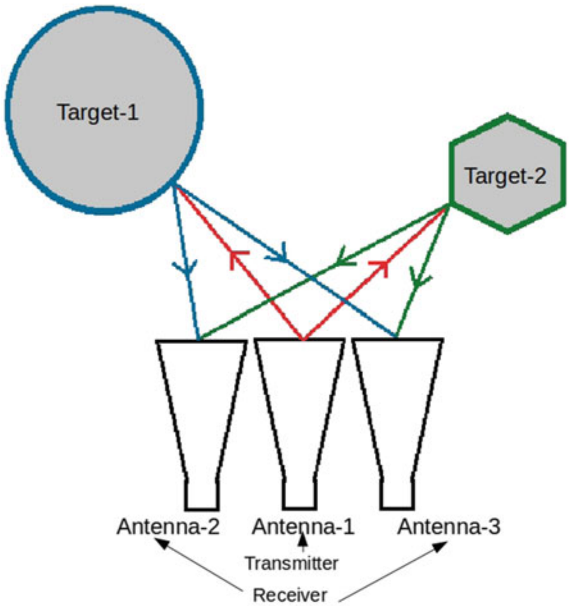
The fundamental assumptions of FastICA algorithm are:

- Statistical independence and non-Gaussian nature of source signals [6].
- Number of mixed signals M is at least equal to the number of source signals N, or $M \geq N$.

The proposed application is depicted in Fig. 1. As seen in Fig. 1, the system comprises three horn antennae of which one is transmitting and other two are receiving antennae. Two targets near simultaneously intercept the signal from the transmitting antenna and scatter the signals resulting in distinct mixtures of signals at each of the receiving horn antennae. The horn antennas were designed using HFSS Electromagnetic simulation tool [18, 19]. Horn antenna was chosen for this experiment as it has better directivity as compared to other antennae.

The two targets are so chosen that they are of different shapes and sizes and kept at different positions and angles from the antennas. Both the targets are taken as solid metallic objects with perfectly conducting surfaces. The signal intercepted by each receiver is a mixture of echoes scattered by both the targets. Since the reflections received are scattered from different objects, they can be assumed to be independent of each other which is a fundamental prerequisite for BSS. The aim of the FastICA algorithm is to separate individual reflected signals from each target from the mixture of scattered signals captured at the receivers. This test setup will give insight into the simulation of airborne targets for radar surveillance application in future.

Fig. 1 Scattered signals of two targets captured by two antennas with single transmitting antenna



The simulations were conducted in four steps. In step 1, the S-matrix of three-horn antenna system without targets was simulated. In steps 2 and 3, the horn antenna system with the center horn antenna as transmitter and the other two antennae on either sides as receivers along with one target at a time were simulated. And finally in step 4 the S-matrix of the horn antenna system along with both targets was simulated as shown in Fig. 1. In HFSS simulations, 6 Adaptive Passes with a delta error of 0.02 at 10 GHz solution frequency were used along with bandwidth setting of 8–11 GHz with a step size of 0.075 GHz.

It is to be noted that in step 4 (with both targets) of simulation, the S-Parameters S_{21} and S_{31} as a function of frequency can be considered as transfer function of the system. However, since there is mutual coupling, to get the exact transfer function, we have to subtract S_{21} and S_{31} of step 1 simulation (without targets) from S_{21} and S_{31} of step 4 simulation (with targets) as,

$$\mathrm{TF}(f) = S_{k1}^{(T)}(f) - S_{k1}(f),\ k = 2, 3 \tag{1}$$

Since FastICA works with time-domain signals, we have to determine echoes from the targets in time domain using proper excitation at antenna 1. This is accomplished by convolving the transfer function with a modulated pulse as,

$$r(t) = \mathrm{IFT}(P(f)\mathrm{TF}(f)) \tag{2}$$

where TF is the transfer function according to (1) and $P(f)$ is the spectrum of the modulated pulse. The input signal for convolution is generated in accordance with the solution frequency of the horn antenna designed such that it is able to capture the desired bandwidth. We have used IT++ Communication and Signal Processing Library [20] for the convolution.

3 Results and Discussion

This section deals with the results obtained for the scattering system described in Sect. 2. Figure 2 shows the baseband raised cosine pulse and the modulated raised cosine pulse at carrier frequency of 9.5 GHz. Figure 3 shows the spectrum of modulated transmitted pulse. As can be seen in Fig. 3, the spectrum of modulated pulse agrees with the bandwidth of the horn antenna designed between 8 and 11 GHz. The mixed scattered signals captured at antenna 2 and antenna 3 are shown in Fig. 4. These signals were fed into the FastICA algorithm to obtain the separated independent signals. Figure 5 shows a comparison between the original and separated signals for Target 1 after BSS using FastICA. Since the targets are at close distances, the individual Target echo is found by averaging the signals at antenna 2 and 3 respectively. From Fig. 5, we can conclude that there is a close agreement between the

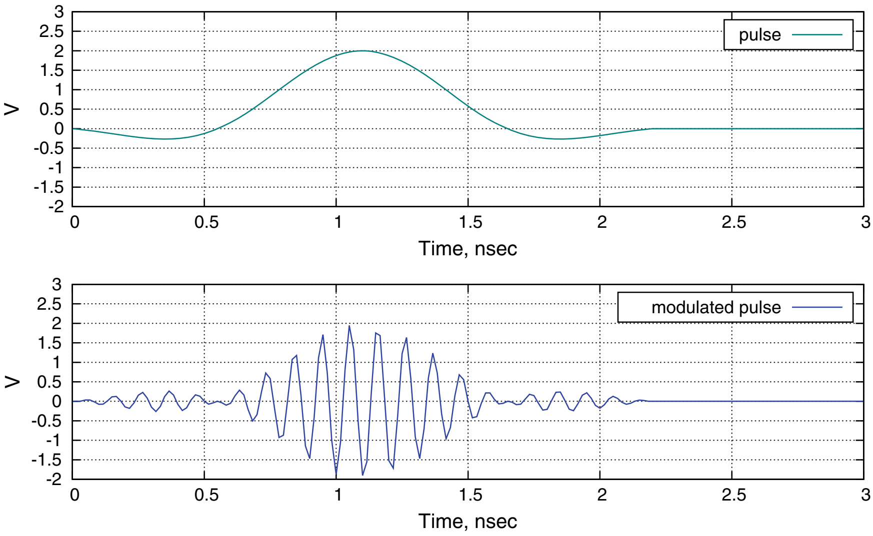


Fig. 2 Baseband (raised cosine) pulse and the modulated pulse at the transmitting antenna, see Fig. 1

Fig. 3 Spectrum of modulated pulse at the transmitting antenna, see Fig. 1

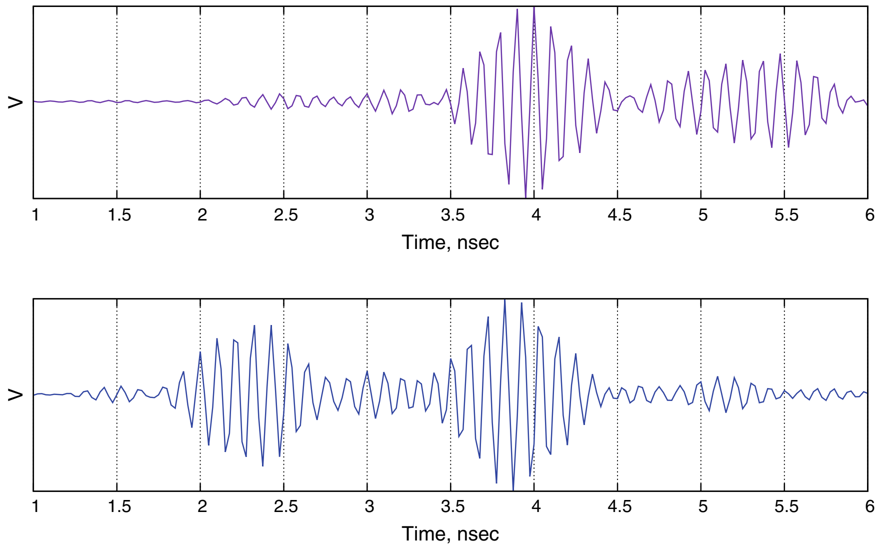


Fig. 4 Mixture of signals received at antenna 2 (top) and antenna 3 (bottom), respectively

original and resolved signal. Figure 6 shows a comparison between the original and separated signals for Target 2 after BSS using FastICA. From Fig. 6, we can conclude that, there is a close agreement between the original and resolved signal for Target 2.

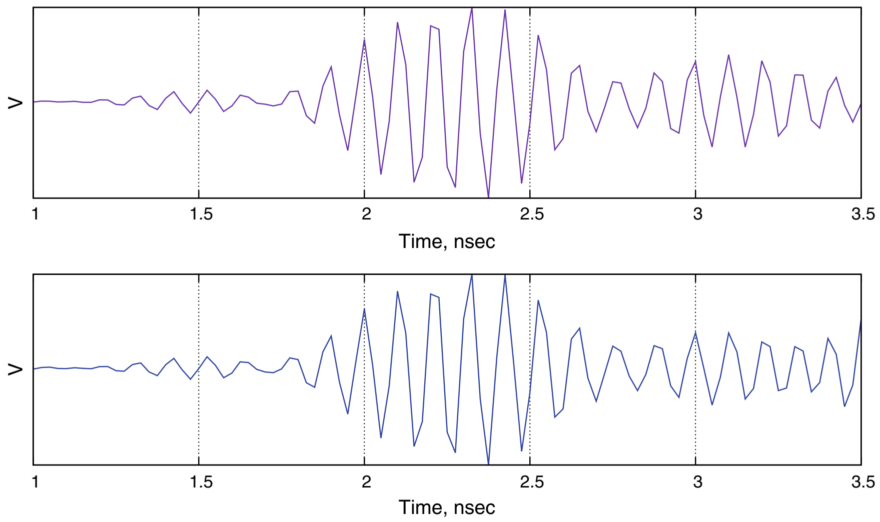


Fig. 5 Original and resolved pulse after source separation using FastICA for Target 1

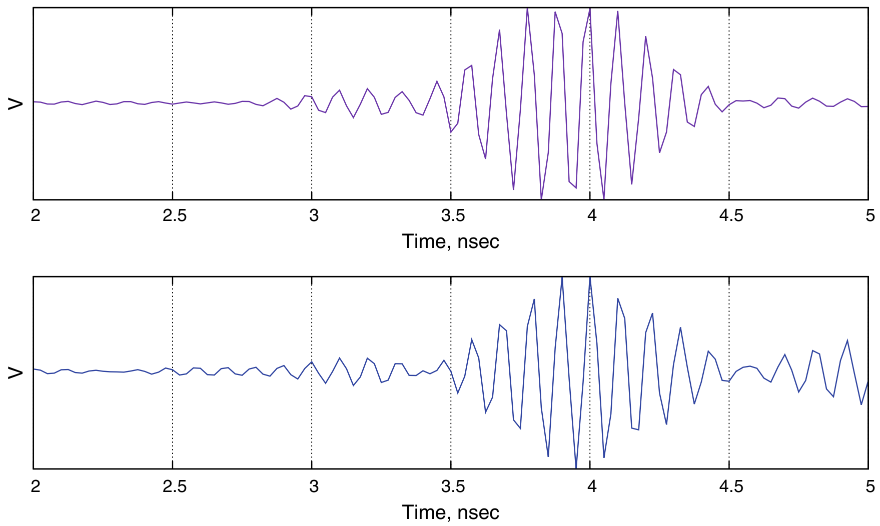


Fig. 6 Original and resolved pulse after source separation using FastICA for Target 2

4 Conclusion

In this article, we explored the applicability of blind source separation technique using FastICA in separating scattered electromagnetic signals from multiple targets in a radar application. Test signals were generated using the HFSS Electromagnetic Tool. The results obtained are very promising for use of FastICA algorithm in many new applications such as radar and bio-medical imaging.

References

1. Yu, X., Hu, D., Xu, J.: Blind Source Separation: Theory and Applications. Wiley, New York (2013)
2. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley Interscience (2001)
3. Hyvärinen, Aapo: Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural New. **10**(3), 626–634 (1999)
4. Saideep, N., Kurup, D.G.,Tripathi, S.: Detection of closely spaced sinusoids in noise using FastICA algorithm. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 3059–309. IEEE (2017)
5. Carlos, E., Takada, J.I.: ICA based blind source separation applied to radio surveillance. IEICE Trans. Commun. **86**(12), 3491–3497 (2003)
6. Comon, P.: Independent component analysis, a new concept? Sig. Process. **36**(3), 287–314 (1994)
7. Aapo Hyvärinen, E.: Oja: independent component analysis: algorithms and applications. Neural Netw. **13**(4–5), 411–430 (2000)
8. Mansour, A., Barros, A.K., Ohnishi, N.: Blind separation of sources: methods, assumptions and applications. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **83**(8), 1498–1512 (2000)
9. Lozano, H.M., Moreno, L.N.O.: Blind source separation of audio signals using independent component analysis and wavelets. In: 2011 21st International Conference on Electrical Communications and Computers (CONIELECOMP), pp. 152–157. IEEE (2011)
10. Vidal, F., Duarte, A., Medeiros, A., Huyart, B.: Blind calibration of five-port receiver based on Independent Component Analysis (ICA). In: Wireless Telecommunications Symposium (WTS), pp. 1–4. Phoenix, AZ (2013)
11. Ichige, K., Imai, M., Arai, H.: Fastica-based blind signal separation and its application to radio surveillance. In: IEEE Workshop on Statistical Signal Processing Proceedings, USA, pp. 546–550 (2007)
12. Wang, S., Jin, G., Jin, G., Wang, Y.: Method to remove the interference in reflected wave of passive radar based on the improved FastICA. In: 9th International Conference on Electronic Measurement and Instruments, Beijing, pp. 27–30 (2009)
13. Ionescu, V., Hnatiuc, M.: Fetal heart rate detection and monitoring from noninvasive abdominal ECG recordings. In: E-Health and Bioengineering Conference (EHB), pp. 1–4. IEEE (2015)
14. Zarzoso, V., Comon, P., Kallel, M.: How fast is Fastica? In: 14th European Signal Processing Confernce, Florence, pp. 1–5 (2006)
15. Candan, C.: Analysis and further improvement of fine resolution frequency estimation method from three DFT samples. IEEE Sig. Proc. Lett. **20**(9), 913–916 (2013)
16. Narayanan, G., Kurup, D.G.: Detection of a real sinusoid in noise using differential evolution algorithm. In: Advances in Intelligent Systems and Computing, vol. 741, pp. 77–83 (2019)

17. Shruthi, N., Mathur, P., Kurup, D.G.: Performance of ultra wideband (UWB) pulsed Doppler radar for heart rate and respiration rate monitoring in noise. In: International Conference on Advances in Computing, Communications and Informatics, ICACCI, pp. 722–725 (2018)
18. Al-Rizzo, H.:Simulation of horn antenna using HFSS (2005). https://doi.org/10.13140/RG.2.1.1551.1202
19. Kaller, D., Schuster, C., Kwark, Y., Altabella, D., Truong, B., Chen, Z., Klink, E.: Using s-parameters successfully in time domain link simulations. In: IEEE 14th Topical Meeting on Electrical Performance of Electronic Packaging, pp. 95–98. IEEE (2005)
20. http://itpp.sourceforge.net

Deep Convolution Neural Network Model for Indian Sign Language Classification

Kruti J. Dangarwala and Dilendra Hiran

Abstract Communication gap between non-hearing and hearing people results in many interaction difficulties across the globe. Indian sign language is the traditional communication alternative in our country. Recently, the integration of deep learning with convolution neural network plays a major role in solving various image classification problems. The proposed deep convolution neural network model is prepared with six convolution layers and three fully connected layers by altering different parameters. This model is evaluated based on 1000 number sign images. These datasets are created in college laboratory with 100 different students. Here, 1–10 number sign images are collected, and the proposed CNN model by six convolutional layers with 1000 epochs is applied. This study highlights the training and validation accuracy analysis as well as training and validation losses. Different performance metrics are calculated to find out the accuracy of the proposed model. The aim of developing the proposed model is to find out classification accuracy for each class. This is multiclass problems as each number signs contain 100 images. This model can predict 10 different signs classes. We have taken 200 signs from 1000 number signs for testing purpose. The results indicate that the accuracy of the proposed method is rapidly increased by increasing the epochs. We have achieved 73% average classification model accuracy.

Keywords Convolutional layer · Validation · Training · Deep neural network

1 Introduction

Non-deaf and normal people use Indian sign language as domestic language in India [1–3]. At present, Indian sign language by using hand gestures is used in almost

K. J. Dangarwala (✉)
Department of Computer Engineering, Pacific Academy of Higher Education and Research University, Udaipur, India

D. Hiran
Department of Computer Science, Pacific Academy of Higher Education and Research University, Udaipur, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_4

all the deaf and dumb schools for teaching, learning and communication purposes. Indian sign language classification is a difficult task for any researchers due to the involvement of large crowd with deaf and dumb people. Indian sign language is made up of various signs with hand gestures as we consider only hand gestures of number signs. Real-time scenarios indicate that solving various image classification problems with deep learning methods leads to grand success with higher accuracy. Deep neural network (DNN) used large number of layers to train the data which require complicated calculations with large amount of parameters [4]. It is difficult to give labeling to all types of data with higher amount [4]. Nowadays, convolutional neural network (CNN) is famous and successful architecture in deep learning methods [5].

Deep neural networks, most preferably known as deep learning, use deep neural network architectures for finding out sequence of features of input data without use of feature extraction methodologies. Deep learning methodologies have very deep architectures which used various hidden layers which indicate that learning methods allow higher levels of accuracy [6]. Deep CNN using AlexNet model is applied to scene image classification [7] which provides better accuracy. Nowadays, pretrained models gain growing success with image classification. Three CNN architectures, namely CifarNet, Alexnet and GoogLeNet, are used with varying various model training parameters which apply to medical images [8]. Detection of computer-aided images occurs with reduction with filter size, stride and pooling parameters [8]. Classification of medical diseases with images is applied to pretrained models VGG16 [3], VGG19 [3], ResNet50 [3] and DenseNet21, and also, comparison of classification is done with data augmentation and without data augmentation [9]. The result shows more accuracy achievement with data augmentation with ResNet50 model [9]. Transfer learning with deep neural network is applied for recognition of Latin and Chinese letters [10]. 3D objects recognition with more accuracy is easily achieved by CNN models [11]. CNN with 13 layers and three different types of data augmentation methods are applied to classify fruit images [12]. CNN approach with data augmentation and without data augmentation analysis is done with fruit classification [12]. State of the art indicates that CNN deep learning approaches are mostly applied for image classification and recognition and handling large image datasets [13]. The pretrained AlexNet model [14] and VGG-16 [3] models are applied to classify tomato crop diseases [15]. 3D convolution operation finds out spatial and temporal features but 2D convolution operation only finds out spatial features [16]. CNN approach is also used for 3D object classification [17, 18].

This paper focuses on Indian sign language digit 1–10 image datasets as input and classifies inputs into 10 different classes as a result of deep learning CNN approach. Big issue is that to collect image datasets for Indian sign language, we have created 1000 image datasets for number signs. Here, image datasets are created by us in our laboratory with different students. Deep CNN proposed model is applied for Indian sign language—number sign classification. Here, we have first found out our training and validation datasets accuracy and losses. As result of CNN proposed model, confusion matrix and classification report are generated for the prediction of actual sign label.

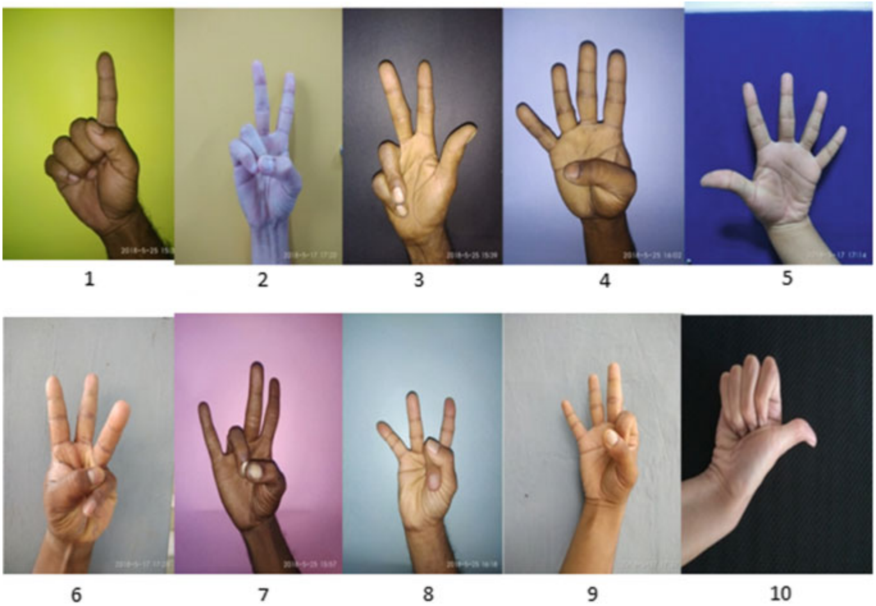


Fig. 1 ISL number sign sample datasets [10]

2 Indian Sign Language: Number Datasets

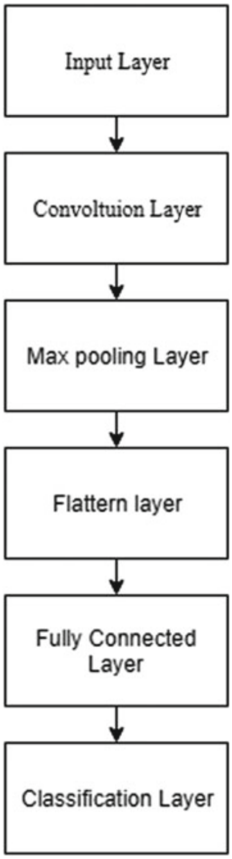
State-of-the-art researches indicate that there are only limited datasets resources for the recognition of Indian sign language [1]. Here, we have created Indian sign language—number datasets. Number datasets with 1–10 digits made up of image signs which are single handed only. These images are captured through our Redmi Note 5 camera with images sizes are 1200 by 1200 RGB pixels. File format JPEG was considered whenever taking the images. Images are resized with 200 by 200 pixels for fasting the execution. These sign datasets are taken from various persons with left handed and right handed. The background of images is considered with different light illumination conditions, i.e., dark and light background. These sign datasets are only created for experiment purposes. Indian has light and dark complexion with skin. So these datasets creation are considered people with skin color complexity. Here, we have created 1000 image datasets for our experiments. A sample dataset of 1–10 digit signs is shown in Fig. 1.

3 Methodologies

Deep learning model execution becomes faster with graphical processing unit (GPU) with CUDA-enabled processor. The study is performed on GPU with CUDA-enabled computer. This section enlightens convolution neural network model and focuses on different variables parameters related with it. Deep learning leads to grand success for classifying images with large dataset of ImageNet. Large image datasets training procedure is time consuming. Training with pretrained deep neural network leads very good performance which extracts features and classifies into multiple classes. CNNs [19, 3] are mostly used to perform image classification task. Image datasets are converted into CSV file. We have split our image datasets into training, testing and validation data. CNN is applied to CSV file and finds out training and loss accuracy on validate data. Then, we have predicted on testing datasets for model analysis. General CNN model is described in Fig. 2.

Convolution layer performs convolve operation with input sign images [20, 19]. It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel. This layer performs convolution operation in which image which is made up of matrix of pixels multiplies with filter matrix (kernel) and performs addition of multiplication values. Then, reach to the next pixel and continue the same process until all pixels are completed. Activation function, rectified linear activation

Fig. 2 General CNN approach [20, 21]



(ReLU), is used here [19]. This activation function is used to find out nonlinearity in our convNet. Here, we have used Python keras library for implementation. We apply a total of six 2D convolution layers with 6 filters, 12 filters, 18 filters, 24 filters, 48 filters and 96 filters sequentially. Filter size is consider as (2, 2) with image width = 200, image height = 200 and depth = 3 as RGB image. 2D convolution generates 2D feature matrix [16].

Max pooling layer generates maximum value of different rectangle areas with input [19]. Max pooling and average pooling are generally used in CNN model [19]. Fruit classification with max pooling has given slightly more achievement than average pooling [12]. Maximum values of rectangle bounded input region are found out by max pooling layer [20]. This model is work with pool size [2, 2]. Max pooling is mainly used to extract subregions of feature map. Consider only maximum value and discards others. Flatten layer converts 2D feature matrix into one dimensional. Dense or fully connected layers are used for classification on the features which are extracted by convolution and pooling layer [19]. This layer is used for combining all features for classification purposes. Three fully connected layers are used here. We have a total of 10 classes with 1–10 digits. Here, classifier recognizes a total of 10 different classes as output. Softmax activation function [17] is used here as we require multiple classification approach. This activation function normalizes the output which consists of some positive numbers that summation is one which is used as probabilities of classification layer. Classification layer is the final layer which solves multiclass classification problems. Here, we apply softmax activation function as multiclass classification approach. It returns the probabilities for each input and assigns classes. It also finds out losses.

Here, we have proposed CNN model with six convolution layers with max pooling, three fully connected layers and classification layer [20]. The model summary is described in Table 1 which indicates number of layers, output size and total parameters used to classify sign images. Convolution layer takes input as image with size 200 pixels by 200 pixels, and 6 filters with size (2, 2) are applied to perform convolve operation. Then, we apply max pooling layer to find out the highest value. So, input size is reduced by half. So, we got output size of images which are half. Others convolutional layers perform same as the previous layers. Parameters in model indicate feature map which varies with size from every convolutional layers. Three dense layers are applied here for classification.

4 Result Analysis

The proposed model implementation is done using Python keras library. Analysis is done with 1000 epochs with batch size 32. Model accuracy on validation data is measured out here. Also, we have generated training and validation accuracy/losses plot for epochs 1–1000. Figure 3 indicates training and validation loss of our datasets. Training loss is decreasing rapidly with variation epochs from 1 to 1000. As indicated that if we increase epochs, training losses are decreasing. But, slight fluctuation

Table 1 Model summary

Types of layer	Output size	Parameters
cconv2d_1 (Conv2D)	(None, 200, 200, 6)	78
activation_1 (Activation)	(None, 200, 200, 6)	0
max_pooling2d_1	(None, 100, 100, 6)	0
conv2d_2 (Conv2D)	(None, 100, 100, 12)	300
activation_2 (Activation)	(None, 100, 100, 12)	0
max_pooling2d_2	(None, 50, 50, 12)	0
conv2d_3 (Conv2D)	(None, 50, 50, 18)	882
activation_3 (Activation)	(None, 50, 50, 18)	0
max_pooling2d_3	(None, 25, 25, 18)	0
conv2d_4 (Conv2D)	(None, 25, 25, 24)	1752
activation_4 (Activation)	(None, 25, 25, 24)	0
max_pooling2d_4	(None, 12, 12, 24)	0
conv2d_5 (Conv2D)	(None, 12, 12, 48)	4656
activation_5 (Activation	(None, 12, 12, 48)	0
max_pooling2d_5	(None, 6, 6, 48)	0
conv2d_6 (Conv2D)	(None, 6, 6, 96)	18,528
activation_6 (Activation)	(None, 6, 6, 96)	0
max_pooling2d_6	(None, 3, 3, 96)	0
flatten_1 (Flatten)	(None, 864)	0
dense_1 (Dense)	(None, 100)	86,500
activation_7 (Activation)	(None, 100)	0
dense_2 (Dense)	(None, 50)	5050
activation_8 (Activation)	(None, 50)	0
dropout_1 (Dropout)	(None, 50)	0
dense_3 (Dense)	(None, 10)	561
activation_9 (Activation)	(None, 10)	0

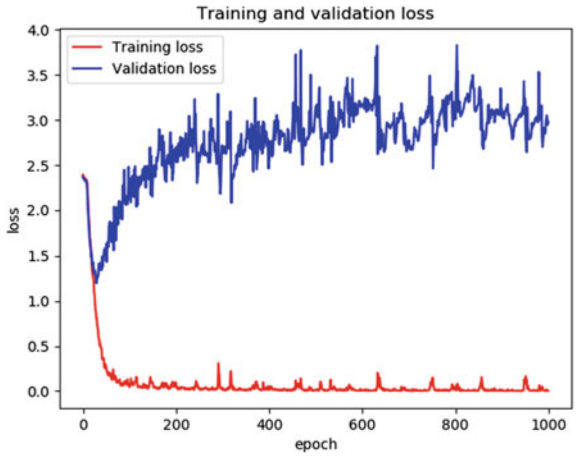


Fig. 3 Model loss curve with sample datasets

arises with validation losses. Figure 4 indicates training and validation accuracy of our datasets. Training accuracy is slightly increasing as we increase epochs from 1 to 1000. The same will arise with validation accuracy. We have achieved validation average accuracy with 72.00%, and execution of this model takes 0:26:50.848767 times.

Classifications of number signs are measured using generation of confusion matrix through implementation. Table 2 indicates confusion matrix with actual label and predicted label. Diagonal value of this Table 1 represents correctly identified total signs. We fit our proposed approach to these datasets, and predictions of classes are done. We have calculated precision, recall and F1 score metrics for each individual class. Individual classes' classification accuracy indicates by recall value in classification report. Precision value indicates that how many times it predicates particular

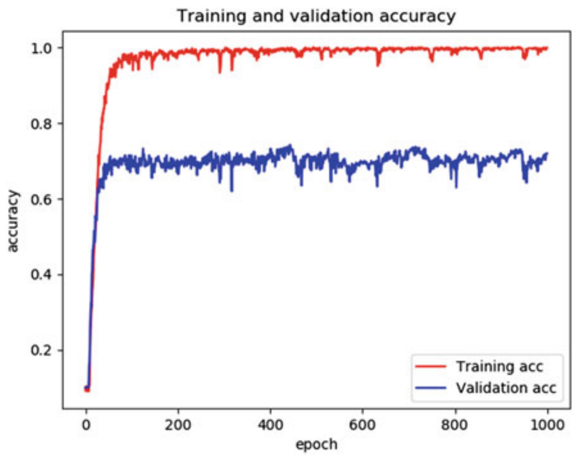


Fig. 4 Model accuracy curve with sample datasets

Table 2 Model—confusion matrix

	Predicted Label										
	S	1	2	3	4	5	6	7	8	9	10
True Label	1	16	1	0	0	0	0	0	1	0	2
	2	2	15	0	1	0	0	0	0	1	1
	3	0	1	19	0	0	0	0	0	0	0
	4	0	1	0	14	0	3	0	1	1	0
	5	1	0	0	1	17	0	1	0	0	0
	6	0	0	0	2	1	13	1	0	3	0
	7	0	1	1	1	0	2	14	1	0	0
	8	1	1	0	2	0	1	0	12	3	0
	9	1	0	0	2	2	0	0	1	13	1
	10	4	0	0	0	0	0	0	0	0	16

Table 3 Model—classification report

Signs	Precision	Recall-individual class accuracy	F1 score	Support
1	0.64	0.80	0.71	20
2	0.75	0.75	0.75	20
3	0.95	0.95	0.95	20
4	0.61	0.70	0.65	20
5	0.85	0.85	0.85	20
6	0.68	0.65	0.67	20
7	0.88	0.70	0.78	20
8	0.75	0.60	0.67	20
9	0.62	0.65	0.63	20
10	0.80	0.80	0.80	20
Average accuracy		0.74(74%)		200

individual signs. Through implementation, we have generated classification report which is summarized in Table 3. We have achieved average classification accuracy that is 74%.

5 Conclusion

Result analysis indicates that we have achieved model accuracy above 70%. We apply this model for testing the datasets of 200 images that are created with different environment conditions. Each class has 20 images. Confusion matrix and classification report indicate 74% average accuracy. Here, we do not have any standard datasets for classification [1]. All researchers applied their own datasets for their proposed approach [1]. Comparison with other researchers' methodologies leads different results. Compared with HOG-SVM approach to our datasets lead to 50% average classification accuracy. In the future, we will enhance more accuracy by adding more image datasets as deep learning approach requires large datasets, but we have limited datasets. With large datasets, we can even achieve better results. Further enhancement is that we can apply this model to the datasets of alphabet and daily communication videos.

References

1. Dangarwala, K., Dilendra, H.: A research gap on automatic indian sign language recognition based on hand gesture datasets and methodologies. Int. J. Comput. Eng. Appl. **12**(3), 46–54 (2018)
2. Karush, S., Gupta, R.: Continuous sign language recognition from wearable IMUs using deep capsule networks and game theory. Comput. Electr. Eng. **78**, 493–503 (2019)
3. Rumi, R.I., Hossain, S.M., Shahriar, A., Islam, E.: Bengali hand sign language recognition using convolutional neural networks. Ph.D. dissertation, Brac University (2019)
4. Gopalakrishnan, K., Gholami, H., Vidyadharan, A., Choudhary, A., Agrawal, A.: Crack damage detection in unmanned aerial vehicle images of civil infrastructure using pre-trained deep learning model. Int. J. Traffic Transp. Eng. **8**(1), 1–14 (2018)
5. Zhong, L., Hu, L., Zhou, H.: Deep learning based multi-temporal crop classification. Remote Sens. Environ. **221**, 430–443 (2019)
6. Schmidhuber, J.: Deep learning in neural networks—an overview. Neural Netw. **61**, 85–117 (2015)
7. Sun, J., Cai, X., Sun, F., Zhang, J.: Scene image classification method based on Alex-Net model. In: 3rd International Conference on Informative and Cybernetics for Computational Social Systems, pp. 363–367. IEEE (2016)
8. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging **35**(5), 1285–1298 (2016)
9. Sato, R., Iwamoto, Y., Cho, K., Kang, D.Y., Chen, Y.W.: Comparison of CNN models with different plane images and their combinations for classification of Alzheimer's disease using PET images. In: Innovation in Medicine and Healthcare Systems, and Multimedia, Smart Innovation, Systems and Technologies, vol. 71, pp. 169–177. Springer (2019)
10. Ciresan, D., Meier, U., Schmidhuber, J.: Transfer learning for Latin and Chinese characters with deep neural networks. In: International Joint Conference on Neural Networks, pp. 10–15. IEEE (2012)
11. Sarkar, K., Varanasi, K., Stricker, D.: Trained 3D Models for CNN based Object Recognition, pp. 130–137 (2017)
12. Zhang, Y., Dong, Z., Chen, X., Jia, W., Du, S., Muhammad, K., Wang, S.H.: Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation. Multimedia Tools Appl. **78**(3), 3613–3632 (2019)
13. Liu, X., Zhang, R., Meng, Z., Hong, R., Liu, G.: On fusing the latent deep CNN feature for image classification. World Wide Web **22**(2), 423–436 (2019)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, pp. 1097–1105 (2012)
15. Rangarajan, A.K., Purushothaman, R., Ramesh, A.: Tomato crop disease classification using pre-trained deep learning algorithm. In: Procedia Computer Science, vol. 133, pp. 1040–1047 (2018)

16. Ji, S., Xu, W., Yang, M., Yu, K.: 3D convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 221–231 (2012)
17. Hegde, V., Zadeh, R.: FusionNet: 3D Object Classification Using Multiple Data Representations. arXiv preprint arXiv:1607.05695 (2016)
18. Socher, R., Huval, B., Bath, B., Manning, C.D., Ng, A.Y.: Convolutional-recursive deep learning for 3D object classification. In: Advances in Neural Information Processing Systems, pp. 656–664 (2012)
19. Sinha, S., Singh, S., Rawat, S., Chopra, A..: Real time prediction of american sign language using convolutional neural networks. In: Singh, M., Gupta, P., Tyagi, V., Flusser, J., Ören, T., Kashyap, R., (eds.) Advances in Computing and Data Sciences, vol. 1045, pp. 22–31. Springer, Singapore (2019)
20. Lumini, A., Nanni, L.: Deep learning and transfer learning features for plankton classification. Ecol. Inform. **51**, 33–43 (2019)
21. Levi, G., Hassner, T.: Age and gender classification using convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition workshops, pp. 34–42. IEEE, USA (2009)

Opinion Mining of Bengali Review Written with English Character Using Machine Learning Approaches

Sabiha Sunjida Ahmed, Sharmin Akter Milu, Md. Ismail Siddiqi Emon, Sheikh Shahparan Mahtab, Md. Fahad Mojumder, Md. Israq Azız, Jamal Ahmed Bhuiyan and M. J. Alam

Abstract In this paper, we have done sentiment analysis for English written Bengali words given in different online shops in Bangladesh. For this work, we have chosen four latest mobile phones popular in Bangladesh. Here, the user reviews were in Bengali words written by English characters. The data was taken from online shopping sites from Bangladesh. Here, we have assumed six different features of mobiles written in the Result section. The main objective of the study was to find out the sentiment of Bengali words written with English alphabets. As it is a trend to write such reviews in Bangladesh, the data was taken and preprocessed to fit in algorithm, and they were compared whether it is positive or negative. Python was used as simulation tool, and Pursehub was used to extract the data set, and the system

S. S. Ahmed · Md. I. S. Emon
Department of CSE, Feni University, Trunk Road, Feni 3900, Bangladesh
e-mail: ahmedsunjida@gmail.com

Md. I. S. Emon
e-mail: emonsahriar0@gmail.com

S. A. Milu
Department of CSTE, Noakhali Science & Technology University, Noakhali, Bangladesh
e-mail: sharminmilu7@gmail.com

S. S. Mahtab (✉) · M. J. Alam
Department of EEE, Feni University, Trunk Road, Feni 3900, Bangladesh
e-mail: mahtabshahzad@gmail.com

M. J. Alam
e-mail: alameee1993@gmail.com

Md. F. Mojumder
Department of ECE, North South University, Dhaka, Bangladesh
e-mail: fahad.mojumder@northsouth.edu

Md. I. Azız
Department of EEE, Dhaka University, Dhaka, Bangladesh
e-mail: israqaziz26@yahoo.com

J. A. Bhuiyan
Department of CE, Feni University, Trunk Road, Feni 3900, Bangladesh
e-mail: jamalrony35@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_5

successfully finds out the positivity and negativity of the reviews. This result was achieved by using confusion matrix and that is making the overall performance of those mobile handsets. Out of 1201 reviews, 599 were found to be negative and 826 were found to be positive. The F1 score was 85.25%, accuracy was achieved 85.31%, and recall rate was 84.95%.

Keywords Sentiment analysis · Machine learning approaches · Natural language processing · Bengali · Naïve · Python

1 Introduction

This universe nowadays relies on rapidly increasing World Wide Web. People from all over the world able to swift and contribute to any types of initiative related from trade to research. This situation clearly indicates that, the importance of general folks' contribution in the area of expertise plays a vital role to make decision in relation to merchandise target and recommendation persuasion. The above-mentioned area is already started to depend on public contribution as public opinion which is defined as "sentiment" in later discussion. A more clear definition of sentiment is as follows: people's observation as exposition regarding any service, a particular brand, or other productive way of serving thyself known as sentiment. In the coming days, sentiment or present customer reviews going to play an essential role in national and international marketplace, where most of them are superior level popular to forecast upcoming market situations. An important strategy for a successful businessman is to calculate the feasibility of future market so analyzing the present market as well as current customer reviews become highly demanding topic to depict future market status. In addition, it also assists a merchant or dealer to decide what to stock in which quantity and what not to stock at all. However, sentiment analysis is not only helpful for sales but also for customer. A customer could easily be understood which product will be good for them. So, all Bangladeshi online buyer and seller as well can easily become beneficiaries whether they are interior or exterior elements of the country. A vast majority of consumer's contributing voluntarily with their outstanding review whenever or wherever they buy something specific. In this situation, tons of reviews were originating in almost every online market day by day and its growing so superfast, where each of them is detached as all of them are generated by random consumer from all over the country. Online market is a very popular place to buy and shop daily products. In this circumstance, people gathering so much experience about everything and their adjudication are so much developed. Additionally, consumers were openly publishing their sentiment in various perspectives. So this area upraises as a captivating ground for us to pursue with other's affection for online product. In our work, to find out opinions written in Bengali written within English words and Bengali words we apply a sentiment analysis system. To collect our required data, we decided to extract data from some vastly popular E-commerce sites of Bangladesh, for example, picaboo.com, chalal.com, rokomari.com, and so on, where Bengali texts are exchanged by their equivalent English words only, and then we calculate summarized review of consumed product.

2 Literature Survey

Our source of inspiration from previously done research on the same topic we are working on. Almost all of them are for gaining knowledge. We emphasize most cases from [1]. Here, they describe language by using various manual and spontaneous accessions where natural language is classified into several classes, for example, emotion-association lexicons and valance. There is another research work we had followed [2]. Researchers of this work gathered short messages and uproot emotions for the collected text messages from specific data set. Moreover, this method s used TF-IDF which stands for “term frequency–inverse document frequency.” Additionally [3] in this paper, they demonstrate a research work on data set which previously collected from twitter. A well-mannered category is generated by them, which is categorized into three parts for each user as follows: positive, negative, and neutral. All of these works were done by using fuzzy classifier and naive Bayes algorithm. Likewise, in paper [4], researchers classified their workflow in two parts, for example, topic-based classification or special categorization method that need to be developed. In this purpose, they apply support vector machine or (SVM in short), maximum entropy classification, Naive Bayes classification. Besides that, another work [5] was completed on data set collected from twitter for Bengali language. They demonstrate the polarity (whether it is positive or negative) of given Bangla text. Also, they show a relatively comparative analysis on various sets of features by SVM and maximum entropy algorithm, likewise using POS tagging by Part-Of-Speech Tagger package. An Indian paper is the last in our list as they analyze and exhibit result from Hindi written review.

3 Methodology

For this work, we did not found any ready data set that is available in Internet or in any library. So that Pursehub was used to extract the reviews from the sites. For showing effectiveness, we did not go for the whole reviews. Rather we went for some popular mobile phone’s review from some online shops. The selected mobiles were used to check the program and comparing among themselves. We went for OnePlus 6 T, iPhone XS, Huawei Mate 20 Pro, Google Pixel 3 as sample. The data set has been shown in Table 1. The data has lot of complex words and joint words like verygood for very good, gd for good, nt gd for not good as well as specified containing numerals like 9ce for nice, as we went for Bengali words which were written in English alphabets such as “onk vlo,” “onek khusi,” “pura taka usul,” “sundor” (written in English script only). So, the data set also contains such words.

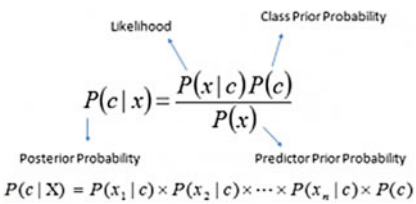
Table 1 Word correcting that bears sentiment

Incorrect words from review	Corrected words
Gud, gd	Good
Khub	Very
Khusi	Happy
Nt	Not
Gr8	Great
Vlo	Good
Onek sundor	Very nice
Osadharon	Fantastic
Darun	Nice
Asolei valo	Really good
Osthir	Awesome
Baje, faltu	Bad
Kharap	Bad
Hudai	Valueless
Deri hoice	Late

Table 1 shows Bengali words written in English along with their English equivalent word. After having the data set ready, authors went for preprocessing the product review which is very much important. First we converted Bengali word written by English alphabets into equivalent English word, and the English incorrect word was also corrected. Then we tagged by Part-of-Speech Tagging. By this, the adjective, noun, and adverb were specified. Figure 1 shows the algorithm of the word.

For classifying the text, we went for naïve Bayes (NB) system or classifier. In machine learning approaches, NB is a system or classifier which has a classification system that is based on Bayes theorem. It has also assumption on prediction that is independent (Fig. 2).

We have listed them and declared the score for them in Table 2. This parameter indicates how much positive or negative the review was. We have divided the scores into three different sections. The scoring method is given in Table 3.



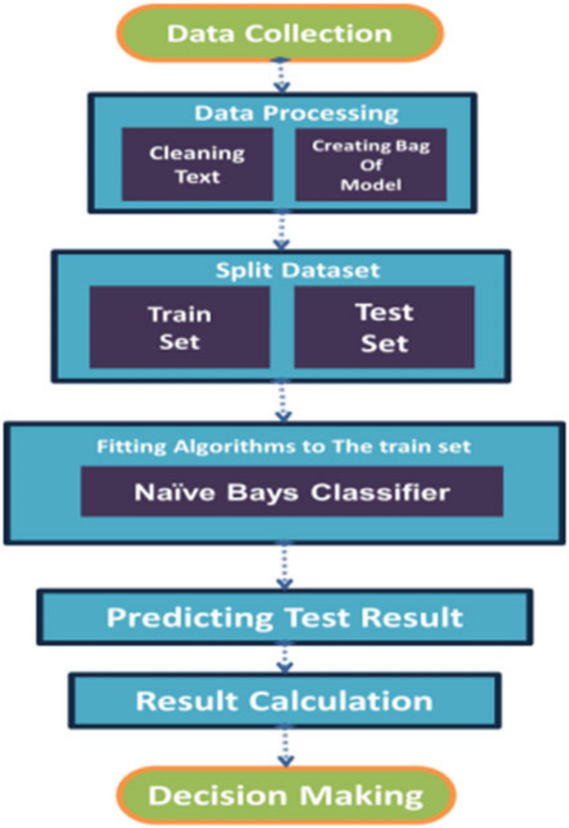


Fig. 1 Flowchart of algorithm

4 Result and Discussions

From Table 9, we can see review sentiment for the specific mobile those released immediately. As shown in column, the table gives the sentiments that specified as positive or negative for these mobiles. We have collected mobile reviews, then counted them, and differentiated using the code implemented in Python. We mean that the positive or negative labels for the calculation were based on the features of those mobile sets like smart cam, strong battery, high resolution of the screen, better sound, outer design, and hardware and software (H/S) performance.

Here, showing the phone reviews was not the aim of the study. Rather the output of this system is the machine that is able to judge as it is in its order of positiveness or negativeness. Precision is the term that is used to positive predictive value, and recall is the term that is used to find sensitivity. Therefore, both of them were used to find and evaluate the relevance (Table 4).

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

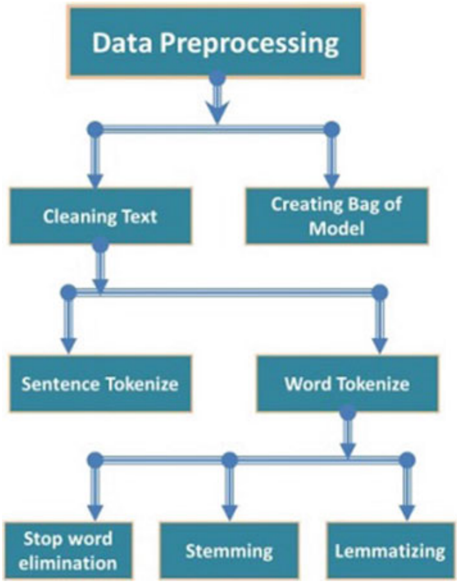


Fig. 2 Flowchart of data preprocessing

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

F_1 score is the calculation of accuracy of tests. It is also called as F-measure.

$$F_1 = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

During the use of units that normalized the space remains under the graphed curve (area under the curve or AUC) is similar to the probability, as we can say the classifier makes a rank which is taken by random choice at instance, that is, positive and not going to take the negative one, and it will rank the positive rank greater than the negative. This can be measured

$$A = \int_{\infty}^{-\infty} \text{TPR}(T)\text{FPR}'(T)\text{d}T$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(T' > T) f_1(T') f_0(T)\text{d}T'\text{d}T = P(X_1 > X_0)$$

Table 2 Sample of data set

Review	Camera	Battery	Screen	Design	H/S performance
OnePlus 6T	atarselfie camera diyeonekvalo pic tola jai	osadharon battery capability	Kenar age tension e cilam, but now i am satisfied	Design ta joss, are cover termotovalo offer r paina	So smart and easy to use
iPhone XS	sobsomoyermotovalo camera	eiphnerbatteray ta khubvalo…	Ekkothay perfect	So far so good	atar quality niyeami satisfied
Huawei Mate 20 Pro	ato outdate phonekivabeato dam a sell korenapnara?	Battery ta xoss, amarjonno best	Sundor screen, darazke thanks atotarataridelivarydeyarjonno	what a price for such design! good	Working so nice and smooth
Google Pixel 3	Front camera kothai!	ata 20 min e full charge hoye jai, just wow	I am amazed how clear and crisp this monitor is, finally	Front color is black but I thought it would be black all over	Recharging is slow but quite good enough

Table 3 Sentiment parameters

Polarity	Degree
0.84–1.00	Excellent
0.67–0.83	Very good
0.50–0.66	Good
0.33–0.49	Bad
0.17–0.32	Very bad
0–0.16	Worst

Table 4 Number of data

Properties reviews	OnePlus 6T		iPhone XS		Huawei Mate 20 Pro		Google Pixel 3	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
Camera	260	40	160	41	167	86	210	20
Battery	140	160	126	75	181	72	198	32
Screen	195	105	147	54	184	69	222	8
Sound	152	148	132	69	27	226	180	50
Design	220	80	146	55	174	79	201	29
H/S performance	234	66	115	86	119	134	186	44
Total	1201	599	826	380	852	666	1197	183

The outcomes, that is found by holdout strategy, are given below:

F_1	85.25%
Accuracy	85.31%
Precision	85.56%
Recall	84.95%
ROC-AUC	85.31%

The confusion matrix is obtained from naïve Bayes which is given below:

True negative (509)	False positive (122)
False negative (90)	True positive (704)

Graphical representation of the outcomes

iPhone XS

Result after analyzing the reviews is graphically represented. It was also taken as O/P (output) from our assumed data set. Below graph is showing the mining reviews for iPhone XS (Fig. 3; Table 5).

For better understanding, we may also need to know the percentage of the positiveness or negativeness. So, we find the percentage (%) of the polarity.

For reviews of iPhone XS, the degree of positivity and negativity is given below

OnePlus 6T

Figure 4 shows the bar chart of the mining reviews for OnePlus 6T.

Here in Fig. 4, we can see customer's review that goes almost positive for this phone, i.e., very high for its camera. But people were not positive with its battery and sound; it was almost 50–50 position. Here in this Table 6, we can see the combine polarity of customer's review of the mobile OnePlus 6T. Despite some negative reactions, it also shows that the customer's review was in their favor.

Huawei Mate 20 Pro

Figure 5 shows the review sentiment in bar chart. This figure gives us the idea that it also can be a very good mobile for buying. But people are not satisfied with sound, and also some dissatisfaction in H/S performance. But it gives a better backup. Figure 5

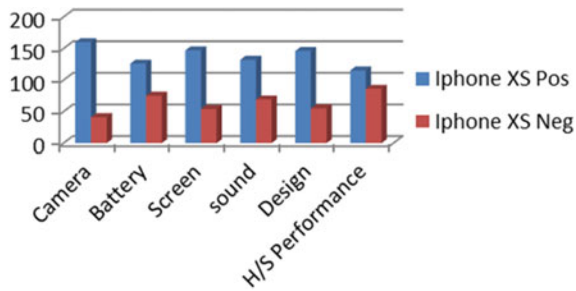


Fig. 3 Bar chart of mining reviews for iPhone XS

Table 5 iPhone degree of sentiment

Degree	Percentage
Excellent	9.95
Very good	36.06
Good	22.47
Bad	9.12
Very bad	15.75
Worst	6.63

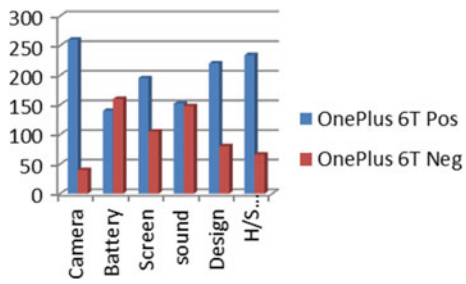


Fig. 4 Bar chart of mining reviews for OnePlus 6T

Table 6 Combine polarity percentage of degree of sentiment

Degree	Percentage
Excellent	9.95
Very good	36.06
Good	22.47
Bad	9.12
Very bad	15.75
Worst	6.63

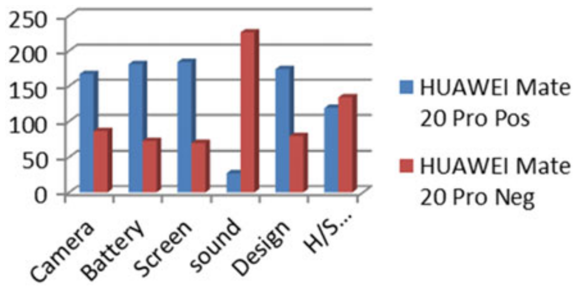


Fig. 5 Bar chart of mining reviews for Huawei

shows the percentage of polarity given by the customers. Table 7 shows the positivity and negativity rate of our experiment data.

Table 7 Positivity and negativity rate of Huawei Mate 20 Pro

Degree	Percentage
Excellent	9.95
Very good	36.06
Good	22.47
Bad	9.12
Very bad	15.75
Worst	6.63

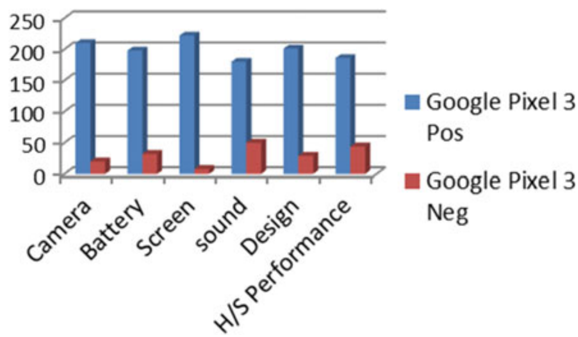


Fig. 6 Bar chart of mining reviews for Google Pixel 3

Table 8 Google Pixel 3 degree of sentiment

Degree	Percentage
Excellent	16.11
Very good	31.55
Good	19.05
Bad	17.77
Very bad	9.94
Worst	5.55

Google Pixel 3: Google Pixel 3 has 64 and 128 GB with Google Lens, Portrait Mode, and Google Assistant that comes in just black, clearly white, and not in pink colors (Fig. 6; Table 8).

Figure shows that it is the phone that customer dream. Here we can see the polarity of customer's review of this phone. Almost all the people were happy with it. This parameter indicates how much positive or negative the review was.

Overall Evaluation

Figure 7 represents the aggregate appraisement percentage of entire reviews of handsets. In addition to this, figure shows whole properties that are dimensioned on the *X*-axis and percentage is constituted on the *Y*-axis. Moreover from this figure, it is easy to discern in all affirmative and contradictory reviews, which covered above there. Furthermore, contradictory and affirmative reviews are frame-up on *X*-axis and percentage is showing on *Y*-axis in the graph.

Evaluated Results

As a matter of fact, a top figure is exhibiting bar diagram regarding four various handsets, which are iPhone XS, Google Pixel, Huawei Mate 20 Pro, and OnePlus 6T. In the same way, the first bar diagram illustrates the actual amount of favorable and non-favorable reviews of six detached features, for instance, camera, battery, screen, sound, design, hardware/software performance, etc. Secondly, the following diagram displays the percentage of positive and negative reviews of every feature that coupled with the third diagram illustrate the combined percentage of affirmative

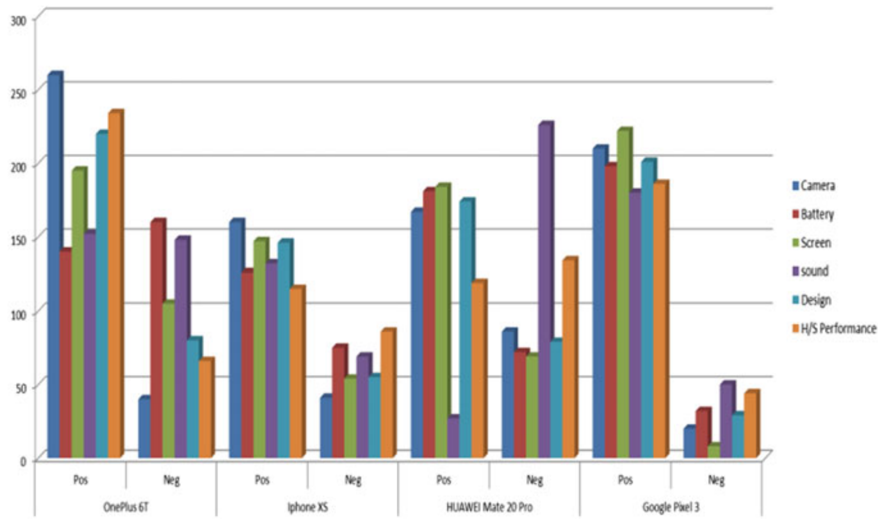


Fig. 7 Overall result

Table 9 Ultimate evaluated results

Mobile	No. of reviews	+ ve review	−ve review	Percentage of positive review
Huawei Mate 20 Pro	1800	1201	599	66.72
Google Pixel 3	1206	826	380	68.49
OnePlus 6T	1518	852	666	56.12
iPhone XS	1380	1197	183	86.73

or positive reviews of the consumer. Likewise as Table 9 mentioned, comparisons of four handsets and consequences establish that iPhone XS is comparatively better than any other handsets or smart phone. In the final analysis, people like Google Pixel, Huawei Mate 20 Pro, OnePlus 6T which is given in percentage, respectively, 68.49%, 66.72%, 56.12%.

5 Conclusion

From this research, we have done sentiment analysis using machine learning approaches with naïve Bayes classifier. Here, the user reviews were in Bengali words written by English characters. The data was taken from online shopping sites from Bangladesh. Here, we have assumed six different features of mobiles that written in the Result section. The main objective of the study was to find out the sentiment of Bengali words written with English alphabets. As it is a trend to write such reviews in

Bangladesh. The data was taken and preprocessed to fit in algorithm, and they were compared whether it is positive or negative. Python was used as simulation tool, and Pursehub was used to extract the data set, and the system successfully finds out the positivity and negativity of the reviews. This result was achieved using confusion matrix and that is making the overall performance of those mobile handsets. Out of 1201 reviews, 599 were found to be negative and 826 were found to be positive. The F_1 score was 85.25%, accuracy was achieved 85.31%, and recall rate was 84.95%.

References

1. Mohammad, S.M.: Sentiment analysis: detecting valence, emotions, and other affectual states from text. In: Emotion Measurement, pp. 201–237. Elsevier (2016)
2. Silva, J.J.D., Haddela, P.S.: A term weighting method for identifying emotions from text content. In: 2013 8th IEEE International Conference on Industrial and Information Systems (ICIIS), pp. 381–386. IEEE (2013)
3. Mehra, R., Bedi, M.K., Singh, G., Arora, R., Bala, T., Saxena, S.: Sentimental analysis using fuzzy and naive bayes. In: 2017 International Conference on Computing Methodologies and Communication (ICCMC), pp. 945–950. IEEE (2017)
4. Pang, K.B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, Vol. 10, pp. 79–86. Association for Computational Linguistics (2002) unpublished
5. Chowdhury, R.S., Chowdhury, W.: Performing sentiment analysis in Bangla microblog posts. In: 2014 International Conference on Informatics, Electronics & Vision (ICIEV), pp. 1–6. IEEE (2014)

Big Data Feature Selection to Achieve Anonymization

U. Selvi and S. Pushpa

Abstract In the age of big data, data is increasing in a tremendous way in many fields and the data shared by the users is in a great risk. To preserve privacy of an individual anonymization-based algorithm like k-anonymity-related algorithm and differential privacy is proposed to make sure that the resulting dataset is free from privacy disclosure. However, majority of these anonymization algorithms are applied in isolated environment, without considering the utility in knowledge task making the dataset less informative. Also the presence of redundant data also decreases the performance and reduces accuracy of anonymization. Hence a preprocessing-based anonymization is required to increase the utility and to achieve accuracy in anonymization. This paper aims to apply the feature selection fast correlation-based filter (FCBF) solution to select the relevant features and remove the redundant data. Then k-anonymity is applied to dataset to achieve data anonymization. Comparisons on real-world dataset were made with anonymized dataset with preprocessing and without preprocessing and result was produced.

Keywords Big data · Anonymization · Data mining · Fast correlation-based feature selection · Data preprocessing · Privacy preservation · MapReduce

1 Introduction

Big data is valuable and contains treasured source of knowledge which is useful for decision making and prediction purposes (e.g., *Product Recommendation* [1]). Privacy of dataset is not preserved as the analytical algorithm proposed tries to expose the knowledge discovery from the data and process huge volumes of big data [2].

Data anonymization is process of hiding sensitive information about an individual to the outsiders. The key reason for users in security in micro-data is that such data contain sensitive information. Anonymization makes the aggregate data to be exposed to the information users for analysis and mining in which privacy is preserved. For example, hospital releases the medical dataset which contains information regarding

U. Selvi (✉) · S. Pushpa
St. Peter's Institute of Higher Education and Research, Chennai, Tamil Nadu, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_6

the personal medical history. Similarly, movie datasets which reveal information regarding background preference and lifestyle are released. When the two datasets are integrated, the personal medical information may be revealed. Anonymity is used in such instance to preserve the privacy of the individual when two different datasets are integrated. To prevent these attack k-anonymity model was proposed [3] to avoid accessing data based on background knowledge. A dataset satisfied k-anonymity [4] which requires that for a data instance, there are at least $k - 1$ unique data instance that share the same feature vector. K-anonymity uses generalization and suppression of attributes to achieve anonymity. But generalization and suppression do not consider the utility of data for classification after anonymity [5]. Later, the extension of k-anonymity related-algorithms like l-diversity, m-variance, differential privacy, and t-closeness was proposed. But anonymizing the real-world dataset generally uses generalization or suppression which reduces the utility of the data. Hence a suitable preprocessing technique is needed to select the suitable features to increase the utility of the data [6].

The FS [7] aims to obtain a subcategory of features from the original problematic that still closely related to it. This subclass is commonly used to train a learning system, with added supports reported in the literature [1, 8]. Feature selection [9] is a preprocessing step to machine learning to reduce dimensionality, removing irrelevant and redundant data to improve the result comprehensibility. It is the process of selecting the subset of features from dataset so that the dimensionality of dataset is reduced. With the increasing size of data, dataset may contain thousands of features with inappropriate and duplicate information which damage the efficiency of learning algorithm. Hence a FS [10] algorithm is necessary for high dimensional data when the number of features used is large in number to apply anonymization algorithm. Dimensionality [19] problem is faced by data mining algorithm and the predictor variable or instance increase with the size of data. The computational cost increases due to huge number of features which seriously affect the operation of data mining algorithm. Feature selection is classified as Wrapper, Filtering and Embedded methods. Wrapper method depends on learning algorithm [11] and Filtering method selection is based on data connected measures [12]. Embedded methods have optimal feature subset within classifier [13].

The research problem in the above task is the selection of optimal feature set for utility preserving entity anoymization, where the utility is considered with respect to classification performance. In existing work, k-anonymity is accomplished by suppression or generalization of cell values, whereas the same can be achieved by selected features, yet ensuring that the selected features retain the classification utility of the dataset.

1.1 Our Contributions

In this work, we consider the task of feature selection by fast correlation-based filtering under anonymity constraint. Normally, feature selection and anonymity algorithm are applied in isolated environment.

Our contribution is summarized as follows:

1. We proposed method for anonymization with preprocessing done by fast correlation-based filtering feature selection. For this, we design a new anonymization metric, named correlation-based k-anonymity which is suitable for high dimensional data.
2. We propose method for solving the above task and show experimental results to validate the effectiveness of these methods.
3. We show that the effectiveness of the proposed methods and show how the privacy-aware feature selection affects the performance of these applications.

The rest of the paper is organized as follows. Section 2 provides some background information about big data, FCBF algorithm, k-anonymity and MapReduce. Section 3 explains the MapReduce framework for feature selection under anonymity constraint. Sections 3.1 and 3.2 cover correlation-based filter solution and FCBF and k-anonymity in MapReduce framework. The experimental results are discussed and examined in Sect. 4. Finally, Sect. 5 summarizes the conclusions of the paper.

2 Background and Key Issues

2.1 Data Preprocessing Dealing with Big Data Scalability Problem

Below is the list of preprocessing techniques applied to big data which does not solve the scaling problem of big data:

- **Instance reduction** [11]: These techniques will have low performance due to the fact of carrying out the same learning algorithm iteratively on the original set of data There are number of techniques for instance reduction to attain subgroups of data from large databases are available but this technique requires high computation capabilities for dealing with big data and follows a iterative procedure.
- **Missing values imputation** [14]: Filling missing values is a hard problem since it is difficult to identify relationship among the data to fill the missed value.
- **Noise treatment** [15]: In order to identify noise and remove noise, noise treatment techniques are applied based on the correlation between datasets and decision taken based on ensembles value.

To get the ideal solution for data preprocessing is still a open issue for all researcher [16]. It will be more complicated when considering scalability issue in big data circumstances. This difficulty is also subjective by other features like data preprocessing technique used, intermediate results in handling large dataset, its capability of handling various volumes of data, parallelization and iterative processing, or the contribution it desires for processing or the productivity it delivers.

2.2 Anonymization Algorithm—K-Anonymity

Hiding identity and sensitive data of individuals of data record is labeled as data anonymization and hence, the privacy of data owners is preserved but still comprehensive data is visible to data analyst intended for data mining and research.

In k-anonymity [10], the info about an individual during exposed information cannot be recognizably identified from a least possible of $k - 1$ individual there in exposed data which means that QID should seem to be appear in at least k records. Generalization and suppression are the two techniques in k-anonymity. Generalization is achieved by replacing QID with less unambiguous values till k similar records are found in the datasets. Suppression is the process of hiding sensitive value (Fig. 1).

3 Big Data: MapReduce

MapReduce is the solution for the big data problem by enabling parallel processing. MapReduce [3] is a framework for data processing enormous volume of data in parallel. Map and Reduce are the two processes in big data. When the input training data is given to Mapper function, it processes the data and generates a key-value pair. The Reducer takes the intermediate key-value pair which further compresses the data to smaller size scalability and cost-effectiveness is achieved in MapReduce framework. MapReduce has become familiar because of it property of handling large dataset. In our approach, we selected MapReduce to handle big data and our feature selection and k-anonymity algorithm are applied in parallel to MapReduce framework to get a reliable anonymized dataset.

Suppression in k-Anonymity: Given t1, t2, . . . , tn ∈ Σ p, an Anonymity parameter k, achieve a k-Anonymous function for suppression x such that c(x) is diminished

Generalization in k-Anonymity: Given t1, t2, . . . , tn ∈ Σ p, an Anonymity parameter k, achieve a k-Anonymous function for generalization y so that cost(y) is diminished

Fig. 1 k-anonymity with generalization and suppression

3.1 Preprocessing Algorithm—A Correlation-Based Filter Approach

Goodness measure of feature for classification is evaluated by correlation in the features. A feature selected is worthy and correlated, if it is extremely relevant and correlated to the group and not redundant and not correlated to remaining feature [17]. One such goodness measure is by using symmetrical uncertainty (SU) [15] which is calculated by correlation analysis of features. SU identifies the relevant features and avoid the redundant features.

$$\mathrm{SU} = 2.\left[\frac{H(X) - H(X/Y)}{H(Y) + H(X)}\right] \tag{1}$$

$$H(X) = -\sum_{x \in X} P(x) \log_2 p(x) \tag{2}$$

$$H(X/Y) = -\sum_{y \in Y} P(y) \sum_{x \in X} P(x/y) \log_2 p(x/y) \tag{3}$$

The FCFS concept uses best-first search as its search strategy. Initially, algorithm selects the empty set of features and starts searching all possible single feature expansion. Equation (1) is used to generate new subsets and added iteratively to a priority queue according to weight. In the successive repetition, the finest subgroup from the queue is selected for further development. If expanding the finest subgroup fails to produce an enhancement in the total quality, it is counted as failure and the subsequent finest subgroup from the line or queue is selected. If five consecutive fails are found, the algorithm considers this as stopping criteria and stops further expansion. In this way, FCFS algorithm reduces the redundancy and selects the highly correlated features. The FCFS [5] can optionally use a heuristic search to enable insertion of all features highly correlated with themselves and with already selected features. FCFS algorithm is explained in Fig. 2.

3.2 Algorithm and Analysis

Based on the approach presented before, we develop an algorithm, named fast correlation-based filter using MapReduce (FCBF).

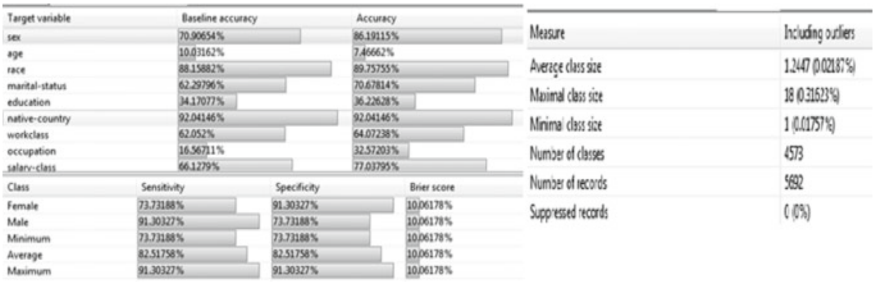
given a data set with
input: $T(S_1;\ S_2;\ :::;\ S_N;C)$ // a training data set
ω – predefined threshold
output: T_{best} // an ideal subset
M mappers one Reducer
Each mapper has FCBF & k-anonymity and proceeds as follows:
iterate j from 1 to *N start*
compute *TUi;c* for S_i;
if (*TUi;c* ¸ ω)
add S_i to T_{0list};
finish;
order T_{0list} in descending *TUi;c* value;
S*p* = *getF irstComponent*(T_{0list});
do start
Sq = *getNextComponent*(T_{0list}; *Sp*);
if (*Sq* <> NULL)
do start
S0 q = *Sq*;
if (*TUp;q* ¸ *TUq;c*)
eliminate *Fq* from T_{0list};
Sq = *getNextComponent* (T_{0list} ; *S0q*);
else *Sq* = *getNextComponent*(T_{0list}; *Sq*);
finish until (*Sq* == NULL);
Sp = *getNextComponent*(T_{0list}; *Sp*);
finish until (*Sp* == NULL);
$T_{best} = T_{0list}$;
Execute k-anonymity
Mapper outputs anonymized set
Reducer collects and merges all data, until k-anonymity is achieved
finish;

Fig. 2 FCBF algorithm with k-anonymity in MapReduce

4 Empirical Study

This section evaluates our projected algorithm with respect to number of relevant selected features, privacy risk, learning accuracy and performance in terms of data utility (Figs. 3 and 4).

A ARX anonymization tool is used to achieve anonymization using k-anonymity algorithm. We use adult dataset as the benchmark for testing FCBF to achieve anonymization [18]. Adult dataset is multivariate dataset with categorical and integer attribute and has 14 attributes, 48,842 instances with missing values. The result shown in Fig. 3 is the analysis of privacy risk of the data when the anonymization is applied to such dataset. Figure 3 shows the analysis of FCBF algorithm and k-anonymity algorithm with the adult datasets.

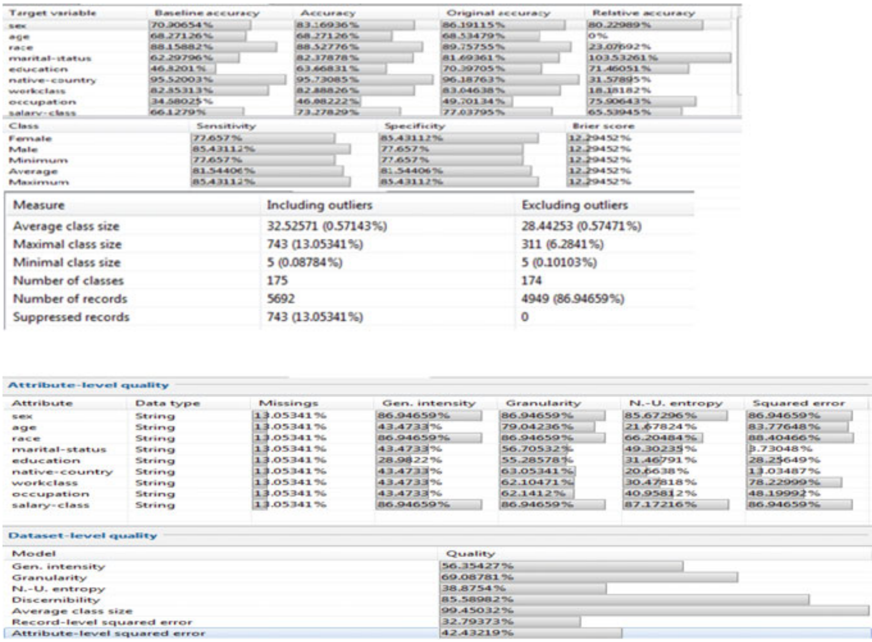


Target variable	Baseline accuracy	Accuracy
sex	70.90654%	86.19115%
age	10.03162%	7.46662%
race	88.15882%	89.75755%
marital-status	62.29796%	70.67814%
education	34.17077%	36.22628%
native-country	92.04146%	92.04146%
workclass	62.052%	64.07238%
occupation	16.56711%	32.57203%
salary-class	66.1279%	77.03795%

Class	Sensitivity	Specificity	Brier score
Female	73.73188%	91.30327%	10.06178%
Male	91.30327%	73.73188%	10.06178%
Minimum	73.73188%	73.73188%	10.06178%
Average	82.51758%	82.51758%	10.06178%
Maximum	91.30327%	91.30327%	10.06178%

Measure	Including outliers
Average class size	1.2447 (0.02187%)
Maximal class size	18 (0.31623%)
Minimal class size	1 (0.01757%)
Number of classes	4573
Number of records	5692
Suppressed records	0 (0%)

Fig. 3 Analysing the privacy risk without FCBF and anonymization algorithm



Target variable	Baseline accuracy	Accuracy	Original accuracy	Relative accuracy
sex	70.90654%	83.16936%	86.19115%	80.22989%
age	68.27126%	68.27126%	68.53479%	0%
race	88.15882%	88.52776%	89.75755%	23.07692%
marital-status	62.29796%	82.37878%	81.69361%	103.53261%
education	46.8201%	63.66831%	70.39705%	71.46051%
native-country	95.52003%	95.73085%	96.18763%	31.57895%
workclass	82.85313%	82.88826%	83.04638%	18.18182%
occupation	34.58025%	46.08222%	49.70134%	75.90643%
salary-class	66.1279%	73.27829%	77.03795%	65.53945%

Class	Sensitivity	Specificity	Brier score
Female	77.657%	85.43112%	12.29452%
Male	85.43112%	77.657%	12.29452%
Minimum	77.657%	77.657%	12.29452%
Average	81.54406%	81.54406%	12.29452%
Maximum	85.43112%	85.43112%	12.29452%

Measure	Including outliers	Excluding outliers
Average class size	32.52571 (0.57143%)	28.44253 (0.57471%)
Maximal class size	743 (13.05341%)	311 (6.2841%)
Minimal class size	5 (0.08784%)	5 (0.10103%)
Number of classes	175	174
Number of records	5692	4949 (86.94659%)
Suppressed records	743 (13.05341%)	0

Attribute-level quality

Attribute	Data type	Missings	Gen. intensity	Granularity	N.-U. entropy	Squared error
sex	String	13.05341%	86.94659%	86.94659%	85.67296%	86.94659%
age	String	13.05341%	43.4733%	79.04236%	21.67824%	83.77648%
race	String	13.05341%	86.94659%	86.94659%	66.20484%	88.40466%
marital-status	String	13.05341%	43.4733%	56.70532%	49.30235%	3.73048%
education	String	13.05341%	28.9822%	55.28578%	31.46791%	28.25649%
native-country	String	13.05341%	43.4733%	63.05341%	20.6638%	13.03487%
workclass	String	13.05341%	43.4733%	62.10471%	30.47818%	78.22999%
occupation	String	13.05341%	43.4733%	62.1412%	40.95812%	48.19992%
salary-class	String	13.05341%	86.94659%	86.94659%	87.17216%	86.94659%

Dataset-level quality

Model	Quality
Gen. intensity	56.35427%
Granularity	69.08781%
N.-U. entropy	38.8754%
Discernibility	85.58982%
Average class size	99.45032%
Record-level squared error	32.79373%
Attribute-level squared error	42.43219%

Fig. 4 Analyzing the privacy risk with FCBF and k-anonymity algorithm

4.1 FCBF and k-Anonymity in MapReduce Framework

As shown in Fig. 5, original datasets are divided into number of subset and given to Mapper phase where the fast correlation-based features solution algorithm is applied to select the relevant features, remove the redundant data and k-anonymity is applied to preserve privacy. Hence the final output of MapReduce framework is anonymized dataset with relevant features and without redundancy without losing data utility.

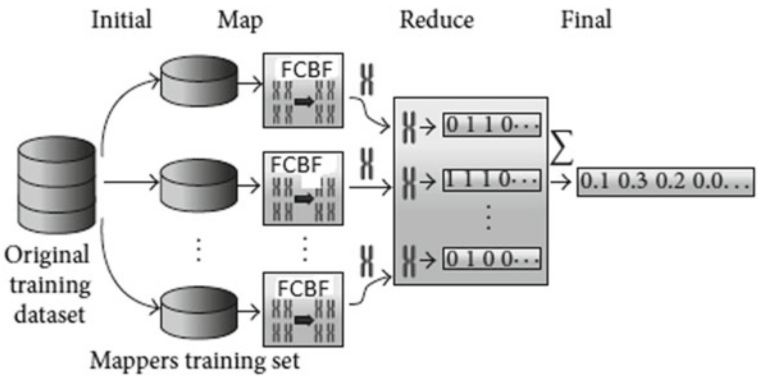


Fig. 5 K-anonymity-based FCBF in MapReduce

5 Conclusion

In this paper, contextual knowledge of big data, feature selection and anonymization has been revisited. Then this paper aims to apply FCBF and k-anonymity algorithm in benchmark dataset like adult dataset, and the results are compared with preprocessing and without preprocessing. Increasing data volume is a challenging issue for data analyst to share, mine and analysis and hence further research is needed in this direction. Alternative track of upcoming work could be research into correlation-based feature selection with k-anonymity for big data with streaming features.

Compliance with Ethical Standards All author states that there is no conflict of interest. We used our own data. Humans and animals are not involved in this research work.

References

1. Kambatla, K., et al.: Trends in Big Data Analytics. Elsevier (2014)
2. Chen, M., Lin, M.: Big Data: A Survey, vol. 19, pp. 171–209. Springer (2014)
3. Sweeney, L.: k-anonymity: a model for protecting privacy. Int. J. Uncertainty Fuzziness Knowl.-Based Syst. **10**(5), 557–570 (2002)
4. Chen, P., et al.: Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data. Information Sciences (2014)
5. Zhang, B., et al.: Feature selection for classification under anonymity constraint. Trans. Data Priv. **10**, 1–25 (2017)
6. Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artif. Intell. **97**(1–2), 273–324 (1997)
7. Hall, M.A.: Correlation-Based Feature Selection for Machine Learning. Waikato University, Department of Computer Science (1999)
8. Chandrashekar, G., Sahin, F.: A survey on feature selection methods. Comput. Electr. Eng. **40**(1), 16–28 (2014)
9. Garcia, S., et al.: Big data preprocessing: methods and prospects. Big Data Analytics (2019)

10. Li, J., et al.: Feature selection: a data perspective. ACM Comput. Surv. **50**(6), Article 94, Publication date (2017)
11. Wald, R., et al.: Comparison of stability for different families of filter-based and wrapper-based feature selection. In: 12th International Conference on Machine and Application (2013)
12. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
13. Saeys, Y., Inza, I., Larra˜naga, P.: A review of feature selection techniques in bioinformatics. Bioinformatics **23**(19), 2507–2517 (2007)
14. De, S., et al.: Bayes Wipe: a scalable probabilistics framework for cleaning big data. J. Data Inf. Q. ACM (2016)
15. Peralta, D., et al.: Evolutionary Feature Selection for Big Data Classification: A Map Reduce Approach. Hindawi (2015)
16. Zhou, B., Pei, J.: The k-Anonymity and l-Diversity Approaches for Privacy Preservation in Social Networks Against Neighborhood Attacks. Springer (2010)
17. Yu, L., et al.: Feature selection for high-dimensional data: a fast correlation-based filter solution. In: Proceeding of the Twentieth International Conference on Machine Learning (2003)
18. Zhao, Z., et al.: Graph regularized feature selection with data reconstruction. IEEE Trans. Knowl. Data Eng., **28**(3) (2016)
19. Raul-Jose, et al.: Distributed Correlation-Based Feature Selection in Spark. Information Science (2018)

Interoperability in Smart Living Network—A Survey

M. Durairaj and J. Hirudhaya Mary Asha

Abstract Embedded systems or embedded devices are the basic hardware needed for Internet of things (IoT). The "Internet" and "Things" are merged together to make them work as a powerful technology called the Internet of Things (IoT). The IoT is the bombardment of the real-world objects into Internet-based things that can exchange massive amount of data with minimal human interventions. However, the security and privacy concepts of infrastructural engineering are highly critical. Smart IoT projects with smart devices are worldwide known. Smart living is a residence filled with technology that processes the information expected to respond to the needs of the occupants. It promotes the locator comfort, convenience, and security through the technology. Smart devices with the IoT services convert the raw data, read from the home sensors and the actuators, and respond with operational commands to control the home network or home appliances technology from anywhere with the help of Internet. Devices such as smartphones, tablets, and personal computers are used to communicate with these technologies. The home technology highly depends on the interoperability among the communication system architecture to achieve security. In this paper, a survey is done on how the interoperability is built with the efficient use of power energy in the existing smart living.

Keywords Internet of Things (IoT) · Energy · Interoperability · Security

1 Introduction

Today, humans are addicted to technologies for their daily chores. Every household has more devices connected to the Internet with a uniquely assigned IP address,

M. Durairaj · J. Hirudhaya Mary Asha (✉)
Assistant Professor, Department of Computer Science, Bharathidasan University, Tiruchirapalli, Tamil Nadu, India
e-mail: hirudhaya_ashaa20@yahoo.co.in

J. Hirudhaya Mary Asha
Research Scholar, Department of Computer Science, Bharathidasan University, Tiruchirapalli, Tamil Nadu, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_7

a combination of numbers used to locate the devices on the network. The operational messages routed to the devices are completely controlled and monitored via the Internet. The Internet follows some standardized protocols designed suitable for information and communication facilities [1, 2]. People can operate the household items from the outdoor location and manage smart communication between devices and application by linking the perfect architecture [3]. In M2H or M2M communication, a number of devices with varying levels of complexity can transmit through a gateway. IoT devices route information with different energy levels to the connected gateway at the same time [4]. Various technologies such as machine learning, pervasive computing, and artificial intelligence are used to recognize and record the activity of the people in their daily life. The research community focuses on the technical possibilities to be done and also the lack of interoperability of smart living devices.

2 Related Works

The home we live in or the place we work can act smart with automation control systems. Smart units of home are connected together with a user interface for interacting with the home appliances [1]. A "smart living" can be defined as a residence filled with technology expected to respond to the needs of the occupants. To promote the home occupant's comfort, convenience, and security, interfaces such as gateway or middleware are built inside smart home. Smart living network is constructed with three types. (1) through wired home electric line network, (2) through separate physical line network for signal transporting, and (3) through a wireless network. The wired networks have their own strength and weakness to connect the source and destination, but the wireless networks provide more advantages than the other two. Smart sensors are designed to sense gas leakages, temperature, recognize the face, energy consumption, humidity, control television, door lock status, doorbell, refrigerator storage sensors, light intensity, smoke, etc., continuously in regular intervals within range and send alerts to the authorized user [2].

A report is released by "TELECOMMUNICATION ENGINEERING CENTRE" that explains the year-wise changes in how the society adopted to the automation of smart living in their everyday life [5]. Using the data sent by the sensors, specialized software or intelligent processing agents trigger actions in the environment by means of actuators. For example, the sensor can smell smoke by air sampling technique, and if any smoke is detected by the sensor, it warns the proprietor via SMS and sends alerts to the nearby fire station [3]. The middleware is used for their inter-communicability [4]. A smart living control can handle enormous heterogeneous devices in the homes and provide user-oriented services like monitoring adult, elders, or kids by using appropriate algorithms [6, 7]. Smart living is constructed with inter-networks of three blocks [8]. They are:

Fig. 1 Model picture of smart living [8]

1. Home Automation System (HAS), an outline that shows the inter-linking of home appliances (e.g., refrigerator, heaters, lights, televisions, and so on)
2. Home Controlling System (HCS), an application, which creates automated commands based on the data sensed by the sensor attached to home appliances and forwards the commands to the actuators to perform some operation.
3. Home Network System (HNS), a complete framework linkage of HAS and HCS communication to exchange information [8] (Fig. 1).

3 Home Framework Designing

The smart living is the projected output gained from the interaction of hardware and software. While designing smart living, there are three main concepts to be considered. They are smart objects participated in the device layer, hardware infrastructure that gives smart capabilities to the object, and finally software layer used to activate the smart capability [9]. Experts are arguing with the smart living functionality that seems to be a better system for managing everyday life. So, the instruments used are goal-oriented part for next-level development [10]. The following chart predicts the dimensional growth of technology in home from 1970 to 2025 [11] (Fig. 2).

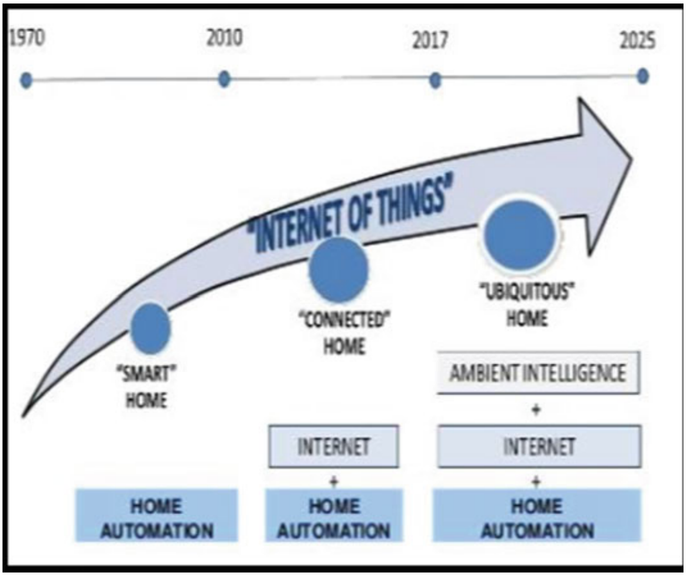


Fig. 2 Growth of smart living automation

3.1 Middleware or Gateway

Middleware is the part of the IoT system architecture which stands between the devices and application layer for providing services by effective communication. It is a software interface designed to transfer messages or queries between the device end and the user application end [12]. The server application detects the client's geographical location and performs concurrent processing, balancing load and managing transactions by sharing client requests to multiple servers and providing fast access to resources using cyber security [13]. Different middleware types are used for connecting applications, web and cloud services for trigger-specific functions. Other components can communicate regardless of their medium using messaging frameworks like Simple Object Access Protocol (SOAP), web services, Representational State Transfer (REST), or JavaScript Object Notation (JSON) [12, 14–19].

3.2 Communication Protocols

The smart living devices are designed to interface with low-bandwidth network than higher bandwidth. The Institute of Electrical and Electronics Engineer (IEEE) working group enhanced the standards supporting M2M communication. In the local home network, different devices are connected with technologies to forward information in order to trigger the action. The wireless sensor networks (WSN) for the home are composed of numerous sensors and connected with different communication technologies [20]. They are as follows:

RFID

Radio frequency identification device (RFID) tags are like wireless microchips attached to the home objects or things. RFID has some in-built power and computational capacity, storage, and an antenna for connecting with the radio signal. The tags are read by the RFID reader. The RFID tags and readers can communicate via radio frequency waves to sense, detect, and communicate with the environment [2, 21, 22] (Table 1).

IEEE 802.11–WiFi

WiFi uses global 2.4 GHz UHF and 5 GHz SHF ISM radio bands. 802.11b, 802.11g, and 802.11n activate on 2.4 GHz ISM band [20, 23, 24] (Table 1).

IEEE 802.15.4—ZigBee

ZigBee supports services like home network start-up, routing of messages in multi-hop model, and management of connection or disconnection of the nodes in the network. Low-power energy consumption, cheap and easy installation procedure has no fixed network size and message routing [14, 20] (Table 1).

Z-Wave

In Z-Wave, each device has an identification code. The controller in the network recognizes the devices and determines its location by an embedded code. Then, the controllers increment the network routing table of representing the arrival. The Z-Wave protocol uses the Source Routing Algorithm (SAR) to route the messages in the fastest route. The sensor gateway/actuator gateway collects internal house data from sensors using Z-Wave. It operates between 868 and 900 MHz [14, 20] (Table 1).

IEEE 802.15.1—The Bluetooth Low Energy (BLE)

BLE is the latest version of the Bluetooth 4.0 specifications. It is highly efficient on low-power sensors. The frequency bandwidth of BLE is same as the classic Bluetooth protocol, but BLE modulation is slightly different which results in maximum signal strength [5, 25] (Table 1).

IEEE802.11ah—HaLow

HaLow is a low-cost WiFi standard specifically designed for IoT applications like smart living and smart grid automation with low cost of implementation with less power consumption. This standard provides greater flexibility of supporting more than 8000 devices on a single application for communication with capacity to cover different distance ranges [12, 14–19] (Table 1).

Table 1 Consolidated overview of communication protocol

Wireless technology	Frequency used	Range	Data rate	Network topology
RFID	100 kHz to 5.8 GHz	0–3 m	640 Kbps	P2P
Bluetooth	2.5 GHZ	Up to 100 m	1–3 Mbps	Star, P2P
BLE	2.5 GHZ	Up to 100 m	1 Mbps	P2P, star
ZigBee	2.5 GHZ	10–100 m	250 Kbps	P2P, star, tree, mesh
WiFi	2.5 GHZ	150–200 m	54 Mbps	P2P, star
Z-Wave	850–950 MHz	100 m	9.6–100 Kbps	Mesh
HaLow	900 MHz	1.5 km	18 Mbps	Mesh, P2P, tree

3.3 Home API

The application programming interfaces or APIs are the points of interaction (POI) between the smart devices and the Internet. Web services are the APIs designed for Internet of things (IoT). A web service is software that any application can parse it. Connected devices access the web services via address of the web API that has developer registration and API key control. The interfaces can be built using different languages such as SOAP, REST, or XML/JSON [25, 26].

3.4 Service Projects

There are numerous services offered by IoT projects. Some of the main services are specification, design, and provision service.

Monitoring home locators
Sensors are fixed for monitoring the physiological and behavioral parameters of the person continuously such as heart rate, blood pressure, mobility, and emotional and mental state of the person residing inside the house. Secure connection is established by decision-making algorithms and transmitting data via wireless body communication networks. Projects such as motion detection, location tracking, facial expression, reading urine, and sweat of the elderly or people with disability are highly possible in IoT environment today. These come under the type of provision services [11, 13, 27].

Surveillance and security systems
In smart living, cameras, alarms and locks are used to block unknown persons entering the home and allowing an authorized person to access the devices inside the home and make decisions of the personalized services available. It is possible to analyze the images through face recognition algorithm and cryptographic keys [15]. The

camera sensor tracks the happenings of home, especially useful for monitoring from far location.

Gesture Recognition Project

Gesture or hand movement models are stored in the middleware database for authorization process [20]. The gestures which are stored for authenticity are recognized using algorithms specially designed for template matching and control home things remotely. These algorithms compare the actual image with the template sign to find the match. In order to recognize the characters, numbers and objects templates are used. The templates are searched from the database based upon the size of the image.

Indoor positioning energy efficiency projects

Devices not linked directly with the middleware use some software proxies for connecting. The power meter is such a proxy used for linking devices done with the user mobile platform and middleware for analyzing, monitoring, and controlling the device. Readymade projects such as semantic smart metering and semantic web technology are used to link power-related data semantically. This project is used for public buildings and private houses. Nowadays, power banks, smart grids, and solar energy are used to replace electricity used for smart living. Environmental pollution could also be avoided [26].

Energy Management

Household energy usage is one of the major impacts on the world's energy consumption [5, 11, 13, 20, 27, 28]. The home system with smart capability consumes more energy than normal houses. The energy management of home is how, when, and where the home appliances are to be activated, deactivated, and controlled [20, 29].

4 Proposed Interoperability in Living Network System

Smart living is an integration of various physical sensors and communication technology. Choosing the right communication protocol between devices improves the overall system performance [26]. Interoperability is the communication trade-off between heterogeneous devices. In the architecture of living network system (see Fig. 3), devices are arranged in the physical layer which forwards the event happens inside the home through the communication protocols and reaches the routers. The routers route the messages to the API. Depending upon the user response, the API triggers the action in the living system through the intermediate middleware [26].

A good interoperability is established by following good protocols in each and every layer of the living system network. No two layers allotted to do same job. A relay of work shuffled in the architecture. The communication between layers takes place through the communication protocols, application programming interface protocols, and the gateway protocols. Uninterrupted power is supplied to the home network to avoid interruption. The devices involved in this network have limited power storage

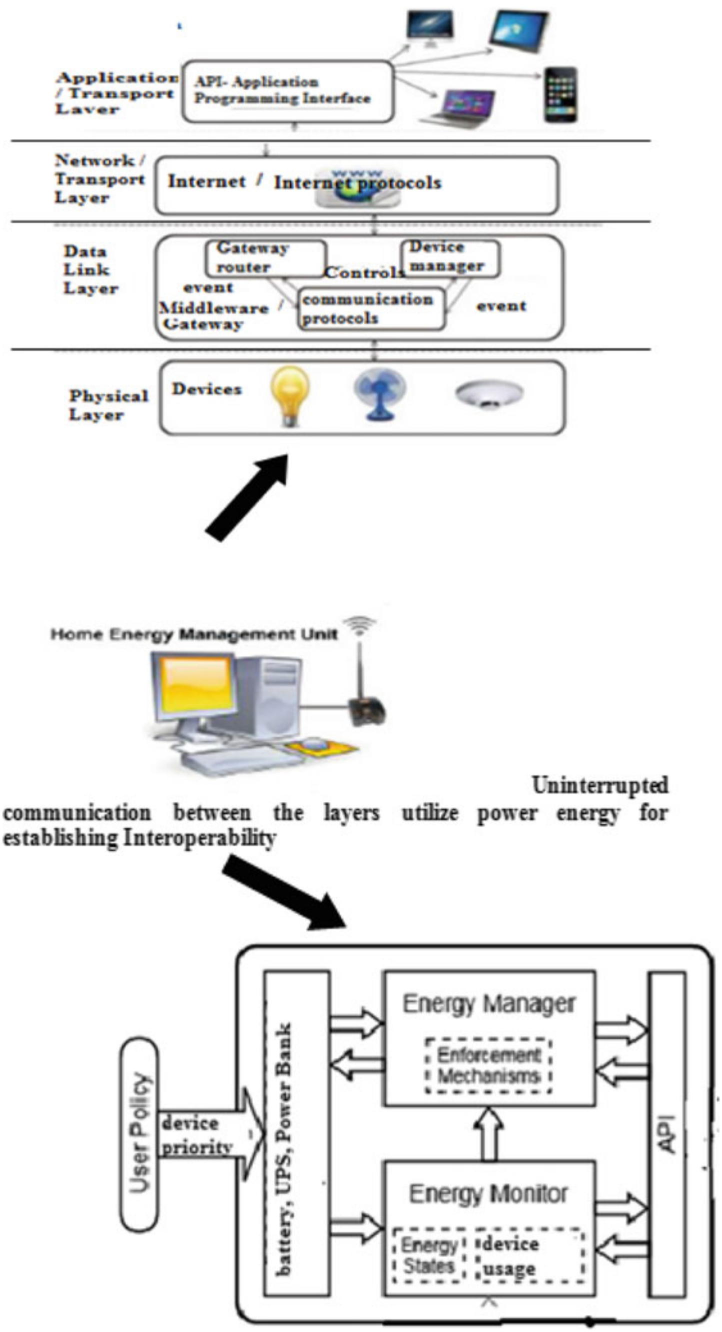


Fig. 3 Proposed smart living network system with efficient use of power energy to avail interoperability

facility. Due to the mobility of the locators, location-based search is helpful, so there is a need of battery backup to avoid interruption from outside. If all the layers are united finely in the network but still has a huge loss of energy, then the entire network is unnecessary one. The energy supplied to every device in the network is based on the preferences of the user. Automatic supply of power energy to the device is dynamic based on the time changes (e.g., door sensor and camera of indoor and outdoor are preferred more than fan and air conditioner in the far location).

The energy manager is a component involved in energy management application. It depends on the communication from the power meter, and the middleware in the energy unit translates the messages between the participating device and the API, which are designed for energy management. If the external power energy is not available, the device can get charge from the artificial power source such as UPS battery. From the energy management API, the user can able to be aware of the energy state and get information based on the device energy consumption, and based on this locator, the devices consume power. It is concluded from the proposed smart living network that power energy is efficient for the working condition of the framework interoperability. Based on the range of communicating distance, suitable protocols such as ZigBee and HaLow are highly used nowadays for smart living mesh network [25].

5 Research Consideration

The sensors are the spies used to gather real-life scenario and trigger action accordingly, to provide security and safety to the locators. It is necessary to find the user-centric, low-cost, innovative, interoperable, and integrated home environment by focusing on industry standards. Middleware or building gateway is the syntactic and semantic perspective of the project. Interfaces involved must improve the interoperability among the participating layers. Designing and controlling home smartly improves the quality of life and avoids thefts by the timely information exchange in a critical situation; intimation of leakage of gas or fire is done automatically by enhancing the routing protocols involved in the network system for interoperability.

6 Conclusion

The invention of the information communication technology has brought major changes in our everyday life. Smart living has a high potential in the present and future business sector platform. Various impacts noted on intelligent living places such as complexity in construction, inflexibility in connection, power energy demands, interoperability problem among the devices, lack of managing appliances are considered as the key barriers to their adoption. Numerous researches are undertaken at IoT projects, but confidence building on smart living is still very low for the users because of the cost, safety, and security risks. Initially and very importantly, home appliances require fine connecting gateway and the server requires high energy harvesting techniques and interoperability features for enhanced continuous flow of interaction more than security constraints. This paper acknowledged the need of power energy to establish interoperability supported by well-designed and flexible protocol standards that can accurately and continuously infer the physiological activities and patterns of sensed data.

References

1. Muntjir, M., Rahul, M., Alhumyani, H.A.: An analysis of internet of things (IoT): novel architectures, modern applications, security aspects and future scope with latest case studies. Int. J. Eng. Res. Technol. (IJERT) **6**(06) (2017)
2. Ashton, K.: That internet of things thing. RFID J. (1991)
3. Malche, T., Maheshwary, P.: Internet of Things (IoT) for building Smart living System. I-SMAC-2017
4. Balevi, E., Al Rabee, F.T., Gitlin, R.D.: ALOHA-NOMA for massive machine-to-machine IoT communication. arXiv preprint arXiv:1803.09323 (2018)
5. Lin, H., Bergmann, N.W.: Iot Privacy and Security Challenges for Smart living Environments. Information **7**, 44 (2016)
6. Telecom ITU.: Standardization sector recommendation of ITU (1996)
7. Barsocchi, P., Calabrò, A., Ferro, E., Gennaro, C., Marchetti, E., Vairo, C.: Boosting a low-cost smart living environment with usage and access control rules. Sensors 2018, Published: 8 June 2018
8. Yun, M., Yuxin, B.: Research on the architecture and key technology of IOT applied on the smart grid", Advances in energy engineering. In: (ICAEE) International Conference, June 2010 [4]
9. Wang, H., Saboune, J., El Saddik, A.: Control your smart living with an autonomously mobile smart phone. In: IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 1, 6, 15–19 July 2013
10. Byun, J., Hong, I., Lee, B., Park, S.: Intelligent household LED lighting system considering energy efficiency and user satisfaction. IEEE Trans. Consum. Electron. **59**(1), 70–76 (2013)
11. Barsocchi, P., Calobro, A.: Boosting a low-cost smart living environment with usage and access control rules. Sensors 2018
12. Muntjir, M., Rahul, M., Alhumyani, H.A.: Architectures, modern applications, security aspects and future scope with latest case studies. IoT in Latest Trends Int. J. Eng. Res. Technol. (IJERT), JERTV6IS060238, **6**(06) (2017)
13. Attaran, M., et.al.: Critical success factors and challenges of implementing RFID in supply chain management. Supply chain Operating Management (2012)
14. Solaimani, S., Keijzer-Broers, W., Bouwman, H.: What we do and don't know about the smart living: an analysis of the smart living literature. Indoor, and Build Environment (2013)
15. Hargreaves, T., Wilson, C.: Analytical Framework for Research on Smart livings and Their Users. Springer, Berlin (2017)
16. Wenbo, Y., Quanyu, W., Zhenwei, G.: Smart living implementation based on internet and WiFi technology. In: Proceedings of the 34th Chinese Control Conference, 28–30 July 2015, Hangzhou, China
17. Li, M., Lin, H.-J.: Design and implementation of smart living control systems based on wireless sensor networks and power line communications. IEEE Trans. Ind. Electron. **PP**(99), 1, 1 Dec. 2014

18. Haradaya, H., Mizutaniy, K., Fujiwaray, J.: IEEE 802.15.4g Based Wi-SUN communication systems. IEICE Trans Commun. **E100–B**(7) (2017)
19. Baños-Gonzalez, V., Afaqui, M.S., Lopez-Aguilera, E., Garcia-Villegas, E.: IEEE 802.11ah: a technology to face the IoT challenge. Sensors **16** (1960)
20. Bitterman, N., Shach-Pinsly, D.: Smart Living- a Challenge for Architects and Designers Architectural Science Review. Taylor & Francis (2015)
21. Vanus, J., Cerny, M., Koziorek, J.: The Proposal of the Smart living Care Solution with KNX Components. 978-1-4799-8498-5/15/$31.00 ©2015 IEEE
22. Fathany, M.Y., Adiono, T.: Wireless protocol design for smart living on mesh wireless sensor network. In: 2015 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) 9–12 Nov 2015
23. Hsu, Y.-L., Chou, P.-H., Chang, H.-C.: Design and implementation of a smart living system using multi sensor data fusion technology. Sensors 17, 1631 (2017)
24. Telecommunication Engineering Centre, Department of Telecommunications, Ministry of Communications, Government of India, Technical Report on M2M enablement in Smart livings, TEC-TR-IoT-M2M-007-01, R1.0 release, 31/03/2017
25. K¨uhn, E., Prellwitz, M., Rohrer, M., Sieck, J.:"A distributed middleware for applications of the internet of things. In: The 7th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 12–14 Sept. 2013, Berlin, Germany
26. Son, H., Tegelund, B., Kim, T., Lee, D.: A distributed middleware for a smart living with autonomous appliances, 0730-3157/15 $31.00 ©. In: 2015 IEEE, 39th Annal International Computers, Software & Applications Conference
27. Mocanu, I., Florea, A.M.: A Multi-Agent System for Human Activity Recognition in Smart Environments. Springer, Berlin (2011)
28. Gajewski, M., Batalla, J.M., et.al.: A Distributed IDS Architecture Model for Smart Living System. Springer, 30 Aug. 2017
29. Fuller, J.D., Ramsey, B.W.: Rogue Z-Wave controllers: a persistent attack channel. In: 40th IEEE Annual conference on Local Area Networks (2015)

Sentiment Analysis of Bengali Reviews for Data and Knowledge Engineering: A Bengali Language Processing Approach

Sharmin Akter Milu, Md. Ismail Siddiqi Emon, Sabiha Sunjida Ahmed, M. J. Alam, Sheikh Shahparan Mahtab, Jamal Ahmed Bhuiyan, Md. Fahad Mojumder and Mahedy Hasan

Abstract Opinion mining is very much attractive field inmachind learning system as it is very much needed for natural language processing. The opinion mining of Bengali written English word has been done successfully using four different classifiers—support vector machine, naive Bayes, logistic regression and random forest. For the work data set was extracted from local online shops using pursehub. The work was done with vital steps—data preparation, classifying reviews according to sentiment score and evaluate the system in all steps. The F_1 score was obtained 85.25%, 88.12%, 88.12%, 82.43% for naïve Bayes, logistic regression, SVM, random

S. A. Milu
Department of CSTE, Noakhali Science and Technology University, Noakhali, Bangladesh
e-mail: sharminmilu7@gmail.com

Md. I. S. Emon · S. S. Ahmed
Department of CSE, Feni University, Trunk Road, Feni 3900, Bangladesh
e-mail: emonsahriar0@gmail.com

S. S. Ahmed
e-mail: ahmedsunjida@gmail.com

M. J. Alam · S. S. Mahtab (✉)
Department of EEE, Feni University, Trunk Road, Feni 3900, Bangladesh
e-mail: mahtabshahzad@gmail.com

M. J. Alam
e-mail: alameee1993@gmail.com

J. A. Bhuiyan
Department of CE, Feni University, Trunk Road, Feni 3900, Bangladesh
e-mail: jamalrony35@gmail.com

Md. F. Mojumder
Department of ECE, North South University, Dhaka, Bangladesh
e-mail: fahad.mojumder@northsouth.edu

M. Hasan
Department of BASc (Marine Engineering), AMC, University of Tasmania, Hobart, TAS, Australia
e-mail: mahedishuvro2020@gmail.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_8

forest, respectively. The accuracy score was obtained 85.31%, 88.05%, 88.11%, 81.82% for naïve Bayes, logistic regression, SVM, random forest, respectively. The precision score was obtained 85.56%, 88.54%, 87.59%, 79.14% for naïve Bayes, logistic regression, SVM, random forest, respectively. The recall score was obtained 84.95%, 88.72%, 88.80%, 85.30% for naïve Bayes, logistic regression, SVM, random forest, respectively.

Keywords Sentiment analysis · Machine learning approaches · Natural language processing · Bengali · Naïve Bayes · SVM · Logistic regression · Random forest

1 Introduction

Nowadays, people are using online platform more and more. Buying things from online shops is more easier, and peoples are doing that frequently. And they also making reviews their thoughts about those product they bought from those sites. Here, in subcontinent, peoples sometime write their opinions mostly with their mother tongue but use English alphabets. In Bangladesh and West Bengal, both parts use Bangla as their mother tongue. But, they also mostly reviewed the thing with English alphabets in Bengali language. So, We have goes for sentiment analysis of that thing.

Over and above, millions of reviews are generating every day in online market review section by random customers of consumed product. Presently, people love to order and buy from online market. Hence, they have various experiences about online market product. In the context of online marketing, buyers are gradually revealing their wonderful type of sentiment entire time of in assorted perspective [1].

So, this segment turns out a fascinating field in the sake of us to explore with others' emotions for online product. Modern civilization depends upon superfast Web throughout the world. Nowadays, folks are able to contribute in any initiative at least with their thought [2]. In this circumstance, individual opinion on fact plays a vital role in modern business strategies and pursuits recommendation already started to rely on sentiment. Sentiment puts a key impact what maximum people or thing thought about a specific brand, product, service or other utility thought about a specific brand, product, service or other utility. Connecting online shops with international and popular accessories in online, the importance of predicting about future market situation is a highly demanding topic based on the present customer reviews. Moreover, reviews help online market owner which product they need to stock more and which stock less as well. In addition, it also provides a valuable solution to new customers which product they will choose to consume or deny [3].

In this state, purchaser from all over the country started to privilege themselves with online market inside and outside of Bangladesh. Most of them willingly deliver their judgment on specific acquired product [4].

In our work, the proposed sentiment analysis system applies for opinions written in Bengali written in English words opinionated texts. We collected possible Bengali opinionated texts of Bengali reviews from the most popular e-commerce sites such as

Daraz.com.bd, Chaldal.com, Rokomari.com, Pickaboo.com and so on. These Bengali texts are preprocessed and substituted by their equivalent English words. The summarized review of the product is then calculated.

Moreover, reviews help online market owner which product they need to stock more and which stock less as well. In addition, it also provides a valuable solution to new customers which product they will choose to consume or deny [3].

In this state, purchaser from all over the country started to privilege themselves with online market inside and outside of Bangladesh. Most of them willingly deliver their judgment on specific acquired product [4].

In our work, the proposed sentiment analysis system applies for opinions written in Bengali written in English words opinionated texts. We collected possible Bengali opinionated texts of Bengali reviews from the most popular e-commerce sites such as Daraz.com.bd, Chaldal.com, rokomari.com, Pickaboo.com and so on. These Bengali texts are preprocessed and substituted by their equivalent English words. The summarized review of the product is then calculated.

2 Methodology

Opinion mining is a system of classifying text relying on orientation of review that sentiment it's carrying [5]. It is very popular nowadays in natural language processing platform.

1. First, it is might to find out term that express sentiment should be extracted. Here, we used ParseHub to extract the data set for this purpose.
2. After that, the orientation of reviews should be determined.
3. Then, we have to find out how much polarity it has. It means finding the strength and intensity.
4. After finding the strength, we have to classify them in positive and negative review based on polarity.

Here, **Maximum-Entropy parts of Speech** were used by us to categorize the reviews with a python coding to give it speeding the system. We have selected four classifiers for our system

1. Naïve Bayes
2. Logistic regression
3. Random forest
4. Support vector machine (SVM).

And for selecting feature, Pang et al. had suggested not to add the reviews which have been objective rather to take subjective. They used Total sent ind which is calculated as

$$\mathrm{TSI} = \frac{p - \frac{t_p}{t_n} \times n}{p + \frac{t_p}{t_n} \times n}$$

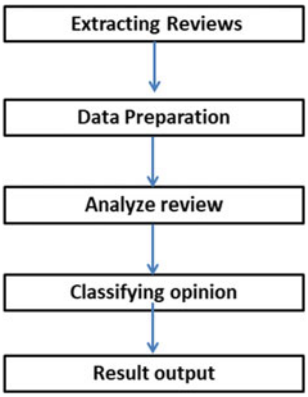


Fig. 1 Flowchart of algorithm

Here, in this formula, p is the count of positive reviews which appear, and negative review count was taken by n. $\frac{t_p}{t_n}$ is the ratio of total positive and negative.

2.1 Objective of the Research

1. Extracting reviews from Bangladeshi and Indian Web sites that contain Banglish review.
2. Analysis of them
3. Categorization of them
4. Finding sentiment
5. Labeling them as positive and negative (Fig. 1).

2.2 Designing System

Used Hardware

1. Core i7 8th generation
2. 4 Giga Ram.

Used Platform

1. Python (Version-3)
2. Anaconda Distribution

3. Natural Language Toolkit.

Data set information
We have collected the data set from Daraz.com.bd, Pickaboo.com, Chaldal.com, etc. Bangladeshi online Web sites as they scraped by pursehub tool. It was collected from January 2015 to January 2019. Almost five years of data has been collected. The data has containing the information like this

a. Reviewer account id
b. Given star rating(generally, there is five star rating system)
c. The actual time of making opinion
d. Reviewed opinion.

2.3 Data Procession, Parts of Speech Tagging and Identify Negative Words (Phrase)

First, the stop word was removed.

Then, the sentences were formed as token. This is called tokenization.

The tokened part was used for parts of speech tagging. It is very effective as we know noun and pronoun do not have any opinion; they were just written to complete sentence.

And, we know verb and adjective also carry different meanings and opinions with the aid of negative prefix. AS an example a sentence can be said "the mobile has nothing great" here, great is a positive word but together ''Nothing great makes it negative feedback.

So, it is very much important to identify such phrases. So, we went for tracking two types of phrases—negation of adjective and verb differently.

2.4 Classifier Description

Naïve Bayes
This classifier uses the basic law of Bayesian theorem where there is a train set D that can be shown by numbering dimensional (n) vector $x_1, x_2, x_3, \ldots, x_n$. Here, N is the number of attributes; if there is M number of classes, then it will e represented as $c_1, c_2, c_3, c_4, \ldots, c_m$. By having the value of X-vector value, the system will make prediction.

$$P(C_i|X) = \prod_{k=1}^{n} P(x_k|C_i)$$

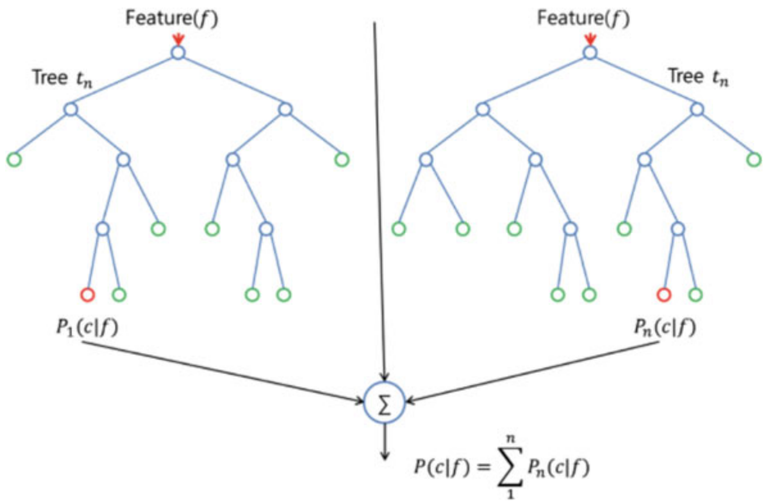


Fig. 2 Random forest mechanism

Random Forest

Random tree was chosen because of its superiority over decision tree. It is a process of system that uses bag of model. When the sample is replaced, it means some tuple from D can be got into D_i, and others can be got several times. Then, the random tree forest has been settled based on D_i. The basic rule is (Fig. 2).

$$\text{Gini} = 1 - \sum_{i=1}^{m} p_i^2$$

Support Vector Machine (SVM)

This method is been for linear and nonlinear data set. When the data is linearly splittable, then support vector machine went for linear hyperlinks. Using mathematical this thing can be written $P.Y + a$=0, where P is the vector weight and $Y = y_1, y_2, y_3, \ldots, y_n$. A is the scalar. This is computable as

$$\sum_{i=1}^{n} \propto_i \ y_i x_{i+1}$$

Logistic Regression

This system can determine prediction of a data set where there are only two values. This predictor system works on the uses of 1 or multiple predictor. It is much better than linear regression system. Cause, this system outputs a curve that is logistic and only in binary system whether it will be zero or one. It work as

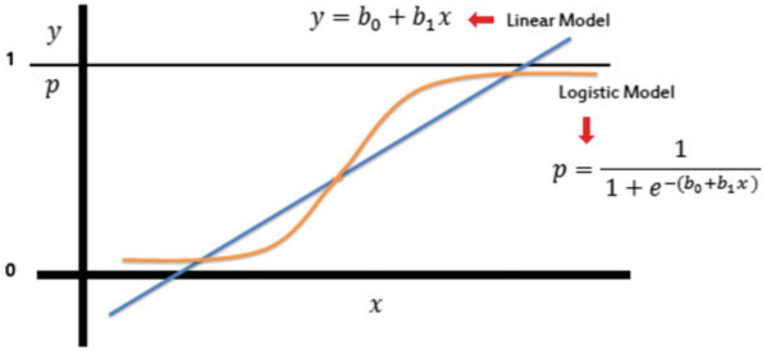


Fig. 3 Logistic regression mechanism

```
X = reviews_df_preprocessed.iloc[:, -1].values
y = reviews_df_preprocessed.iloc[:, -2].values

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 42)
```

Fig. 4 Data preprocessing

$$\beta^1 = \beta^0 \left[X^T W X\right]^{-1} . X^T (y - \mu)$$

β is a vector of the logistic regression coefficients
W is a square matrix of order N with elements $\pi_i n_i (1 - \pi_i)$ on the diagonal and zeros everywhere else
μ is a vector of length N with elements $\mu_i = n_i \pi_i$ (Fig. 3).

2.5 Implementation of the Work

Data preprocessing part can be seen from Fig. 4.

The using of natural language processing toolkit processor can be seen from Fig. 5.

3 Result and Discussions

The final output of this work is to make the machine enable of finding whether the review is positive or negative.

Precision is the term that is use to positive predictive value and Recall is the term that is used find sensitivity. Therefore, both of them are used to find and evaluate the relevance.

```python

class NltkPreprocessor:

    def __init__(self, stopwords = None, punct = None, lower = True, strip = True):
        self.lower = lower
        self.strip = strip
        self.stopwords = stopwords or set(sw.words('english'))
        self.punct = punct or set(string.punctuation)
        self.lemmatizer = WordNetLemmatizer()

    def tokenize(self, document):
        tokenized_doc = []

        for sent in sent_tokenize(document):
            for token, tag in pos_tag(wordpunct_tokenize(sent)):
                token = token.lower() if self.lower else token
                token = token.strip() if self.strip else token
                token = token.strip('_0123456789') if self.strip else token
                # token = re.sub(r'\d+', '', token)

                if token in self.stopwords:
                    continue

                if all(char in self.punct for char in token):
                    continue

                lemma = self.lemmatize(token, tag)
                tokenized_doc.append(lemma)

        return tokenized_doc

    def lemmatize(self, token, tag):
        tag = {
            'N': wn.NOUN,
            'V': wn.VERB,
            'R': wn.ADV,
            'J': wn.ADJ
        }.get(tag[0], wn.NOUN)

        return self.lemmatizer.lemmatize(token, tag)

```

Fig. 5 Python coding using NLTK

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

F_1 score is the calculation of accuracy of tests. It is also called as F-measure.

$$F_1 = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

During the use of units that normalized, the space remains under the graphed curve (area under the curve or AUC) which is similar to the probability. As we can say, the classifier makes a rank which is taken by random choice at instance that is positive and not going to take the negative one. As it will rank the positive rank greater than the negative. This can be measured

Table 1 Output of the different classifiers

Classifier	F_1 score	Accuracy (%)	Precision (%)	Recall (%)	ROC
Naïve Bayes	85.25	85.31	85.56	84.95	85.31
Logistics regression	88.12	88.05	87.54	88.72	88.05
Support vector machine	88.12	88.11	87.59	88.80	88.11
Random forest	82.43	81.82	79.74	85.30	81.83

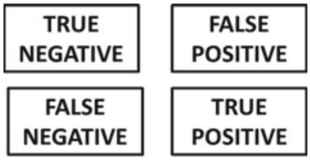


Fig. 6 CM organization

$$
\begin{aligned}
A &= \int_{\infty}^{-\infty} \mathrm{TPR}(T)\mathrm{FPR}'(T)\mathrm{d}T \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I\left(T' > T\right) f_1\left(T'\right) f_0(T)\mathrm{d}T'\mathrm{d}T = P(X_1 > X_0)
\end{aligned}
$$

The outcomes that are found by hold-out strategy is given in Table 1.

For finding better result, we can use confusion matrix (CM). The formation is given in Fig. 6.

The output result is given for all the classifier one by one as the format of Fig. 5 is given in Figs. 7, 8, 9 and 10.

The sample output result can be seen from Fig. 11.

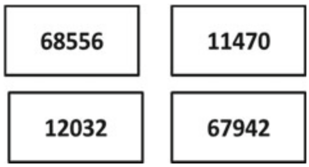


Fig. 7 Naïve Bayes

Fig. 8 Logistic regression

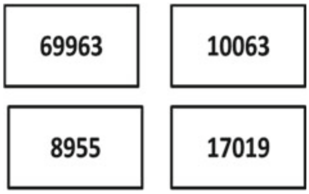


Fig. 9 Support vector machine

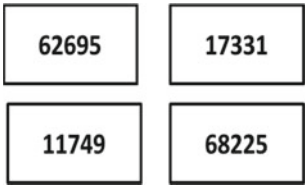


Fig. 10 Random forest

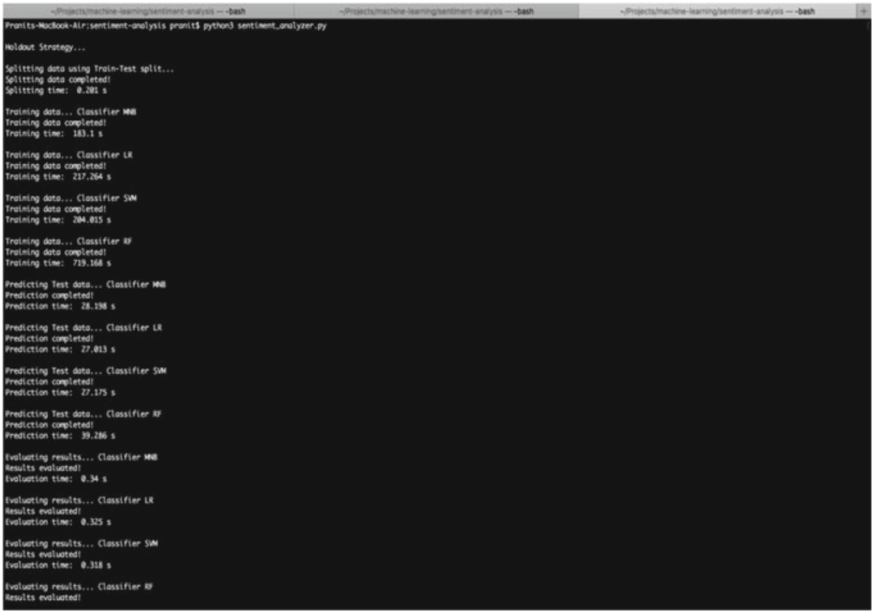


Fig. 11 Sample output after completion

4 Conclusion

Opinion mining is a very much attractive field in machine learning system as it is very much needed for natural language processing. The opinion mining of Bengali written English word has been done successfully using four different classifiers—support vector machine, naive Bayes, logistic regression and random forest. For the work, data set was extracted from local online shops using PurseHub. The work was done with vital steps—data preparation, classifying reviews according to sentiment score and evaluate the system in all steps. The F_1 score was obtained 85.25%, 88.12%, 88.12%, 82.43% for naïve Bayes, logistic regression, SVM, random forest, respectively. The accuracy score was obtained 85.31%, 88.05%, 88.11%, 81.82% for naïve Bayes, logistic regression, SVM, random forest, respectively. The precision score was obtained 85.56%, 88.54%, 87.59%, 79.14% for naïve Bayes, logistic regression, SVM, random forest, respectively. The recall score was obtained 84.95%, 88.72%, 88.80%, 85.30% for naïve Bayes, logistic regression, SVM, random forest, respectively.

References

1. Mohammad, S.M.: Detecting valence, emotions, and other affectual states from text. In: Emotion Measurement, pp. 201–237. Elsevier (2016)
2. Silva, J.J.D., Haddela, P.S.: A term weighting method for identifying emotions from text content. In: ICIIS, pp. 381–386. IEEE xplore (2013)
3. Mehra, R., Bedi, M.K., Singh, G., Arora, R.: Sentimental analysis using fuzzy and naive bayes. In: ICCMC, pp. 945–950. IEEE xplore (2017)
4. Pang, K.B., Lee, L.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the ACL- 02, vol. 10, pp. 79–86 (2002) (unpublished)
5. Chowdhury, R.S., Chowdhury, W.: Performing sentiment analysis inbangla microblog posts. In: ICIEV, pp. 1–6. IEEE xplore (2014)

Imbalanced Dataset Analysis with Neural Network Model

M. C. Babu and S. Pushpa

Abstract The comprehensive real-time dataset is skewed. The dataset is often imbalanced and hard to classify with the existing balanced dataset. The dataset may be skewed in the range of 10:1 ratio. Due to the data imbalance and errors, the data classification accuracy rate can be reached only to 90%. The classification accuracy rate can be increased if the number of errors in dataset is minimized. Hence, in this paper, we propose to identify error in SCRUM dataset by linear neural network model. For analysis, two datasets with errors and without errors are taken and analyzed with neural network model. The model train to determine error in dataset. With the proposed method, the model determines error in dataset with 98% accuracy.

1 Introduction

In machine learning and datasets, classification plays a major role. The classification is used to learn about new data and analyze patterns in data. The datasets are primarily categorized into balanced and imbalanced dataset. The real-world dataset is mostly imbalanced due to the nature of data availability. Hence, to classify imbalanced data and learn about errors in datasets, different algorithms are employed. The conventional algorithms at times consume more power and are complex due to size and nature of datasets.

The software defect prediction model creates imbalanced dataset. The dataset becomes more imbalanced when the number of software employed on cross platform increases in numbers. The imbalanced source leads to classify normal software as defective instance. The cross platform defective instance solves with subclass discriminant analysis (SDA). The SDA learns defective features with increased classification accuracy. The SDA improves balanced subclass with semi-supervised transfer component analysis. The method makes source and target data consistent and increases prediction [1].

M. C. Babu (✉) · S. Pushpa
Department of Computer Science and Engineering, St. Peter's Institute of Higher Education and Research, Chennai, Tamil Nadu, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_9

The data stream with multiple classifiers classifies the single and multiple class instances. However, when new data is applied, the classifier fails to identify them. The new instance classifies with k classifiers. The one versus all (OVA) is a classifier which differentiates from other classes. The OVA performs better by minimal error correlation and produces high OVA classification. The approach improves OVA accuracy. Furthermore, the OVA incorporates new class that is present in data streams. The OVA is an inefficient classifying model for component classification. The OVA does not perform effective classification for imbalanced dataset [2].

The imbalanced data classifies with Takagi–Sugeno–Kang (TSK) fuzzy classifier. The TSK improves the classifier accuracy for imbalanced dataset. The TSK works by incorporating Bayesian fuzzy clustering algorithm and imbalance learning algorithm. The data interpretability increases by BF3C which determines fuzzy rules with probabilistic model. In addition, the number of clusters can be estimated with Bayesian interface. The BF3C with repulsion among clusters occur due to antecedent parameter in imbalanced dataset. The classification performance increases with imbalanced dataset via fuzzy rules derived through misclassification error weights [3].

In imbalanced class dataset, the data is skewed toward one class compared to the other. The convectional learning class algorithm fails to determine numerous classes in dataset. The minority dataset has minimal classes. The misclassification rate minimize with in imbalanced dataset eliminate with statistical decision theory. In class with unknown distributions, the stochastic gradient algorithm applies to minimize misclassification error [4].

Gravitation is the fundamental technique applied to solve imbalanced data classification. The data gravitation classifies data by gravitational class comparison. The gravitation and weight computation for different classes are influenced due to irrelevant attributes and noisy classes. The gravitational classification improves with DGC+ which works based on weight. The weights provide information of attributes, which help predict data samples distances. The local and global data information improves classification accuracy and decision boundary [5].

The concept drift and class imbalance problem solve with class imbalance and online active learning paired drifting steams. The ensemble comprises dynamic and stable classifier which solves gradual and sudden concept drift. In addition, a hybrid labeling strategy with uncertainty criterion and dynamic adjustment threshold helps learn about minority category instances. The approach helps understand and captures data in decision boundary [6].

The imbalanced dataset classifies with support vector machine (SVM) algorithm. The SVM trains with imbalanced data to improve classification accuracy. However, the SVM classifies false data during training. The false occurrence eliminates by adjusting kernel matrix boundary for imbalanced data distribution [7].

2 Related Works

The class imbalance creates bias in data during training. The bias influences dataset classification by lowering sensitivity and minority class false positive. The multivariate data is analyzed with Mahalanobis–Taguchi system (MTS). The MTS system has built-in classifier which measures the scale of data. Hence, the MTS model does not depend on data distribution to solve class imbalance problem. The MTS classified imbalanced data with probabilistic thresholding [8].

In data mining and machine learning, classifying imbalanced dataset is hard. The imbalanced data was linearly classified with ROC and AUC curves. The method is not appropriate to eliminate heterogeneity and nonlinearity in data. Hence, kernalized online imbalanced learning (KOIL) algorithm applies to classify data by increasing AUC score and minimizes functional regularizer. The model performance improves by applying buffers to collect decision boundary information. In addition, the fluctuations caused due to data a smooth updating function apply to confine support vectors. The information loss eliminates by involving compensation scheme. Finally, the data similarity eliminates with kernel learning framework [9].

The biomedical data often suffers from imbalanced data problem and results in poor prediction and minimal performance. Since, the trained classifiers data are obtained from majoring class of data. Hence, an ensemble learning framework comprises selection algorithm, ensemble learning method and incremental learning to solve imbalanced dataset problem [10].

The imbalanced data classification solves data learning with imbalanced sample of data in class. Conventional classifiers do not perform well in classification since imbalanced data is not taken into account. The imbalanced data problem solves with modifications to learning algorithm, data distribution, data samples and ensemble methods. The imbalanced dataset classification improves by fuzzy rough set theory and weighted average aggregation [11].

The imbalanced dataset classification solves with decision tree ensemble and divergence properties. The splitting criteria with divergence scheme employ to solve c4.5 and CART criteria. The diversity among tree increases with AUROC values with respect to class priors. In addition, alpha trees apply with lift aware criteria to increase tree growth. The ensemble thus produces interpretable rules which increases lift values [12].

In visual imbalance task, machine learning applies for supervised learning. In imbalanced dataset, a particular dataset outnumbers the other dataset. Conventional machine learning dataset often biases toward majority class data resulting in poor accuracy [13].

The class imbalance problem exists in real-world dataset. The classifiers are biased toward majority class dataset due to classifiers trained with imbalanced data. Hence, cluster synthetic data generation model applies to solve imbalanced dataset classification. The cluster synthetic data model works based on oversampling region of data. The model further optimizes data generation and minimizes computational complexity [14].

The support vector machine (SVM) is a machine learning tool employed to learn balanced dataset. However, the SVM does not perform the same when dealing with imbalanced dataset. The SVM, furthermore, is sensitive to noise in datasets. The (class imbalance learning (CIL) makes SVM less sensitive to noise in dataset. Hence, fuzzy SVM applies to eliminate dataset noise and outliners. The FSVM assigns with different fuzzy function values, and the membership functions solve SVM sensitivity to noise. The FSVM performance for CIL improves to solve imbalance class problem due to noise and outliers [15].

3 Methodology

The imbalanced dataset with errors affects the performance of a system. The errors in network alter the system outcome in classification. Hence, the errors in dataset should be identified with mapping. However, identifying error cells in dataset are hard since the dataset is huge and requires more computation time. The computation time and complexity to identify errors in cell identify with neural network linear regression. The regression help identify whether the given dataset correlates with other datasets. The level of correlation and positive matching cells in network can be identified. A SCRUM database survey across different employees is acquired, and error in dataset is identified with the neural network model. The error in dataset neural network works by adjusting weights of cells in neural network model. The input data to the neurons is given by

$$y_k, \quad \text{where } k = 1, 2, \ldots k.$$

The input neuron values are multiplied with respect to cell weights w_{ki}. The weights are summarized to produce a constant bias term N. The N term activates the functions g in neural network model as shown in Fig. 1.

The activation g replaces with tangent function for simplicity as represented by

$$\text{Tanh}\,(x) = 1 - \text{e}^{-x}/1 + \text{e}^{-x}$$

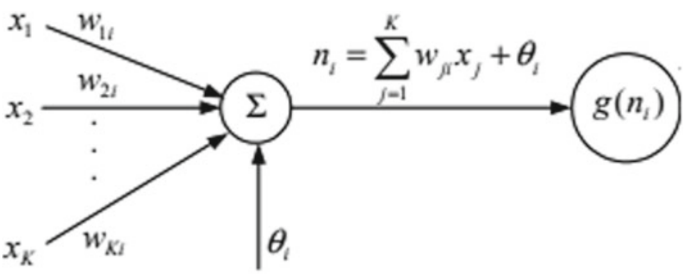


Fig. 1 Neural network cell bias model

The output node $g(n_i)$is given by

$$g_i = g\left(\sum_{j=1}^{K} w_{ji}x_j + \theta_i\right)$$

Multiple nodes are combined to form a series and parallel neural network. The output function of a multiple neural network model is given by

$$\begin{aligned} y_i &= g\left(\sum_{j=1}^{3} w_{ji}^2 g(n_j^1) + \theta_j^2\right) \\ &= g\left(\sum_{j=1}^{3} w_{ji}^2 g\left(\sum_{k=1}^{K} w_{kj}^1 x_k + \theta_j^1\right) + \theta_j^2\right) \end{aligned}$$

The neural network model was trained for different classes of imbalanced dataset. The neural network model was trained for validation and test phase for datasets. The training and testing phase was done to evaluate data variability and factors in neural network training. Furthermore, the imbalanced dataset created errors during training and loss functions. The error function was expressed with respect to loss in training function. The training performance of an optimal neural network was achieved by error minimization.

4 SCRUM Database

SCRUM is a model used to carry out agile methodology. The SCRUM framework helps carry out changes to software development cycle. The SCRUM framework is flexible, enables others to collaborate and provides continuous feedback to users regarding changes carried out during development cycle. The SCRUM framework is selected based on different project sizes and according to industry scale such as small industries, medium industries and large industries. However, the SCRUM users prefer certain SCRUM framework irrespective of project size and industry scale. The user feedback of different frameworks such as SAFe, DAD and LeSS is collected for SCRUM usability analysis based on certain prerequisites. The rating of SCRUM users differs for different people who meet the prerequisite criteria and people who do not meet the prerequisite criteria. The dataset are analyzed for errors created by user ratings of people who do not meet the prerequisite criteria to extract meaningful data. The error in the dataset is analyzed by comparing two datasets via neural network linear regression model.

5 Results and Discussion

Figure 2 shows balanced dataset analysis of user rating who meets the prerequisite criteria (dataset 1). Figure 2 shows the end of training when the error level has reached six iterations. Figure 2 shows the error comparison between train and test dataset. The mean squared error of the dataset is set at 10^1.

Figure 3 shows the combination of training data, and test data. From Fig. 3 is it evident that most of the error between the training and test data occur at −1.193. The outlier errors occur at −2.993 and 5.557 which shows that the dataset suffer minimal errors.

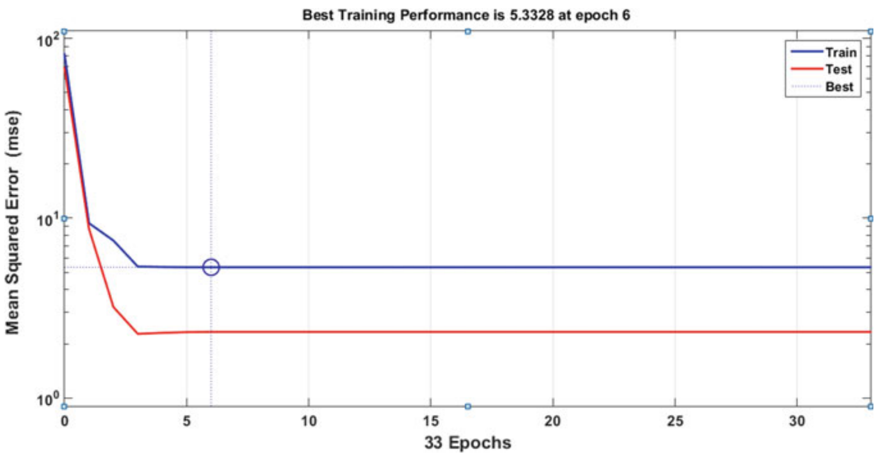


Fig. 2 Training performance of dataset 1

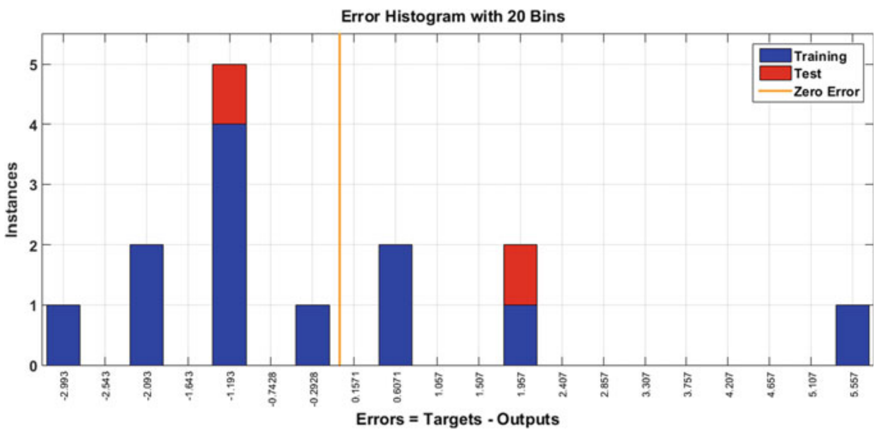


Fig. 3 Error histogram of dataset 1

Figure 4 shows regression analysis of data points and training data. The data point correlates with train data with 0.30247 factor. Similarly, the data points and train data compare to evaluate the relation among the dataset. The data fits with no errors; hence, the *R* value is set to 1 by network model as shown in Fig. 5. The test data and train data are combined to form all data to evaluate the relationship between data points. The data points correlate with 0.33899 factor as shown in Fig. 6. Figure 7 shows the best fit for data point with respect to model. The data points are in complete correlation with *R* values of 1.

Figure 8 shows the comparison of test and test imbalanced dataset 2 user rating who does not meet the prerequisite criteria with error. The training ends at epoch 5 for dataset 2 compared to dataset 1. The mean squared error reaches at 100. The train data achieves above the test data.

Figure 9 shows the error histogram comparison for dataset 2. The error level reaches a maximum extent of 6 compared to Fig. 3 which is only 5. The error occurs at outliner, namely −1.477 and 1.388.

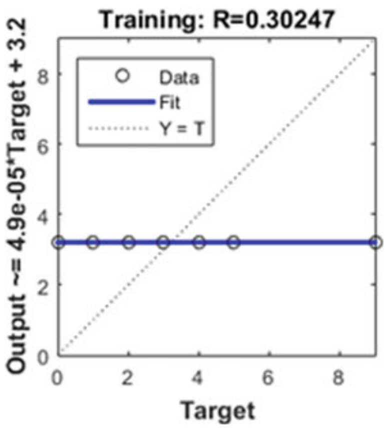


Fig. 4 Train data and data point correlation for dataset 1

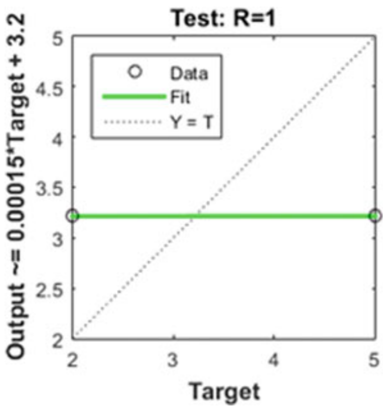


Fig. 5 Test data and data point correlation for dataset 1

Fig. 6 All data and data point correlation for dataset 1

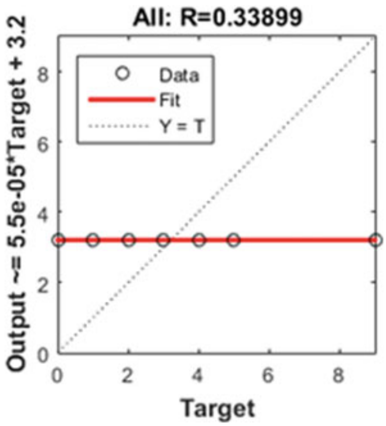


Fig. 7 Best fit for dataset 1

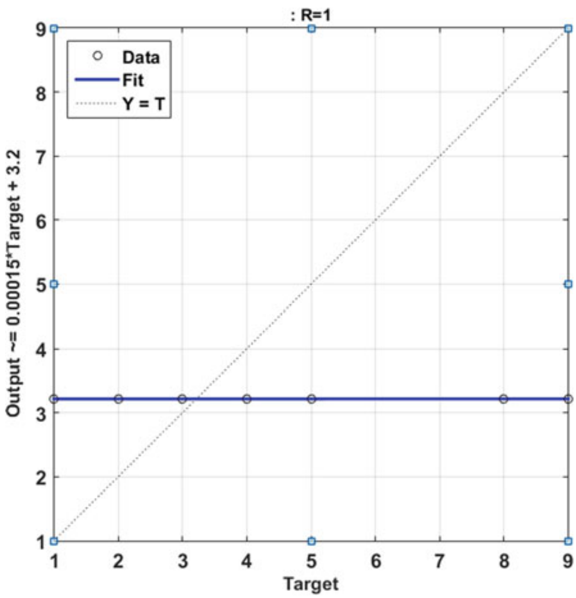


Figure 10 shows the correlation between data points and trained data points. The data points correlate with 0.15733 factor. Figure 11 shows data points comparison with test data. The data correlates with R value of 1. Figure 12 shows the all data point correlation with dataset. The data points and values from dataset show a correlation with r value of 0.25859. Compared to Fig. 6, the correlation value is low. Figure 13 shows the best fit comparison and fit value. The fit value correlates with r value of 0.99999.

The neural network model predicts and determines the error in dataset with respect to data points. For dataset 1, the error rate is minimum; hence, the R value is set to

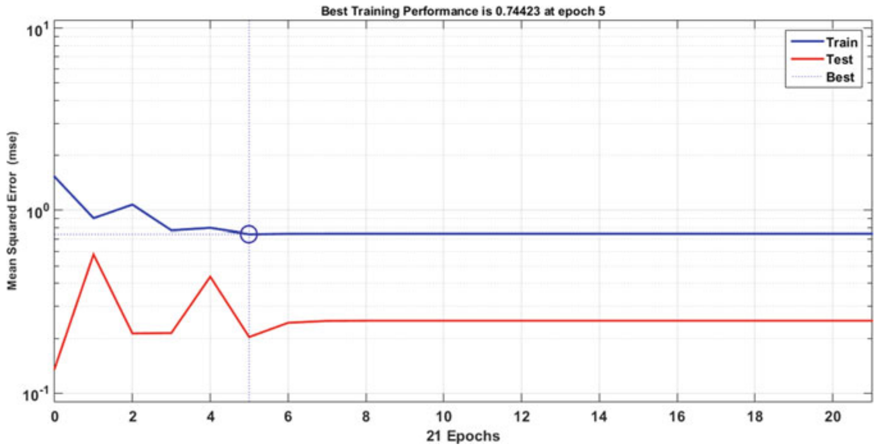


Fig. 8 Training performance of dataset 2

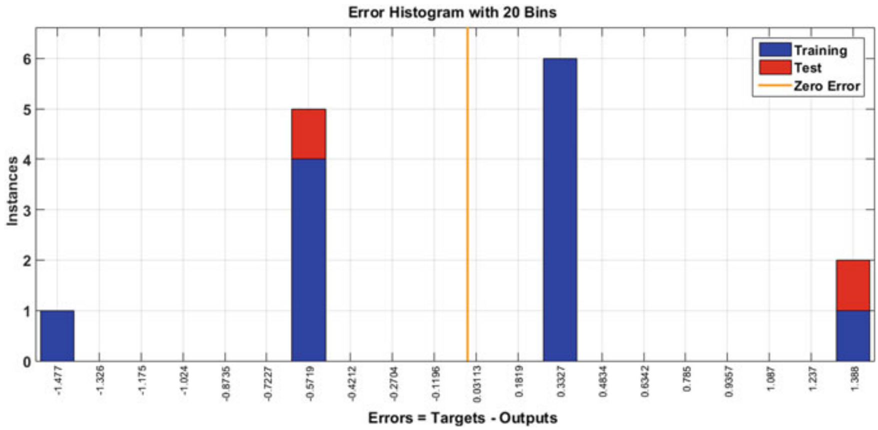


Fig. 9 Error histogram of dataset 2

1 by neural network model. However, for dataset with errors, the *R* value is set to 0.99999 by neural network model.

6 Conclusion

The neural network linear model applies to determine error in dataset. Initially, the neural network model trains with balanced dataset analysis of SCRUM user rating who meets the prerequisite criteria (dataset 1). The error histogram and linear model are derived for dataset 1. Similarly, the dataset 2 with user rating who does not

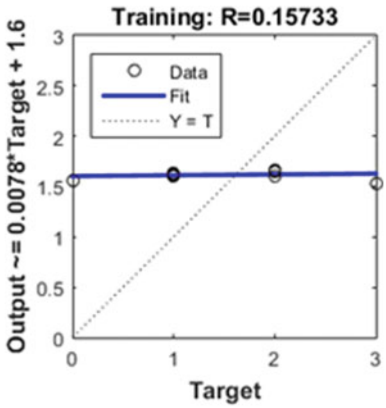


Fig. 10 Train data and data point correlation for dataset 2

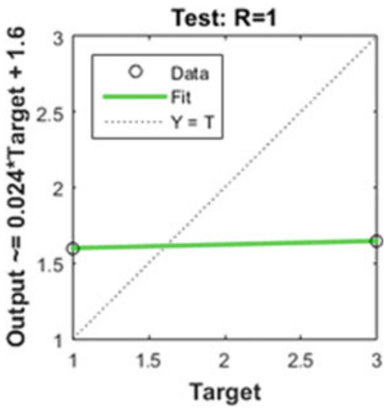


Fig. 11 Test data and data point correlation for dataset 2

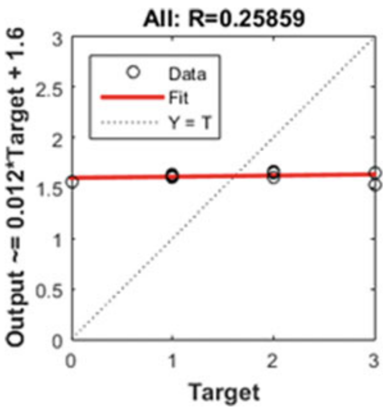


Fig. 12 All data and data point correlation for dataset 2

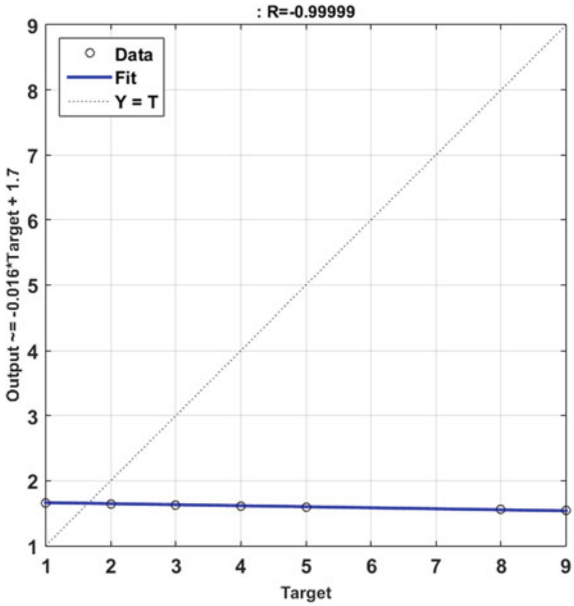


Fig. 13 Best fit for dataset 2

meet the prerequisite criteria for data is applied to neural network model to analyze error histogram. The dataset 2 and dataset 1 are applied to neural network model to determine the errors in dataset. The neural network model predicts error in dataset with 98% accuracy.

Compliance with Ethical Standards All author states that there is no conflict of interest. We used our own data. Humans and animals are not involved in this research work.

References

1. Jing, X., et al.: An improved SDA based defect prediction framework for both within-project and cross-project class-imbalance problems. IEEE Trans. Softw. Eng. **43**(4), 1–20 (2016)
2. Hashemi, S., Yang, Y., Mirzamomen, Z., Kangavari, M.: Adapted one-versus-all decision trees for data stream classification. IEEE Trans. Knowl. Data Eng. **21**(5), 624–637 (2009)
3. Gu, X., Chung, F.L., Ishibuchi, H., Wang, S.: Imbalanced TSK fuzzy classifier by cross-class Bayesian fuzzy clustering and imbalance learning. IEEE Trans. Syst. Man Cybern. Syst. **47**(8), 2005–2020 (2017)
4. Diamantini, C., Potena, D.: Bayes vector quantizer for class-imbalance problem. IEEE Trans. Knowl. Data Eng. **21**(5), 638–651 (2009)
5. Cano, A., Zafra, A., Ventura, S., Member, S.: Weighted data gravitation classification for standard and imbalanced data. IEEE Trans. Cybern. **43**(6), 1–16 (2012)
6. Zhang, H., Liu, W., Shan, J., Liu, Q.: Online active learning paired ensemble for concept drift and class imbalance. IEEE Access **3**(2), 92–104 (2018)
7. Wu, G., Chang, E.Y., Member, S.: KBA: Kernel boundary alignment considering imbalanced data distribution. IEEE Trans. Knowl. Data Eng. **17**(6), 786–795 (2005)

8. Su, C., Hsiao, Y.: An evaluation of the robustness of MTS for imbalanced data. IEEE Trans. Knowl. Data Eng. **19**(10), 1321–1332 (2007)
9. Hu, J., Yang, H., Lyu, M.R., King, I., Member, S., So, A.M.: Online nonlinear AUC maximization for imbalanced data sets. IEEE Trans. Neural Netw. Learn. Syst. **3**(2), 1–14 (2016)
10. Oh, S., Lee, M.S., Zhang, B.: Ensemble learning with active example selection for imbalanced biomedical data classification. (CISTI)EEE/ACM Trans. Comput. Biol. Bioinform. **8**(2), 316–325 (2011) (Iber. Conference Information System Technologies)
11. Ramentol, E., Vluymans, S., Verbiest, N., Bello, R., Cornelis, C.: IFROWANN: imbalanced fuzzy-rough ordered weighted average nearest neighbor classification. IEEE Trans. Fuzzy Syst. **5**(1), 1–15 (2014)
12. Park, Y., Member, S., Ghosh, J.: Ensembles of trees for imbalanced classification problems. IEEE Trans. Knowl. Data Eng. **26**(1), 131–143 (2014)
13. Mera, C., Branch, J.W.: A survey on class imbalance learning on automatic visual inspection. IEEE Lat. Am. Trans. **12**(4), 657–667 (2014)
14. Lim, P., Member, S., Goh, C.K., Member, S., Tan, K.C.: Evolutionary cluster-based synthetic oversampling ensemble (ECO-ensemble) for imbalance learning. IEEE Trans. Cybern. **5**(3), 1–12 (2016)
15. Batuwita, R., Palade, V.: FSVM-CIL: fuzzy support vector machines for class imbalance learning. IEEE Trans. Fuzzy Syst. **18**(3), 558–571 (2010)

Review of Parallel Processing Methods for Big Image Data Applications

K. Vigneshwari and K. Kalaiselvi

Abstract The coexistence of technologies, like big data application, cloud computing, and the numerous images in the Web has paved the need for new image processing algorithms that exploit the processed image for diverse applications. There arises a need for new image processing algorithms to utilize the processed image for diverse applications though many techniques with variations exist. Ultimately, the enduring issue is to enhance the effectiveness of huge image processing and to maintain the combination of the same with recent works. This paper presents a review of the newest progress in researches on parallel processing methods for the processing of big data. Initially, the reviews about the parallel processing methods were carried out by highlighting some promising parallel processing methods in recent studies, such as the representation of MapReduce (MR) framework, distributed, parallel methods, and Hadoop framework. Subsequently, focus on analysis and deliberations about the challenges and promising solutions of parallel computing methods on big data in various applications and on image processing applications were made and concluded with a summary of number of open problems and research areas.

Keywords Big data · Image processing · MapReduce (MR) framework · Hadoop · Parallel processing methods · Distributed system and applications

1 Introduction

Due to the frequent use of image in web pages, social sites, goods in shopping Web sites, etc., there arises a rapid need for processing large files in big data analysis. These types of pictures should be generally used for various categories of applications such as content-based image retrieval (CBIR), image explanation and labeling,

K. Vigneshwari (✉)
Department of Computer Science, VELS Institute of Science, Technology and Advanced Studies, Chennai 600117, India

K. Kalaiselvi
Department of Computer Science, School of Computing Sciences, Vels Institute of Science Technology and Advanced Studies (VISTAS), Formerly Vels University, Chennai 600117, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_10

and image comfortable identification. Fitting to the huge number of images and maintaining the computation complexity of images processing algorithms is very difficult, therefore it is essential to make use of distributed environments through accelerators to development of these huge pictures. From the work [1], they are four advantages of distributed systems when compared to remote system are (i) data sharing, that permits multiple users to access a general database; (ii) machine sharing, that permits multiple users to allocate their devices; (iii) communications, with the purpose of facilitates communication of machine with each other more easily; (iv) flexibility, that enables a distributed computer to deploy the workload over the connected machines in an successful manner. Better flexibility, reliability, and efficiency are found in distributed environments when compared with single [2].

For effective data processing in distributed systems, a number of existing environments like MapReduce (MR) [3], Spark [4], Storm [5], and Hadoop [6] are used. These models are based on open-source environment and are appropriate for diverse areas. MR [3] exhibits effectiveness in processing and creation of large datasets. A map function which is user-defined, and it is used to process a value pair which generates a set of intermediary value pairs, and a reduce function that combines each and every one intermediary value related with the same intermediary key. As illustrated in this model, it is expressible for several real-world tasks. This function is used to parallel and distributed environment to effectively make use of the resources of a huge distributed structure.

Spark is a computer programming language depending on the Ada programming language and designed for the improvement of high reliability software. It is used in systems where predictable and highly consistent operation is vital and also assists demand safety, security applications. It constantly development unbounded streams of information designed for real-time processing, as Hadoop performs designed for batch processing and is suitable for real-time analytics, online machine knowledge, continuous separation, and more [5, 6].

In this review work, the facts of a distributed image processing system which is developed depending on Hadoop [5] derived from MR [3] are studied. In comparison with the traditional image processing methods, parallel processing is able to unquestionably attain high-tech improvements. Given the distributed resources restricted to a single machine, until now in the meantime, this effort is a challenging. In recent years, researchers have put effort to propose image processing algorithms, considering the high competence with the purpose of parallel processing convey, with the purpose of can be experimented in parallel, between image classification, extraction of features, and matching know how to serve as representative instances. As the algorithms are capable of running parallel on multiple nodes and hence considerably improves the time effectiveness. Currently, some common existing frameworks for image processing algorithms should have attained improved performance for their parallelism. In spite the fact, the purpose of these frameworks have effectively experimented the capable processing of text data and stream data and they have done small part to image processing applications.

2 Literature Review

In the recent years, the big data has successfully developed into an essential topic that is purposeful to large-scale computational problems undoubtedly. The details of existing parallel processing techniques under general applications, medical applications, and image processing applications are reviewed in this section.

2.1 Review of Parallel Processing Methods—Many Applications

Pavlo et al. [7] performed a comparison that shows Hadoop is 2.50 times slower when compared to comparable database management system (DBMS) except in the case of data loading [8]. A benchmark including a group of tasks is defining with the purpose of run on an open source of MR and also on two parallel DBMSs. The efficiency of the system is measured for diverse degrees of parallelism on a group of 100 nodes for every job. The attractive trade-offs are revealed from the results. The observation on performance analysis of these DBMSs was found to be was remarkably enhanced results for higher load data and better execution time when compared to MR system.

Anderson and Tucek [9] disparaged with the purpose of the Hadoop system is scalable; however, it provided lower effectiveness per node, less than 5 MB/s processing rates with the purpose of recent studies. Hadoop solves scalability problem in distributed environment excluding missing efficiency [9]. The proposed data-intensive scalable computing (DISC) systems contain not overcomes an inaccuracy of previously existing better results by focusing on scalability without consideration of effectiveness. So there is an urgent necessitates altering the plan of potential information-intensive computing and cautiously deeming the direction of upcoming study.

Li et al. [10] and Jiang et al. [11] designed a distributed system based on the MR. It uses hash tables intended for enhanced effectiveness and moreover incremental processing [10]. Li et al. [10] introduced a new enhanced MR framework with the purpose of making use of hash techniques to allowing quick in-memory processing. Experimentation of Hadoop-based prototype with real-world workloads demonstrates with the purpose of the proposed enhanced MR framework that considerably increases the improvement of map tasks, permits the diminishing improvement to keep up with the map improvement, with up to three orders of size reduction of internal information spills, and better results in the direction of be returned constantly throughout the task.

Jiang et al. [11] conducted an evaluation study of MR on a 100-node cluster of Amazon Elastic Compute Cloud (EC2) among diverse stages of parallelism. Five design factors are identified, and results reveal with the purpose by cautiously fine-tuning these factors, the higher performance of Hadoop be able to be enhanced by a factor of 2.5–3.5. It is similar to parallel database systems. The results show that it is therefore potential to build a cloud data processing system that is both elastically scalable and efficient.

MR implements a processing and generates huge datasets to a many of real datasets [3]. Mohammed et al. [12] studied and reviewed the MR programming framework, which is experimented on clinical big data and medical healthcare data. The usage of MR and Hadoop on a distributed system provides a new opportunity in the rising period of big data analytics.

Wang et al. [13] reported on a technique with the purpose of automatically finds whether a particular adverse event (AE) is basis on an accurate drug depending on the essence of PubMed citations. An adverse drug event (ADE) taxonomy is introduced to identify neutropenia depending on a preselected set of drugs, which was then experimented to various sets of 76 drugs in the direction of confirm whether they caused neutropenia. The verification of results for the purpose of AUROC was 0.93 and 0.86 correspondingly.

Nguyen et al. [14] proposed new approach to processing and storing clinical signal data depending on the Apache HBase system and the MR programming model by means of an incorporated Web-based information visualization layer. The results are evaluated to estimate upwards of 50 TB of clinical signal data designed for a 200-bed medical center within the next 5 years.

Sweeney et al. [8] developed a new Hadoop Image Processing Interface (HIPI) with the purpose of aspires in the direction of constructing an interface designed for computer vision between MR model. The major aim is constructing a new tool with the purpose of creating an improvement in huge scale image processing and visualization projects mainly to get in hopes with the purpose to authorize researchers to easily generate the applications.

A new high-performance computing (HPC) platform is proposed designed for the quick investigation of tissue microarrays' (TMAs) virtual slides [15]. Upon real-time examination of over 90 TMAs, multiplex biomarker experiments were speeded up extremely.

Wei et al. [16] presented an OpenCluster distributed computing framework in the direction of support hastily introducing better image processing pipelines of astronomical big information. It is used to handle complex information processing problems which were demonstrated for the developing a pipeline for the Mingantu Ultrawide Spectral Radio heliograph. It provides a higher fault tolerance results and flexible manner for considerably reducing development cost.

Wiley et al. [17] focused on Hadoop, a popular open-source framework, in which the information is segmented among storage attached directly in the direction of hand nodes, and the parallel scheduling of workload is on the nodes with the purpose of hold the necessary input data. Developing a scalable image processing pipeline for the Sloan Digital Sky Survey (SDSS) imaging database using Hadoop was accounted. From the results, it concludes that the Hadoop becomes a good test bed for multi-terabyte imaging dataset.

2.2 Review of Parallel Processing Methods—Image Processing Applications

Hadoop MR, MR, Spark platform of processing frameworks are occurring such as addressing the difficulty of rendering a scheme designed for computationally intensive information processing and distributed storage space. Some of the works are discussed as follows.

Kohlwey et al. [18] presented a new system designed for an extensive search of cloud-scale biometric images. Here, the implementation details of human iris matching inside this prototype framework are provided. Finally, opportunities for future research are discussed. Vemula and Crick [19] presented a Hadoop-based library in the direction of support large-scale image processing by developing the Hadoop Image Processing Framework. It focuses permitting the developers using image processing applications to control the Hadoop MR framework and not including the master nodes and develops an added source of difficulty and error in their MR.

Hadoop Distributed File System (HDFS) framework has been also developed to handle and development the big remote-sensing data applications [20]. These algorithms are implemented in remote-sensing images which are able to be straightforwardly processed in distributed environment. The results demonstrate that these frameworks are efficiently managing and processing of big remote-sensing data.

Almeer [21] used a Hadoop platform for huge remote image applications. This platform is applied to Tag Image File Format (TIFF), Joint Photographic Experts Group (JPEG), Bitmap (BMP), and Graphics Interchange Format (GIF) formats. In addition, it also clearly measures the difference among the single PC runtime and the Hadoop runtime.

Kune et al. [22] identified the requirements of the overlapped information organization and proposed a two-phase expansion to Hadoop Distributed File System (HDFS) and MR framework, called XHAMI, to address them. It is applied to image processing area. From the results, it concludes that the proposed XHAMI has lesser storage overhead and improves the system performance especially for overlapped data.

Ryu et al. [23] presented a new video processing framework in Apache Hadoop in the direction of parallelize video processing tasks in a distributed environment. In an 8-core environment by means of two quad-core systems, the video processing framework provides 75% of scalability. Kim et al. [24] developed a new MR framework with the purpose of runs on a Hadoop platform, and the media processing library Xuggler. Consequently, the encoding time in the direction of transcode huge amounts of video substance is exponentially decreased, using a transcoding function. The experimental results were analyzed, for providing optimal framework for video transcoding. It concludes that proposed distributed video transcoding framework gives a better performance in terms of rapidity and quality.

Moise et al. [25] focused on Hadoop-based application as case study; the image similarity search obtains an experience with the parallel processing framework for processing terabytes of data. The findings are common as the greatest practices and suggestion in the direction of the big information researchers and practitioners.

Sozykin and Epanchintsev [26] presented a MapReduce Image Processing framework (MIPr), which presented the capability in the direction of make use of distributed computing for image processing. Moreover, the MIPr comprises of the high-level image processing API designed for developers. This API permits the creation of sequential functions with the purpose of process of one image or a group of associated images. Additionally, MIPr comprises of MR achievement of well-liked image processing methods, which are able to be used for distributed image processing exclusive of some software improvement. This framework significantly performs better image processing in Hadoop distributed environment.

Yamamoto and Kaneko [27] discussed on how to introduce a parallel distributed environment of a video database by means of using the computational store in a cloud environment. But still the performance for video processing applications remains an open issue. This can be solved by introducing a parallel implement using MR on Hadoop platform.

Liu et al. [28, 29] developed a System for Efficient Image Processing (SEIP), which is constructing on Hadoop, and make use of extensible in node structural design in the direction of maintaining different types of image processing methods on distributed platforms by means of Graphics Processing Unit (GPU) accelerators. This pipeline-based framework is used in the direction of speed up of huge image file processing (Table 1).

3 Inference from the Review

Because of huge collection sizes and higher computational time of current image processing methods, current image collections should not be processed capably on one computer. So, the image processing applications require a distributed computing environment. The distributed computing seems to be a difficult task in deep technical information and often not used by researches that's when the purpose of introducing the image processing algorithms comes in picture. The MR, Hadoop framework is required and with the purpose of allowed researches in the direction of focus on image processing methods and hides from the difficult details of distributed computing.

4 Solutions

The usage of the computational resource in a CC environment for processing image database, in parallel, is an emerging research topic. Currently, the open-source cloud computing for Apache Hadoop distribution is available from MR which offers a common framework designed for processing images that are capable of experimenting in parallel. Therefore, to attain a raise in time effectiveness by not including the

Table 1 Inferences of the review work

Author name and year of Publication	Proposed work	Advantages	Disadvantages	Tools	Applications
Asish Kumar Sen [3]	MR paradigm	Clusters efficiently in the case of hardware failures	It is not appropriate designed for online transactions	Experimented internally on Google over the past four years	Google applications
Mohammed et al. [12]	MR programming	Handle massive huge of medical information in medical health informatics associated fields	Nil	Java	Medical applications
Wang et al. [13]	MR paradigm	Provides better results	An important challenge in this is to build precise and computationally efficient classifier for medical applications	Ontology	Adverse drug events (ADEs)
Nguyen et al. [14]	Apache HBase distributed system and the MR programming paradigm	HBase in favor of data management and MR processing algorithm	Nil	Signal archiving and computation system	Store and processing clinical signals
Sweeney et al. [8]	Hadoop Image Processing Interface (HIPI)	It is flexible manner in computer vision literature	Nil	JAVA	Facebook
Wang et al. [15]	High-Performance Computing (HPC) platform	Provides high throughput analysis platform for TMA biomarker discovery and performs quicker	Nil	Aperio ScanScope CS scanner	TMAs

(continued)

Table 1 (continued)

Author name and year of Publication	Proposed work	Advantages	Disadvantages	Tools	Applications
Kohlwey et al. [18]	MR paradigm	Higher results for human iris matching	It is not suitable for other images	SDK	Face matching, iris recognition, and fingerprint recognition
Vemula and Crick [19]	Hadoop image processing framework	Lesser complexity and error	Does not support multimedia processing framework	Hadoop, Cloudera CDH5, and Java 1.7	Flickr search
Wang et al. [20]	HDFS based framework	Proficiently handle and process of big remote sensing data	Higher computation complexity and computational network load	Orfeo toolbox	Remote sensing images
Almeer [21]	Hadoop-based framework	It is scalable in processing multiple large images Runtime is clearly obvious	But the images have to be resized before processing to fit the Java heap size limitation	Hadoop	Remote sensing images
Kune et al. [22]	XHAMI	Lesser storage overhead, it greatly increases the system performance It easily applied to any application development process	Issues concerned in managing the series of executed map functions	virtualization setup-based Xen hypervisor	Large-scale canvas images
Ryu et al. [23]	MR paradigm	The system shows 75% of scalability	However, hyper threading doesn't seem to increase performance	C and C++ environment	Face tracking system

(continued)

Table 1 (continued)

Author name and year of Publication	Proposed work	Advantages	Disadvantages	Tools	Applications
Kim et al. [24]	Cloud transcoding system	Easily applied to video data	Cartoonizing engine, and is not intended for other video processing extensions	Java interfaces of FFmpeg library	Video data
Powell et al. [30]	SIPF	It is capability to execute a diversity of scalable image operations	Great multi touch display and CAVE augmented truth venues	Kakadu Java native interface	JPEG 2000 images

performance results. Dong et al. [29, 31] developed a successful processing structure selected Image Cloud Processing (ICP) toward effectively deal with the information explosion in image processing application. The results are measured in ImageNet dataset which are chosen to confirm the ability of ICP structure over the conventional methods in terms of time complexity and performance of results. The typical ICP framework includes two image processing mechanisms, i.e., Static Image Cloud Processing (SICP) and Dynamic Image Cloud Processing (DICP) are illustrated in Fig. 1; SICP is designed to process individual's significant image information with the purpose to store it in the distributed structure. DICP is introduced for the dynamic requests, where it begins from the clients and should be capable to return the results instantaneously.

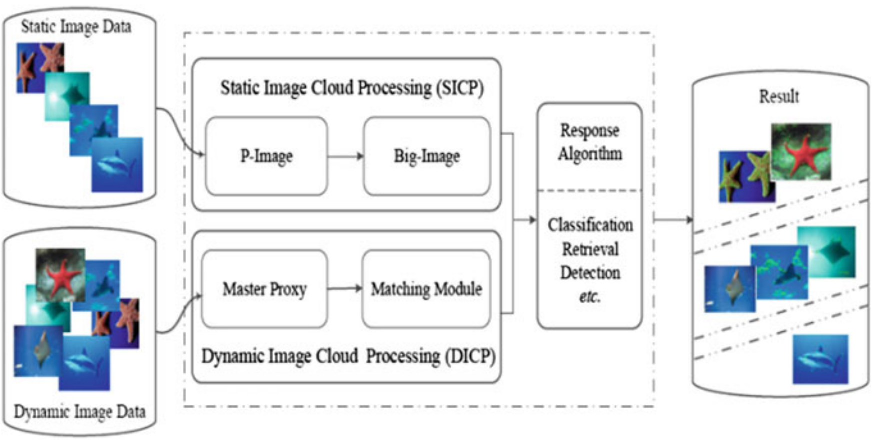


Fig. 1 ICP framework

5 Conclusion and Future Work

In the previous decade, Hadoop and MR frameworks have become a usual structure designed for big data in the business. Even though from the review work it is apparent that the Hadoop and MR frameworks at present are mainly functional to several applications, it is able to be able to practice binary images. Several numbers of frameworks have been developed in the direction of increasing efficiency of implementing Hadoop-related solutions. This review work shows how such a parallel processing framework is able to be used toward building a brief and capable solution toward a big data image processing issues of processing image applications and compares them to Hadoop application programming interface (API)-based implementation. While most recent research works are majorly focused on optimizing the image processing methods toward the increase in higher efficiency, these works will not be implemented in parallel environment so still it has scalability issues. So, in future work new algorithms are implemented, which could be experimented in parallel to increase higher time efficiency without reducing system results together with the rising image scale. The aspiration given by the existing work achievements and restrictions allows the future work to focus on the development of a successful parallel processing framework designed for big image data by the use of the CC capability. In the future work, it develops a successful processing structure selected Image Cloud Processing (ICP) toward effectively deal with the information explosion in image processing applications.

References

1. Tanenbaum, A.S., Van Steen, M.: Distributed Systems: Principles and Paradigms, pp. 7–8. Prentice Hall, Upper Saddle River, NJ (2007)
2. Fleischmann, A.: Distributed Systems: Software Design and Implementation, pp. 4–5. Springer, Berlin, Heidelberg (2012)
3. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. Commun. ACM **51**(1), 107–113 (2008)
4. Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., Stoica, I.: Spark: cluster computing with working sets. HotCloud **10**(10–10), 1–7 (2010)
5. White T.: Hadoop: The Definitive Guide, 1st edn. O'Reilly Media (2009)
6. Shvachko, K., Kuang, H., Radia, S., Chansler, R.: The hadoop distributed file system. In: IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), pp. 1–10 (2010)
7. Pavlo, A., Paulson, E., Rasin, A., Abadi, D.J., DeWitt, D.J., Madden, S., Stonebraker, M.: A comparison of approaches to large-scale data analysis. In: Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, pp. 165–178 (2009)
8. Sweeney, C., Liu, L., Arietta, S., Lawrence, J.: HIPI: a Hadoop Image Processing Interface for Image-Based Mapreduce Tasks. University of Virginia, Chris (2011)
9. Anderson, E., Tucek, J.: Efficiency matters! ACM SIGOPS Operating Syst. Rev. **44**(1), 40–45 (2010)
10. Li, B., Mazur, E., Diao, Y., McGregor, A., Shenoy, P.: A platform for scalable one-pass analytics using Map reduce. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of data, pp. 985–996 (2011)

11. Jiang, D., Ooi, B.C., Shi, L., Wu, S.: The performance of mapreduce: an in-depth study. Proc. VLDB Endowment **3**(1–2), 472–483 (2010)
12. Mohammed, E.A., Far, B.H., Naugler, C.: Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends. BioData Min. **7**(1), 1–23 (2014)
13. Wang, W., Haerian, K., Salmasian, H., Harpaz, R., Chase, H., Friedman, C.: A drug-adverse event extraction algorithm to support pharmacovigilance knowledge mining from PubMed citations. In: AMIA annual symposium proceedings: 2011. American Medical Informatics Association, Bethesda, Maryland, USA, pp. 1464–1471 (2011)
14. Nguyen, A.V., Wynden, R., Sun, Y.: HBase, MapReduce, and integrated data visualization for processing clinical signal data. In: AAAI Spring Symposium: Computational Physiology (2011)
15. Wang, Y., McCleary, D., Wang, C.-W., Kelly, P., James, J., Fennell, D., Hamilton, P.: Ultra-fast processing of gigapixel tissue microarray images using high performance computing. Cell. Oncol. **34**(5), 495–507 (2011)
16. Wei, S., Wang, F., Deng, H., Liu, C., Dai, W., Liang, B., Mei, Y., Shi, C., Liu, Y., Wu, J.: OpenCluster: a flexible distributed computing framework for astronomical data processing. Publ. Astron. Soc. Pac. **129**(972), 024001 (2016)
17. Wiley, K., Connolly, A., Gardner, J., Krughoff, S., Balazinska, M., Howe, B., Kwon, Y., Bu, Y.: Astronomy in the cloud: using mapreduce for image co-addition. Publ. Astron. Soc. Pac. **123**(901), 366–380 (2011)
18. Kohlwey, E., Sussman, A., Trost, J., Maurer, A.: Leveraging the cloud for big data biometrics: meeting the performance requirements of the next generation biometric systems. In: IEEE World Congress on Services (SERVICES), pp. 597–601 (2011)
19. Vemula, S., Crick, C.: Hadoop image processing framework. In: IEEE International Congress on Big Data (BigData Congress), pp. 506–513 (2015)
20. Wang, C., Hu, F., Hu, X., Zhao, S., Wen, W., Yang, C.: A Hadoop-based distributed framework for efficient managing and processing big remote sensing images. ISPRS Ann. Photogram. Remote Sens. Spat. Inf. Sci. **2**(4), 63–67 (2015)
21. Almeer, M.H.: Cloud Hadoop map reduce for remote sensing image analysis. J. Emerg. Trends Comput. Inf. Sci. **3**(4), 637–644 (2012)
22. Kune, R., Konugurthi, P.K., Agarwal, A., Chillarige, R.R., Buyya, R.: XHAMI–extended HDFS and MapReduce interface for big data image processing applications in cloud computing environments. Softw. Pract. Exp. **47**(3), 455–472 (2017)
23. Ryu, C., Lee, D., Jang, M., Kim, C., Seo, E.: Extensible video processing framework in apache Hadoop. In: IEEE 5th International Conference on Cloud Computing Technology and Science (CloudCom), vol. 2, pp. 305–310 (2013)
24. Kim, M., Cui, Y., Han, S., Lee, H.: Towards efficient design and implementation of a hadoop-based distributed video transcoding system in cloud computing environment. Int. J. Multimedia Ubiquitous Eng. **8**(2), 213–224 (2013)
25. Moise, D., Shestakov, D., Gudmundsson, G., Amsaleg, L.: Terabytescale image similarity search: experience and best practice. In: IEEE International Conference on Big Data, pp. 674–682 (2013)
26. Sozykin, A., Epanchintsev, T.: MIPr-a framework for distributed image processing using Hadoop. In: 9th International Conference on Application of Information and Communication Technologies (AICT), pp. 35–39 (2015)
27. Yamamoto, M., Kaneko, K.: Parallel image database processing with MapReduce and performance evaluation in pseudo distributed mode. Int. J. Electron. Commer. Stud. **3**(2), 211–228 (2012)
28. Epanchintsev, T., Sozykin, A.: Processing large amounts of images on Hadoop with OpenCV. In: CEUR Workshop Proceedings, vol. 1513: Proceedings of the 1st Ural Workshop on Parallel, Distributed, and Cloud Computing for Young Scientists (Ural-PDC 2015), pp. 137–143 (2015)
29. Liu, T., Liu, Y., Li, Q., Wang, X.R., Gao, F., Zhu, Y.C., Qian, D.P.: SEIP: system for efficient image processing on distributed platform. J. Comput. Sci. Technol. **30**(6), 1215–1232 (2015)

30. Powell, M., Rossi, R., Shams, K.: A scalable image processing framework for gigapixel mars and other celestial body images. In: IEEE in Aerospace Conference, pp. 1–11 (2010)
31. Dong, L., Lin, Z., Liang, Y., He, L., Zhang, N., Chen, Q., Cao, X., Izquierdo, E.: A hierarchical distributed processing framework for big image data. IEEE Trans. Big Data **2**(4), 297–309 (2016)
32. Bajcsy, P., Vandecreme, A., Amelot, J., Nguyen, P., Chalfoun, J., Brady, M.: Terabyte-sized image computations on hadoop cluster platforms. In: IEEE International Conference on Big Data, pp. 729–737 (2013)

Exploring the Potential of Virtual Reality in Fire Training Research Using A'WOT Hybrid Method

El Mostafa Bourhim and Abdelghani Cherkaoui

Abstract Virtual reality (VR) is a creative methodological approach dedicated to preparing individuals in an intelligent VR condition. Recently, expanding consideration has been attracted to VR for the provision of fire evacuation knowledge and behavioral assessment, as they are profoundly captivating and advance driving psychological learning. The motivation behind this investigation is to characterize and organize the threats, weaknesses, strengths and the opportunities (SWOT) and their sub-factors of the VR as a predictive and effective research tool on human fire behavior. However, SWOT analysis includes no means of analytically estimating the weights defining the intensity of the factors. The proposed framework A'WOT (AHP-SWOT) integrates the analytic hierarchy process (AHP) and SWOT analysis. AHP's connection to SWOT permits precisely and systematically determined priorities for the factors contained in SWOT analysis and makes them measurable.

Keywords Virtual reality · Human behavior in fire · Fire training · SWOT-AHP

1 Introduction

A fire is an emergency in which appropriate individual intervention is essential to limit physical, psychological and material damage. Since an emergency situation in case of fire is also stressful [1] and can therefore hinder performance during emergency response [2], firefighters or ordinary citizens must undergo exercises and training to educate the intervention procedures and apply them under conditions of high psychological and physical stress [3]. The designing of the virtual environments (VEs) is done to provide a few trainings. The present article proposes an integrated framework to evaluate the SWOT of virtual reality-based fire training utilizing the

E. M. Bourhim (✉) · A. Cherkaoui
EMISYS: Energetic, Mechanic and Industrial Systems, Industrial Engineering Department, Mohammadia School of Engineers, Engineering 3S Research Center, Mohamed V University, Rabat, Morocco
e-mail: elmostafabourhim@research.emi.ac.ma

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_11

AHP in combination with the SWOT. SWOT tool is utilized for the strategic management and strategic planning and has been applied to VR in a variety of contexts such as fire evacuation research [4] and rehabilitation research [5].

The paper is organized as follows. Section 2 presents a literature review on VR in fire training. Section 3 discusses the methodology and how to integrate the AHP model with SWOT analysis (hybrid method). Section 4 contains the results obtained by applying the proposed method and discussion, while Sect. 5 presents our conclusion.

2 Fire Training in a VR

The increasing use of VR devices clearly establishes its high possibilities in different areas, like the human behavior during a fire [4, 6] and the safety measure. Few current researches on VR are presented below. One of the benefits of using VEs for fire training is that they generally attract users, improving their involvement in the process compared to conventional methods. VR has also been used for fire emergency training. Some of these are also very useful in the fire training process [6, 7] and many other emergency situations tackling. For example, VR has been used to teach people to evacuate emergency situations in case of fire [8]. One of the main benefits of using VR technology for fire training is that it provides you to train in dangerous conditions. For example, in a study on the use of VR to teach fire evacuation techniques [9], VR could be most efficient compared to the evacuation drills to motivate people to train in personal fire safety.

The purpose of the comprehensive study is to describe SWOT groups and their sub-factors of VR as an efficient and practical tool for training in fire emergency situation and prioritize them.

3 Research Methodology

3.1 SWOT Analysis

SWOT analysis is considered a powerful used architecture that assesses the threats and opportunities and analyzes the internal factors such as the strength and the weakness of a service or product [10]. In general, the SWOT analysis lists the factors with a description that are associated with the existing and future trend of the internal and external environment. The expressions of the individual factors are imprecise and brief, which describes the subjective opinions. However, SWOT analysis is a powerful and simple tool for assessing the situation and allows a focus group to determine the strengths, opportunities, weaknesses and threats of a given method or product.

- A strength can be considered as capacity that allows VR to achieve its defined goals.
- A weakness can be viewed as a limitation, defect or fault of VR to achieve a specific aim.
- An opportunity comprises external factors that the VR technology can capitalize on or use to its advantage.
- A threat can be any unfavorable external situation that may have a negative impact by presenting a boundary or limit to achieve objectives.

Despite the broad application of SWOT, it has few limitations that are the qualitative measurement of the decision making is impossible and there is no mechanism to rank the one factor versus the other. This would be made possible by combining the SWOT with the AHP.

3.2 The AHP Method

AHP [11] is very appropriate for the weighing problem that is multiobjective as it integrates both the qualitative and the quantitative analyses. It structures alternatives into a hierarchical order through a series of pairwise comparisons. This reference [12] is intended to explain in more detail the general AHP framework.

3.3 A'WOT (SWOT-AHP) Methodology

The point of using AHP strategy inside SWOT examination is to systematically survey factors of the SWOT and compare their powers. Extra esteem from SWOT strategy can be cultivated by doing match insightful examinations between SWOT factors and breaking down them through eigenvalue procedure as applied in AHP approach. This provides a good basis for analyzing the current or anticipated situation, or new strategy alternative deeply and completely [13]. The AHP could convert the intangible criteria into values (numbers) and evaluate the weights of the factors using the pairwise evaluation [11]. The hybrid method A'WOT (AHP-SWOT) proceeds as follows:

Step 1. Evaluation based on the SWOT is done.
Step 2. Pairwise comparisons between SWOT factors are accomplished within every SWOT group.

(1) Structuring the model.
(2) Construct judgment matrix A.

Table 1 1–9 point scale used for the pairwise comparisons [14]

Strength of significance	Synonyms
1	Equally significant
3	Moderately significant
5	Strongly significant
7	Very strongly significant
9	Extremely significant
2, 4, 6, 8	In-between values in the two adjacent judgments

$$A = \left(a_{ij}\right) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix} \tag{1}$$

where

$$\begin{gathered} a_{ij} > 0, \quad i, j = 1, 2, \dots, n \\ a_{ii} = 1, \quad i = 1, 2, \dots, n \\ a_{ij} = 1/a_{ji}(i \neq j), \quad i, j = 1, 2, \dots, n \end{gathered} \tag{2}$$

Saaty's scale is shown in Table 1.

(3) Evaluating the relative weights of decision elements $\lambda_{\max}$
(4) Consistency check:

$$AW = \lambda_{\max} W \tag{3}$$

The consistency check index C.I.:

$$\text{C.I.} = \frac{\lambda\text{max} - n}{n - 1} \tag{4}$$

The consistency ratio C.R.:

$$\text{C.R.} = \frac{\text{C.I.}}{\text{R.I.}} \tag{5}$$

where R.I. is the average random consistency index (Table 2).

The inconsistency level is found to be acceptable if $CR \leq 0.10$. Otherwise, the preferences need to be reviewed and verified.

Table 2 SWOT analysis

Strengths (S)	Weaknesses (W)
• S1: Improved ecological validity • S2: Safe and controlled testing and training environment • S3: Immersive low-cost training environment • S4: Engaging interactive training environment • S5: Great number of trials • S6: Accuracy and error-free training environment • S7: Real-time performance feedback • S8: Multimodal interaction	• W1: Studies in fire training remain infrequent in the literature • W2: Technical challenges • W3: Compatibility with other hardware • W4: Side effects
Opportunities (O)	**Threats (T)**
• O1: The emergence and growth of new technological fields • O2: The commercial availability of VR devices • O3: The adoption of virtual reality (VR) technology in fire drills by professionals • O4: Intuitive navigation platform	• T1: Human sensory limitations • T2: The potential health risks of this technology (nausea, disorientation and so on) • T3: Ethical challenges

Step 3. Comparisons between the four groups in the SWOT are made pairwise.

The detailed framework of the proposed methodology is provided in Fig. 1.

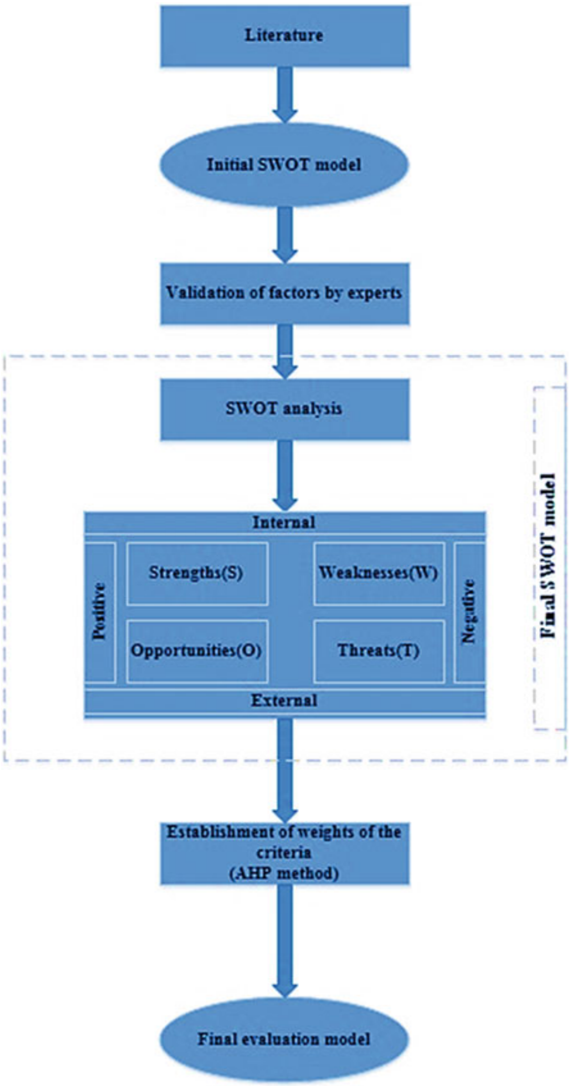
4 Results and Discussion

The first part is comprised of a meeting section to generate factors for each SWOT class that are indispensable to the selection. First, all factors determined by individual members of the focus group were selected. At the end of this session, the focus group participants and interviewees determined the most crucial factors in each SWOT category (see Table 3).

AHP results uncovered that specialists gauge qualities to be the most significant consideration pursued by circumstances, shortcomings and dangers (see Table 3).

Under strengths, improved ecological validity was rated as the most influential factor. This would support the fact that ecological validity takes vital role in the VR fire training research. The central and crucial question is ecological validity of VR when it is used as a predictive tool. It is also the case in VR fire training applications where VR is used to assess human behavior. Actually, in such predictive uses, the closeness of human response within the VE to the equivalent real life is mandatory.

Fig. 1 Process flowchart



Under the weakness category, side effects were the highest rated factor. Based on the expert panel reviews, side effects are more important than the technical challenges. The emerging world of VR is not only providing a new, fully immersive fire simulation, but also creating new ways for people to see the world. The problem, however, is that we do not know its long-term effects on either the body or the mind.

Under opportunities, the adoption of VR technology in fire drills by professionals is the highest rated factor. According to expert group reviews and analysis, it looks that the growth will really be seen in the professional space. Based on various VR

Table 3 Factor priority scores and global priority scores of SWOT factors

SWOT groups	Weight	SWOT factors	Local priority	Global priority
Strengths (S)	0.478	S1: Improved ecological validity	0.413	0.197
		S2: Safe and controlled testing and training environment	0.171	0.082
		S3: Immersive low-cost training environment	0.078	0.037
		S4: Engaging interactive training environment	0.121	0.058
		S5: Great number of trials	0.039	0.019
		S6: Accuracy and error-free training environment	0.079	0.038
		S7: Real-time performance feedback	0.048	0.023
		S8: Multimodal interaction	0.050	0.024
Weaknesses (W)	0.138	W1: Studies in fire training remain infrequent in the literature	0.135	0.019
		W2: Technical challenges	0.271	0.037
		W3: Compatibility with other hardware	0.070	0.010
		W4: Side effects	0.524	0.072
Opportunities (O)	0.256	O1: The emergence and growth of new technological fields	0.142	0.036
		O2: The commercial availability of VR devices	0.214	0.055
		O3: The adoption of virtual reality (VR) technology in fire drills by professionals	0.550	0.141
		O4: Intuitive navigation platform	0.094	0.024
Threats (T)	0.128	T1: Human sensory limitations	0.286	0.037
		T2: The potential health risks of this technology (nausea, disorientation and so on)	0.571	0.073
		T3: Ethical challenges	0.143	0.018

statistics, the VR technology is going to be a huge industry in a few years, in both software and hardware.

Lastly, in threat category, the potential health risks of the VR technology are the highest rated factor. Due to the rapid growth of VR technology, such profound advances come with equally serious risks to our physical and emotional well-being. Consider that a high percentage of participants experience stress or anxiety after wearing a VR headset for more than a few minutes. Other unfavorable physical side effects can include severe eyestrain, motion sickness and nausea. This can be viewed as a dubious point for VR fire training research. Other external factors that affect VR in fire training are human sensory limitations and ethical challenges.

Table 4 shows the values of the global and local priority of the SWOT analysis factors.

5 Conclusion

This article presents a SWOT-AHP hybrid method for the field of VR fire training. Numerous VR strengths are depicted that will keep giving a justification for developing existing VR fire applications and making new ones. Weaknesses exist in the VR, especially with certain limitations in areas of human sensory technology and health risks, but do not compromise the reliability of the VR field. With thoughtful system design that targets fire research applications, it is anticipated that VR in fire research will keep on exponentially developing and gain acknowledgment as a standard tool. Threats to the field do exist; however, none are unsafe and all are likely addressable with the high inspiration that seems to be with the many scholars of this area. The current study was realized by combining SWOT with the AHP to generate the internal and the external factors qualitative values. The results of the hybrid method were promising. Making pairwise examinations powers the specialists to consider the heaviness of the elements and to analyze the circumstance in more profundity and detail. This methodology does not consider interdependencies between levels because AHP lacks this functionality. In our future work, this can be avoided by replacing AHP with the analytic network process (ANP) [14]. Likewise, the positioning of SWOT factors with pairwise examinations can account for the inconsistency of subjective opinions.

Appendix

SWOT factors—pairwise comparison of matrices.

Strengths	S1	S2	S3	S4	S5	S6	S7	S8	Local weight
S1	1	3	5	4	9	6	7	8	0.413
S2	1/3	1	3	2	4	2	3	3	0.171
S3	1/5	1/3	1	1/2	2	1	2	2	0.078
S4	1/4	1/2	2	1	3	2	2	3	0.121
S5	1/9	1/4	1/2	1/3	1	1/2	1	1/2	0.039
S6	1/6	1/2	1	1/2	2	1	2	2	0.079
S7	1/7	1/3	1/2	1/2	1	1/2	1	1	0.048
S8	1/8	1/3	1/2	1/3	2	1/2	1	1	0.050

Weaknesses	W1	W2	W3	W4	Local weight
W1	1	1/2	2	1/4	0.135
W2	2	1	4	1/2	0.271
W3	1/2	1/4	1	1/7	0.070
W4	4	2	7	1	0.524

Opportunities	O1	O2	O3	O4	Local weight
O1	1	1/2	1/4	2	0.142
O2	2	1	1/3	2	0.214
O3	4	3	1	5	0.550
O4	1/2	1/2	1/5	1	0.094

Threats	T1	T2	T3	Local weight
T1	1	1/2	2	0.286
T2	2	1	4	0.571
T3	1/2	1/4	1	0.143

References

1. Proulx, G.: A stress model for people facing a fire. J. Environ. Psychol. **13**(2), 137–147 (1993). ISSN 0272-4944. https://doi.org/10.1016/S0272-4944(05)80146-X
2. Spitzer, W.J., Neely, K.: Critical incident stress: the role of hospital-based social work in developing a statewide intervention system for first-responders delivering emergency services. Soc. Work Health Care **18**(1), 39–58 (1993)
3. Williams-Bell, F.M., Kapralos, B., Hogue, A., Murphy, B.M., Weckman, E.J.: Using serious games and virtual simulation for training in the fire service: A review. Fire Technol. (in press). https://doi.org/10.1007/s10694-014-0398-1
4. Kinateder, M., et al.: Virtual reality for fire evacuation research. In: 2014 Federated Conference on Computer Science and Information Systems, Warsaw, pp. 313–321 (2014). https://doi.org/10.15439/2014F94
5. Rizzo, A.S., Kim, G.J.: A SWOT analysis of the field of virtual reality rehabilitation and therapy. Presence: Teleoperators Virtual Environ. **14**, 119–146 (2005)
6. Bourhim, E.M., Cherkaoui, A.: Simulating pre-evacuation behavior in a virtual fire environment. In: 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bangalore, pp. 1–7 (2018). https://doi.org/10.1109/ICCCNT.2018.8493658

7. Cha, M., Han, S., Lee, J., Choi, B.: A virtual reality based fire training simulator integrated with fire dynamics data. Fire Saf. J. **50**, 12–24 (2012)
8. Xi, M., Smith, S.P.: Simulating cooperative fire evacuation training in a virtual environment using gaming technology. In: 2014 IEEE Virtual Reality (VR). IEEE, pp. 139–140 (2014)
9. Padgett, L.S., Strickland, D., Coles, C.D.: Case study: using a virtual reality computer game to teach fire safety skills to children diagnosed with fetal alcohol syndrome. J. Pediatr. Psychol. **31**(1), 65–70 (2006)
10. Rauch, P.: SWOT analyses and SWOT strategy formulation for forest owner cooperations in Austria. Eur. J. Forest Res. **126**(3), 413–420 (2007)
11. Saaty, T.L.: The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. McGraw-Hill International Book Co., New York (1980)
12. Bourhim, E.M., Cherkaoui, A. Selection of optimal game engine by using AHP approach for virtual reality fire safety training. In: Abraham, A., Cherukuri, A., Melin, P., Gandhi, N. (eds.) Intelligent Systems Design and Applications. ISDA 2018 2018. Advances in Intelligent Systems and Computing, vol. 940. Springer, Cham (2020)
13. Kurttila, M., Pesonen, M., Kangas, J., Kajanus, M.: Utilizing the analytic hierarchy process (AHP) in SWOT analysis—a hybrid method and its application to a forest-certification case. Forest Pol. Econ. **1**(1), 41–52 (2000). ISSN 1389-9341. https://doi.org/10.1016/S1389-9341(99)000040
14. Saaty, T.L.: Analytic Hierarchy Process. Encyclopedia of Biostatistics (2005). https://doi.org/10.1002/0470011815.b2a4a002

Two-Way Sequence Modeling for Context-Aware Recommender Systems with Multiple Interactive Bidirectional Gated Recurrent Unit

K. U. Kala and M. Nandhini

Abstract For modeling the user behavior in recommender systems, the task of combining the contexts of interactions corresponds to the sequential item history has inevitable role in improving the quality of recommendations. The resort of existing recommendation models is the left-to-right autoregressive training approach. While training a certain model at a specific time step, both future (right) context/data along with the past (left) is always available in the given training set sequences. It is intuitive that the current behavior of the user has certain connections with their future actions too. Future behaviors of users can boost the quality of recommendations. In this paper, two-way sequence modeling technique is proposed for concatenating both left-to-right (past) and right-to-left (future) dependencies in a user interaction sequence. Inspired from the text modeling techniques, a Multiple interactive Bidirectional Gated Recurrent Unit (MiBiGRU) architecture is proposed to model the two-way dependencies in recommender systems. Modeling future contexts along with past contexts is an auspicious way for attaining better recommendation accuracy.

Keywords Sequence-aware recommender system · Context-aware recommender system (CARS) · Deep learning · Gated recurrent unit (GRU) · Two-way sequence modeling · Bidirectional GRU

1 Introduction

Sequence-aware recommender systems (SARSs) are becoming an emerging topic in the recommender system (RS) domain, which can predict the next few items of the sequence based on the observed historical sequence of the user [1]. Users' historical sequences provide valued inklings for predicting future behavior. Markov Chains [2] and in combination with matrix factorization techniques [3] solved the historical behavioral pattern matching problems for recommender systems [RSs]. Recent advancements in deep neural networks [4–9] have led to promising approaches in modeling the distribution of user interest in sequences. For capturing patterns from

K. U. Kala (✉) · M. Nandhini
Department of Computer Science, Pondicherry University, Puducherry 605014, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_12

the sequential history, recurrent neural networks (RNNs) are more effective. These techniques are deficient in modeling the temporal gaps between events explicitly.

Contextual features like location, time of the day, user mood, etc. have the ability to control the user preferences; thus, it provides another dimension of information to the RSs. Context-aware recommender systems [CARSs] are utilized such additional information for better personalization of the recommendations [10]. For example, sales statistics may differ with the related contextual variables such as location, weather. Liu et al. [11] explore the scope of contextual modeling (modeling contexts along with historical sequence simultaneously) in their context-aware RNN (CARNN) model. In this model, for every user-specified context, a separate parameter matrix is learnt. This makes the count of parameters increases along with the increase in the number of additional contexts. It makes the training process more difficult.

In literature, it is very common to develop SARS by predicting the next item based on the previous behaviors of the user [1, 11, 12]. This approach is reasonable for the short-range SARS scenario; it is not that much satisfied with relatively long sequence data. The reason for this is internal dependencies are not considering while modeling such long sequences. While training a certain model at a specific time step, both future (right) context/data along with the past (left) is always available in the given training set sequences. It is intuitive that the current behavior of the user has certain connections with their future actions too. Hence, it reasonably believes that the current SARS scenario can be improved by leveraging the right context appropriately during training. Fajie et al. [13] investigated several augmentation strategies to incorporate future contexts along with and past contexts. Intuitively, SARS may achieve further improvements if both future and past preferences of the user are considered during training in terms of both input and item sequences.

In long sequential histories of users, the patterns are found to be changed even with the changes in temporal contexts. Rakkapan et al. [14] designed a new context-aware SARS for modeling using stacked RNNs that model the dynamics of input and temporal contexts. Here, a single parameter matrix is used for learning both input and contexts; it effectively solved the complexity of the training with increased number of parameters.

In this paper, a Multiple interactive Bidirectional Gated Recurrent Unit [MiBiGRU] architecture is proposed for two-way context-aware sequential recommendations. MiBiGRU consists of two BiGRUs, one is for modeling input sequence (Item-BiGRU) and the other is for modeling contextual sequences (Context-BiGRU). BiGRU consists of two GRU cells: one is forward cell and the other is backward cell. The input sequence is organized from the first item to the last item in the forward GRU structure, and the model computes the forward hidden states. The input sequence is arranged in reverse order from last to first in backward GRU, and the model computes backward hidden states [15]. The output of GRUs in both directions is concatenated and applying linear transformation for generating the output of Input-BiGRU. Context-BiGRU also works in the same way to generate the output on the contextual input sequences. For modeling the recurrent representation of the user, the signals from both the BiGRUs are used in the MiBiGRU architecture. To compute the final prediction, output of the last hidden layer is passed to the softmax prediction unit. The objectives of the proposed approach are

- To model the past and future behavior of the user for achieving further improvements in SARS.
- To achieve better context awareness in SARS two interacting BiGRUs, one for item and the other for corresponding contexts, respectively.

2 Related Work

Recommender systems (RSs) assist the users in choosing the items of their interest in a personalized way from the wide and varied collections of items. Recent developments in the area of RS focused on SARS, which is the most recent and relevant approach in this area. SARS is satisfying the need of multiple user-item interactions in various application domains, whereas the traditional RS does not support It is an efficient way of finding further evident patterns through richer user modeling techniques like deep learning and machine learning. Quadrana et al. [1] efficiently surveyed various SARS approaches for modeling user behavioral patterns. In the application domains like time-series prediction, natural language processing, DNA modeling, where the data is inherently maintaining a sequential nature, sequence learning methods are very useful [1]. Inspired from these domains, sequence learning based on RNNs is most successfully used for SARSs.

Gated recurrent unit (GRU) and long short-term memory (LSTM) are the two variations of RNN, which are most commonly used for sequence-aware recommender systems. By reparametrizing the RNN, the vanishing gradient problem is effectively solved in LSTM and GRU [16–18]. The input gate is used to multiply the input to the LSTM cell by the activation function of the gate, and the forget gate multiplies the previous values to the cell, i.e., the interactions of the LSTM cells are only through the gates. GRU reduces the number of gates in the LSTM architecture by an update gate which combines input gate and forget gate and the memory cell state is combined with the hidden state. GRU already leaves behind LSTM on a set of applications. [19–21].

Contextual information such as time of the day, location, user mood, etc. has the ability to control the user preferences; thus, it provides another dimension of information to the RSs. Context-aware recommender systems [CARSs] are utilized such additional information for better personalization of the recommendations [10]. GRU has already proved his role in handling such additional contexts in SARS [11, 22–25]. The most prominent GRU architectures for handling context-aware SARS are shown in Table 1.

The basic RNNs are unidirectional in nature, i.e., output of every layer depends only on the previous layer output. Typically, the user preferences depend on the past and future contexts in both directions in the sequences, capturing future contextual history leads to the improvement of accuracy in recommendations. For solving this

Table 1 Context-aware GRU architectures for SARS

Features	GRU Architectures				
	STGRU [22]	CAGRU [11]	CGRU [23]	Latent Cross [24]	CARA [25]
Sequential-based	✓	✓	✓	✓	✓
Context-aware	✓	✓	✓	✓	✓
Static/dynamic contexts	Only dynamic	One context	One context	✓	✓
Bidirectional	×	×	×	×	×

problem, the concept of bidirectional RNN (BiRNN) came into action, where two separate recurrent nets called forward and backward to learn the sequence, both are connected to the same output layer [26–28]. In the area of RSs, for modeling past and future user actions, Yuan et al. [13] designed a neural network (NN) architecture, which combines two individually trained unidirectional objective functions based on NextItNet [4], and then, he introduced a CNN architecture, which utilizes the blank-masking technique inspired from the gap-filling techniques used in NLP along with two-way objective function training. It is complex structure, inspired from the RNN architectures used in NLP [26–28]; the proposed model utilizes the scope of BiGRU, a variation of BiRNN in two-way sequence modeling.

Along with BiRNN, stacked-RNN also showed remarkable performance for the success of NLP tasks. In stacked-RNN, more than one RNN are stacked one above the other as layers. The output of each hidden states is forwarded to the above-hidden state layer as input. In stacked architecture, the short terms interactions are capture by the lower-level layer, while the high-level layers aggregate these interactions to global level [4]. Kiewan et al. [29] experimented stacked-GRU along with layer normalization technique for RSs, and the results show adding more layers shows slight improvements in the result. Liu et al. [15] evaluated the performance of different RNN structure (GRU, LSTM, stacked-GRU, BiGRU) on the process of producing recommendations from the reviews and the experiments show good result on stacked and BiGRU architectures. Rakkapan et al. [14] designed a new context-aware SARS using stacked RNNs that model the dynamics of input and temporal contexts. Here, a single parameter matrix is used for learning both inputs and contexts; it effectively solved the complexity of the training with increased number of parameters.

Inspired from the literatures, we are proposing a Multiple interactive Bidirectional Gated Recurrent Unit [MiBiGRU] architecture for two-way context-aware sequential recommendations. MiBiGRU consists of two BiGRUs, one is for modeling input sequence (Input-BiGRU) and the other is for modeling contextual sequences. BiGRU is an efficient way of modeling past and future actions of a user and it has the ability to find better sequential patterns. For contextual modeling, a variation of stacked-GRU, i.e., two interactive BiGRUs instead of stacking them, is used. In short, MiBiGRU is a combination of two multiple interactive bidirectional GRUs.

3 Proposed Two-Way Sequence Modeling Approach

GRU is a RNN variation that works well in a sequence prediction and recommendation problem. In this paper, we are proposing an architecture, which has the ability to model the sequence based on both historical and future user actions. Inspired from the reviewed literatures, our architecture utilizes the power of BiGRU in modeling a sequence bidirectionally. For modeling input contexts along with the item sequence, an additional BiGRU is incorporated. These two interactive GRUs can effectively model the sequence along with the strategies for incorporating the right context from both data and context perspective.

3.1 Problem Definition

Let $X = \left\{x_1^u, x_2^u, x_3^u, \ldots\right\}$ represents the history of items that the user u is interacting with (purchase, click, view, etc.) in the timestamp order. An input context c_t^u is associated with each interaction x_t^u of a user. Input can represent physical or social contexts or other temporal contexts. For a user u, the task of sequential recommendation is for predicting x_{t+1}^u from the given historical data, comprising the past interactions $\left\{x_1^u, x_2^u, \ldots, x_t^u\right\}$ together with their input contexts $\left\{c_1^u, c_2^u, \ldots, c_t^u\right\}$.

3.2 The MiBiGRU Model

The proposed model utilizes both future (right) context/data along with the past (left). Both are always available in the given training set sequences, while training the model at a specific time step. It is intuitive that the current behavior of the user has certain connections with their future actions too. Hence, it is reasonable to believe that, in order to improve the current SARS scenario, during the training process, appropriate leveraging of the right context will be helpful. A major issue with all of the basic RNN networks such as GRU and LSTM is that they learn representations from previous time steps. Sometimes, it is necessary to learn the representations from future time steps for the better understanding of the context and eradicate uncertainty. BiRNN can effectively learn representations from both past and future time steps.

The major part of our proposed architecture is two BiGRU networks, one corresponds to input sequences and the other corresponds to the context sequences, respectively. These two interactive BiGRUs learn the patterns from left to right and right to left in the item and context sequence at a particular time step. Figure 1 shows the MiBiGRU architecture for two-way context-aware RSs. The repeating module in a bidirectional RNN could be a conventional RNN, LSTM, or GRU. GRU is used in the proposed model. There are two types of connections, one going forward in time, which helps us learn from previous representations and another going backward in

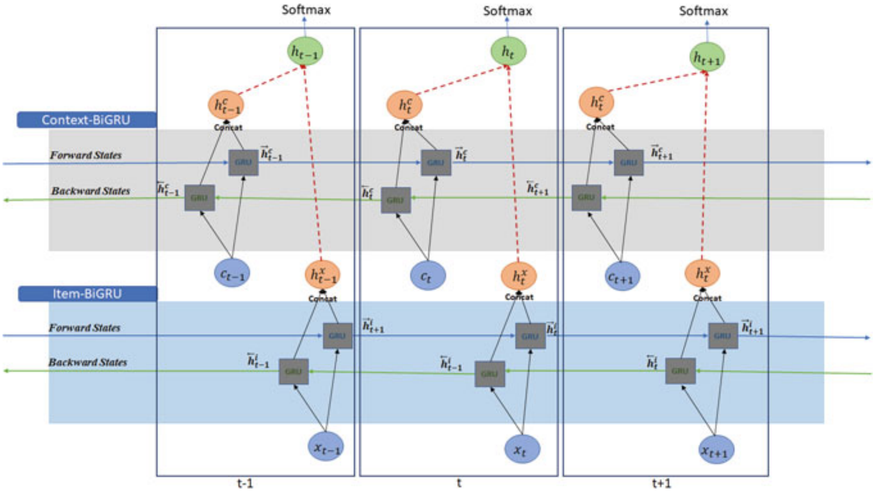


Fig. 1 Proposed MiBiGRU architecture

time, which helps us learn from future representations. Forward propagation is done in two steps: We move from left to right, starting with the initial time step the values are computed until reach the last time step and we move from right to left, starting with the last time step we compute the values until we reach the initial time step.

3.3 Back Ground

This section deals with the basic processing of GRU for sequential modeling applications and how GRU is modified to BiGRU for modeling sequential data bidirectionally by considering both forward and reverse order of the data at the same time.

The basic GRU. We are focusing on the efficacy of GRU method from among the both gated RNN networks. GRU has an update and reset gate in the network, which deals with the updated degree of each hidden state, that is, it decides which information has to pass to the next state and which are not needed to be passed [9, 12]. GRU calculates hidden state h_t at time t from the output of the update gate z_t, reset gate r_t, current input x_t, previous hidden state h_{t-1}, and the activation function $\hat{h}_t$ calculated from the reset gate as:

$$h_t = (1 - z_t)h_{t-1} + z_t\hat{h}_t, \tag{1}$$

The equations for z_t, r_t and $\hat{h}_t$ are given by

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{2}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1+} b_r) \tag{3}$$

$$\hat{h}_t = \tan h(W x_t + U(r_t \odot h_{t-1})) \tag{4}$$

where σ a logistic is function and $\odot$ is an elementary multiplication operation.
Bidirectional GRU. The BiGRU model is designed using two independent interacted GRUs. To the first GRU, the sequences are fed in forward time order, and into another GRU, in backward order.

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t \tag{5}$$

where symbol $\oplus$ represents concatenation operation of two hidden representation vectors. Both $\vec{h}_t$ and $\overleftarrow{h}_t$ contain the information of the whole sequence. Hidden state $\vec{h}_t$ encodes the information of the item (from x_1 to x_t), while $\overleftarrow{h}_t$ captures the information in the sequence (from x_t to x_1).

3.3.1 MiBiGRU Model

The major part of our proposed architecture is two BiGRU networks, one corresponds to input item sequences and the other for the corresponding context sequences, respectively. These two interactive BiGRUs learn the patterns from left to right and right to left in the item and context sequence at a particular time step. Figure 1 represents the MiBiGRU model for two-way context-aware RSs. The figure depicts the working of MiBiGRU at the time steps $t - 1$, t and $t + 1$.

Item-BiGRU is an individual BiGRU, which has the capability of storing both the previous and later information as the current time basis of the item-sequence data, and thus, the good performance is achieved. With both forward and backward directions taken care of in one network, input information from the past and future of the currently evaluated time frame can be directly being used to minimize the objective function. Forward state and backward state can be executed the same way as in the basic GRU and the hidden states' outputs at time t of each pass, $\vec{h}_t^i$, $\overleftarrow{h}_{t-1}^i$ is concatenating to find the output hidden state h_t^x of the item sequences. Context-GRU is also a BiGRU, but it is working on corresponding contextual sequences to find the two-way contextual dependency along with the item dependency in the item-GRU. The predictions using two-way sequence data for CARS are modeled with the two interactive BiGRUs (Item-BiGRU and Context-BiGRU). The final hidden states from both BiGRUs $\left(h_t^x, h_t^c\right)$ at time t are used to form the output hidden state h_t which is used for computing the probabilities using softmax normalization.

A schematic representation of the proposed model is depicted in Fig. 1. A recurrent GRU cell is used in each block and depicts the working at three time steps: $t - 1$, t and $t + 1$ for better understanding of data flow in the modeling process. The architecture models the users evolving representation (h_t) that depends on current interaction x_t, the hidden state h_t^x (evolved from past and future interactions using Item-BiGRU), and the hidden state h_t^c of the evolving input context c_t (evolved from the corresponding past and future input contexts with the help of Context-BiGRU). These two output hidden states are concatenated again for getting the desired output representation (h_t) at time step t.

3.3.2 Training and Prediction

The proposed model utilizes a softmax layer which takes hidden representation of the user h_t at each time step for generating the probability vector.

4 Conclusion and Future Work

In this paper, we are proposing a Multiple interactive Bidirectional Gated Recurrent Unit [MiBiGRU] architecture for two-way context-aware sequential recommendations. The major part of our proposed architecture is two BiGRU networks, one corresponds to input sequences (Item-BiGRU) and the other corresponds to the context sequences (Context-BiGRU), respectively. These two interactive BiGRUs learn the patterns from left to right and right to left in the item and context sequence at a particular time step. The ability of modeling future contexts along with past contexts is an auspicious way for attaining better recommendation accuracy. As a continuation of the research, the proposed model has to be implemented and the performance has to be evaluated in comparison with other relevant GRU models for CARS. Further research can be done by stacking the MiBiGRU model with the expectation of multi-stacked BiRNN out-perform the single stack BiRNN.

References

1. Quadrana, M., Cremonesi, P., Jannach, D.: Sequence-aware recommender systems. ACM Comput. Surv. **51**(4), 36 Article 66 (2018)
2. Chen, S., Moore, J.L., Turnbull, D., Joachims, T.: Playlist prediction via metric embedding. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 714–722. ACM (2012)
3. Rendle, S., Freudenthaler, C., Schmidt-Thieme, L.: Factorizing personalized markov chains for next-basket recommendation. In: Proceedings of the 19th International Conference on World Wide Web, pp. 811–820. ACM (2010)
4. Kang, W.-C., McAuley, J.: Self-attentive sequential recommendation. In: 2018 IEEE International Conference on Data Mining (ICDM), pp. 197–206, IEEE (2018)
5. Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T., Ma, J.: Neural attentive session-based recommendation. In CIKM. ACM, pp. 1419–1428 (2017)
6. Tang, J., Wang, K.: Personalized top-N sequential recommendation via convolutional sequence embedding. In: ACM International Conference on Web Search and Data Mining (2018)

7. Tuan, T.X., Phuong, T.M.: 3D convolutional networks for session-based recommendation with content features. In: Rec Sys. ACM (2017)
8. Yuan, F., Karatzoglou, A., Arapakis, I., Jose, J.M., He, X.: A simple convolutional generative network for next item recommendation. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp. 582–590. ACM (2019)
9. Kala, K.U., Nandhini, M.: Applicability of deep learning techniques in recommender systems. IIOABJ **10**(1), 11–20 (2018)
10. Kala, K.U., Nandhini, M.: Scope of context awareness in cross domain recommender system—a brief review. Int. J. Eng. Technol. **7**(4), 5570–5579 (2019)
11. Liu, Q., Wu, S., Wang, D., Li, Z., Wang, L.: Context-aware sequential recommendation. In: ICDM, pp. 1053–1058 (2016)
12. Yu, F., Liu, Q., Wu, S., Wang, L., Tan, T.: A dynamic recurrent model for next basket recommendation. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 729–732. ACM (2016)
13. Yuan, F., He, X., Guo, G., Xu, Z., Xiong, J., He, X.: Modeling the Past and Future Contexts for Session-based Recommendation. CoRR, abs/1906.04473v2., (2019)
14. Rakkappan, L., Rajan, V.:. Context-aware sequential recommendations with stacked recurrent neural networks. In: Liu, L., White, R. (eds.) The World Wide Web Conference (WWW '19), pp. 3172–3178. ACM, New York, NY, USA (2019)
15. Liu, D.Z., Singh, G.: A Recurrent Neural Network Based Recommendation System. Available online: http://cs224d.stanford.edu/reports/LiuSingh.pdf. Accessed on 16 May 2018
16. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 309, 1735–1780 (1997)
17. Gers, F., Schraudolph, N., Schmidhuber, J.: Learning precise timing with LSTM recurrent networks. J. Mach. Learn. Res. **3**, 115–143 (2002)
18. Cho, K., van Merrienboer, B., Gulcehre, C., Bougares, F., Schwenk, H., Bengio, Y.: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, pp. 315 . arXiv preprint arXiv:1406.1078 (2014)
19. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical Evaluation of Gated Re-current Neural Networks on Sequence Modeling. arXiv preprint arXiv:1412.3555 (2014)
20. Jozefowicz, R., Zaremba, W., Sutskever, I.: An empirical exploration of recurrent network architectures. In: Proceedings of the 32nd International Conference on Machine Learning (ICML-15), pp. 2342–2350 (2015)
21. Bahdanau, D., Cho, K., Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate. arXiv preprint arXiv:1409.0473 (2014)
22. Liu, Q., Wu, S., Wang, L., Tan, T.: Predicting the next location: A recurrent model with spatial and temporal contexts. In: AAAI, pp. 194–200 (2016)
23. Smirnova, E., Vasile, F.: Contextual sequence modeling for recommendation with recurrent neural networks. In: Proceedings of ACM Recommender Systems conference, Como, Italy, (RecSys '17), ACM (2017)
24. Beutel, A., Covington, P., Jain, S., Xu, C., Li, J., Gatto, V., Chi, H.: Latent cross: making use of context in recurrent recommender systems. In: Proceedings of WSDM (2018)
25. Manotumruksa, J., Macdonald, C., Ounis, I.: A contextual attention recurrent architecture for context-aware venue recommendation. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 555–564. ACM (2018)
26. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. IEEE Trans. Signal Process. **45**, 2673–2681 (1997)
27. Baldi, P., Brunak, S., Frasconi, P., Soda, G., Pollastri, G.: Exploiting the past and the future in protein secondary structure prediction. BIOINF: Bioinf. **15** (1999)
28. Graves, A., Schmidhuber, J.: Framewise Phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. **18**(5–6), 602–610 (2005)
29. Villatel, K., Smirnova, E., Mary, J., Preux, P.: Recurrent Neural Networks for Long and Short-Term Sequential Recommendation. ArXiv:abs/1807.09142 (2018)

Stage Audio Classifier Using Artificial Neural Network

M. S. Arun Sankar, Tharak Sai Bobba and P. S. Sathi Devi

Abstract Perceptual quality of audio signals at the receiver and transmission data rate are the major concerns for the speech codec developers. But both these parameters are inversely proportional in general. In the era of 4G, 3GPP launched Enhanced Voice Services (EVS) codec which can operate in multiple data rates with a six-stage speech classifier using threshold-based GMM statistical model. In this work, we propose a seven-stage audio classifier for voiced, unvoiced, transition, multi-speaker, silence, background noise and music signals using neural network by employing Levenberg Marquardt (LM) algorithm. In comparison with conventional statistical approach that requires determination of manual thresholds, the neural network method can simplify the categorization process especially while using a large number of parameters. The categorization is done by using extracted seven features that constitute to a 32-dimensional vector. TIMIT and NOIZEUS databases are used to generate the dataset and a classification accuracy of 94% is obtained. As the network model can perform efficiently using lesser number of neurons, the complexity is also less.

Keywords Speech coding · LPC · Neural network · Speech classifier · CELP · Speech

1 Introduction

The speech signal is an air pressure wave that emerges from a speaker's mouth and nose as a result of the modulation done by the vocal tract to the air expelled from lungs [1]. The transmission requires 64 kbps [2] which is not affordable in many situations and a new area has been opened called speech coding and optimization. For various categories of speech, voiced, unvoiced and silence regions, the production mechanism varies and hence, another method to lower the bit rate is to use coding

M. S. Arun Sankar (✉) · T. S. Bobba · P. S. Sathi Devi
Department of Electronics and Communication Engineering, National Institute of Technology Calicut, Calicut, Kerala, India
e-mail: arun_p150036ec@nitc.ac.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_13

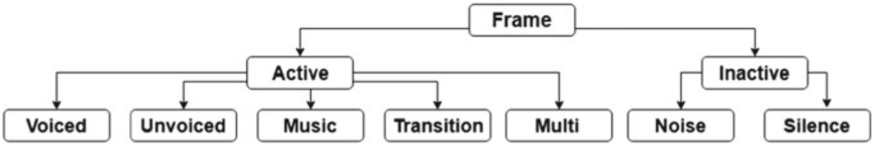


Fig. 1 Classification using the proposed method

scheme that suits best for each category of speech [3]. This method is adopted in various codecs developed including the latest Enhanced Voice Service (EVS) codec [4]. EVS supports both Algebraic Code Excited Linear Prediction (ACELP) and MDCT coding modes and does content-driven switching in between speech and audio frames.

The performance of speech/music classifier of EVS is optimized in [5] by using neural network approach. In [6], a feedforward neural network is used for voiced/unvoiced/silence classification using five features and a similar classification using unsupervised learning is done by Huiqun and O'Shaughnessy [7]. In this article, we propose a seven-stage audio classifier as shown in Fig. 1 which includes the six categories of EVS audio codec and an additional multi-speaker category for teleconferencing applications. The neural network is developed with Levenberg Marquardt (LM) algorithm [8] using selected seven features in which two features are selected from the SAD module of EVS codec.

The subsequent sections are arranged as follows: Sect. 2 describes about the present categorization model in EVS audio codec, Sect. 3 explains the development of the proposed model for classification and the experimental results are discussed in Sect. 4 followed by conclusion given in Sect. 5.

2 Paper Preparation

2.1 Description of Speech Classification

EVS codec is mainly used to support voice quality and network capacity. The SAD module [4] is present in the front end to classify the signal as active (containing some useful information) or inactive (pause and background noise). SAD module has three sub-modules for the signal processing and the final output of the SAD module is obtained by the combined output of these three modules which is used as classifier decision.

SAD 1 It is a sub-band SNR-based signal activity detector and the spectral analysis is performed twice per frame. Using SNR per critical band ($SNR_{CB}(i)$), higher average SNR $\left(SNR_{av_H}\right)$ and average SNR (SNR_{av}) are calculated as,

$$\mathrm{SNR_{av}} = \log_{10} \sum_i \begin{cases} \mathrm{SNR_{CB}}(i); & \mathrm{SNR_{CB}}(i) > \mathrm{thr} \\ \mathrm{min_{snr}}; & \text{otherwise} \end{cases} \quad (1)$$

$$\mathrm{SNR_{av_H}} = \log_{10} \sum_i \begin{cases} \mathrm{SNR_{CB}}(i); & \mathrm{SNR_{CB}}(i) > \mathrm{thr_H} \\ \mathrm{min_{snr}}; & \text{otherwise} \end{cases} \quad (2)$$

where thr and $\mathrm{thr_H}$ are comparison thresholds. The signal activity is detected by comparing the updated values of $\mathrm{SNR_{av_H}}$ and $\mathrm{SNR_{av}}$ to a threshold which is statistically calculated.

SAD 2 This is also a sub-band-based signal activity detector that makes an activity decision for each frame by measuring modified segmental SNR ($\mathrm{MSSNR_{CB}}$) given as,

$$\mathrm{MSSNR_{CB}} = \sum_i \log_{10}\left(\left(\mathrm{SNR_{CBlog}}(i) + \alpha(i)\right)^{\beta(i)}\right) \quad (3)$$

The values of α(i) and β(i) are fixed based on critical bands and are mentioned in the standard. Unvoiced signal is detected if $\mathrm{SNR_{CB}}(18) > 5$ and $\mathrm{SNR_{CB}}(19) > 5$. Hangover process is followed next, SAD2 module uses two hangover processes to classify active and inactive frames.

SAD 3 The audio signal is decomposed into complex values sub-bands using complex low-delay filter bank (CLDFB) that generates a time-frequency matrix of 16 time slots. Spectral flatness features and tonality features are computed by using spectral amplitude $\left(\mathrm{A_{sp}}\right)$ computed as follows,

$$\mathrm{A_{sp}}(8k + j) = \sqrt{X_{F_p}(k, j)^2 + X_{F_p}(k, 15 - j)^2}, k \text{ is even} \quad (4)$$

$$\mathrm{A_{sp}}(8k + 7 - j) = \sqrt{X_{F_p}(k, j)^2 + X_{F_p}(k, 15 - j)^2}, k \text{ is odd}$$

After the classification, the inactive frames such as silence frames and noise frames are encoded with lower data rates and active speech frames are encoded with high data rates.

3 Proposed Model

3.1 Scope of Neural Networks

Neural networks can model nonlinear and complex relationships, which makes it much suitable for speech processing, since all the relations we have discussed till now are mostly nonlinear [9, 10]. From the neural network illustration shown in Fig. 2, the parameters extracted are given as inputs which are further processed by the hidden layer(s) to give the outputs that serve as the decision. Main advantage

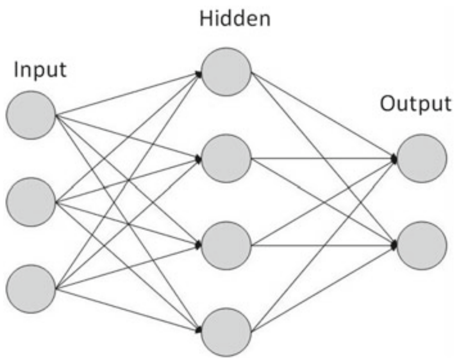


Fig. 2 Schematic representation of artificial neural network

is unlike any other prediction models or any other threshold techniques; it does not make any assumptions about the input such as linearity assumption or second-order Gaussian assumption.

3.2 Feature Selection

After a careful review through the features of SAD module of EVS codec, the following features are selected,

Energy of the critical bands Both SAD 1 and SAD 2 sub-modules are completely dependent on these critical band energies. But since we are only considering narrow-band signals, only the first 17 critical bands are considered.
Energies of the sub-bands In the SAD 3 sub-module, we are considering bands of 400 Hz bandwidth and hence, energies of initial nine sub-bands are considered as another feature.

The performance accuracy of the classifier using the abovementioned two features is not satisfactory and the details are given in next section. G.729 [11]. The features that have significant impact on the performance of model are selected and are as follows;

The Zero Crossing Rate (ZCR) is a simple measure that differentiates between voiced and unvoiced regions is given as,

$$\mathrm{ZCR} = \sum_{i=0}^{158} \mathrm{sign}(x(i)) - \mathrm{sign}(x(i+1)) \tag{5}$$

$$x = \frac{\sum_{n=1}^{3}\sum_{i=1}^{117}\left|S^{[m-n]}(i) - S^{[m]}(i)\right|}{\sum_{n=1}^{3}\sum_{i=1}^{117}\left|S^{[m-n]}(i) + S^{[m]}(i)\right|} \tag{6}$$

where $S^{[m]}(i)$ represents the present frame's ith spectral coefficient and $S^{[m-n]}(i)$ represents the nth previous frame's ith spectral coefficient.

Spectral Peakiness Measure: In the power spectrum, the first two peaks are measured. Power spectrum peaks are found by searching for samples that are higher than their immediate neighboring samples in the power spectrum [12] and helps in detecting the music frames.

$$P_1 = \frac{\sqrt{\frac{1}{K}\sum_{i=1}^{K} S(i)^2}}{\frac{1}{K}\sum_{i=1}^{K}|S(i)|}; \quad P_2 = \frac{\max|S(i)|}{\frac{1}{K}\sum_{i=1}^{K}|S(i)|} \tag{7}$$

where P_1, P_2 are the spectral peakiness measures, K highest-value spectral peaks, denoted as $S(i)$ and $K < 6$.

RMS value: Root mean square (RMS) value for the signal is chosen as a feature because, for the silence frame the RMS value will be very low and thus it helps in the detection of silence frames with ease.

$$\text{RMS} = \sqrt{\frac{1}{160}\sum_{i=0}^{159} x(i)^2} \tag{8}$$

Normalized Auto-Correlation: Normalized Auto-Correlation (NAC) [11] is chosen as a feature because for noise frames the auto-correlation value will be very high and helps in the detection of noise frames. For voiced frames, the values will be in the vicinity of one and near zero for unvoiced frames. A unit delay normalized auto-correlation is defined as,

$$C = \frac{\sum_{i=0}^{158} x(i)x(i+1)}{\left(\sum_{i=0}^{158} x(i)\right)\left(\sum_{i=1}^{159} x(i+1)\right)} \tag{9}$$

The selected seven features add up to 32-dimensional input to neural network as shown in Table 1 and the trained neural network predicts the class accordingly as shown in Fig. 3.

Table 1 Feature set of the proposed method

Feature	Dimension
Critical Band Energy (CBE)	17
Sub-Band Energy (SBE)	9
Spectral peakiness measure	2
ZCR, SF, RMS and NAC	1 × 4

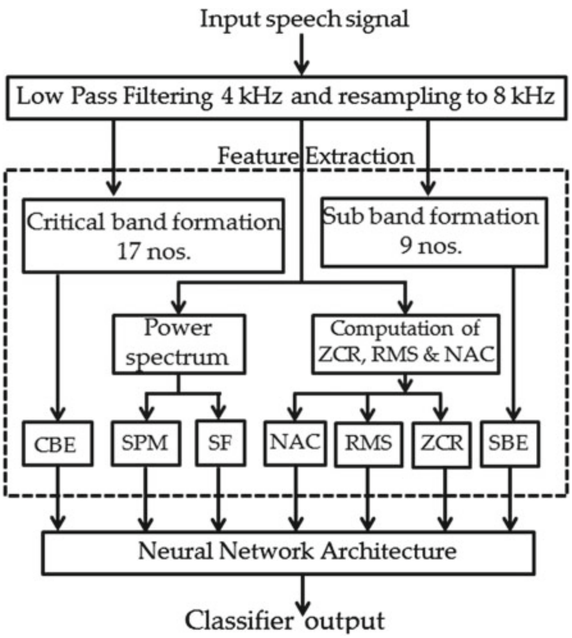


Fig. 3 Structure of proposed neural network model

4 Results and Discussion

4.1 Database

The development of neural network is done using speech samples taken from TIMIT [13] and NOIZEUS [14] databases.

4.2 Training Set

The speech samples from TIMIT and NOIZEUS database are band-limited to 4 kHz and re-sampled at 8 kHz. Using PRAAT [15] and SONIC VISUALIZER [16] software, the signal is manually converted to desired frame categories that add up to 100,000 samples dataset. Two sets of features are extracted from training set which are described as follows;

Set A: This set of features is same as that used in the SAD module of the EVS codec, which include 17 critical band energies used by SAD1 and SAD2 modules, and nine sub-band energies in SAD3 module. This 26-dimensional feature vector is extracted from speech samples of abovementioned database that adds up to 100,000 training sets.

Table 2 Regression values of NN_{Prop} and NN_A neural networks

No. of neurons	Training	NN_{Prop} testing	Validation	Training	NN_A testing	Validation
5	0.935	0.905	0.9343	0.7174	0.719	0.7015
10	0.936	0.913	0.9344	0.7174	0.719	0.7118
15	0.936	0.9135	0.9373	0.7475	0.75	0.7273
20	0.94	0.9135	0.9383	0.7636	0.76	0.7361
25	0.94	0.914	0.9448	0.7627	0.76	0.7653
30	0.941	0.9167	0.9533	0.7698	0.761	0.7684

Set B: This training set consists of features selected for the proposed model. The seven features mentioned above are extracted from speech samples of database that form an input training matrix of dimension 100,000 × 32.

4.3 Evaluation

The training of neural network is done using LM algorithm which is a multi-layer feedforward backpropagation algorithm. The training set is split up into 60% for training, 20% for testing and 20% for validation.

Initially, the neural network is trained with feature set A called as Neural Network A model NN_A, and feature set B that corresponds to the proposed model NN_{Prop} and the variation of the network performance are tabulated in Table 2. It is evident that there is improvement in the regression values as the number of neurons is increased from 5 to 30 for NN_A model. But while comparing it with the ideal regression value of 1, the performance of the network is not good.

There is an exponential growth in the performance of NN_{Prop} model in comparison with that of the NN_A model as evident from the results. The performance improvement with increase in number of neuron is not that significant and hence, the network can be modeled with less number of neurons. A comparison of the mean square error (MSE) and the regression values of the testing phase for NN_{Prop} and NN_A is illustrated in Fig. 4 against number of neurons.

The MSE is scaled by 10 for both models to increase the visibility and it is evident from the results that MSE for NN_A is nearly five times as that for NN_{Prop}. The deviation is slightly reduced at higher number of neurons due to the reduction in MSE value of NN_A which is significant in comparison to the minute reduction of MSE for NN_{Prop}. The similar trend can be observed for the regression values of both models where the average raise of 0.2 for NN_{Prop} is lowered slightly at higher number of neurons in comparison with NN_A due to its improved performance. The higher regression value close to 1 and the functioning of the NN_{Prop} model at lower number of neurons makes it suitable for audio classification.

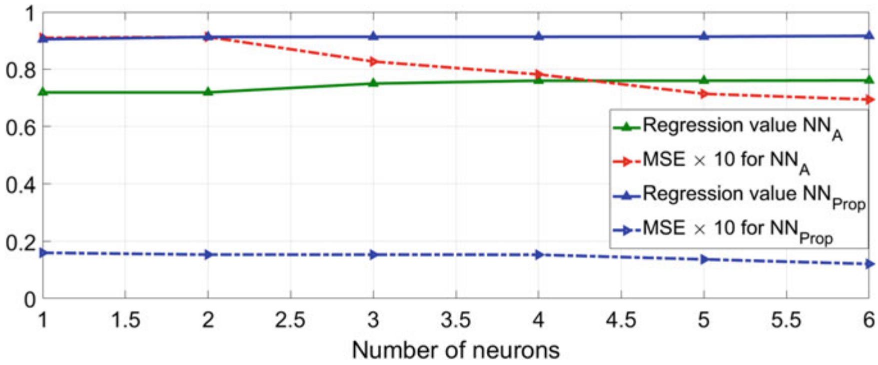


Fig. 4 Variation in performance of NN_{Prop} and NN_A as a function of number of neurons

5 Conclusion

In this work, a seven-stage audio classifier is proposed for voiced, unvoiced, transition, multi-speaker, music, background noise and silence signal classification using neural network by employing LM algorithm that uses a set of 32 features as input. Results are encouraging as an accuracy of 94% is obtained with the proposed model. In addition to the easiness in developing neural network model compared to the classifier based on statistical methods, the capability of proposed model to function at reduced number of neurons lowered the complexity. The inclusion of multi-speaker class will improve the design of the speech codec especially under teleconferencing mode. Our future work focuses on deep learning approach for classification to improve the accuracy further.

References

1. Atal, B.S.: The history of linear prediction. IEEE Signal Process. Mag. **23**(2), 154–161 (2006)
2. Spanias, A.S.: Speech coding: a tutorial review. Proc. IEEE **82**(10), 1541–1582 (1994)
3. Chu, W.C.: Speech Coding Algorithms: Foundation and Evolution of Standardized Coders. Wiley (2004)
4. Recommendation 3GPP TS 26.441 Codec for Enhanced Voice Services (EVS): General Overview, 3GPP, Sept 2014
5. Li, Z., Xie, Z., Wang, J., Grancharov, V., Liu, W.: Optimization of EVS speech/music classifier based on deep learning. In: 14th IEEE International Conference on Signal Processing (ICSP), pp. 260–264 (2018)
6. Ghiselli-Crippa, T., El-Jaroudi, A.: Voiced-unvoiced-silence classification of speech using neural nets. In: IJCNN-91-Seattle International Joint Conference on Neural Networks, vol. 2, pp. 851–856 (1991)
7. Huiqun, D., O'Shaughnessy, D.: Voiced-unvoiced-silence speech sound classification based on unsupervised learning. In: 2007 IEEE International Conference on Multimedia and Expo, IEEE (2007)

8. Basterrech, S., Mohammed, S., Rubino, G., Soliman, M.: Levenberg—Marquardt training algorithms for random neural networks. Comput. J. **54**(1), 125–135 (2011)
9. Atal, B., Rabiner, L.: A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition. IEEE Trans. Acoust. Speech, Signal Process. ASSP-24, 201–212 (1976)
10. Qi, Y., Hunt, B.R.: Voiced-unvoiced-silence classifications of speech using hybrid features and a network classifier. IEEE Trans. Speech Audio Process. **1**(2), 250–255 (1993)
11. ITU-T: G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP) (2007)
12. Recommendation, I. T. U. T. G.: 720.1: Generic sound activity detector, ITU-T (2010)
13. Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S., Dahlgren, N.L., Zue, V.: TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Linguistic Data Consortium, Philadelphia (1993)
14. Hu, Y., Loizou, P.: Subjective evaluation and comparison of speech enhancement algorithms. Speech Commun. **49**, 588–601 (2007)
15. Boersma, P., Weenink, D.: Praat: Doing Phonetics by Computer, Version 6.0.40. (2018)
16. Cannam, C., Landone, C., Sandler, M.: An open source application for viewing, analysing, and annotating music audio files. In: Proceedings of the ACM Multimedia 2010 International Conference, Firenze, Italy, Oct 2010, pp. 1467–1468 (2010)

Predicting Short-Term Electricity Demand Through Artificial Neural Network

Amelec Viloria, Jesús García Guliany, Noel Varela, Omar Bonerge Pineda Lezama, Hugo Hernández Palma, Lesbia Valero and Freddy Marín-González

Abstract Forecasting the consumption of electric power on a daily basis allows considerable money savings for the supplying companies, by reducing the expenses in generation and operation. Therefore, the cost of forecasting errors can be of such magnitude that many studies have focused on minimizing the forecasting error, which makes this topic as an integral part of planning in many companies of various kinds and sizes, ranging from generation, transmission, and distribution to consumption, by requiring reliable forecasting systems.

Keywords Primary feeder · Demand short-term electricity prognosis · Neural networks · Forecast accuracy

A. Viloria (✉) · N. Varela · F. Marín-González
Universidad de la Costa, St. 58 #66, Barranquilla, Atlántico, Colombia
e-mail: aviloria7@cuc.edu.co

N. Varela
e-mail: nvarela2@cuc.edu.co

F. Marín-González
e-mail: fmarin1@cuc.edu.co

J. G. Guliany
Universidad Simón Bolivar, Barranquilla, Colombia
e-mail: jesus.garcia@unisimonbolivar.edu.co

O. B. Pineda Lezama
Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras
e-mail: omarpineda@unitec.edu

H. Hernández Palma · L. Valero
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

L. Valero
e-mail: lvalero@ul.edu.co

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_14

1 Introduction

Electricity demand forecasting is basically defined as the science or art of predicting the future load on a given electric system for a specific future time period. This prediction can range from a few minutes for operational purposes to years for planning purposes [1]. Due to the importance of predicting daily load, different forecasting methods have been developed, including methods based on artificial intelligence, particularly neural networks. This research is developed using the information of demand measured by the SCADA system [2, 3] of a feeder in the city of Medellin in Colombia, taking readings every 15 min in order to create the database for designing and building an artificial neural network (ANN) that will forecast the short-term energy demand through the MATLAB tools.

2 Development

2.1 The Training Set

The training set is the set of data used by the neural network to learn the patterns present in the data, defined as the weights of the network. In network development, this dataset will represent a total of 3,125 consumption measurements of 3,850 available data.

The selected data corresponds to 48 months of the analyzed set of 50 months, corresponding to a period of 5 years of measurements [4].

2.2 Validation Set

The validation set is used for the final checking of the network, where the data used are the most recent consequent to the last value of the sample.

2.3 Test Set

The test set is selected from the database. In this case, it corresponds to the data remaining after the training patterns are selected. This selected dataset, corresponding to the last 48 consumption records in the database under study (last three months), is used to assess the network accuracy [5].

2.4 Selection of the Neural Network Architecture

According to the literature review, there are several ways to determine the architecture to develop a predictive neural network, in most cases, using networks with back-propagation training algorithm [6, 7]. In order to create a neural network, the following elements must be selected.

2.4.1 Number of Input Neurons

The number of input neurons corresponds to the measurements recorded by the SCADA system at the exit of the circuit under analysis in a time interval of one year, which in the database is equivalent to 12 months, and the consumption of each selected month would be equivalent to 24 data. In this case, the number of input neurons is 24 data.

For the development of this research, a single hidden layer was considered, which is sufficient to ensure the ability to generalize the network, so the network would have a total of three layers: the input layer, the hidden layer, and the output layer.

The number of neurons making up the hidden layer will be 75% of the total inputs [8]. The number of network entries is 24, which is equivalent to the number of hours of the day with the highest consumption of a month, of the 49 months of the database, so that the number of neurons contained in the hidden layer will be 18 neurons.

In order to decide the number of neurons to be contained in the output of the neural network, it was established to use just one output neuron, because multiple output neural networks, especially if these outputs are widely spaced, will produce lower results compared to a single output network. It is recommended to have a specialized network for each of the desired outputs in each prediction.

In this study, the number of neurons in the network output layer will be one, because the research only intends to predict the value of the selected variable for the following month [9].

2.4.2 Transfer Function

This function is intended to prevent the exits from reaching very high values, which can paralyze the network and stop its training. The sigmoidal function (whose output range is between -1 and $+1$) was used as the transfer function, which is preferable for prediction tasks according to [7].

2.5 Evaluation Criteria

For the evaluation criteria to measure the network efficiency, the mean-squared error (MSE) was considered. The calculated mean-squared error is defined as the difference between the network output and the desired response, which is used as the completion factor of the training. For this phase, a parameter of 300 iterations was set, and the termination factor used from the MSE has a threshold value of 0.5 [10].

The network training uses the descending gradient technique immersed in the back-propagation algorithm. The neural network training will stop when any of the following variants occur:

1. The programmed number of iterations is reached.
2. If the evaluation function (MSE) falls below the stated goal.
3. If the error measured by the evaluation function is increased for a specific number of iterations (the latter case requires the existence of the validation set). In any case, the weights obtained are those found in the minimum error measured by the evaluation function [11].

2.6 Implementation of the Prediction Model with Neural Networks

In order to implement the proposed model, the procedure summarized in Table 1 [2] was carried out. In the table, the structure of the implementation and functioning of the neural network developed in MATLAB can be interpreted.

Figure 1 shows the training of the network once iteration 300 was reached, which

Table 1 Summary of data from the developed neural network

Pre-processed database	3125 measurements
Variable	kW
Data frequency	Monthly
Number of iterations	400
Number of data for neural network training	3.850
Number of data for neural network test	48
Number of input neurons	24
Number of hidden layers	1
Number of neurons hidden in each layer	18
Number of output neurons	1
Transfer function	Sigmoidal function
Error function	Quadratic mean error

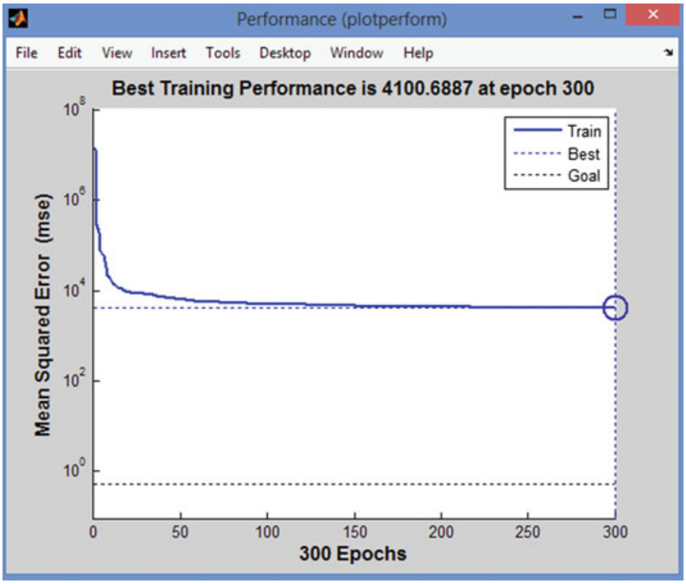


Fig. 1 Training performance

was the maximum number of iterations allowed, as the training stop condition. It is necessary to save the already-trained network to use it in the calculation of the error level. As the final data on the implementation of the neural network model, the following graphical results that measure the efficiency of the network are presented in Fig. 1.

Figure 2 shows that the correlation coefficient is quite high ($R = 0.99014$), which is very close to 1, so it can be said that there is a high correlation between real values (abscissa) and forecasts (oriented). The blue line indicates the points at which the real values coincide with or are equal to the predicted values [12].

2.7 *Calculation of the Error of the Proposed Model*

For the calculation of the neural network error, the remaining 48 records of electricity consumption from the database are used, which correspond to the last 2 months. These data will be used to determine the error level of the forecast of demand in the medium term (months). The prediction of the next 2 months (48 data) of demand, using the selected neural network, is compared with actual demand data to validate the developed neural network model. Tables 2 and 3 present the demand values for the last two months of the database and the values obtained with the prediction made by the implemented neural network [8].

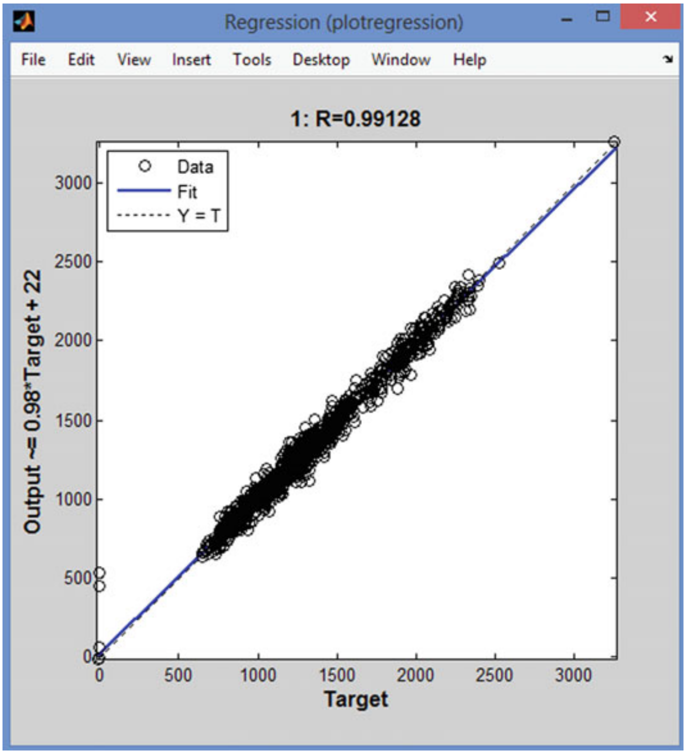


Fig. 2 Correlation between actual values and consumption forecasts

With the results obtained presented in Table 2, the behavior of the implemented neural network was evaluated and validated with a set of test data. The comparison is made between the predicted values and the actual consumption, the last two months of the database were taken as reference. Table 2 shows the accumulated forecast error for each case with a limit of 300 iterations in the network training.

3 Results

Through the validation of the neural network, it is possible to have an acceptable certainty margin to make the prediction of the possible electricity demand of the circuit under analysis for the following month by using the developed neural network. The procedure is described below.

In order to complete the forecasting, it is necessary to create a new database that includes the creation of an Excel sheet where the last month of consumption is located, in this case, the day of highest consumption of the month, which is equivalent to 24 data previous to those that will be predicted [13–16].

Table 2 Prediction of the last 48 electricity consumption data, applying the developed ANN

Month	Time	Actual consumption (kW)	Prediction (kW)	Relative error level (%)
August	1	2035	2014.41	−2.36
	2	1020	1002.23	−3.35
	3	952	958.30	+1.1
	4	952	968.11	+2.85
	5	987	1020.03	+4.36
	6	1254	1325.02	+3.58
	7	1520	1482.32	−2.8
	8	1352	1299.14	−2.5
	9	1254	1262.21	+0.68
	10	1140	1204.35	+3.5
	11	1547	1458.21	−4.3
	12	1685	1750.36	+3.85
	13	1585	1498.47	−3.54
	14	1895	1785.24	−2.4
	15	1547	1578.41	+0.9
	16	1547	1585.14	+1.1
	17	1458	1495.24	+1.3
	18	1487	1498.36	+0.5

Table 3 Next month's neural network forecast

Time	Feeder demand (kW)
1	2015.71
2	1012.33
3	959.40
4	978.71
5	1035.03
6	1347.12
7	1474.92
8	1289.14
9	1272.32
10	1214.41
11	1447.21
12	1796.33
13	1447.47
14	1736.24
15	1574.41
16	1569.15

When executing the developed code lines, the consumption forecast of the following month of the circuit under analysis is obtained. Table 3 shows the results obtained.

4 Conclusions

A forecast of daily-load consumption allows considerable money savings for the supplying companies, by generation and operation costs. In order to obtain a good result, it is necessary to reduce the forecast error to the minimum possible.

The results show the performance and accuracy of the neural networks for forecasting the demand in the short term, with minimal prediction errors.

Future research should evaluate the behavior of electricity demand considering data grouping techniques, in particular the cluster technique, to form a more compact and reliable database by eliminating atypical data and looking for a metrics of characteristic consumption curves to form the neural network database.

References

1. Hu, C., Du, S., Su, J., et al.: Discussion on the ways of purchasing and selling electricity and the mode of operation in China's electricity sales companies under the background of new electric power reform. Power Netw. Technol. **40**(11), 3293–3299 (2016)
2. Xue, Y., Lai, Y.: The integration of great energy thinking and big data thinking: Big data and electricity big data. Power Syst. Autom. **40**(1), 1–8 (2016)
3. Wang, Y., Chen, Q., Kang, C., et al.: Clustering of electricity consumption behavior dynamics toward big data applications. IEEE Trans. Smart Grid **7**(5), 2437–2447 (2017)
4. Rong, L., Guosheng, F., Weidai, D.: Statistical Analysis and Application of SAS (China Machine Press, 2011)
5. Sanchez, L., Vásquez, C., Viloria, A., Meza-Estrada, C.: Conglomerates of Latin American countries and public policies for the sustainable development of the electric power generation sector. In: Tan, Y., Shi, Y., Tang, Q. (eds.) Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science, vol. 10943. Springer, Cham (2018)
6. Sánchez, L., Vásquez, C., Viloria, A., Rodríguez Potes, L.: Greenhouse gases emissions and electric power generation in Latin American countries in the period 2006–2013. In: Tan, Y., Shi, Y., Tang, Q. (eds.) Data Mining and Big Data. DMBD 2018. Lecture Notes in Computer Science, vol. 10943. Springer, Cham (2018)
7. Perez, R., et al.: Fault diagnosis on electrical distribution systems based on fuzzy logic. In: Tan, Y., Shi, Y., Tang, Q. (eds.) Advances in Swarm Intelligence. ICSI 2018. Lecture Notes in Computer Science, vol. 10942. Springer, Cham (2018)
8. Perez, R., Vásquez, C., Viloria, A.: An intelligent strategy for faults location in distribution networks with distributed generation. J. Intell. Fuzzy Syst. Preprint 1–11 (2019)
9. Ghia, A., Rosso, A.: Análisis de respuesta de la demanda para mejorar la eficiencia de sistemas eléctricos, 2nd edn. Camara Argetina de la Construccion, Buenos Aires (2009)
10. Pérez Arriaga, J.I., Sánchez de Tembleque, L.J., Pardo, M.: La gestión de la demanda de electricidad vol. I, no. I (2005)

11. Silva, V., Jesús, A.: Indicators systems for evaluating the efficiency of political awareness of rational use of electricity. In: Advanced Materials Research, vol. 601, pp. 618–625. Trans Tech Publications (2013)
12. Perez, R., Inga, E., Aguila, A., Vásquez, C., Lima, L., Viloria, A., Henry, M.A.: Fault diagnosis on electrical distribution systems based on fuzzy logic. In: International Conference on Sensing and Imaging, pp. 174–185). Springer, Cham (2018)
13. Ozger, M., Cetinkaya, O., Akan, O.B.: Energy harvesting cognitive radio networking for IoT-enabled smart grid. Mob. Netw. Appl. **23**(4), 956–966 (2017)
14. Bradley, P., Fayyad, U., Mangasarian, O.: Mathematical programming for data mining: formulations and challenges. Informs J. Comput. **11**, 217–238 (1999)
15. Rahmani, A.M., Liljeberg, P., Preden, J., Jantsch, A.: Fog Computing in the Internet of Things. Springer, New York (2018). ISBN 978-3-319-57638-1, ISBN 978-3-319-57639-8 (eBook)
16. Abualigah, L.M., Khader, A.T., Al-Beta, M.A., Alomari, O.A.: Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. Expert Syst. Appl. **84**, 24–36 (2017)

Detection of Tomatoes Using Artificial Intelligence Implementing Haar Cascade Technique

Pabbisetty Nikhitha, Palla Mohana Sarvani, Kanikacherla Lakshmi Gayathri, Dhanush Parasa, Shahana bano and G. Yedukondalu

Abstract The twenty-first century is consumed with the automation of the world around us. Almost everything from things as easy as walking to driving a car has been automated. But according to our study, one of the least researched areas for automation has been agriculture. There are endless possibilities to the possibilities of automating the agricultural field yet many chose not to walk down this path. Machine learning is a sub-branch of artificial intelligence, it has been applied to image processing and this intelligence is demonstrated by machines in contrast to the natural intelligence displayed by humans. The artificial intelligence can be used in many sectors like transportation, finance, health care and banking and also it can be used in image processing and it helps us to implement object detection to detect and recognize the objects from our given input images and video. This is why we have dedicated this paper to help make the life of many farmers far easier. We decided on constructing low-cost agricultural robot architecture. This robotic architecture reduces the work upon farmers and helps them survey their farm area in a matter of minutes. The robotic architecture can survey the fields and assess every tomato to identify the perfect and ripe tomatoes that can be harvested. Hence, the farmers won't have to waste their time searching and can go only to the designated areas of their farm for the harvest. We were able to accomplish this with the help of Haar Cascade. We developed a model that can be run upon various servers such as Windows, Unix or even a Linux Server. We were able to train our model by identifying the images with a 'positive' and 'negative' label. All those images which contained a background without a tomato were labelled as 'negative' images, while those that contained a tomato were labelled as 'positive' images. The goal of our model implementing Haar

P. Nikhitha · P. Mohana Sarvani · K. Lakshmi Gayathri (✉) · D. Parasa · S. bano
Department of Computer, Science and Engineering, Koneru Lakshmaiah Educational Foundation, Vaddeswaram, India

S. bano
e-mail: shahanabano@icloud.com

G. Yedukondalu
Department of Mechanical Engineering, Koneru Lakshmaiah Educational Foundation, Vaddeswaram, India
e-mail: yedukondalu@kluniversity.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_15

Cascade was to create a negative image far larger than the positive images. We were able to make this possible by implementing large data sets that consisted of nearly 200 images of tomatoes. These large data sets are then further clearly segregated to identify the state of every tomato. Haar Cascade classifier provides high accuracy even the images are highly affected by the illumination. The Haar Cascade classifier has shown superior performance with simple background images.

Keywords Tomato detection · Haar Cascade · Artificial intelligence

1 Introduction

Automation has taken the world by surprise. We can see how quickly the world has welcomed automation with open hands and implemented it into various fields. Automation is simply the process of reducing human efforts in any sort of processor work. Our project focuses on implementing automation into the field of agriculture to reduce the work and efforts of many farmers. We came up with an agricultural architecture which will scan the complete farming area to identify which tomatoes are ripe enough to be harvested. On average, a farmer will have to spend more than 3–4 h a day to survey his field and look at each and every tomato to identify whether it is ready to be harvested. With our architecture, it will take the farmer no longer than 10–15 min to survey his entire farms and complete the identification process. To accomplish this, we used Cascade GUI Trainer along with PyCharm which is a part of OpenCv. Cascade GUI Trainer is an open-source software that allows people to train, test and improve their cascade classifier models. This trainer uses a graphical interface that sets the required parameters for classification. It is commonly used alongside the tools of OpenCv as they both work hand in hand to make classification far easier. OpenCv is another open-source software that is used mostly for the various libraries and packages it contains. It is famous for its ability to train the computer in order for it to gain a high-level understanding of digital images or videos.

2 Flow Chart

The GUI full form is the graphical user interface, and it was designed to integrate many functions in image processing. At first, we take the set of images as an input folder. We have to recognize the images, and they are sent for the filter for filtering the images based on their pixel size. Then, we have to train the images using the Haar Cascade method, and the images consist of red and green colour tomatoes. By using Haar Cascade technique and generate the data XML. This Haar Cascade is applicable for detecting tomatoes in both videos and images. In detecting tomatoes in images, first we will take all the images which consist of both positive and negative images. The positive images consist of images containing tomatoes, and negative

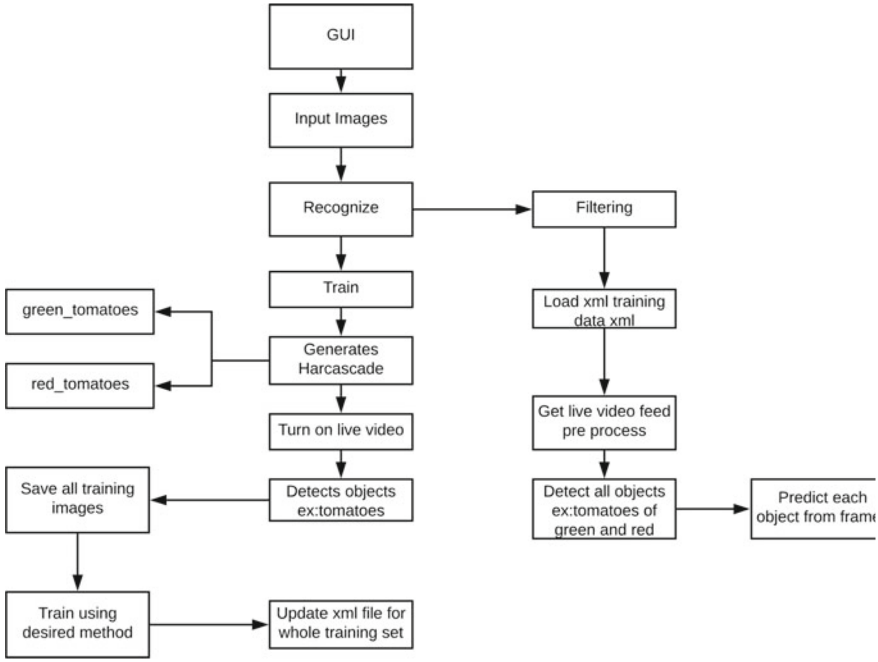


Fig. 1 Flow of diagram

images consist of images rather than tomatoes we need to save all the images and train an XML using all images by using Haar Cascade we can detect the tomatoes from a bunch of vegetables. We can predict the red and green tomatoes. This process is also applicable to live. In live video it can detect the tomatoes from vegetables and in videos we have to give input as a video which contains a different kind of red and green tomatoes and by using Haar Cascade XML generated we can detect the tomatoes flowing in the video or presentation in the video the detection of tomatoes is represented by squares around the tomatoes and by using this method we can detect all the green and red tomatoes in the video and each tomato is identified by square (Fig. 1).

3 Procedure

To create our agriculture architecture, we needed to accomplish three major stages. The first stage was to collect and asses the required data sets. Once we had gathered all the data sets that we required, we would then have to organize them as required. The second stage is to generate and train our model using Haar Cascade for it to identify tomatoes. Once we have successfully generated our model, we will move on to the final stage in which we use OpenCv to implement our model on various videos and images. Let us take a further look at each stage.

3.1 Collection of Data Set

Data sets are the most important part of training our model. Without these, it would nearly impossible to train our model. We are using a supervising method to train our model. A supervising model is a process in which the model trains upon learning from the images which already hold our target object. It assesses all these images to learn more and more about our target object from its height structure and its features such as edges and lines. Without these, our model would be clueless about what we are trying to identify.

3.2 Collection of Images

The first step is to collect as many images as we can of our target image (in our project tomatoes). We will then take the collected images and further segregate them into two different folders. We will consider the labelling of these two segregations to be ‘positive’ and ‘negative’. The positive labelled images will contain all the images that contain our target image, in which our project is the tomatoes. These images will be used to build a vector file which will be further used for detection. The negative labelled images will consist of all those images that do not contain our target image. They can be of anything such as vehicles, toys and objects as long as they do not contain our target image. For you to get a clear understanding, let us take a look at a positive and negative image. The positive image is on the left followed by negative images on the right.

We were able to observe that our model achieved a higher accuracy when we segregated these images in the ration of 1:2 for positive to negative. This is said there should be twice as many negative images as the positive images in our data set (Figs. 2 and 3).

3.3 Image Classification

Before we go further into our project, let us take a look at a few basic things on image classification. We all know that the machine’s perception of an image is completely different from what we can see. In fact, a machine can only understand numbers, therefore making it merely impossible for it to understand an image like the human mind. For a machine to understand and classify an image, it first turns the image into a two-dimensional array of its choice. From here on, each pixel in the image is

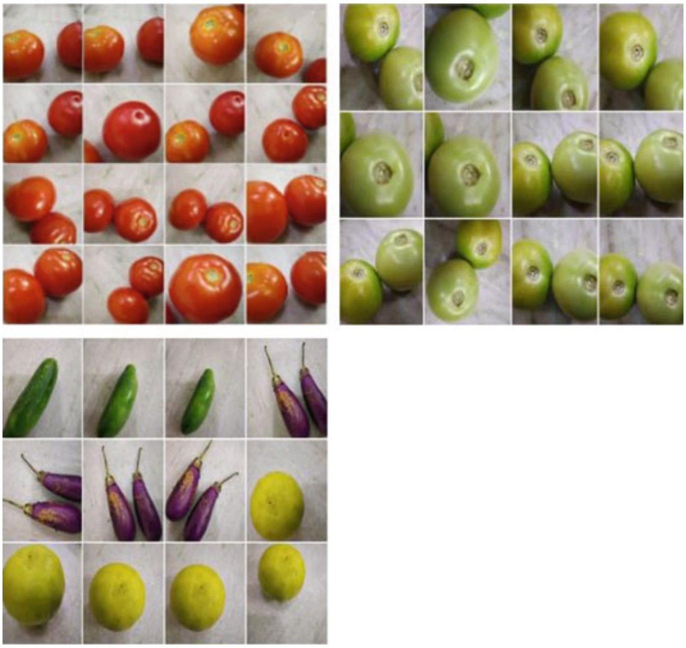


Fig. 2 Positive images

Fig. 3 Negative images

Fig. 4 What the computer sees from an image

given a value between 0 and 255 and plotted in our array. Thus, for the machine to classify an image, it requires some pre-processing to achieve features and attributes that distinguish an image.

The image classification aim is to predict a single label for the given image. The images contain a three-dimensional array of integers starts from 0 to 255, and the size of positive images should be greater than negative images. The image consists of width × height ×3; here, the 3 represents the three colour channels, i.e. red, green, blue, and these three colours are represented as RGB colour channels (Fig. 4).

3.4 Haar Cascade

In our project, we will be using the Haar Cascade GUI Trainer to generate as well as train our model. Haar Cascade is an open-source software that is used to detect objects using a machine learning technique called cascading. It can train models by using positive and negative images. As we have discussed before, our positive images will include our target object, while our negative images will include everything except our target object. Once the model is trained, it is now able to identify any object of our choice. Haar Cascade uses three major stages to train as well as generate our model. Let us take a look at these three stages within Haar Cascade.

3.5 Haar Features

Haar features are the black and white boxes inside the target window that you can observe. These windows are known as the frame checkers which check each and every frame of the image to identify whether our object is present within the frame or not. Haar feature algorithm needs a lot of positive and negative images to train the classifier. It contains rectangular regions in the detection window. It sums up the

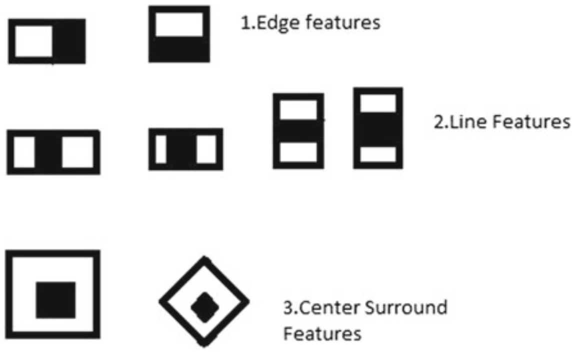


Fig. 5 Haar features

pixel intensities in each and every region and calculates the difference between the sums. It consists of edge features, line features and four rectangle features. Here are a few of the windows that haar features use (Fig. 5).

3.6 Algorithm: AdaBoost Classifier

In this project, we have used the Haar Cascade technique which works on AdaBoost algorithm, and in this, we have trained a data set that takes the images as an input and processes them with the help of AdaBoost classifier, which has high detection rates. It is an effective learning algorithm that helps us to produce a classifier by selecting the required features in the image.

The AdaBoost classifier plays a major role within the cascading process. It analyses the frames that contain our target object in order to further learn about it. With this analysis, it can select all the important features of our target object and reject any of the unnecessary features that it may include. This algorithm created a 'strong' classifier by creating a linear combination of weighted 'simple' classifiers. It can create this linear combination of weak classifiers until our model is able to identify our target images with a high accuracy rate.

3.7 Integral Images

When we run our Haar Cascade to train our model, we are required to input image size. This is the size of the frame that will be used to compare each and every possible frame in our data set. This frame will be moved over each and every image until the complete image is covered. Depending on the size of our image, we will take into consideration the size of our integral image. The smaller the target image the smaller our frame should be and vice versa.

3.8 Generating Model Using Haar Classifier

Now that we have gone through the three major stages used within Haar Cascade, let us take a look at the procedure to create our own model using this handy tool. Our model is constructed and saved in the form of an XML file. To begin our process, we must first note down the path of our data set. Keep in mind that the images in our data set should be neatly segregated into two folders based on their ‘positive’ or ‘negative’ label. We then place this path within the Haar Cascade software and set the parameters to common. The software will now ask us to enter the number of stages we would like to follow in creating and training our model. The number of stages we select the more accurate our model will become. However, the downside to selecting more stages is that it will take a far longer time to train our model. We observed that our accuracy did not increase with much difference after we trained it with—therefore, we selected—as the number of stages to train our model. Before our software begins to train our model, we will be left with one final step.

It will ask us to input the width and height of the frame that you would like to use while training your model. The height and width of the frame are decided upon the size of the target object you are dealing with. As tomatoes are fairly medium in size we decide to take the height and width of our frame as 32 × 32. Once you have entered the window size, Haar Cascade will begin training your model and create a log file to show you each and every iteration (Fig. 6).

```
numPos: 94
numNeg: 1000
numStages: 15
precalcValBufSize[Mb] : 1024
precalcIdxBufSize[Mb] : 1024
acceptanceRatioBreakValue : -1
stageType: BOOST
featureType: HAAR
sampleWidth: 32
sampleHeight: 32
boostType: GAB
minHitRate: 0.995
maxFalseAlarmRate: 0.5
weightTrimRate: 0.95
maxDepth: 1
maxWeakCount: 100
mode: BASIC
Number of unique features given windowSize [32,32] : 510112
```

Fig. 6 Log file

3.9 Using PyCharm to Demonstrate Classifier

In this project, we were able to implement the software known as PyCharm which is one of the platforms of OpenCv developed by Google. PyCharm comes with many inbuilt libraries and packages that allow developers to enhance their models. PyCharm is an integrated development environment. It is a software environment to write programmes. It combines all the features of the software needed by the developer. In PyCharm, we use Python since all libraries are open source. In PyCharm, we use Python since all libraries are open source, PyCharm can be used as cross-platform IDE for Python. It is compatible with all operating systems like Windows, Linux and Mac OS. PyCharm helps developers to write the programmes in Python more effectively and quickly.

3.10 Important Packages in Our Programme

There are two important packages that we have imported and used within our programme. These two packages are:

I. **CV2**

The first library that we will be using is CV2. This package includes everything that we require by returning objects such as NumPy objects (like an array) and native Python objects (like lists, tuples, dictionary, etc.). As NumPy is included within our CV2 library, this allows us to perform various mathematical functions on matrices. NumPy is a highly stable and fast array processing library. For example, if you load an image, an array is returned in which array[i, j] gives you the pixel value at (i, j) position.

II. **HSV**

HSV is a colour picker, in which it describes the way the colours are combined to create the spectrum we see. Using this method, an object with certain specified colour is detected and it reduces the influence of intensity of light from outside it is closer to the way in which how humans perceive colours.

HSV (hue, saturation, value),where Hue in HSV is the colour representation part of the model and all the colours fall in between the range of (0–360). Saturation is that it indicates the amount of grey present in a particular colour in a range of 0–100% reducing it towards zero increases the level of greyness it can also be represented using 0 and 1, where 0 is grey and 1 as a primary colour. Value is nothing but brightness, and it works with saturation and describes the intensity or brightness level in the range of 0–100, where 0 is completely black and 100 is the brightest colour.

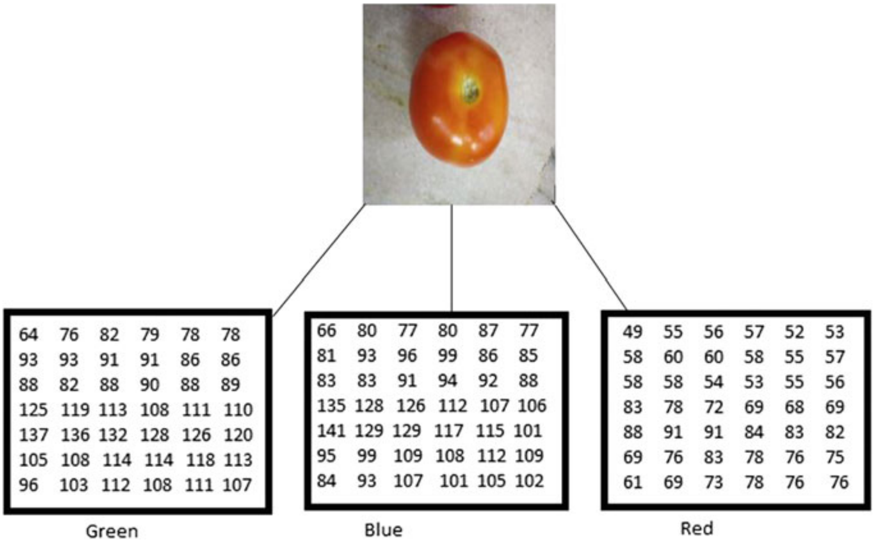


Fig. 7 RGB values

3.11 Important Function Calls in Our Programme

I. **RGB**

An image is usually represented using discrete values. A digital image can be represented in the form of a two-dimensional matrix as $f(x, y)$ in which it contains n rows and m columns in RGB, and the red, green and blue colours were added together in varying proportions which produce an extensive range of the colours. RGB method is widely successful mostly used in sensor and image processing applications (Fig. 7).

II. **BGR TO GREY**

To convert an image to BGR, we use imread function which has channels stored in BGR order by default. Next, to convert the colour of an image we use cvt. colour function. At first, it will receive an input image. To convert BGR to GREY we use COLOUR_BGR2GRAY function. Greyscale is simply reducing complexity from 3D pixel value to a 1D value. Greyscale images are more suitable for certain applications like image processing when the RGB format and our application need a greyscale image, then we use the RGB to grey conversion (Figs. 8 and 9).

Gray = cv2.cvtcolour(image,cv2.Colour_BGR2GRAY(1)

By using BGR to RGB, we can get the output as follows (Figs. 10 and 11).

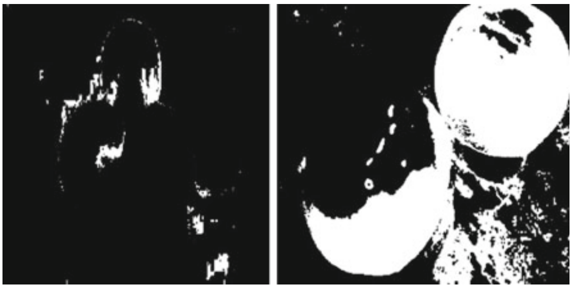
In our project, we capture the video frame by frame to check whether the desired object is present within the given image or video. To do this, we are using live

Fig. 8 Original image

Fig. 9 Converted image

Fig. 10 Input images

Fig. 11 Detection using edge features



detection in this project in which it checks the object from the live video and detects the object to check from the live video we need to use the command

$$\text{Cap} = \text{cv2.videocapture}(0)$$

where the '0' indicate the camera of your system, and you can even use images or source video like giving the video or image as an input in the code.

$$\text{Cap} = \text{cv2.videocapture(video-src)}$$

III. **IMREAD**

To read the given input source image, the imshow function can be used, it helps the user to read the original image and generate the output based on this which is also part of cv2 package and the channels of imshow were stored in BGR (blue, green and red)

$$\text{Image} = \text{cv2.imread('c:/user/Desktop')}$$

IV. **IMSHOW**

The imshow function helps us to display the images. This function contains two inputs in which the first contains a string which is the name we want to assign our window followed by the name of the image to show.ds

$$\text{Cv2.imshow}('\text{originalimage}',\ \text{image})$$

V. **WAITKEY**

The waitkey function is one of the most useful functions, and it helps the user to delay the output window for n milliseconds; if the time is declared as (0), it stays on the screen for an infinite amount of time until a keyboard event occurs if it is given as (10), it stays on the screen for 10 ms.

$$\text{Cv2.waitkey}(\text{delay} \Leftarrow 0)$$

4 Results

See Figs. 12 and 13.

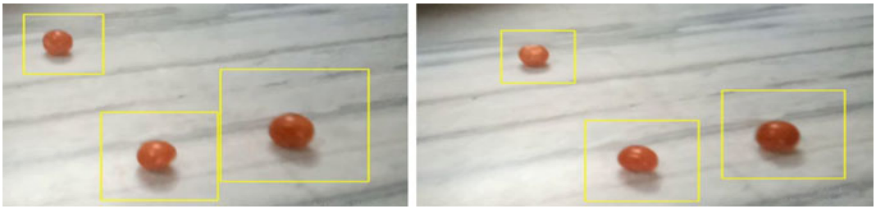


Fig. 12 Detection of tomatoes in live video

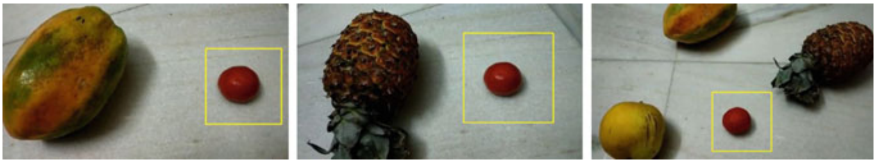


Fig. 13 Detection of tomatoes from different objects in live video

5 Conclusion

In this project, we have used the method of Haar Cascade in the field of computer vision in which it trains the model by giving a pre-processed data set consisting of positive and negative images. It produces an XML cascade, as the model gets ready it can detect the specified object we need by checking the frames using AdaBoost classification technique in this, we have created our custom data set which can be used in real-time applications. We have given the video sequence consisting of different objects, in which we need to track the required object. It was developed with the scope that it can be useful for the farmers in the field of agriculture, and as the workload can be reduced, it is just an initial stage we would like to extend the process using robotic vision techniques.

6 Future Scope

By using this Haar Cascade technique which implements using artificial intelligence, we can predict the required object we need. If we implement this to a robot, then we can train the robot in such a way that it can detect the ripe tomatoes from the plant and pluck it in large amounts within a less interval of time. With this technique, it reduces the stress and effort to the farmers.

Passive Safety System for Two- and Four-Wheeled Vehicles

Vatsal Mehta, R. Ujwal, Rakshith Narun, Savan Vachhani and Babu Rao Ponangi

Abstract Multiple passive safety systems such as helmet and seatbelt protection on two- and four-wheeled vehicles are one of the primary concern in today's life. In this work, physical sensors have be embedded in the helmet and seat belt which will detect the human presence and control the ignition of the engine through a microcontroller. This system is connected to a microcontroller using a short-distance wireless module to transmit a signal to the receiver connected to the ignition system of the vehicle. Ignition in the engine takes place only if the sensor detects a human being wearing the helmet in two-wheelers or seatbelt in four-wheelers. It has an in-built lock system, placed near the engine which would inform the rider to wear the helmet for riding the two-wheeler or to the driver to wear the seatbelt for a four-wheeler. In case of accidents, these systems would prevent the driver from succumbing to injuries due to the absence of a helmet or the seatbelt. The helmet is wireless hardware, which is powered by a rechargeable battery. By using these systems in vehicles, the safety of the rider/driver can be improved by a great margin.

Keywords Seatbelt · Helmet · Microcontroller · Sensor

1 Introduction

In recent times, there have been a lot of fatalities due to motor vehicle accidents. These were caused because of ignorance of rules or lack of protective gear worn by the rider/driver. Hence, a few systems have been developed to improve the safety of the rider/driver such as smart helmet [1] where features like alcohol detection, accident identification, location tracking, use as a hands free device, fall detection

V. Mehta (✉) · S. Vachhani
Department of Mechanical Engineering, PES University, Bangalore, India

R. Ujwal · R. Narun
Department of ECE, PES University, Bangalore, India

B. R. Ponangi
Department of PG Studies, Mechanical Engineering, PES University, Bangalore, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_16

have been implemented; a GSM-based helmet which sends a text message to family members of the rider in case of an accident [2]; helmet which detects if the rider is wearing a helmet and if the rider is drunk [3]; a four-wheeled vehicle with driver-assist systems, collision warning and collision avoidance systems to increase safety of driver and pedestrian has also been researched [4]; a smart helmet with GSM and GPS technology to inform the loved ones about the accident of the rider with the location of the rider [5]. Currently, there are not many effective and redundancy-free systems which would prevent the ignition from turning on when a helmet/seatbelt is not worn. This paper aims to do the same and hence increase the safety of an automobile driver with this system, by reducing injuries during accidents.

2 Literature Survey

Nowadays, wearing helmets in India has become mandatory. There has been an increase in the number of traffic accidents year by year. According to section 129 of Motor Vehicle Act 1988, it a compulsory for every individual who is riding a bike to wear helmet following standards of the Bureau of Indian Standards. Riding a bike without a helmet in India is an offence according to Act of 1988. Helmet acts as a safety device during accidents. On multiple occasions, a person tends to forget his/her helmet in his/her homes. To avoid this, the system proposed does not let the vehicle start if the helmet is not worn. Rewanth et al. [6] explain about multiple active safety systems like ABS in four-wheelers and stability program being used in two-wheelers. To prevent the rider from forgetting or neglecting wearing a helmet, this system makes it compulsory for a two-wheeler, while riding past a specified speed limit to wear it. Also, the system helps prevent drunk and driving cases. Mohd et al. [7] had worked on a project which demonstrated that motorcycle's engine ignition will turn on when a helmet is worn, and the belt has been buckled. They used LED which would flash when the vehicle's speed exceeds 100 kmph, to alert the rider. Chitte et al. [8] had worked on a smart helmet and intelligent bike system, which uses an RF module as a wireless link for communication between transmitter and receiver. From their experimental set- up, they concluded that ignition of the bike's engine will start if the helmet is worn. Using Arduino lily pad they had controlled system and sensors. They also highlighted that when an accident occurs, using GSM message will be sent to their registered number and current location will be shared by the GPS module. Grassi et al. [9] have worked on a concept for powered two-wheeled vehicles where the injuries to a rider during an accident are drastically reduced with the help of a belted safety jacket.

There are many safety systems in a four-wheeled vehicle such as seatbelt pre-tensioners, adaptive cruise control (ACC), traction control system (TCS), alarm system, and blind-spot warning system. The seatbelt pre-tensioners are used for tightening or reducing slack of seatbelt, to protect a person from sudden movement during hard braking. Four-wheeler safety alarm system provides complete protection against lifting or unlocking of the vehicle. Adaptive cruise control or ACC is used to detect the distance and speed of the preceding vehicle to maintain an appropriate distance. Optimization of grip and stability of car on the road during acceleration is done by traction control system (TCS) by measuring the wheel rotations. Blind-spot warning system helps in detection and warning of vehicles which cannot be seen by the driver. Dong et al. [10] have studied on driver inattention monitoring system for intelligent vehicles which discusses the use of an active safety system to check for the driver's inattentiveness. The method adopted for doing so is image-based inattention detection by using visible spectrum camera, IR camera, and a stereo camera. They have also discussed about a warning interface that would be able to draw the driver back from an inattentive state to an attentive state. Chen et al. [11] have proposed a system on Differential Braking 3 based Rollover Prevention for Sport Utility Vehicles with Human-in-the-loop Evaluations by using an anti-rollover control algorithm based on the Time-To-Rollover (TTR) in real-time with steering and direct yaw moment control inputs. The results they achieved were that the anti-rollover control system significantly improved performance for all test tracks but in a human-in-the-loop experiment, because of the steering pattern of the real driver being much noisier than that of the UMTRI driver model, the performance degraded. Jeppsson et al. [12] have worked on a system to reduce car-to-pedestrian contact with the help of rapid deceleration using Vacuum Emergency Braking. Using projection profiling, the presence of a helmet on the motorcyclist's head was checked. Waranusast et al. [13] have proposed a real-time system on Machine Vision Techniques for Motorcycle Safety Helmet Detection. This paper presents a system which would automatically detect motorcycle riders and determine whether the rider is wearing safety helmets or not. They have adopted K-Nearest Neighbor (KNN) classifier to differentiate a motorcyclist from a non-motorcyclist. Using projection profiling, the presence of a helmet on the motorcyclist's head was checked. Sivaraman et al. [14] have introduced a basic framework based on an active learning algorithm for on-road robust vehicle tracking and recognition. The system is passively trained using supervised learning for a recognition system. The next round of learning is based on active vehicle recognition system. This framework is used to detect vehicles and people around the vehicle. This data is used for having a safe vehicle journey even on a busy road. Puthan et al. [15] have discussed about active and passive safety passenger car technologies, which could be used to save lives in India. They have also discussed about the effectiveness of each safety technology.

3 Technical Studies

3.1 Arduino Nano Microcontroller

Arduino Nano is a small, complete and breadboard-friendly board based on the ATmega328 or ATmega168 equipped with digital and analog I/O pins that can be interfaced to various expansion boards as seen in Fig. 1. It can be powered via the

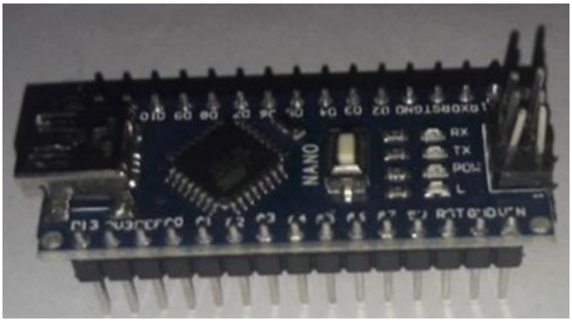


Fig. 1 Arduino Uno microcontroller

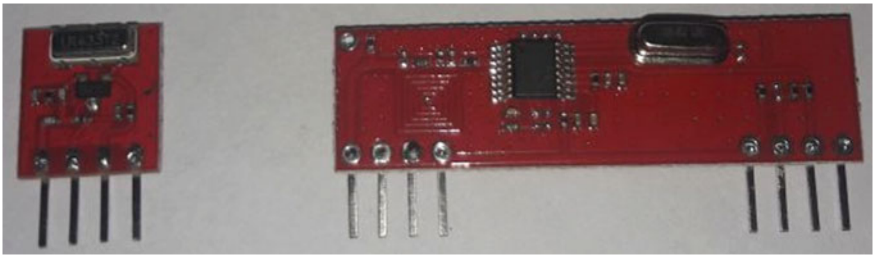


Fig. 2 RF transmitter and receiver

Mini-B USB connection, 6-20V unregulated external power supply or 5V regulated external power supply.

3.2 RF Communicator

RF module is a small electronic device which is used to transmit and/or receive radio signals between two devices. RF communications incorporate a transmitter and a receiver where a transmitter is used to transmit a RF signal and a receiver is used to receive a RF signal as seen in Fig. 2.

3.3 LM35 Temperature Sensor

The LM35 is an integrated circuit sensor which is used to measure temperature with an electrical output proportional to the temperature (in °C) as seen in Fig. 3. Precision of this sensor, in measuring temperature, is more than that of a thermistor. The LM35 maintains a very good accuracy. LM35 draws few micro amps from its supply and possesses a low self-heating capability, therefore making it efficient.

Fig. 3 LM35 temperature sensor

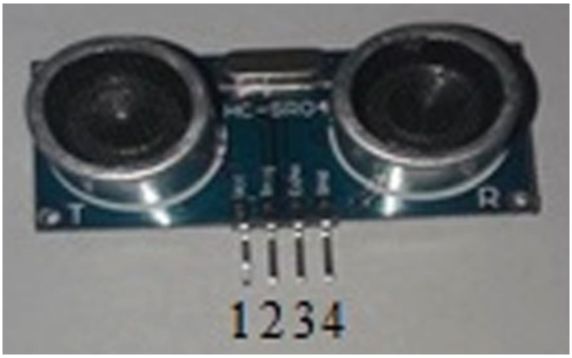


Fig. 4 HC-SR04 ultrasonic sensor

3.4 Ultrasonic Sensor

HC-SR04 Ultrasonic sensor is used to measure the distance to an object using sound waves as seen in Fig. 4. This is achieved by measuring the elapsed time between the sound wave being generated and the sound wave bouncing back. This sensor emits an ultrasound at 40,000 Hz, which travels through the air and if there is an object or obstacle on its path, the sound waves will bounce back to the module. Taking into account the travelling time and the speed of the sound waves, the distance is calculated.

3.5 Radio Frequency Identification (RFID)

RFID uses an electromagnetic field to automatically identify and track tags, contain electronically-stored information attached to the objects as seen in Fig. 5. Passive tags collect energy from a nearby RFID reader's interrogating radio waves. Unlike

Fig. 5 Radio frequency identification

a barcode, the tag need not be within the line of sight of the reader, hence it is embedded in the seat. An RFID system consists of a transponder which is located on the seatbelt that is used to identify the signal and a transceiver, which is fixed on the seat. When the seatbelt is near the electromagnetic field of the seat, due to induction, a voltage is generated in the reader's antenna coil and this voltage serves as power for the microchip.

4 Construction

4.1 Two-Wheeler Unit

There are two sensors on the helmet, which are used to detect if the helmet is completely worn by the rider. The sensors are connected a microcontroller for reading the data. An RF transmitter is connected to the microcontroller for wireless communication of the microcontroller with the bike. This is complete wireless device from the bike; hence, a battery is connected the power the entire circuit. On the bike

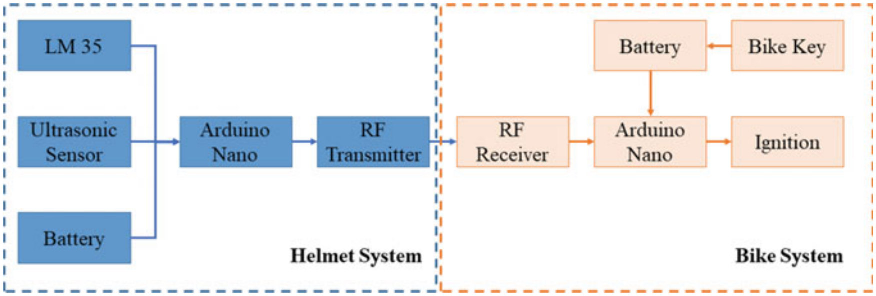


Fig. 6 Block diagram of two-wheeler passive safety system

side, the microcontroller is connected to RF receiver to read the data sent from the helmet. The controller is connected to the ignition system to control it based on the data received from the helmet. This circuit is connected to the battery unit of the bike (Fig. 6).

On the helmet, a temperature sensor and a distance sensor are also fixed. The temperature sensor is present to read the presence of a human by measuring the temperature of the person wearing the helmet. The distance sensor is necessary to understand how far the helmet is worn by the user. The distance sensor and temperature sensor in unison are used to detect if a person has completed wearing the helmet. The sensors will give the data to the microcontroller. The microcontroller will act as the deciding system to understand the data and deciding if someone is wearing the helmet or not. The microcontroller will then send OK/Not OK signal based on the sensor data to the bike using RF communication. The bike system after reading the data from the helmet will decide if it has to allow the bike to start or not. If the OK signal is read, then the bike system allows the ignition to happen, otherwise it will disable the ignition system (Fig. 7).

4.2 Four-Wheeler Unit

There are two sensors in this system to detect if the user is wearing the sensor. Microcontroller is used to understand the sensor data and it is powered by the car's battery unit. The controller is connected to the ignition system in the ECU (Fig. 8).

In this safety system, a proximity sensor is used to detect if the seatbelt is in proximity to the seat. The microcontroller will first detect if the seatbelt is locked in by checking the seatbelt lock integrated in the car. If the seatbelt is locked in, then the system will detect the seatbelt's proximity to the seat. The proximity detector is a RFID reader. There is a RFID tag on the belt. If the belt is in close proximity to the seat, which means that RFID reader is able to read the tag, which would mean that the belt is not worn over the body. The microcontroller will disable the ignition system of the four-wheeler until the seatbelt is worn by the user (Fig. 9).

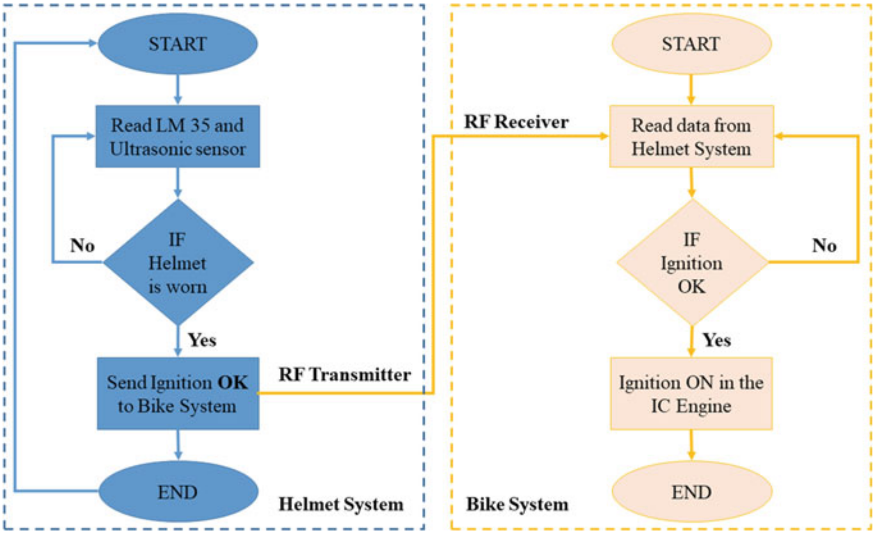
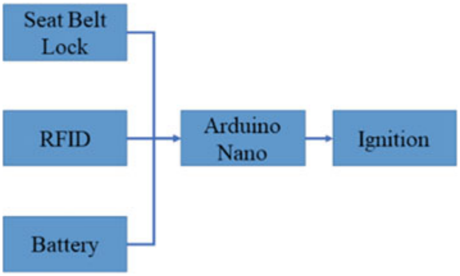


Fig. 7 Flowchart of two-wheeler passive safety system

Fig. 8 Block diagram of four-wheeler passive safety system



5 Results and Discussion

As seen in the pictures below, the safety system gets activated when the helmet is not worn by the rider or the seatbelt is not worn by the driver. Hence, the aim of this project is achieved theoretically and experimentally.

Case 1. Two-wheeler passive safety system on a helmet (Figs. 10 and 11).
Case 2. Four-wheeler passive safety system on a seatbelt (Figs. 12 and 13).

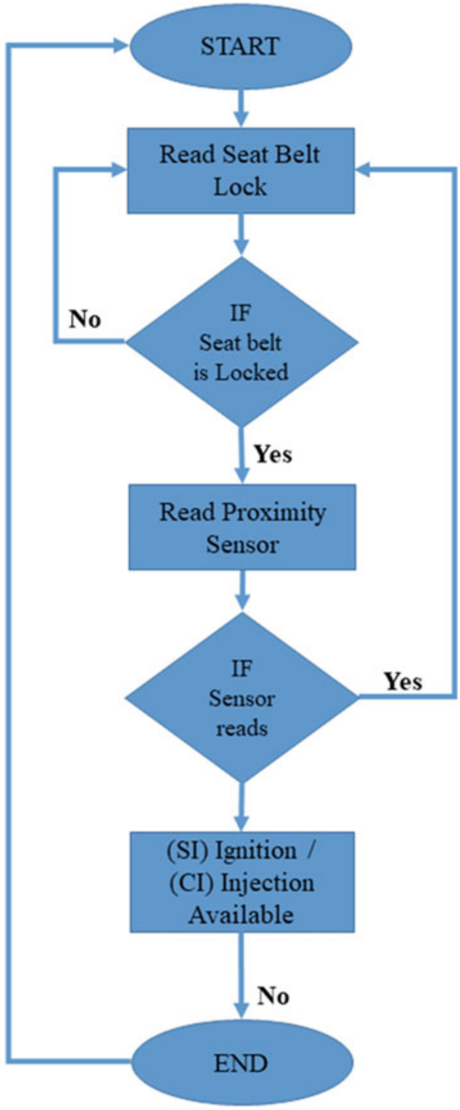


Fig. 9 Flowchart of four-wheeler passive safety system

6 Conclusion and Future Scope

In this paper, the issue of safety of the rider/driver during an accident has been approached with a self-developed algorithm, which is optimized to work efficiently and redundant stages have been eliminated. There have not been currently many approaches to invent a system as such, whereas there has been a requirement for one for a very long time. As discussed in sub-sections, the flowcharts above represent the approach taken to resolve this modern-day issue. By displaying the approach to

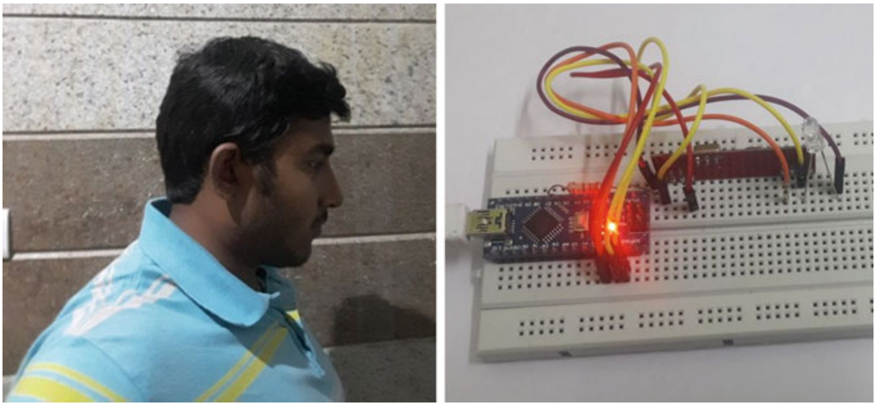


Fig. 10 Helmet not worn by the rider. Therefore, no ignition command is given to the bike's ECU

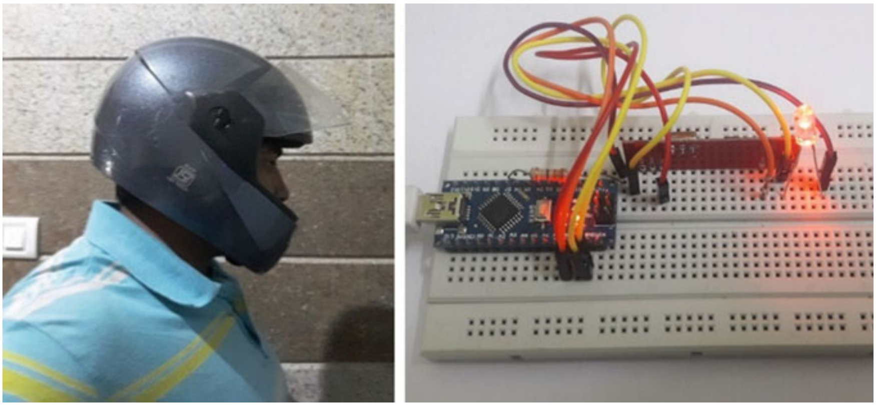


Fig. 11 Helmet worn by the rider. Therefore, ignition command is given to the bike's ECU, denoted by glowing LED

the redundancy with these concepts, it can be concluded that it can be very effective and has the potential to save many lives.

Based on the current specifications of a helmet, fitting such a circuit would be challenging unless it is miniaturized. Therefore, that would be the next aim. Testing it in the real world to verify its functionality would also have to be taken into consideration.

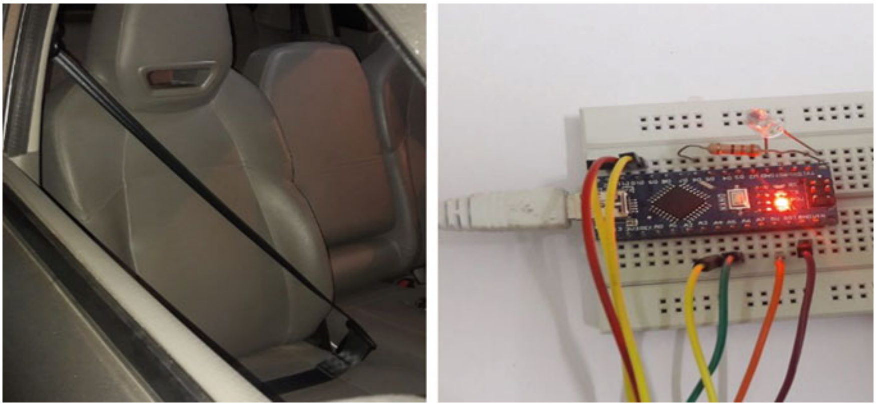


Fig. 12 Seatbelt locked in but no driver present between seat and seatbelt. Therefore, ignition command is not given to the ECU (If seatbelt is not locked in, same results are obtained)

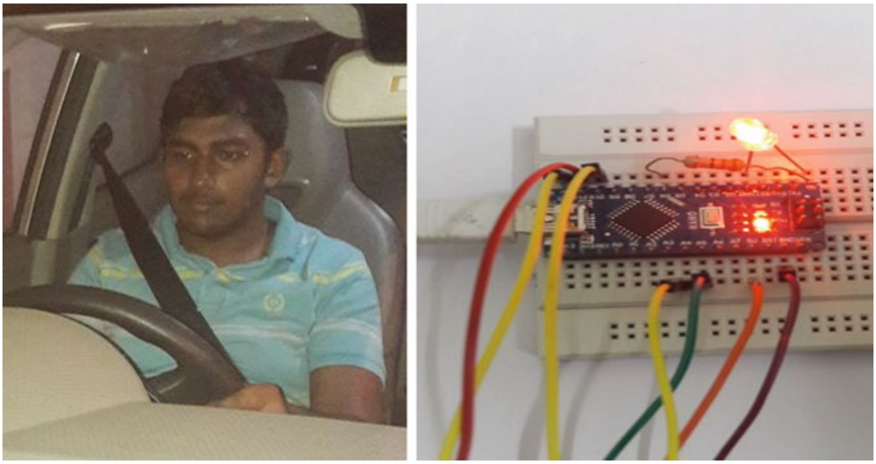


Fig. 13 Seatbelt locked in and driver present between seat and seatbelt. Therefore, ignition command is given to the ECU, represented by glowing LED

Compliance with Ethical Standards All author states that there is no conflict of interest. We used our own data. In result section, Author's photograph is only used.

References

1. Shravya, K., Yamini, M., Keerthi, D., Harika, K., Ranjan, K.: Smart helmet for safe driving. In: E3S Web of Conferences (2019)

2. Abhilash, D., Sarthak, A.: Smart helmet. Major project for the completion of B.Tech, submitted at VNS Group of Institutions (2015)
3. Tintu, G., Rajashree, M.K.: Intelligent Helmet for Road Safety. Major project for the completion of B.Tech, submitted at Dr. M V Shetty Institute of Technology
4. Smita, D., Shreya, D.: Smart vehicle automation. Int. J. Comput. Sci. Mobile Comput. (IJCSMC) **04**(09) (2017)
5. Vinod, G.V., Sai Krishna, K.: Smart helmet. Int. J. Eng. Sci. Res. Technol. (IJESRT) **07**(04) (2018)
6. Rewanth, R., Sakthi Ganesh, M.: Passive safety systems for bike via helmet. Int. Res. J. Eng. Technol. (IRJET) **04**(03) (2017)
7. Mohd, R.M.K.A.M., Madzhi, N.K., Johari, J.: Smart helmet with sensors for accident prevention. In: International Conference on Electrical, Electronics and System Engineering (2013)
8. Chitte, P.P., Salunke Akshay, S., Thorat Aniruddha, N., Bhosale Nilesh, T.: Smart helmet & intelligent bike system. Int. Res. J. Eng. Technol. (IRJET) **03**(05) (2016)
9. Grassi, A., Barbani, D., Baldanzini, N., Barbieri, R., Pierini, M.: Belted safety jacket: a new concept in powered two-wheeler passive safety. In: AIAS 2017 International Conference on Stress Analysis (2017)
10. Dong, Y., Hu, Z., Uchimuraand, K., Murayama, N.: Driver inattention monitoring system for intelligent vehicles: a review. IEEE Trans. Intell. Transp. Syst. **12**(2) (2011)
11. Chen, B.-C., Peng, H.: Differential-Braking-Based Rollover Prevention for Sport Utility Vehicles with Human-in the-loop Evaluations
12. Jeppsson, H., Östling, M., Lubbe, N.: Real life safety benefits of increasing brake deceleration in car-to-pedestrian accidents: simulation of vacuum emergency braking. Accid. Anal. Prev. (2018)
13. Waranusast, R., Pattanathaburt, P.: Machine vision techniques for motorcycle safety helmet detection. In: 28th International Conference on Image and Vision Computing, New Zealand (2013)
14. Sivaraman, S., Trivedi, M.M.: A general active-learning framework for on-road vehicle recognition and tracking. IEEE Trans. Intell. Transp. Syst. **11**(02) (2010)
15. Puthan, P., Thalya, P., Lubbe, N.: Active and passive safety passenger car technologies: potentials to save lives in India. IRCOBI Asia (2018)

Measuring the Financial Performance of MSMEs Through Artificial Neural Networks

Jesus Silva, Lissette Hernandez, Ana Emilia Hernandez, Noel Varela, Hugo Hernández Palma, Osman Redondo Bilbao, Nadia Leon Castro, Ronald Prieto Pulido and Jesús García Guliany

Abstract Given the importance of micro, small and medium-sized enterprises (MSMEs) in Colombia, both in terms of the number of enterprises and the generation of employment, it is important to know their nature, as well as the main determinants of their financial performance. In this sense, this paper aims to provide relevant information for the formulation of strategies, programs and public policies that promote practices within companies and thus improve the performance of this segment of organizations.

J. Silva (✉)
Universidad Peruana de Ciencias Aplicadas, Lima, Peru

L. Hernandez
Universidad del Atlántico, Puerto Colombia, Colombia
e-mail: lissettehernandez@mail.uniatlantico.edu.co

A. E. Hernandez · N. Varela
Universidad de la Costa (CUC), Calle 58 # 5566, Atlántico, Barranquilla, Colombia
e-mail: ahernand48@cuc.edu.co

N. Varela
e-mail: nvarela2@cuc.edu.co

H. H. Palma · O. R. Bilbao · N. L. Castro
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

O. R. Bilbao
e-mail: oredondo@ul.edu.co

N. L. Castro
e-mail: nleon@ul.edu.co

R. P. Pulido · J. G. Guliany
Universidad Simón Bolivar, Barranquilla, Colombia
e-mail: rprieto1@unisimonbolivar.edu.co

J. G. Guliany
e-mail: jesus.garcia@unisimonbolivar.edu.co

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_17

Keywords Financial performance · Organizational practices · Microfirms · Artificial neural network (RNA)

1 Introduction

In recent years, important researches [1–5] analyze the role of adopting administrative practices in business performance; however, the vast majority has concentrated on the study of large companies, characterized by shareholding, accounting, managerial and organizational structures with a certain level of development. Few studies analyze the impact of the adoption of certain types of administrative practices in microenterprises and even less in emerging economies such as Colombia.

The objective of this paper is to analyze whether some practices and variables (financial planning, capital stock, marketing, control of purchases and inventories, recording of costs and maintenance and the age and education of the owner) have a determining role in the performance of microenterprises (1–10 employees). The information used consists of microdata obtained from the Business Practices in Small Firms from a developing countries survey developed by the World Bank between 2011 and 2017 [6]. The empirical application includes both econometric techniques and artificial neural networks (ANN).

The hypothesis to be proved states that the adoption of certain types of administrative practices plays a relevant role in understanding the financial performance of this segment of companies. The study acquires special relevance since the subject has been sparsely studied for this segment of companies and above all in developing economies [7].

Empirical evidence reveals a differentiated importance in the adoption of accounting and financial practices by the analyzed companies. The only practice that shows an important impact on the financial performance (approximated by income, measured in sales) of the organizations is the financial planning, while marketing practices, cost recording and maintenance and control of purchases and inventories are not statistically significant to explain variations in sales, or their impact on them is limited [8]. The capital stock and the age and education of the owner are variables that also contribute to explain the financial performance of this segment of organizations.

2 Methodology

2.1 Sample Data

The survey gathers information on the adoption of 4 business practices, mainly accounting and financial practices. Each practice is made up of a set of questions, aimed at knowing whether the organization performs specific tasks or not. Table 1 shows the set of tasks that make up each specific practice [9].

Table 1 Average sample indicators

	Total	Trade	Services	Manufacturing
Observations	7,587	42,587	2,968	985
Monthly sales (USD)	16,478	16,473	20,014	18,447
Age of owner (years)	46.3	47.3	44.5	42.24
Education of owner (years)	8.6	9.7	8.1	7.3
Years in the market (years)	4.57	3.21	4.44	3.53
Capital stock (USD)	20,125	20,478	20,365	18,147
Profit margin (%)	33.3	34.6	36.3	35.1

Given that each organizational practice is made up of a different number of tasks, in order to make the information available for each comparable practice, the data were standardized by dividing the number of tasks performed by each company for each practice by the total number of tasks available in that practice. The results provide values ranging from 0 to 1 in each practice.

Table 1 shows some of the main characteristics of the companies interviewed. On average, they are organizations that sell $17,841 dollars per month, with owners who are 46 years old and have a third year of secondary schooling, organizations that have an average of 3.5 years in the market and a capital stock equivalent to 1.3 months of sales. Its gross profit margin is just over a third of its sales [10].

Comparatively, companies in the service sector present the highest average sales, capital stock and level of education, while those in the commerce sector (65% of the total number of companies in the sample) show the highest profit margin, even though their sales are, on average, 25% lower than those in the manufacturing and services sectors. Manufacturing companies (4.9% of the sample) are the oldest.

2.2 Methods

The characteristics of the available information allow the use of econometric methods, which permit establishing a functional relationship based on pre-established theoretical models, but also, given the amount of information available, it is possible to develop non-parametric estimation models, as in the case of ANN models [11]. Based on the above, both approximations are made, allowing comparison between results.

3 Results and Analysis

3.1 Econometric Application

By using the developed econometric model, the function presented in (1) [12] is estimated:

$$Y/L_t = C + \beta_1 \mathrm{ED} + \beta_2 \mathrm{EduD} + \beta_3 \mathrm{Ant} + \beta_4 \mathrm{Stock} + \beta_5 \mathrm{pracMark} + \beta_6 \mathrm{pracCont} + \beta_7 \mathrm{pracReg} + \beta_8 \mathrm{pracPlan} \quad (1)$$

where ED is the age of the owner, EduD are the years of education received by the owner, Ant is the years of the company in the market, Stock is the capital stock, PracMark are the marketing practices, PracCont are the accounting practices, PracReg are the cost recording and maintenance practices, PracPlan represents the financial planning practices, and $\beta_1, \ldots, \beta_8$ are the estimated coefficients for each of the variables, see Table 2.

The econometric analysis shows a differentiated importance of the adoption of administrative practices for the companies under study; although, as a whole, they show to be important to explain the growth of sales, the practice that turns out to be more important is financial planning.

The owner's age and education variables are statistically significant at a level of 5% and show the expected signs. According to the coefficients, the owner's age allows him to accumulate certain types of experience and knowledge that impact sales; however, the coefficients are low, indicating that the impact, although positive, is small. On the other hand, the years in the market, although statistically significant,

Table 2 Results of the econometric estimation

	Standardized coefficients (Beta)	Statistical *t*	Significance	Standard error
Constant		−18.985	0.000	1458.145
Age of owner	0.055	5.147	0.000	20.147
Education of owner	0.074	6.685	0.000	62.354
Years in the market	−0.045	−4.123	0.000	152.354
Stock capital	0.3254	38.471	0.000	0.010
Marketing Practices	−0.0247	−1.98	0.01	185.698
Financial planning practices	0.785	64.200	0.000	324.147
Practical records of costs and mantle	−0.050	−5.125	0.000	124.254
Purchasing and inventory control practices	−0.021	−2.214	0.001	124.147

do not show an expected sign. Theoretically, a positive sign would indicate that the longer the organization's time in the market, the more experience it tends to accumulate that allows it to know its customers and competitors and increase sales and therefore revenues. The fact that time in the market shows a negative sign may be the result of the high mortality of microorganizations in Colombia. So, given that the average age of the companies in the sample is only three years, as they accumulate years, sales tend to be lower, and by the third year, they become so low that they are forced to close.

The capital stock is statistically sound to explain variations in sales, and its impact is the second-highest range of the set of analyzed variables. It shows the expected sign and is congruent with the idea that the greater the physical capital available to workers, their productivity tends to be higher, and this is reflected in sales.

The most important variable to explain the behavior of sales is the adoption of financial planning practices, which include timely reviews of budgets, expenses and profits and losses of organizations, as well as future sales planning. The coefficient is statistically significant and positive, a fact that reinforces the idea that the adoption of certain types of organizational practices can have an impact on the company performance [13].

The practices of recording purchases and maintaining and controlling purchases and inventories do not show the expected sign. The latter is not statistically significant at the level of 5%. The model as a whole is statistically significant ($F = 0.00$) and has a goodness-of-fit coefficient, measured by the determination coefficient R^2 of 0.621.

3.2 The ANN Model

Once the econometric model was estimated, an artificial neuronal network model was made. Of the 7,587 observations considered valid, 63.4% was used for neuron training, 30.1% for correct training tests, the remaining 6.5% was used to validate how close the values predicted by the network are to the actual values.

The type of ANN used is a multi-layer perceptron [14]. The input layer receives information on the variables or factors of the owner's age, the owner's years of education, the company's time in the market, the capital stock and the degree of adoption of four administrative practices (marketing, purchase and inventory control, cost recording and maintenance and financial planning), see Table 3.

The results show a hidden layer composed of four units, which are shown in Table 3, whose activation function is the hyperbolic tangent 4, and the output layer has an activation function for the income per employee, which relates the weighted sum of units of a layer, with the values of units in the correct layer returned unchanged (identity function) by the function $\gamma(c) = c$.

Analysis of the reserve sample indicates that the ANN results are solid for estimating the behavior of monthly sales. Graph 1 shows the actual and estimated values for the reserve sample. For 90% of the observations (companies with monthly sales of

Table 3 Summary of case processing

		N	Percentage
Input layer variables: age owner, education owner, time in market, capital assets, degree of adoption of marketing practices, control of purchases and inventories, cost recording and maintenance and financial planning	Training	4810	63.4
	Tests	2284	30.1
	Reservation	493	6.5
Valid		7,587	100.0
Hidden layers	Number of hidden layers		1
	Number of units in the hidden layer 1a		6
	Activation function		Hyperbolic tangent
Output layer	Dependent variables	1	Monthly sales
	Number of units		1
	Scale change method for the scale dependent		Standardized
	Activation function		Identity
	Error function		Sum of squares

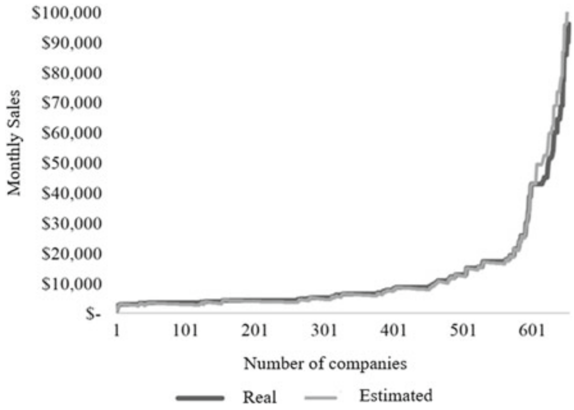


Fig. 1 Graphical representation of the estimation results

less than $42,528 dollars), ANN estimates, with errors of less than 2%, the behavior of sales per company. However, for those organizations with monthly sales of more than 42,528 dollars (10% of the reserve sample), the neural network systematically overestimates the value of sales. This could be an indicator that the higher the sales, the more complex the factors influencing their determination (Fig. 1).

The results obtained coincide, in a certain sense, with those obtained by [14] since, in order for the activities indicated by these authors (participation in the credit market, payment of taxes, access to training and membership in business associations) to be carried out, it is necessary to have certain practices within the organizations. [15] refers to the fact that the characteristics of organizations persist over time as they are correlated with unobserved entrepreneurial skills and capabilities. Some of the unobserved characteristics could be the lack of training in financial and accounting practices, derived from the educational deficiency in Colombian entrepreneurs.

4 Conclusions

An extensive literature shows the importance of adopting management practices in the performance of small, medium-sized and large enterprises; however, research on microenterprises is very limited. Examining the case of microenterprises in Colombia, this study identifies a differentiated importance in their adoption by these enterprises. To measure its impact, an econometric analysis was applied and another one through artificial neural networks of about 9,000 companies from the World Bank's Business Practices in Small Firms in Developing Countries, and the results are compared.

In terms of the preliminary analysis of the survey, the average indicators reflect an alarming situation of educational backwardness on the part of Colombian entrepreneurs since the average level of education is third in secondary school, which implies that some of them know how to read, write and perform basic mathematical operations but do not possess skills and knowledge related to the practices suggested in this research, nor the basic tools to be able to acquire them.

Empirical results show that financial planning practices, which include: monthly review of sales (actual and estimated), preparation of annual cost and sales budgets and preparation of accounting records of financial performance are those that show an impact on the financial performance of the company, approximated from sales revenue. In contrast, the practices of control of purchases and inventories and registration of purchases and maintenance do not show a significant impact on the financial performance of the companies. On the other hand, the capital stock has a significant impact on monthly income, validating the theory that a greater ratio of capital per worker impacts productivity and subsequently sales.

The ANN analysis is able to explain, with errors at less than 2%, the monthly sales of companies with sales less than 850,000 pesos per month. However, for organizations with incomes above that amount, the network tends to systematically overestimate the value of sales, indicating that for larger companies, there are other factors that influence their performance.

The results contribute to the literature by providing information on key aspects of the financial performance of Colombian microorganizations, including financial planning practices and the variables of capital stock, age and educational level of the business owner. The first factor implies that public and private sector support must be accompanied not only by resources and technical training but must also include financial advice to promote the success of the organization. The variables of capital stock and educational level of the owner of the company, second and third factor, point to the need to implement a comprehensive investment strategy in education that allows to increase the level of knowledge and skills of the Colombian population and develop technological innovations that allow small companies to intensify their level of capital used.

References

1. Alvarez, A.T.V., Carrasco, L.V.M., y Córdova, Z.A.F.: Estrategia organizacional y la rentabilidad en em-presas del sector automotriz de la Zona Central del Ecuador. Revista Eniac Pesquisa, **5**(2), 181–192 (2016). https://doi.org/10.22567/rep.v5i2.399
2. Alsaaty, F., Zenebe, A., Sengupta, S.: The Influence of Some Macroeconomic Factors on the Growth of Micro Firms in the United States. Available at SSRN: https://ssrn.com/abstract=2775339 or http://dx.doi.org/10.2139/ssrn.2775339. Accessed on 12 June 2017
3. Bechara, J.E.A., Cruz, J.C.T., Ceballos, H.V.: Predicciones de modelos econométricos y redes neuronales: el caso de la acción de SURAMINV. Semestre Económico Universidad de Medellín, **12**(25), 95–109 (2009). Available from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120634620090003000007&lng=en&nrm=i-so. Accessed on 07 Aug 2017
4. Bizarrón, M.E.B., Palacios, E.M.C., Bobadilla, L.I.Z., García, N.L.A.: El desarrollo de la Mipyme y la vinculación universitaria en Puerto Vallarta, Jalisco. Eur. Sci. J. ESJ **10**(19) (2014). Available at http://www.eujournal.org/index.php/esj/article/view/3788/3604. Accessed on 18 June 2017
5. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: 1994. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pp. 487–499 (1994)
6. Guzmán, M.G., Guzmán, M.M., Fuentes, M.R.: Análisis del uso de las TIC en las pymes de Guayaquil en el año 2015. Revista OIKOS **20**(41), 109–119 (2016). Available at https://dialnet.unirioja.es/descarga/articulo/5841091.pdf, ISSN 0718–4670. Accessed on 13 May 2017
7. Hahsler, M., Karpienko, R.: Visualizing association rules in hierarchical groups. J. Bus. Econ. **87**, 317–335 (2017)
8. Silverstein, C., Brin, S., Motwani, R., Ullman, J.: Scalable techniques for mining causal structures. Data Min. Knowl. Discovery **4**(2–3), 163–192 (2000)
9. Amelec, V., Carmen, V.: Relationship between variables of performance social and financial of microfinance institutions. Adv. Sci. Lett. **21**(6), 1931–1934 (2015)
10. Viloria, A., Lezama, O.B.P.: Improvements for determining the number of clusters in k-means for innovation databases in SMEs. Procedia Comput. Sci. **151**, 1201–1206 (2019)
11. Kamatkar, S.J., Kamble, A., Viloria, A., Hernández-Fernandez, L., Cali, E.G.: Database performance tuning and query optimization. In: International Conference on Data Mining and Big Data, pp. 3–11. Springer, Cham (2018)
12. Viloria, A., et al.: Integration of data mining techniques to PostgreSQL database manager system. Procedia Comput. Sci. **155**, 575–580 (2019)
13. Lanzarini, L., Villa Monte, A., Aquino, G., De Giusti, A.: Obtaining classification rules using lvqPSO advances in swarm and computational intelligence. In: Lecture Notes in Computer Science, vol. 6433, pp. 183–193. Springer, Berlin (2015)
14. Amelec, V., Carmen, V.: Validation of a model for productivity evaluation for microfinance institutions. Adv. Sci. Lett. **21**(5), 1610–1614 (2015)
15. Jiménez, L.R.G., Rodas, M.F.G., Quiroz, M.Q.G.: Opción de Financiamiento a Pymes ubicadas en la Provincia del Guayas enfocadas en la Búsqueda de Capitales mediante la Emisión de Títulos a través del Mercado de Valores Ecuatoriano. Empresarial **10**(40), 21–30 (2017). Available at https://dialnet.unirioja.es/descarga/articulo/5924579.pdf, ISSN No. 1390-3748. Accessed on 18 May 2017

Automation of Admission Enquiry Process Through Chatbot—A Feedback-Enabled Learning System

M. Samyuktha and M. Supriya

Abstract Chatbots are existing since few years and recently it has started acquiring popularity. Earlier to chatbots, people use help desk as the enquiring medium and hence people working at help desks have to work all the days and answer all the questions. Most of the queries are repetitive in nature and answers are given from a structured database. In order to reduce the effort of humans, we can have a chatbot deployed for the same activity. This work focuses on a chatbot which has been developed to provide a faster human-like interaction for admission enquiry system. The chatbot is capable of handling negative or irrelevant scenarios and responds to the queries in faster manner. Decision making by the chatbot on choosing the right set of sentences is done using LSA algorithm and cosine similarity. In addition to answering, the chatbot also maintains data of questions which is not being answered. This data can be used for future analysis for retrieval-based system. The chatbot also takes the feedback from the customers and this data can be analyzed using the feedback category report generated by the chatbot using LDA algorithm

Keywords Chatbot · LSA algorithm · LDA algorithm · Natural language processing (NLP) · Cosine similarity · Term frequency-inverse document frequency (TF-IDF)

1 Introduction

Nowadays, there is rise of interest in automatic conversational agents, such as chatbots which are being applied across various applications and has opened new ways for research. This is due to the recent advancements in natural language processing (NLP) and machine learning (ML). As per a report by Robert Toth, the real credit

M. Samyuktha (✉) · M. Supriya
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

M. Supriya
e-mail: m_supriya@blr.amrita.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_18

for bringing chatbots further into the spotlight goes to Facebook Messenger, due to the fact that Facebook made it possible to integrate chatbots into the platform.

Chatbots or chat agents are automated programs that are capable of interacting with users or customers using natural language processing technique and artificial intelligence in a manner that the user feels that he/she is having conversation with alive human being. The basic architecture for a chatbot is very similar to that of a search engine, i.e., whenever an input is entered each time, a new search is done to produce a meaningful and an appropriate output. Based upon the generated response, the chatbots can be classified as retrieval-based and generative-based chatbots. Retrieval-based models have predefined responses that will be used at the later levels while in case of generative-based model, the responses generated may be new and would not have appeared prior. This paper focuses on the first approach and takes up the feedback in order to enable to go with the second one to. However, the current implementation only considers the feedback for later processing.

Chatbots are widely used in e-commerce sites, handling daily task such as news, weather reports, and daily quotes. Chatbots can be ones' personal helpers that help to manage ones' health, grocery list, social life, helping to connect with locals and tourist guides, etc. Chatbots are great promise for customer time savings and have greater advantages when it comes to costs.

2 Related Work

Literature survey was done for the better understanding of various types of chatbots and its applications. Some of the papers which are quite relevant to this project have been summarized in this section. The very first chatbot was named as Saya whose aim was to enable computer to given answers like human. The answers were based on set of questions that was feed in the database [1]. ELIZA, a computer program was developed as conversational agents [2]. After this, the chatbots have evolved a long way which has been discussed below. The main objective of the paper proposed by Abdul-kader et al. [3] was to find the lexical match for the given query. Additionally, the work also evaluates the cosine similarity and Jaccard's coefficient and concludes that the former model performs better. The dataset used was TREC8 type questions on David Beckham. The idea to have multiple feature extraction methods in order to retrieve best response is relatively new and it can be used to train newborn bot. Currently, the bot was tested on a particular query; hence future work needs to be done on extending across various domains. The problem of keyword matching has been overcome by using Automatic Speech Recognition System (ASR) in [4]. Here, the responses are stored after preprocessing using Ripple Down Rules and the decision tree method is used to find the lexical match. The test was carried out in 41 sessions by the undergraduates across various domains and it was found that it responded correctly for at least 80% of the requests.

The authors claim that current conversation system does not have personality type and hence there is a need for human-like responses. This is achieved by using Machine

Translation and Text Summarization in [5]. The Kaggle dataset containing Twitter Customer Support data has been used in this work. The conversation agent is of sequence-to-sequence type with one bidirectional long short-term memory (LSTM) layer. It also has a transformer with two identical layers for encoder and decoder. The model was evaluated on the bases of semantic match and word overlapping. The major advantage of LSTM is that it can stop or pass on the information depending on the importance as it adopts word embedding method [6]. Also, it was found that LSTM outperformed IR in aspects of t-test and bilingual evaluation understudy (BLEU) measures. This idea can be used across other social networks as well.

The support vector machine (SVM) algorithm was used to extract the knowledge pair and construct the knowledge base using Artificial Markup Language (AIML) [7]. It was found that user satisfaction is at a higher rate but however system triggered correct responses only of 65%. Another limitation is that SVM is done through quadratic programming and hence cannot be implemented on large scale. A method proposed by Amazon is capable of chatting with humans on the topics in trend [8]. This study focuses on the implementation of reinforcement learning. The system consists of natural language (NL) generation and retrieval models. An Indonesian language model chatbot [9] was developed by extracting the conservations from movie subtitle. Since movie subtitle mimic the conversation of the humans, feeding the same to knowledge database is considered to be a very good idea. The vector representation for words has been done for word embedding as a part of deep learning. On the other hand for Chinese question generation model, children's tales were used [10]. Another approach that was used in conversational model was by including user profiles. This was implemented using neural networks [11]. Another study was made on chatbots or conversational agents and it was found that a lot of retrieval-based models are in place and generative models are just upcoming [12]. Chatbots need not be always be used for customer support. It can also be implemented as tutor based on the frequently asked questions by the students [13].

From the study, it has been understood that there are several challenges in building the chatbot. One of the major challenges is the NLP semantic, i.e., sometimes the bot may fail to answer if the question portrayed in different manner. Another challenge is the training of the neural network as there is no controlled measure to monitor what the bot has learned.

3 Proposed Methodology

The proposed system architecture is shown diagrammatically in Fig. 1. The following subsection explains the architecture in detail.

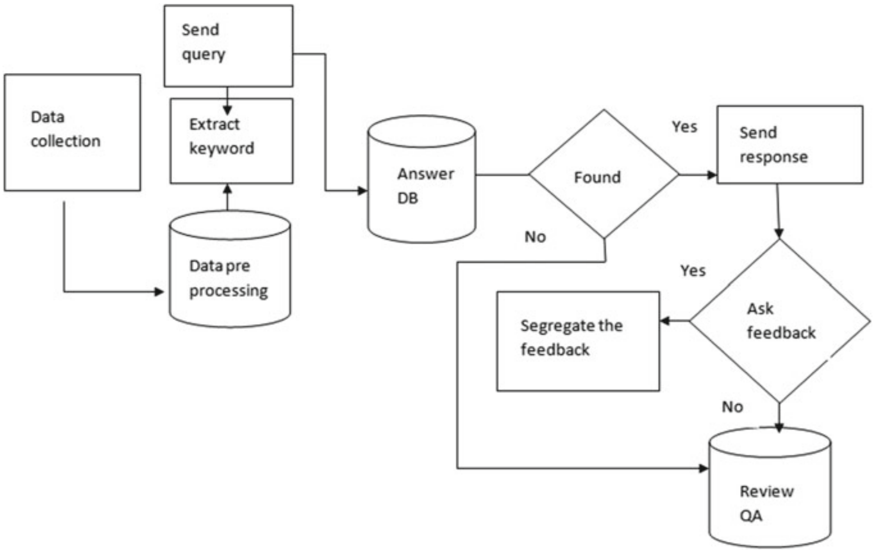


Fig. 1 System architecture

3.1 Preprocessing

The problem with textual data is that it is in form of text (i.e.) in strings, but the machine learning systems need data in the form of numeric as has to be represented as feature vectors. First step is to convert the entire text and the user input into lowercase. Tokenization is done for converting the normal text or strings into a list of words. Stop word Removal is also done by removing the commonly used words that adds little value to the data. Lemmatization is used instead of stemming because stemming often reduces to non-existent words, whereas lemmas reduce to actual words in the dictionary.

3.2 TF-IDF Approach

Once preprocessing is done, we create a function for greeting the user by the bot. Instead of using Bag-of words for adding weights to the words; term frequency-inverse document frequency (TF-IDF) approach is used. Here, technically we have only one document in the corpus, the second document is the user's input. TF-IDF is used to get two vectors in vector space. Then we calculate Cosine **similarity** using the formula given below.

$$\text{cosinesimilarity }(d1, d2) = \frac{\text{dotproduct}(d1, d2)}{\|d1\| * \|d2\|}$$

where $d1$, $d2$ are the documents.

3.3 Identifying the Search Text

Once the question is asked by the user, the bot has to identify the corresponding answer. In order to perform this search, the bot has to go through the entire text document. This can be eased out by reducing the dimensions if the search space is done using LSA algorithm [14]. Singular Value Decomposition (SVD) is applied to form the matrix. Initially, the number of features or topics is specified. Then matrix transformation is done to the original matrix. Finally, the dot product of the matrices is taken to form a reduced matrix. The equation given below is used to compute the SVD.

$$M = U\varepsilon V^*$$

where

M is $m \times m$ matrix,
U is $m \times n$ left singular matrix,
ε is $n \times n$ diagonal matrix and
V is $m \times n$ right singular matrix.

3.4 Generating Response and Analyzing the Feedback

To get the appropriate response from chatbot for the user's input questions (Fig. 2), a function is defined that searches the user's utterance. The feedback given by the users is great source of information in order to analyze the performance of the chatbot. However, it is time-consuming task if we have to go through the comments manually. Hence in order to find the key insights from the data, Latent Dirichlet Allocation (LDA) algorithm has been used to extract the key topics and get the users feedback without having to go through all of them.

LDA is a model to classify text present in the document into different topics. It is built using Dirichlet distributions. The Dirichlet distribution comes under the continuous multivariate probability distributions. Here it builds a topic per model and words per topic model, which are later modeled as Dirichlet distribution. The system also maintains the data of user's input that are not being answered. This data can be used for future analysis.

4 Experimental Setup

The input dataset containing FAQs about 100 questions was stored in a word document. These questions were collected from the websites of the deemed universities.

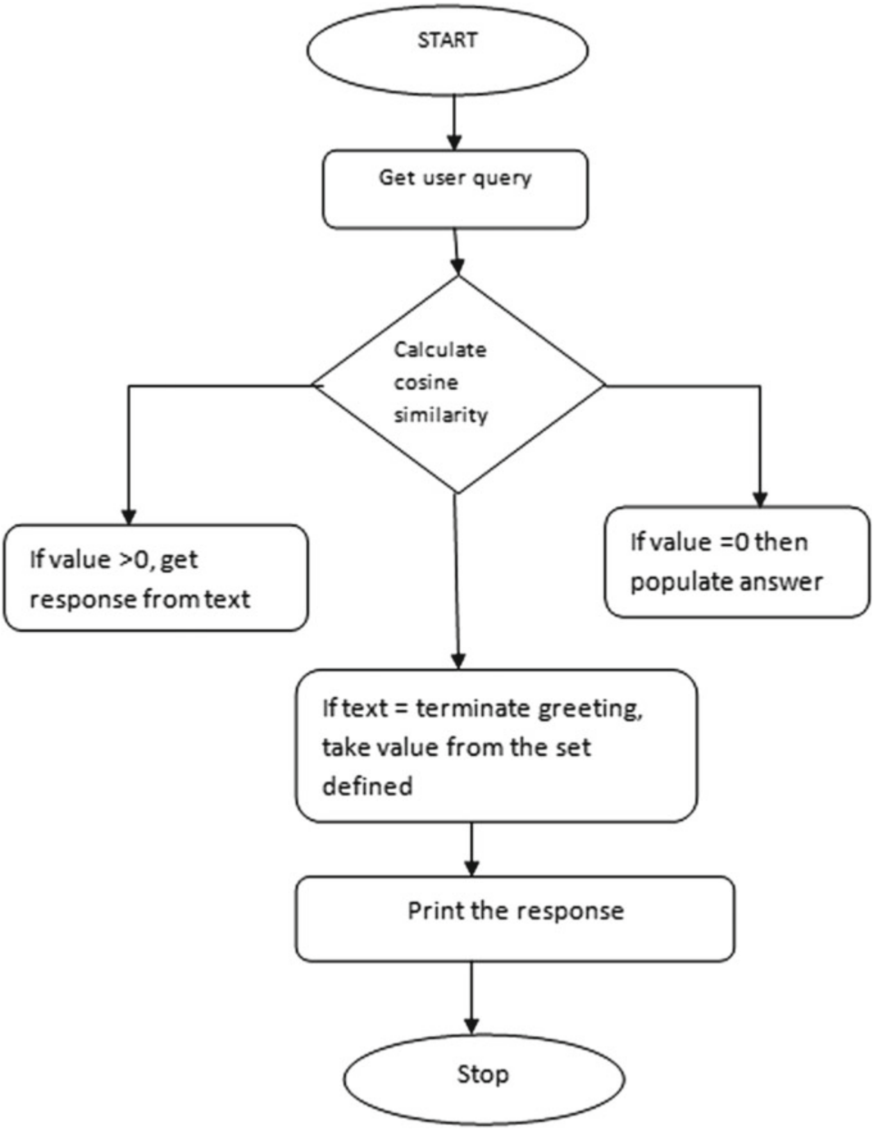


Fig. 2 Generating response

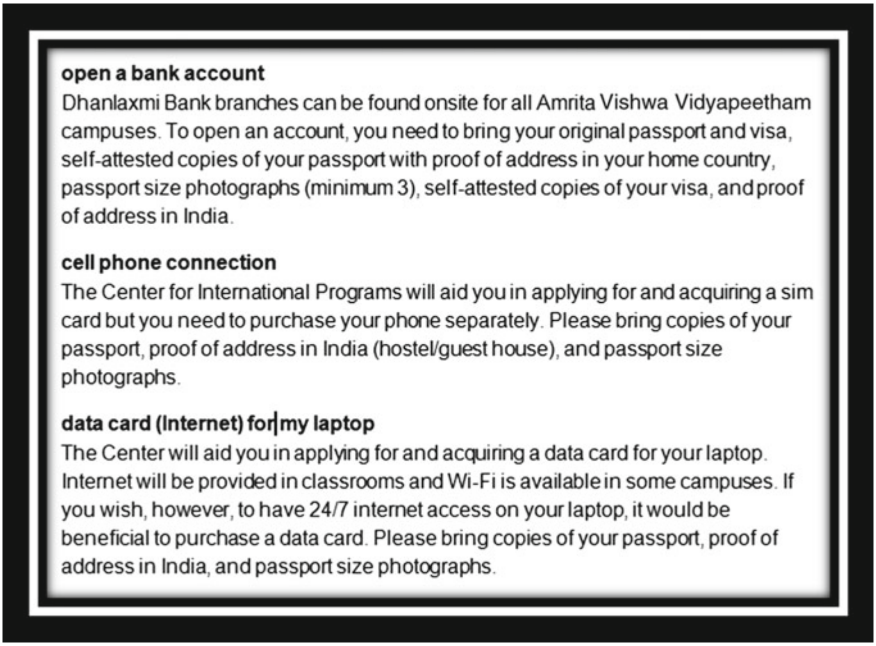
The data includes text and hyperlinks. The feedback results are stored in an excel sheet containing both the comments and the rating. For training purpose, most of the feedbacks were taken from the customer reviews of various institutions.

The chatbot was implemented using Python with the help of Natural Language Toolkit (NLTK) packages and UI was done with tkinter package. About 60 plus questions across various categories were tested. The chatbot is also able to store the unanswered questions in a separate file. The feedback comments were also taken from the user for each question. For testing purpose, online feedback from various institutions was collected in order to generate the classification report.

5 Experimental Results and Analysis

To evaluate the chatbot, the chatbot was tested with 60 plus questions across all the domains as mentioned in the previous section. Figure 3 shows a sample of the dataset from where the answer is fetched. This dataset has been picked from the Web site of the author's institution page. Based on the results, a confusion matrix was generated to analyze the accurate response of the chatbot thereby we can calculate the performance measures such as accuracy, recall, precision, and also the F-measure.

From the results given in Fig. 4, one can clearly understand that F-measure is nearer to the value of precision or recall. Also, recall and high precision indicate that high False Negative and low False Positive. High recall and low precision indicate that low False Negative but high False Positive. Accuracy of the chatbot is 66%. However, this could be improved with the increase in the testing dataset and appropriately testing the same with multiple test cases.



open a bank account
Dhanlaxmi Bank branches can be found onsite for all Amrita Vishwa Vidyapeetham campuses. To open an account, you need to bring your original passport and visa, self-attested copies of your passport with proof of address in your home country, passport size photographs (minimum 3), self-attested copies of your visa, and proof of address in India.

cell phone connection
The Center for International Programs will aid you in applying for and acquiring a sim card but you need to purchase your phone separately. Please bring copies of your passport, proof of address in India (hostel/guest house), and passport size photographs.

data card (Internet) for my laptop
The Center will aid you in applying for and acquiring a data card for your laptop. Internet will be provided in classrooms and Wi-Fi is available in some campuses. If you wish, however, to have 24/7 internet access on your laptop, it would be beneficial to purchase a data card. Please bring copies of your passport, proof of address in India, and passport size photographs.

Fig. 3 Dataset

```
Confusion Matrix :
[[19  9]
 [ 8 14]]
Accuracy Score : 0.66
Report :
              precision    recall  f1-score

           0       0.70      0.68      0.69
           1       0.61      0.64      0.62
```

Fig. 4 Confusion matrix results

After a session gets completed, the user will be asked to rate the system with a rating value between 1 and 5, 1 being a low rating and 5 being a high rating. These feedbacks will get stored in the database and can be used to retune the chatbots performance. Also, as has been mentioned earlier, the unanswered questions have been stored in a separate file which can be considered to retrain the chatbot to improve the accuracy.

6 Conclusions

The primary objective of this work was to build a chatbot that can act as an admission enquiry agent. However, it is known fact that the chatbot cannot handle all the questions as the primary source of its answer is based on the data that is given in its backend. Hence to improvise the data and learning capability of the chatbot, a feedback system is introduced. The user comments are collected and evaluated using topic modeling techniques. This will reduce the efforts involved in manually reading all the reviews and identifying the key topics. Also, the chatbot answers as per the segregated topic without having to go through all the answers present in the backend. Future work must be done on analyzing the data which the bot has not answered. Using machine learning algorithms, we can implement the self-learning bot. Also, we must analyze on the threshold to be kept for cosine similarity as currently it returns all possible answers.

References

1. Aleksandra (Sasha) Kugel Shira David, "Saya ChatBot", technical paper, 2008
2. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM **9**(1) (1966)
3. Abdul-kader, S.A.: Question Answer System for Online Feedable new born Chatbot, pp. 863–869 (2017)
4. Herbert, D., Kang, B.H.: Intelligent conversation system using multiple classification ripple down rules and conversational context. Exp. Syst. Appl. **112**, 342–352 (2018)
5. Shukhin, L.M., Borzunov, E.E.: A new Chatbot for customer service on social media. Farm. Zh. **5**, 89–91 (1975)
6. Hardalov, M., Koychev, I., Nakov, P.: Artificial intelligence: methodology, systems, and applications **6304** (2010)
7. Hussain, S., Athula, G.: Extending a conventional chatbot knowledge base to external knowledge source and introducing user based sessions for diabetes education. In: Proceedings of 32nd IEEE International Conference on Advance. Information Network Application Work. WAINA 2018, vol. 2018, pp. 698–703 (2018)
8. Serban, I.V., et al.: A Deep Reinforcement Learning Chatbot (Short Version), pp. 1–9. ACM Publications, no. Nips (2018)
9. Chowanda, A., Chowanda, A.D.: Generative Indonesian conversation model using recurrent neural network with attention mechanism. Procedia Comput. Sci. **135**, 433–440 (2018)
10. Lee, C.-H., Chen, T.-Y., Chen, L.-P., Yang, P.-C., Richard T.-H.: Automatic question generation from Children's stories for companion Chatbot. In: IEEE International Conference on Information Reuse and Integration for Data Science. 978-1-5386-2660-3, (2018)
11. Liu, Z.X., Sun, C., Wang, B., Wan, X.: Content-oriented user modeling for personalized response ranking in Chatbots. ACM Trans. Audio, Speech, Lang. Proces. **26**(1) (2018)
12. Io1, H.N., Lee, C.B.: Chatbots and conversational agents: a bibliometric analysis (2017). IEEE 978-1-5386-0948-4/17/2017
13. Niranjan, M., Saipreethy, M.S., Kumar, T.G.: An intelligent question answering conversational agent using Naïve Bayesian classifier. In: Proceedings of 2012 IEEE International Conference on Technology. Enhancement Education. ICTEE 2012 (2012)
14. Thomas, N.T.: An e-business chatbot using AIML and LSA. In: International Conference on Advances in Computing, Communications and Informatics (ICACCI) (2016)

Hardware-Assisted QR Code Generation Using Fault-Tolerant TRNG

Yaddanapudi Akhileswar, S. Raghul, Chitibomma Meghana and N. Mohankumar

Abstract True random number generator (TRNG) is used to generate a purely random sequence in key generation. Real-world applications use key bits or strings as a passcode to secure systems. The security of a system depends on a robust design and the ambiguity of the keys that are used so that they are unpredictable. In this proposal, TRNGs are designed using reversible gates and fault-tolerant circuits. So the chance of the TRNG hardware produces faulty output is avoided. The inputs for this system are obtained from CPU usage. The generated true random number sequence is used in generating QR codes due to the uniqueness of the generated sequence. The proposed TRNG design highlights the effectiveness of using reversible fault-tolerant gates for TRNG application over the conventional logic implementation and reversible gate design in terms of power, area, and randomness. The proposed design in 90 nm technology consumes only 25.26 μW of power.

Keywords Hardware security · QR code · TRNG · Fault-tolerant reversible gates

1 Introduction

In reversible gates, one-to-one mapping exists between input and output vectors by which input vectors can be reconstructed from the output vectors. They have an equal number of inputs and outputs which pave us the way for determining the outputs for a given input. In non-reversible gates for every computation, energy is dissipated. Energy dissipation implies to the loss of information. For every bit of information being lost, $kT \ln 2$ of energy is lost where k is Boltzman constant and T is temperature. The common reversible gates are Feynman gate, Peres gate, Toffoli gate, and Fredkin gate.

Y. Akhileswar · S. Raghul · C. Meghana · N. Mohankumar (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: n_mohankumar@cb.amrita.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_19

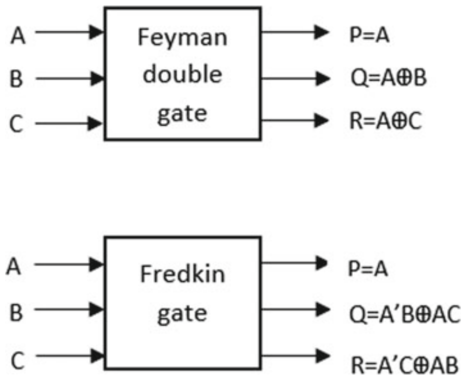


Fig. 1 Parity preserving gates

1.1 Reversible versus Fault-Tolerant Reversible Gates

Parity preserving gates are one type of reversible gates which are fault-tolerant. Parity checking is one of the most efficient and easiest approaches for exposing the fault in digital systems. A gate is said to be parity preserving where parity is matched between the input and output vectors. For example, for a 3 × 3 reversible gate $I_1 \oplus I_2 \oplus I_3 = O_1 \oplus O_2 \oplus O_3$ and also $I_1 = O_1$. Fredkin and Feynman Double gates in Fig. 1 are parity preserving gates [13].

The parity check for Fredkin gate and Toffoli gate being fault intolerant is shown in Table 1.

From the above table, we can see that the Toffoli gates have not passed the parity test for the inputs 110,111. So it is becoming tedious to unmask the fault detection at the gates.

TRNGs are designed using visible spectrum for secured communication [1]. A technology-independent TRNG (TI-TRNG) architecture was proposed with bias detection mechanisms to self-calibrate which enhances power supply noise for older technologies [2]. Random numbers are generated by designing a random number generator which uses optical theremin [3]. The DBG system proposed is able to generate random bit sequence with the help of LFSR and CA [4]. Phase noise of a ring oscillator is used to design TRNG [5]. Self-timed rings are replaced by RO to design TRNG [6]. Cross-coupled NAND gates are used to generate random numbers [7]. Multiple-scroll chaotic systems are used for designing TRNG [8]. Bio-related signals are used to generate random numbers [9] Low-cost EEG [10] and the output characteristics of the solar panel are used to design random number circuits [11]. It is also used in mobiles for OTP generation [12]. It plays a vital role for secured systems for number generation [14] and security systems for smartphones [15]

Table 1 Parity check for Fredkin and Toffoli

Input sequence			Output sequence (Fredkin)			Output sequence (Toffoli)			Parity (Fredkin)	Parity (Toffoli)
A	B	C	D_1	E_1	F_1	D_2	E_2	F_2		
0	0	0	0	0	0	0	0	0	✓	✓
0	0	1	0	0	1	0	0	1	✓	✓
0	1	0	0	1	0	0	1	0	✓	✓
0	1	1	0	1	1	0	1	1	✓	✓
1	0	0	1	0	0	1	0	0	✓	✓
1	0	1	1	0	1	1	0	1	✓	✓
1	1	0	1	1	0	1	1	1	✓	✗
1	1	1	1	1	1	1	1	0	✓	✗

2 Proposed Design

In this proposal, true random number generator (TRNG) was initially designed using the reversible gates and later fault-tolerant gates were opted to design the system for efficient results. The decision of using fault-tolerant gates was decided based on its characteristics.

The fault-tolerant gates were chosen to design the TRNG due to its feasibility in its approach toward error detection. When the fault-tolerant gates are used, we can use its parity preserving design to detect the error at the latest. The defected gate in the designed TRNG system can also be found.

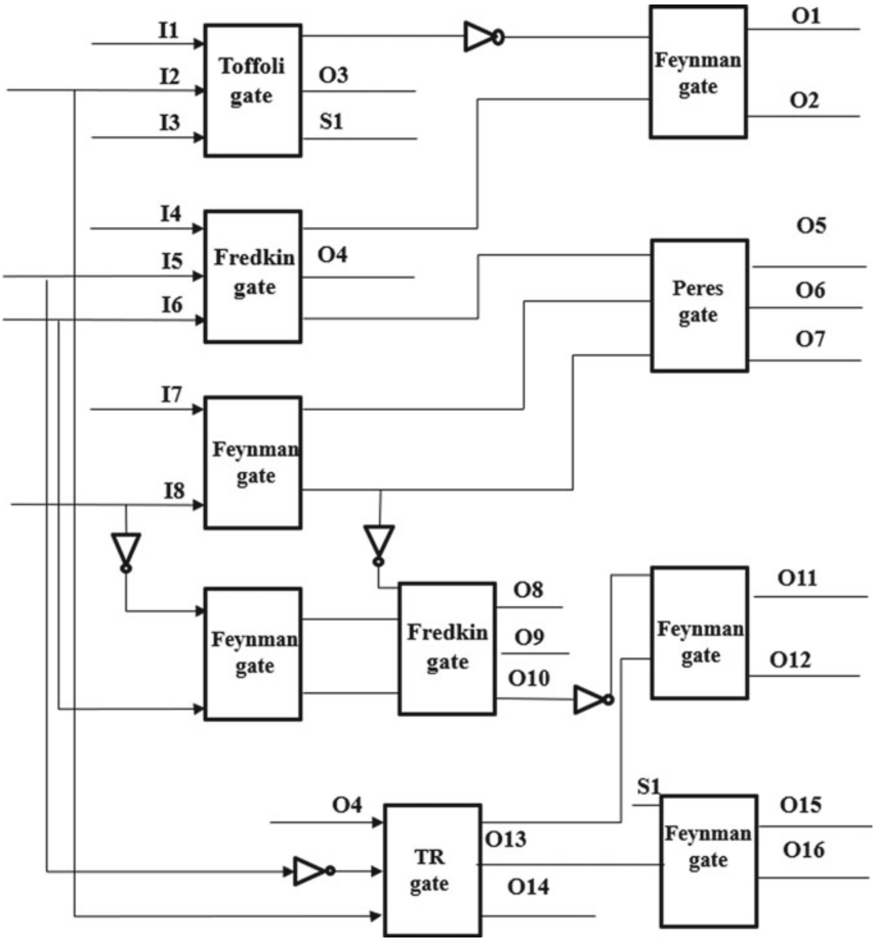


Fig. 2 TRNG system designed using reversible logic

The TRNG system in Fig. 2 has been designed using reversible logic. Tofolli, Feynman, Peres, and Fredkin gates were used randomly to design the TRNG system. We have achieved the randomness through the design combined with the natural random inputs. The CPU utilization signal from the computing system is chosen as physical input to the TRNG and CPU utilization percentage ranges from 0 to 100.

The design in Fig. 2 takes 8 binary bits as an input which in turn results out 16 binary bits. The CPU usage percentage is represented as 8-bit binary data for processing.

The availability of fault-tolerant gates gives us an advantage over reversible gates. For example, the design with a Feynman double gate and Fredkin gate in Fig. 3 is considered.

The design in Fig. 3 takes 18 binary bits as an input which in turn results out 18 binary bits. The intake of 18 bits contains 8 actual bits which are a binary form of CPU usage percentage as in the design from Fig. 2 and 10 garbage inputs which are chosen randomly for the reversible logic. The output also has two additional garbage bits with only 16 bits considered as output.

The fault-tolerant design was verified with its parity preserving logic. In system from Fig. 3, XOR of all the input was equal to the XOR of all outputs which prove its parity preserving design. For estimating the parity preserving characteristics, garbage bits are also considered.

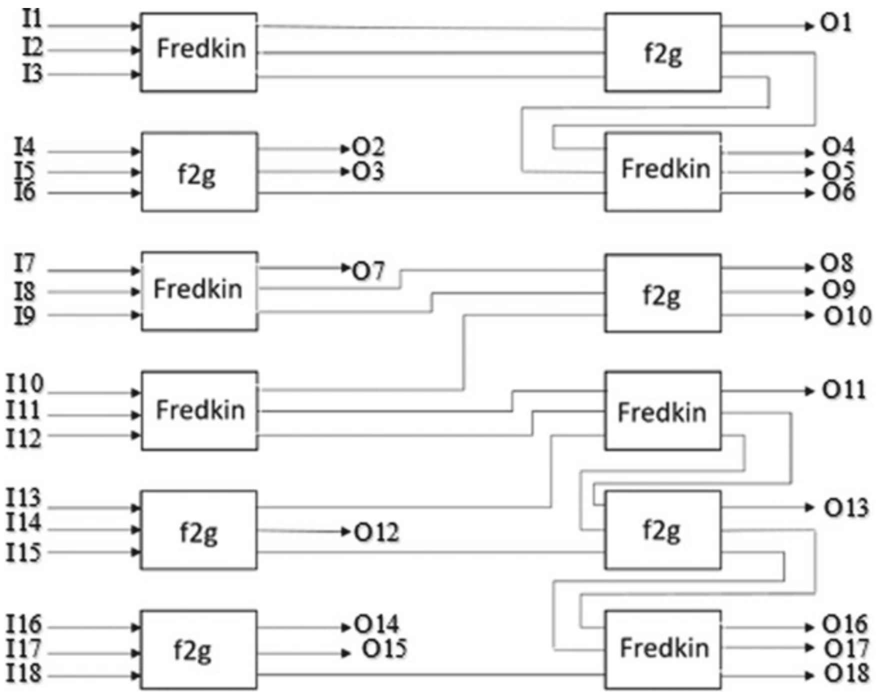


Fig. 3 TRNG design using fault-tolerant gates

Table 2 Verification of fault-tolerant design

Input (I_1, I_2 … I_{18})	Output (O_1, O_2 … O_{18})	$I_1 \oplus I_2 \oplus \cdots \oplus I_{18}$	$O_1 \oplus O_2 \oplus \cdots \oplus O_{18}$
000000110000000000	000000101000000000	0	0
000111110000000000	010000101000000000	1	1
001010110000000000	001010101000000000	0	0
010001110000000000	000110101000000000	0	0
010110110000000000	010110101000000000	1	1
011000110000000000	000101101000000000	0	0

Table 3 Gate level parity preserving check for Feynman double gate

ABC	*LMN*	$A \oplus B \oplus C$	$L \oplus M \oplus N$
001	001	1	1
101	110	0	0
111	100	1	1

From Table 2, it can be observed that the $I_1 \oplus I_2 \oplus I_3 \oplus I_4 \oplus \cdots \oplus I_{18} = O_1 \oplus O_2 \oplus O_3 \oplus O_4 \oplus O_5 \cdots \oplus O_{18}$ which proves the parity preserving the logic of the circuit. The identification of fault in the system is verified by matching the input parity with output parity of the system.

If the parity mismatch occurs, it is a sign of system being faulty. The gate-level fault can be deduced from the mismatch of input parity with output parity which is given in Table 3. In Table 3, some sample inputs and outputs of the Feynman double gate with *ABC* as inputs and *LMN* as outputs are considered.

The design from reversible logic gets being fault intolerant is replaced with fault-tolerant gates in the design. The damaged physical gate can be replaced if the parity mismatch is found thus making the circuit agile. The main characteristic of this is any tampering of intermediate bits or output bits which can be identified easily with the parity discrepancy thus making it more secured and transparent system.

3 Results and Analysis

This TRNG can be used in real-world application like OTP generation, user ID generation, booking or order ID, and many more which requires randomness at each instance. Here, OTP generation is considered which is specifically used for bank transaction authorization. In general, OTP number is received as a text for the corresponding requested user. To make this system more robust, the output of the TRNG along with time stamp is stored in QR code. So, the system will send a QR code to the requested user. The output of the TRNG is stored in the QR code. Following the output of TRNG second, month, year, day and hour are stored in the

QR code. Hence, it becomes more tedious to unmask the pattern or sequence of the OTP generation. The added advantage is OTP or QR code can be made available for use within a time frame. The detailed process of QR code generation is given in Fig. 4

When the OTP is requested, the system will take the current CPU usage as the input for the fault-tolerant TRNG and the output is generated. For instance, during the OTP request if the CPU usage is 25%, the input to the TRNG will be 000011010000000000 in which 00001101 will be the CPU usage (i.e., 25%) and 0000000000 will be the garbage inputs. The corresponding output Fig. 5 is 001001011100000000 which in decimal is 38,656.

The time stamp during the OTP request is appended with the TRNG output and saved in the QR code. So at each OTP request, the time stamp and the TRNG output vary creating distinct OTPs. The sample QR code generated is shown in Fig. 6.

Scanning the above QR code, the data contained is 14386565201914191049. The pattern of the OTP sequence is shown in Table 4.

Thus by scanning the QR code, the transactions can be completed in a secured manner. The user himself is not aware of the OTP that is sent to him as it is stored in the QR code.

Sample outputs for the reversible and fault-tolerant design are shown in Figs. 7 and 8 respectively.

The same has been synthesized using 90 nm technology node dissipates low power. Moreover, the area occupied is also less.

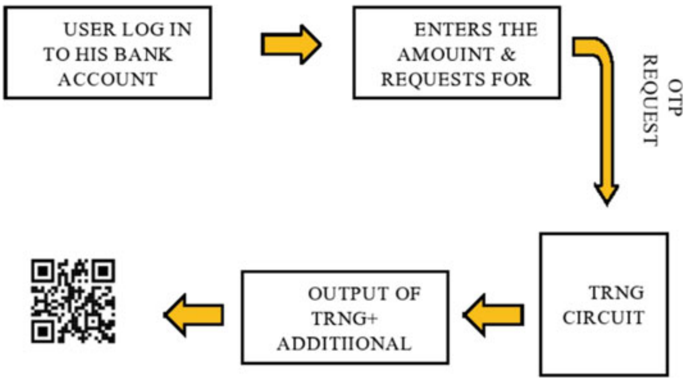


Fig. 4 Overall process of QR code generation

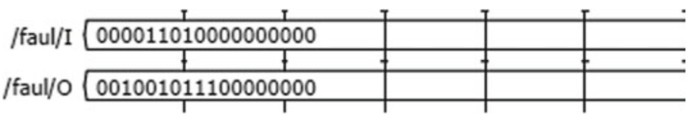


Fig. 5 Bitstream generation of TRNG

Fig. 6 Sample QR code generated by the proposed system

Table 4 Pattern of OTP sequence

Seconds	TRNG	Month	Year	Seconds	Date	Hour	Minute
14	38,656	5	2019	14	19	10	49

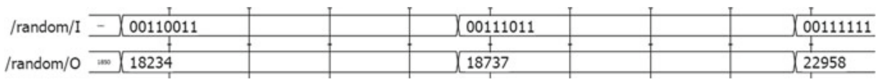


Fig. 7 Sample outputs for the reversible design

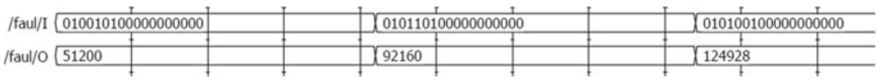


Fig. 8 Sample outputs for the fault-tolerant design

From Table 5, we can infer that the power and area of the fault-tolerant design are greater when compared to the reversible design. This is due to the number of gates are more in fault-tolerant design. The fault-tolerant design has 12 3 $\times$ 3 reversible gates, and reversible design has 5 2 $\times$ 2 and 5 3 $\times$ 3 reversible gates. Hence, the fault-tolerant design has a considerate increase in area.

Table 5 Power and area reports

Design	Internal power	Switching power	Leakage power	Total power	Total area
Reversible	13.7238 μW	2.7262 μW	882.1771 nW	17.3312 μW	230.274937
Fault-tolerant	20.6150 μW	3.3193 μW	1.3329 μW	25.26 μW	311.246624

4 Conclusion

The TRNG system was thus designed using both reversible logic gates and fault-tolerant logic gates. The TRNG system designed from fault-tolerant gates was verified using a parity preserving technique. Hence, TRNG systems modeled from fault-tolerant gates are more secure and transparent than TRNG systems designed using reversible logic. Further TRNG system from reversible logic being more power-efficient than TRNG system from fault-tolerant logic. Thus, the systems designed can be chosen as suitable for the concerned application. The system output has been used to generate the QR code. The generated QR code can be used in different applications such as allotment of user id, OTP. TRNG system design [made of fault-tolerant gates] can be custom made using the inputs required by the user with the help of AI-based technology can be considered as a future development of this work.

References

1. Lee, K., Lee, S., Seo, C., Yim, K.: TRNG (True Random Number Generator) method using visible spectrum for secure communication on 5G network. IEEE Access **6**, 12838–12847 (2018)
2. Rahman, M.T., Xiao, K., Forte, D., Zhang, X., Shi, J., Tehranipoor, M.: TI-TRNG: technology independent true random number generator. In: 2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC), San Francisco, CA, pp. 1–6 (2014)
3. Sharma, R., Ullagaddimath, R., Roy, A.B., Halder, A., Hegde, V.: Optical theremin based True Random Number Generation (TRNG) system. In: 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, pp. 571–575 (2015)
4. Rajagopalan, S., Rethinam, S., Lakshmi, G., Mounika, P., Vani, R., Chandana, D.: Diffused bit generator model for TRNG application at CMOS 45 nm technology. In: 2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS), Vellore, pp. 1–5 (2017)
5. Wieczorek, P.Z.: Lightweight TRNG based on multiphase timing of bistables. IEEE Trans. Circuits Syst. I Regul. Pap. **63**(7), 1043–1054 (2016)
6. Martin, H., Peris-Lopez, P., Tapiador, J.E., San Millan, E.: A new TRNG based on coherent sampling with self-timed rings. IEEE Trans. Industr. Inf. **12**(1), 91–100 (2016)
7. Li, C., Wang, Q., Jiang, J., Guan, N.: A metastability-based true random number generator on FPGA. In: 2017 IEEE 12th International Conference on ASIC (ASICON), Guiyang, pp. 738–741 (2017)
8. Guo, C., Zhou, Y.: A true random number generator based on the multiple-scrolls chaotic system. In: 2018 3rd International Conference on Information Systems Engineering (ICISE), Shanghai, China, pp. 98–103 (2018)
9. Yu, H., Kim, Y.: True random number generator using bio-related signals in wearable devices. In: 2018 International SoC Design Conference (ISOCC), Daegu, Korea (South), pp. 231–232 (2018)
10. Gavas, R.D., Navalyal, G.U.: Fast and secure random number generation using low-cost EEG and pseudo random number generator. In: 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Bangalore, pp. 369–374 (2017)
11. Ritter, S., Pigg, T., Brown, C. and Ray, B.: True random number generator using solar output characteristics. In: 2018 6th IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE), Huntsville, AL, USA, pp. 224–226 (2018)

12. Özkaynak, F., Özdemir, H.İ., Özer, A.B.: Cryptographic random number generator for mobile devices. In: 2015 23nd Signal Processing and Communications Applications Conference (SIU), Malatya, pp. 1733–1736 (2015)
13. Parhami B.: Fault-tolerant reversible circuits. In: 2006 40th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, pp. 1726–1729 (2006)
14. Prasad, R.S., Siripagada, A., Selvaraj, S., Mohankumar, N.: Random seeding LFSR-based TRNG for hardware security applications. In: Studies in Computational Intelligence, vol 771, Springer, New York (2019)
15. Mohan, A.K., Devi, N., Sethumadhavan, M., Santhya, R.: A selective generation of hybrid random numbers via android smart phones. Int. J. Pure Appl. Math. **118**(8), 311–317 (2018)

Classification of Digitized Documents Applying Neural Networks

Amelec Viloria, Noel Varela, Omar Bonerge Pineda Lezama, Nataly Orellano Llinás, Yasmin Flores, Hugo Hernández Palma, Carlos Vargas Mercado and Freddy Marín-González

Abstract The exponential increase of the information available in digital format during the last years and the expectations of future growth make it necessary for the organization of information in order to improve the search and access to relevant data. For this reason, it is important to research and implement an automatic text classification system that allows the organization of documents according to their corresponding category by using neural networks with supervised learning. In such a way, a faster process can be carried out in a timely and cost-efficient way. The criteria for classifying documents are based on the defined categories.

Keywords Text categorization · Artificial neural networks · Multilayer perceptron

A. Viloria (✉) · N. Varela · F. Marín-González
Universidad de la Costa, St. 58 #66, Barranquilla, Atlántico, Colombia
e-mail: aviloria7@cuc.edu.co

N. Varela
e-mail: nvarela2@cuc.edu.co

F. Marín-González
e-mail: fmarin1@cuc.edu.co

O. B. P. Lezama
Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras
e-mail: omarpineda@unitec.edu

N. O. Llinás · Y. Flores
Corporación Universitaria Minuto de Dios—UNIMINUTO, Barranquilla, Colombia
e-mail: nataly.orellano@uniminuto.edu

Y. Flores
e-mail: yasmin.flores@uniminuto.edu

H. H. Palma · C. V. Mercado
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

C. V. Mercado
e-mail: cvargas@ul.edu.co

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_20

1 Introduction

The automatic classification of documents has gained increasing research interest in recent times, since the exponential increase of available information in digital format over the last few years, and expectations of future growth make it necessary to organize all this content in order to improve the search and access to information, which has become a difficult task by means of the manual classification [1]. To this end, the development of an intelligent system for the automatic classification of texts is necessary.

Since the development of science and technology presents an accelerated advance, the information in each knowledge area increases exponentially, and its treatment and storage become more complex. The explosive growth of information available in digital documents in the area of information technology and systems has made it necessary to develop new tools and instruments that facilitate the conduct of search processes in an efficient and effective way, as well as the management of these resources. In order to facilitate the search for information, documents are often categorized into a limited set of classes or categories. These classes represent specific areas of knowledge and are generally consolidated by experts [2, 3].

The context of this study aims to create an intelligent system that allows documents to be automatically categorized using expert systems [4] in such a way that a faster process is carried out with less time and cost. For the development of the research, data available from the University of Mumbai repository was used in different formats (Microsoft Word, PDF, plain text) that have been used to train and subsequently evaluate the results obtained in each training group. It is important to indicate that the classification of the documents will be categorized into 14 groups which are physics, mathematics, social sciences, natural sciences, art, economy, education, engineering, environment, medicine, juridical, psychology, language and diverse.

2 Method

2.1 Sample Data

The data was collected from the repository of Scientific Journals of the University of Mumbai in India [5]. This repository has a large number of digitized documents such as 1,254,325 documents composed of: full text scientific articles, scientific journals and fascicles.

As stated above, one of the reasons for using this repository is that its metadata has a complete structure such as author, title, keywords, publication, URL, among others. Each of these data provides more information about the document that is why when selecting the repository, the corpus metadata was considered since it will be of great help for classification purposes [6].

Table 1 Metadata used for a document

Authors
Title
Description
Date of publication
Keywords
Language
Url

Metadata is structured and coded data describing characteristics of instances containing information to help identify, discover, evaluate and manage. In other words, metadata is data about the data, and they will be extracted and used in the identification of documents after the analysis of the repositories. Table 1 shows these data.

2.2 *Network Architecture Design*

Since the study requires the construction of a previously supervised multilayer network, the network outputs must be previously known. This network is applied to the classification of documents into 14 categories which are: physics, mathematics, social sciences, natural sciences, art, economics, education, engineering, environment, medicine, law, psychology, language and diverse. These categories are mentioned after an exhaustive study and review of each of the documents in the repository [7].

In order to determine the belonging of a document to a certain category, a vocabulary had to be assigned to each of them, that is, each category has subcategories that best describe the category. For this purpose, the process includes the Dewey Classification System [8], which constitutes structured lists of terms (concepts) that represent, in a univocal way, the conceptual content of the documents and is a system that quantifies the relevance of a term to describe a category.

To conclude the category, the category–glossary relationship is used, i.e., the network entry is compared with each glossary or vocabulary of each category to assign weights to each category to determine the category to which it corresponds; see Fig. 1.

Figure 2 shows the structure of the network and its connections, the input layers will be represented by title data, keywords and description, and the output layer will be made up of the 14 categories. The final model corresponds to a network with 14 inputs and 14 outputs.

```
public String[] neuralNetworkInput(List<String> words){
      String[] input = {"0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0"};

      for (String word : words) {
         Glossary gosa = Glossary List Get Glossary by Word (word);
         if (gosa!=null) {
           List<Categories> cats = gosa.getCategories();
         for (Categories categories : cats) {
            int index = Integer.parseInt(""+Categories.getId())-1;
            float aux = Float.parseFloat(input[index]);
            if (aux!=0) {
               aux = (float) (aux + 0.1 - 0.01);
            }else{
               aux = (float) (aux + 0.1);
            }
            input[index]= ""+aux;
         }
         }
      }
      return input;
   }
```

Fig. 1 Algorithm for weight allocation

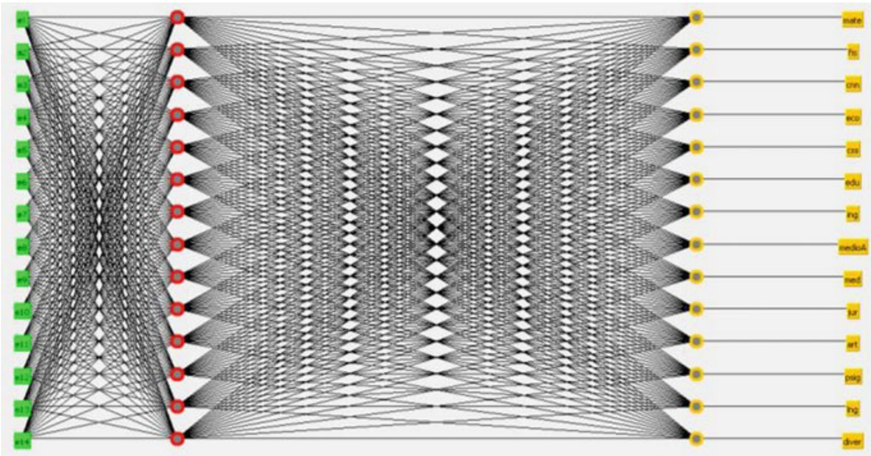


Fig. 2 Result of the network in Weka

2.3 Learning from the Network

Supervised learning is characterized by knowing how the network should respond to a given input. In this way, the desired output is compared with the mains output, and if there are discrepancies, the weights are iteratively adjusted. Thus, the learning stage aims to minimize the error between the output provided by the network and the desired or true output [9, 10].

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances        1152               97.4619 %
Incorrectly Classified Instances        30                2.5381 %
Kappa statistic                          0.9717
Mean absolute error                      0.011
Root mean squared error                  0.0629
Relative absolute error                  8.5963 %
Root relative squared error             24.8401 %
```

Fig. 3 Sample 1 result

3 Results

These results are based on the two training sets of 2000 and 4000 documents. Then, the results of each are presented and discussed. Weka (Waikato Environment for Knowledge Analysis) [11, 12] was used to develop the corresponding training tests to classify the texts according to the categories, which allows the verification and constancy of the results of the correct and bad classification that may be generated.

3.1 Sample 1

Figure 3 shows the first set of data that consists of 2000 documents, which results present a high percentage of correct classification, indicating that there is a good classification. Then, their results are shown in percentages of classification versatility and margin of error.

When analyzing the sample 1, the instances of correct classification show 97% of well-classified documents and a minimum error margin of 3%, indicating a greater range of correct classification of documents.

In Fig. 4, the confusion matrix shows the type of correct and incorrect predictions about the set of documents. It makes it possible to understand how the network makes a mistake when trying to classify the new set of documents. In the graph of this matrix, the correct predictions are represented on the diagonal [13].

3.2 Sample 2

Figure 5 shows that the second set of data included 4.000 documents. As the number of documents increased, it was necessary to add new words to the vocabulary in order to keep the assertiveness margin from decaying.

```
=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   j   k   l   m   n   <-- classified as
 164   0   0   0   0   0   0   0   0   0   0   0   0   0 |   a = mate
   0 198   0   0   0   0   0   0   0   0   0   0   0   0 |   b = fis
   0   0  98   0   0   0   0   0   0   0   0   0   0   0 |   c = cnn
   0   0   0 120   0   0   0   0   0   0   0   0   0   1 |   d = eco
   0   0   0   1 141   0   0   0   0   0   0   0   0   0 |   e = css
   0   1   0   0   0 109   0   0   0   0   0   0   0   0 |   f = edu
   0   2   1   0   0   0  13   0   0   0   0   1   0   0 |   g = ing
   1   0   1   0   0   0   0  69   0   0   0   0   1   1 |   h = medioA
   0   1   1   0   0   0   0   0  24   0   2   0   0   0 |   i = med
   0   0   0   0   1   1   0   0   0  51   0   0   0   2 |   j = jur
   0   0   1   0   2   0   0   0   0   0  20   0   0   0 |   k = art
   2   2   2   0   0   0   0   0   1   0   0  62   0   0 |   l = psig
   0   0   0   0   0   0   0   0   0   0   0   0  21   0 |   m = lng
   1   0   0   0   0   0   0   0   0   0   0   0   0  62 |   n = diver
```

Fig. 4 Confusion matrix of sample 1

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances        2223               89.3489 %
Incorrectly Classified Instances       265               10.6511 %
Kappa statistic                          0.8819
Mean absolute error                      0.023
Root mean squared error                  0.1109
Relative absolute error                 17.8743 %
Root relative squared error             43.7063 %
```

Fig. 5 Results of sample 2

The result of this new training group shows a value of 90% of correct instances and with a margin of error of 10%. Unlike the first group, the margin of assertiveness was reduced by a minimum percentage due to the large number of documents to be analyzed, that is, as the number of documents to be analyzed increases, the vocabulary should be increased.

As shown in Fig. 6, the confusion matrix identifies the type of correct and incorrect predictions about the data set. It makes it possible to understand in what sense the network is mistaken when trying to classify the new texts [14].

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	<-- classified as
153	0	4	1	2	4	0	0	0	0	1	0	0	0	a = mate
1	198	2	4	0	3	6	1	0	1	5	0	2	0	b = fis
2	0	121	3	2	5	1	8	0	0	2	1	1	0	c = cnn
1	0	1	113	2	4	0	0	1	2	2	1	0	0	d = eco
3	0	0	2	146	8	1	2	0	3	7	1	1	0	e = css
0	0	1	1	1	357	2	1	1	3	4	0	2	0	f = edu
4	4	1	1	2	9	354	3	1	2	7	1	5	0	g = ing
2	2	3	2	3	4	4	267	3	0	2	0	6	1	h = medioA
6	0	3	1	0	3	2	2	165	0	1	0	0	0	i = med
4	0	0	2	2	4	1	0	0	104	4	1	1	0	j = jur
3	1	0	0	0	2	1	2	0	0	101	0	2	0	k = art
5	1	3	0	0	2	1	1	0	2	1	22	1	0	l = psig
0	0	1	1	0	1	0	0	0	0	0	0	120	0	m = lng
1	0	1	1	0	1	0	0	0	0	1	0	0	2	n = diver

Fig. 6 Confusion matrix of sample 2

4 Conclusions

The use of the Weka tool allows the execution of training tests of neural networks with the purpose of predicting the area of belonging of a text [15]. To ensure a good classification of documents, it was necessary to increase a greater number of words to the vocabulary in sample 2 in order to give a better accuracy in the classification of documents.

Correctly defining the vocabulary of each of the categories makes the classification to have a good percentage of successes, correctly assigning the document to its category, in such a way that the performance of the network accuracy improves according to the size of the vocabulary. The use of metadata helped achieving better results in the representation, localization and retrieval of electronic resources.

References

1. Vásquez, A.C., Lazo, O.R., Agnelli, R.C.: Categorización de Textos mediante Máquinas de Soporte Vectorial. Revistas Signos, pp. 1–24 (2011)
2. Mendoza, M., Ortiz, I., Rojas, V.: Categorización de texto en bases documentales a partir de modelos computacionales liviano. Revista de investigación de Sistemas e Informática. **10**(1), 2–12 (2013)
3. Pérez, P.M., Colarte, J. (Feb, 2007) Multimedia para discapacitados. Presentada en: Congreso y Feria Internacional Informática 2007 (en línea). Disponible en: http://www.informaticabana.cu/eventovirtual/educacion/discapacitados.pdf
4. Bechara, J.E.A., Cruz, J.C.T., Ceballos, H.V.: Predicciones de modelos econométricos y redes neuronales: el caso de la acción de SURAMINV. Semestre Económico Universidad de Medellín. **12**(25), 95–109 (2009). Available from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120634620090003000007&lng=en&nrm=i-so. Access on 07 Aug 2017

5. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pp. 487–499 (1994)
6. Hahsler, M., Karpienko, R.: Visualizing association rules in hierarchical groups. J. Bus. Econ. **87**, 317–335 (2017)
7. Silverstein, C., Brin, S., Motwani, R., Ullman, J.: Scalable techniques for mining causal structures. Data Min. Knowl. Discov. **4**(2–3), 163–192 (2000)
8. Amelec, Viloria, Carmen, Vasquez: Relationship between variables of performance social and financial of microfinance institutions. Adv. Sci. Lett. **21**(6), 1931–1934 (2015)
9. Viloriaa, A., Lezamab, O.B.P.: Improvements for determining the number of clusters in k-means for innovation databases in SMEs. Procedia Comput. Sci. **151**, 1201–1206 (2019)
10. Kamatkar, S.J., Kamble, A., Viloria, A., Hernández-Fernandez, L., Cali, E. G.: Database performance tuning and query optimization. In: International Conference on Data Mining and Big Data. Springer, Cham, pp. 3–11 (2018)
11. Viloria, A., et al.: Integration of data mining techniques to postgreSQL database manager system. Procedia Comput. Sci. **155**, 575–580 (2019)
12. Lanzarini, L., Villa Monte, A., Aquino, G., De Giusti, A.: Obtaining classification rules using lvqPSO In: Advances in Swarm and Computational Intelligence. Lecture Notes in Computer Science. vol. 6433, pp. 183–193. Springer, Berlin, Heidelberg (2015)
13. Borja-Borja, M.G.: Algoritmo de Entrenamiento de una Neurona Artificial Perceptrón para Reconocimiento de Caracteres, Aplicando Principios Heurísticos. Revista ECIPerú. **6**(1), 4 (2009)
14. Alonso, M.Á.L.: Las estructuras conceptuales de representación del conocimiento en internet. Scire: representación y organización del conocimiento. **6**(1), 107–123 (2000)
15. Torrez Torrez, E.D.: Sistema inteligente para la detección de conversaciones con posible contenido pedofílico basado en redes neuronales, Doctoral dissertation (2018)

Plant Leaf Diseases Recognition Using Convolutional Neural Network and Transfer Learning

J. Arunnehru, B. S. Vidhyasagar and H. Anwar Basha

Abstract In field of modern agriculture, artificial intelligence plays a major role in crop protection. Plant diseases have always been a cause of great concern to plant growth and crop cultivation around the globe. Plant diseases can affect plants from day-to-day activities. These diseases not only have serious consequences on plants health but also on human health affecting in various ways such as spreading viruses, bacteria, and fungi causing infections. The improvement in computer vision and increasing smartphone penetration have paved the way for deep learning possible through smartphone-assisted diagnosis. Deep learning is used on a large amount of data and it is a self-learning technique. We propose an additional method to classify the diseased leaves using the transfer learning on top of convolutional neural network model to improve the efficacy of image processing while applying deep learning.

Keywords Artificial intelligence · Leaf diseases · Image segmentation · Object detection · Convolutional neural network · Transfer learning

1 Introduction

Development in technologies has given us the facility to grow enough eatables to meet the requirements of more than 7.5 billion people. However, the quality of the raw food still remains questionable because of the number of variables such as climate change, plant diseases, the decline in pollinators, and others. Leaf diseases are a red flag for food security at the international level. They also have grave consequences for small-scale farmers whose livelihood depends on the good health of these plants. More than eighty percent of the food production is generated by small-scale farmers and it is noted that more than 50% of this total production is lost to the pests. This makes the small-scale farmers a community that's particularly affected by pathogen-derived disruptions in food supply.

J. Arunnehru (✉) · B. S. Vidhyasagar · H. Anwar Basha
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani, Chennai, Tamilnadu, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_21

Numerous measures have been put in places to stop plant loss to these pathogen-derived diseases. Widespread usage of pesticides has in the past ten years increasingly been supplemented IPM approaches. Identifying a disease correctly in its early development period is a crucial step for efficient diseases management. In the past, these disease identification has been performed by some agricultural wing organizations or other institutions. In recent times, such efforts have additionally been performed by organizations by sending data and using online diagnosis. In addition, organizations are eyeing mobile phone technology which is rapidly growing by building tools based on smartphones.

The deep neural network has been successful in many diverse domains in recent years, for example, end-to-end learning. It provides a mapping between inputs to an output that is an image of a diseased plant mapped with crop disease pair. The mathematical function that takes input from the incoming edge and provides a mathematical output as an outgoing edge is called nodes in a neural network.

2 Related Work

To develop an application like disease-specific chemical and fungicide applications for the appropriate management strategy using pesticide and fertilizers which leads to control and to maintain crop health from diseases. This can be a great help in keeping the disease under control and boosting production. In [1], the paper discusses the dependable monitoring system for crop cultivation and facilitate advancement in cost-efficient. In addition to that, plant disease detection uses volatile profiling-based, spectroscopic, image-based technologies are been used. In the papers [2–5], RNA analysis, microscopy, and nucleic acid approach were used for disease recognition. In the paper [6], the color feature extraction method was used for the disease classification. Plant disease can also be detected using shape feature extraction as discussed in the paper [7]. Similar to the paper [7], the plant disease can also be recognized using texture feature extraction according to paper [8]. Paper [9], combines the process followed in the paper [7, 8] forming a more robust method for better classification method. In paper [10], the authors proposed a remedial measures approach with neural network for disease management in agricultural crops. In paper [11], the author incorporates the feature extraction by particle swarm optimization [12] and the infected leaf spot of cotton is determined by the feed-forward neural network for improving the accuracy. Support vector machine algorithms are been used by the author in the paper [13], for the plant disease detection and differentiation. In paper [14, 15], authors present an approach for recognizing tea leaf disease using neural network-based ensemble learning for feature extraction and obtained testing accuracy of 91%. In paper [16], the author presents the ANNs technique along with the k-means algorithm for automatic detection of diseases.

3 Feature Extraction

To extract the features from the leaf using its color, the leaf is first converted into a grayscale image by applying the average method Eq. 1, where I_R is image red channel, I_G is image green channel, and I_B is image blue channel, this then converts the image to gray scale and then is converted into binary sequence through binarization, and the contour is then extracted. The features are extracted with the help of the characteristic of this contour line is shown in Fig. 1.

$$Grayscale = (I_R + I_G + I_B)/3 \tag{1}$$

In addition, plant diseases identification can also be done using texture features extraction, converting the image to gray scale and by segmenting the images into parts and then training the neural net on those images. The segmentation process is used to get rid of any background in the images that get trained which may lead to false true while prediction because of the background part of different images was also trained as a part of the leaf image.

3.1 Image Pre-processing and Labeling

While collecting the different images for training, there is a possibility that images are of different sizes and resolutions and formats which may affect the training process, so the images are converted into a common size, resolution, and format then trained. Same images of different orientation are trained so that while classification the model can recognize the diseased leaf if the leaf image is in a different orientation of the trained ones. Furthermore, all the images are pre-processed which involves manually duplicating all the images and shrinking the background to the extent the leaves part of the image get highlighted. Images which have low resolution and dimensions less than 550 px were excluded. Images were resized to 256 × 256 because very large image size adds up to the training time. This can be done automatically using the Python script using the OpenCV framework.

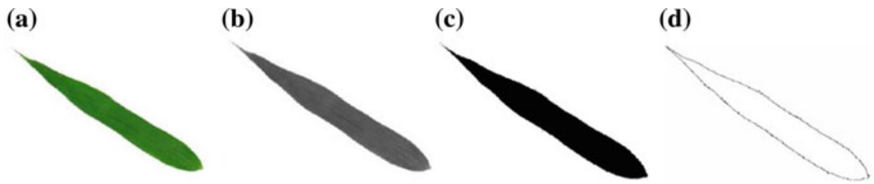


Fig. 1 Process of extracting Contour from Leaf image **a** Input data image. **b** Converted into gray scale. **c** Binary conversion. **d** Contour extraction

3.2 Dataset

Choosing an appropriate dataset is very important as it is required at every stage of the process from the training phase to the evaluating phase. Images are collected from various sources and places and are grouped into fifteen different classes, each class representing a disease type. We need a large, verified dataset of healthy plants and diseased plants to accurately classify the diseased plant. To make sure the neural network has enough data to train itself we have used about 55000 images classified over 38 classes. The performance of our models is measured with the ability to accurately predict the correct leaf-diseases pair.

3.3 Augmentation Process

Next process is to enrich the dataset with augmented images. By adding more augmented images to the datasets, ability of the neural network to learn the apt features increases. A dataset of almost 60,000 images is created spread across 38 different classes. The disease name of that particular leaf is used as the label name for its respective classes so that when the classifier finds its corresponding class, we can also know the name of the disease along with the leaf name. Augmentation process is done to increase the size of the dataset and to introduce a small amount of distortion on some images which help to keep overfitting on the check during the training stages. The image augmentation can be performed by perspective transformation, affine transformation, and many more. Simple image rotation is also a very effective way of augmentation. To express rotations and translations, affine transformation is used. The transformation matrix is formed with three points from the original image and their corresponding output image location. A 3×3 transformation matrix is required for the perspective matrix. The sample images are shown in Fig. 2.

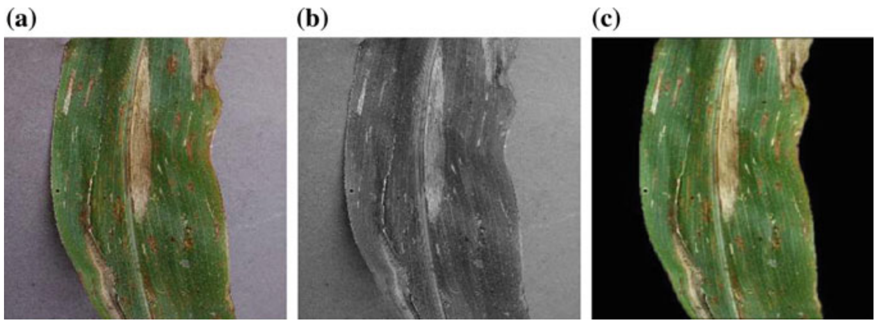


Fig. 2 Augmentations of a leaf image **a** Input data image / Raw. **b** Converted into gray scale. **c** Segmentation

4 Convolutional Neural Network Training

Artificial intelligence coupled with an appropriate deep learning framework can amplify the overall scale that can be achieved within those domains. The machine-learning field is constantly evolving. The challenge is to make models that can run on mobile platforms. Today, we have a wide range of frameworks available at our disposal. Each is built differently for different purposes. Some of the well-known frameworks are Tensorflow, caffe, PyTorch, Chainer, Keras, and Deeplearning4J. TensorFlow can be used on both desktop and smartphones. It is very flexible and supports various languages such as R, Python, and C++ to create deep learning models along with wrapper libraries. TensorFlow is widely used for these two reasons.

1. TensorBoard for effective data visualization of network modeling and performance
2. TensorFlow Serving for instant deployment of new workflows while retaining the same server architecture and APIs. It provides integration with other models which are different from the conventional practices and can be extended to serve different model and data types.

Figure 3 represents the architecture of the pre-trained neural network, which, in this case, is the $Inception\_V3$, trained on the $ImageNet$ dataset. As it is clear from the architecture that there various layers which do the feature extraction process, bottleneck creation, region of interest identification, and much more. We use all the layers but the last one. That's where the role of transfer learning starts. We remove the last layer of this neural net and append it with our own custom layer which has been trained on the given dataset of leaf images. This layer is then responsible for the classification and it is the one which determines the accuracy of the model. Every convolutional layer has maps of equal size (m), M_a, and M_b and a kernal of size K_a and K_b is replaced over the certain regions of the original image. The hopping factor / skipping factor S_a and S_b represents how many pixels have to be skipped in respective x and y directions between subsequent convolutions [8].

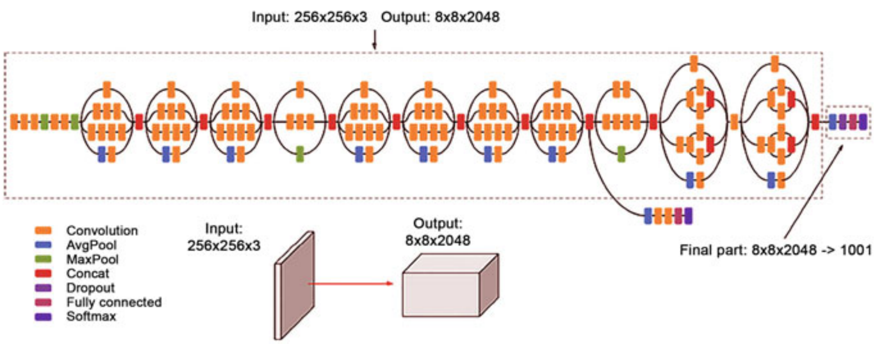


Fig. 3 Architecture of the Inception_V3 CNN

The softmax function is used as an activation method for all the back propagations in our recurrent neural network. Softmax is handy because it converts the result of the final layer in your neural network into a probability distribution. Softmax is used to optimize the values of the weight supplied to the nodes of the neural net in order to find the global maxima or minima each iteration of the recurrent neural network.

4.1 Transfer Learning

Transfer learning is a machine-learning method where a pre-existing trained model is reused as the input for a model on a second task. It is mostly used with a deep learning process where pre-trained models are used as the initial point on NLP tasks and computer visions. This allows you to rapidly improve performance when modeling the second task. Note that it works with deep learning only if the model features learned from the first task are general and this type of transfer learning is called an inductive transfer. There are two types of transfer learning: (1) Develop model approach, (2) Pre-trained model approach. We have used pre-trained model approach which consists of three subsections: (1) Select source model, (2) Reuse model, and (3) Tune model. The reuse model under the pre-trained model is the approach used here. It helps us reduce training time significantly.

4.2 Performed Tests

The well-known way to test the performance is to split the data into a test set and training set using the test set data for prediction. Now since the original output for the testing set and our predicted outcome are known, the accuracy of our prediction can be calculated. To measure this accuracy, we use tenfold cross-validation technique. It was constantly been replaced after every thousand training iteration.

5 Experimental Results and Discussions

The results presented here are based on the training related to the whole database containing both pre-processed image and augmented images. It is well known that convolutional network learns features when trained on a large database so the results supporting the original images are not explored here. The model was trained repeatedly and it got an average overall accuracy of 96%. The training model was tested on all the classes individually.

A test was performed from all the images in the validation set. It was noted that the accuracy was slightly less for classes which contained less training data images. The dataset was trained on three different types of datasets such as color, grayscale, and segmentation. Different Python functions were used to convert the images into the required input type and it was noted that images trained on segmentation type of dataset had the highest prediction value. It was also noted that there was on an average 3% improvement in results just but training the images on a segmented dataset compared to the color database. From Table 1, the model performance has been evaluated and shown. The reason for the high consumption of time in training is because it was done on a CPU. But either spinning up a VM or increasing the core/RAMs or running it on a GPU/TPU, the time for training can drastically reduce.

Table 2 denotes the prediction results based on the performance of all the three models, namely raw, gray scale, and segmented. All the test images were of a single class and the accuracy denotes the probability of that image being that "class." It is observed that when the augmentation of the images are done and given for predictions for all the three models, the best performance is done by the raw model no matter what the type of augmentation is performed on the test images. But if you are aiming for high accuracy, it is highly recommended that the augmentation should be a segmentation and the model to be the one trained on raw images. Figure 4 shows the accuracy graph obtained for training and validation with maximum iteration of 4000 epochs.

Table 1 Evaluation of raw, gray scale, and segmented model performance

	Raw_Model	Grayscale_Model	Segmented_Model
Image size	256 × 256	256 × 256	256 × 256
Image count	54307	54307	54307
Training time	5 h 10 m	4 h 3 m	5 h 30 m
Accuracy	92%	88%	90%

Table 2 Accuracy rate (%) obtained for raw, gray scale, and segmented models on five test images

Test images	Raw_Model	Grayscale_Model	Segmented_Model
Image 1	98.68	97.34	86.17
Image 2	95.48	90.33	90.21
Image 3	93.48	58.03	91.83
Image 4	84.47	81.6	51.53
Image 5	89.29	73.48	74.46
Average	92.28	80.156	78.84

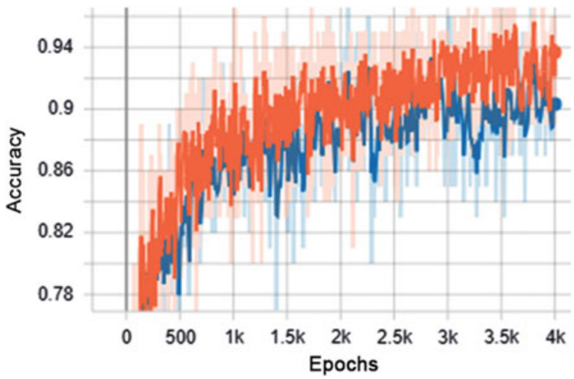


Fig. 4 Accuracy of the fine-tuned CNN

6 Conclusion

There are several strategies in automatic or computer-based leaf disease identification and classification method. However, there are no business products available in market, excluding those that are managing plant species identification on the basis of leaf images. This work explores a novel methodology using deep learning and transfers learning to automate the classification and detection of leaf diseases from the images of plant leaf. The proposed methodology is equipped to detect the presence of the leaf disease and differentiate the healthy leaves from the diseased ones which can be diagnosed visually. The model was executed by collection of the images to use for training and validation, pre-processing and augmentation of images, and eventually the procedure of coaching the deep CNN and fine-tuning.

In this work, we collected 60,000 leaf disease images from various Internet sources and performance was evaluated with different test cases. The experimental results achieved accuracy between 92 and 97% In addition to that, augmentation approach had bigger influence in achieving respectable higher results. In comparison with alternative techniques used and bestowed, comparable or maybe higher results were achieved. An extension of this study is to collect more images of leaf disease plants for enriching the information and getting better accuracy of the model using different techniques.

References

1. Sankaran, S., Mishra, A., Ehsani, R., Davis, C.: A review of advanced techniques for detecting plant diseases. Comput. Electron. Agric. **72**(1), 1–13 (2010)
2. Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L.R., Davis, C.E.: Advanced methods of plant disease detection. A review. Agron. Sustain. Dev. **35**(1), 1–25 (2015)

3. Mahlein, A.K., Rumpf, T., Welke, P., Dehne, H.W., Plumer, L., Steiner, U., Oerke, E.C.: Development of spectral indices for detecting and identifying plant diseases. Remote. Sens. Environ. **128**, 21–30 (2013)
4. Xiuqing, W., Haiyan, W., Shifeng, Y.: Plant disease detection based on near-field acoustic holography. Trans. Chin. Soc. Agric. Mach. **2** (2014)
5. Mahlein, A.K., Oerke, E.C., Steiner, U., Dehne, H.W.: Recent advances in sensing plant diseases for precision crop protection. Eur. J. Plant Pathol. **133**(1), 197–209 (2012)
6. Chaudhary, P., Chaudhari, A.K., Cheeran, A.N., Godara, S.: Color transform based approach for disease spot detection on plant leaf. Int. J. Comput. Sci. Telecommun. **3**(6), 65–70 (2012)
7. Patil, S.B., Bodhe, S.K.: Leaf disease severity measurement using image processing. Int. J. Eng. Technol. **3**(5), 297–301 (2011)
8. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D.: Deep neural networks based recognition of plant diseases by leaf image classification. Comput. Intell. Neurosci. (2016)
9. Reed, T.R., Dubuf, J.H.: A review of recent texture segmentation and feature extraction techniques. CVGIP: Image Underst. **57**(3), 359–372 (1993)
10. Babu, M.P., Rao, B.S.: Leaves recognition using back propagation neural network-advice for pest and disease control on crops. IndiaKisan. Net: Expert. Advis. Syst. (2007)
11. Revathi, P., Hemalatha, M.: Identification of cotton diseases based on cross information gain deep forward neural network classifier with PSO feature selection. Int. J. Eng. Technol. **5**(6), 4637–4642 (2014)
12. Trelea, I.C.: The particle swarm optimization algorithm: convergence analysis and parameter selection. Inf. Process. Lett. **85**(6), 317–325 (2003)
13. Rumpf, T., Mahlein, A.K., Steiner, U., Oerke, E.C., Dehne, H.W., Plümer, L.: Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Comput. Electron. Agric. **74**(1), 91–99 (2010)
14. Hansen, L.K., Salamon, P.: Neural network ensembles. IEEE Trans. Pattern Anal. Mach. Intell. **10**, 993–1001 (1990)
15. Karmokar, B.C., Ullah, M.S., Siddiquee, M.K., Alam, K.M.R.: Tea leaf diseases recognition using neural network ensemble. Int. J. Comput. Appl. **114**(17) (2015)
16. Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M. and ALRahamneh, Z.: Fast and accurate detection and classification of plant diseases. Int. J. Comput. Appl. **17**(1), 31–38 (2011)

A Shape-Based Character Segmentation Using Artificial Neural Network for Mizo Script

J. Hussain and Vanlalruata

Abstract This paper presents a new character segmentation algorithm by using shape-based-oriented feature vectors for offline printed Mizo script document. The approach in this technique involves area estimation for each isolated blobs which forms a different type of shape, finding the required amount of morphological dilation based on each estimated area, training the artificial neural network (ANN) to map between the area of the shape with the required amount of morphological dilation. This experiment is performed on a font size ranging from 8 to 72, on a dataset containing 70 printed documents each having approximate 200 characters, forming a total of 14,000 characters for single font size. In total, an approximate of 14,000 $\times$ 16 = 224,000 characters is considered. The experimental results show that using feed forward back propagation (FFBP) algorithm for both segmentation and classification achieved the best result with a 97% accuracy rate of recognition.

Keywords Mizo character segmentation · Mizo OCR · Artificial neural network · Character segmentation techniques · Feature extraction · Image processing · Character recognition · Image processing

1 Introduction

Mizo optical character recognition (OCR) translates images of Mizo printed script into a machine-readable format. Segmentation plays a vital role, and an effective segmentation technique enhances the performance of OCR [1]. In the Mizo script, complications in segmentation arise due to multiple disjoint characters' shapes.

Mizo scripts are mostly derived from Latin character, and hence, it exhibited similarity with English character, except that special characters are known as vowel are included in the Mizo script such as Â, â, Ê, ê, Î, î, Ô, ô, Û, û, Ṭ and ṭ. These vowel characters with dot at bottom and circumflex at the top have separate meanings and different tone. Base on the tone, the meaning also changes. Hence, these vowel characters play a crucial role while reading, writing and speaking Mizo language.

J. Hussain · Vanlalruata (✉)
Department of Mathematics and Computer Science, Mizoram University, Aizawl, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_22

Fig. 1 Isolated shape (Blobs) segmentation problem

The Mizo vowel segmentation problem using a direct bounding box [2] technique is depicted in Fig. 1. To counter this segmentation problem, the initial step is to merge these separate blobs into single blobs so that the character can be segmented as one entity. The overview of this technique is represented in Fig. 2.

In this experiment, the document image under consideration is duplicated, and named as a foreground and background image. The proposed segmentation technique is performed in the background image. A bounding with similar size from the background image is drawn on the foreground image on which the actual segmentation is to be performed. Therefore, in this case, the background image only acts as a reference point to determine the bounding box size. This technique of segmentation works well only when considering a static font size. However, when there is a variation in the font size, there is a fault in this segmentation. Since the amount of dilation apply is static, the isolated blobs may not merge as the font size varies. The problem is depicted in Fig. 3. Therefore, to handle this problem, a new technique which adapts to font variation is proposed by making the amount of dilation variable dynamic. To achieve this, an ANN has been trained with multi-font size to handle character segmentation irrespective of its font size.

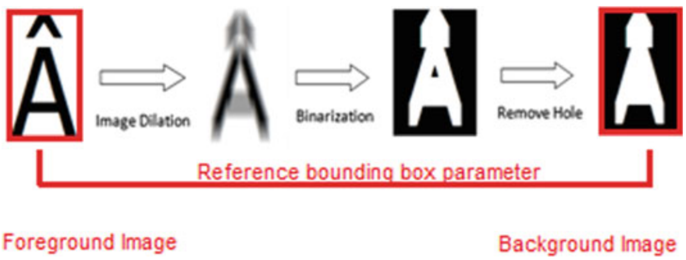


Fig. 2 Stages of character segmentation

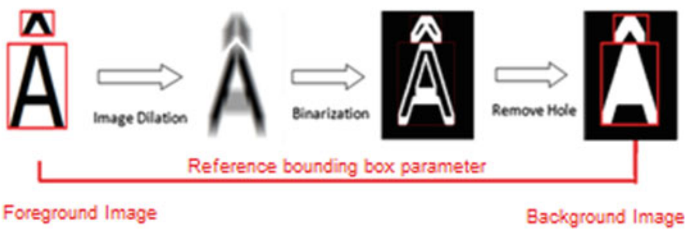


Fig. 3 Multi-font size character segmentation problem

2 Related Work

Several segmentation techniques have been proposed in order to achieve high recognition accuracy. However, based on different writing style and script diversity, a unified segmentation method is an open problem.

Upreti et al. [3] have works with handwritten Hindi character segmentation in which the header lines are extracted and eliminated from the rest of the script. Inspite of clamming an accuracy of 93%, may problem have not been address including, overlapping character, broken letter etc. Saba et al. [4] paper presents a survey on touching character segmentation, in which both number and letter in English are consider. He concludes that no technique is found to be accomplished with 100% accuracy. Kumar et al. [5] paper deals with upper zone touching character for Gurmukhi script. This proposed technique achieved an accuracy of 91%, where parameter tuning can be performed to enhance. Kumar et al. [6] again work with segmentation of touching character including isolated character. The proposed technique is based on water reservoir base and fills the empty spaces with black dots. The paper claimed an accuracy of 92%. Siddharth et al. [7] work with segmentation enhance with skew detection and correction. The work concludes that the limitation is the noise present in the document, which corresponds to mis-calculation in width and height. The paper claimed an accuracy of 94%. The work of Almuallim and Yamaguchi [8] performed word segmentation using strokes, geometrical and topological features. A recognition rate of 91% was achieved. This paper also concludes that the failure of recognition is mainly due to the segmentation error. Erlandson et al. [9] implement the shape of the word in Arabic, based on this, a unique feature of each shape is retrieved, and this retrieved information is then matched with a database which stores similar feature vector. Amin [10] again used the same attributes (feature vector) obtained from words. These vectors are then fed into an ANN and try to obtain the segmentation information. Khorsheed and Clocksin [11] implement the segmentation-free approach to counter the existing projection method. The order of the word in terms of sequence is normalized. Template features are used to match the recognition patter. Tse et al. [12] work on a cursive character. To counter the overlapping or joint character, erosion technique is implemented. Feature extraction is performed using connected component. Combination of this at times generates the best accuracy in terms of recognition. Ramsis et al. [13] focus on the problem of overlapping by isolating the baseline in between the connected characters. However, variation in font size and the pattern of the input account for decrease in performance. The authors did not report the accuracy percentage. Broumandnia et al. [14] work on a technique known as edge detection. The baseline known as horizontal edge is detected using horizontal projection on the valley point in the projected graph which indicates the separation points, and ANN is used to distinguish the correct valley point for performing segmentation. Garain et al. [15] worked for machine printed touching character segmentation using factors like transitions, blob thickness, area of blobs and the centric point.

3 Proposed Work

The shape base character segmentation can be decomposed into several stages as depicted in Fig. 4.

3.1 Find the Area of Isolated Shape (Blobs)

Areas of the circumflex (ˆ) and dot (.) present in the Mizo script for different font size are calculated in this stage. The training data in consideration is a noise-free image document; therefore, the detected smallest area of isolated blobs is considered to be the area of dot (.), and the second smallest area is considered to be the area of circumflex (ˆ). Proceeding in this pattern, we find all the areas of the shape of circumflex and dot for different font size. The area of the shape is to identify using connect component analysis (CCA) [16].

All the areas of this blobs present in the document image are calculated and stored in an array name *bwsize[]*. Since, the purpose is only to find the area of circumflex and dot, which is considered to be the smallest and second smallest from *bwsize[]*. The two values are assigned to T_1 and T_2, respectively. All areas smaller or equal to T_1 are considered to be dot, and area greater than T_1 but smaller or equal to T_2 is consider to be the area of circumflex. In this process, all the areas for shape of circumflex and dot are calculated with a font size ranging from 12 to 71 points.

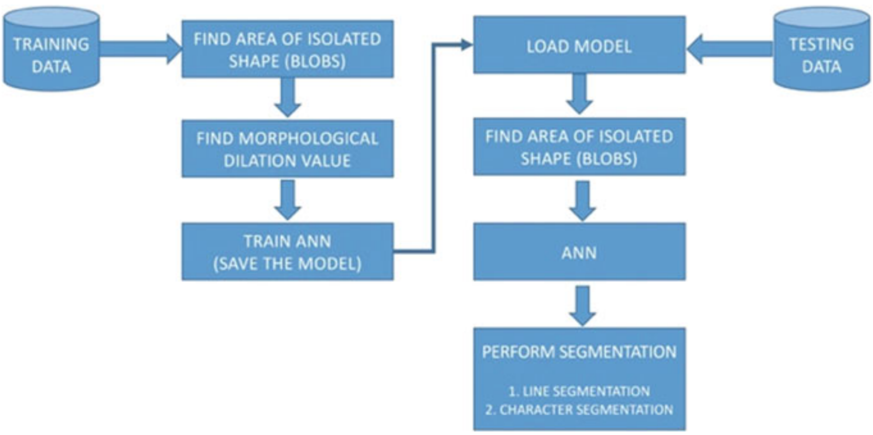
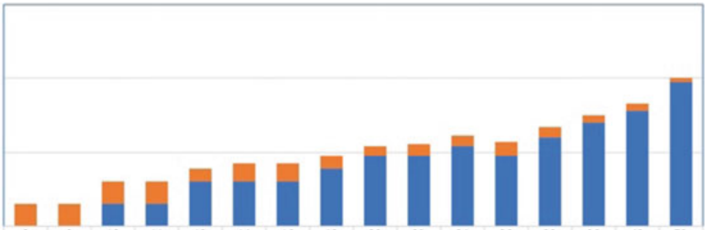


Fig. 4 Diagramatic representation of the proposed methodology



Font Size	8	9	10	11	12	14	16	18	20	22	24	26	28	36	48	72
Morhpological Dilation Value	1	1	2	2	2	3	3	3	3	4	5	5	6	7	9	13
Area of Dot	1	1	2	2	4	4	4	6	9	9	12	9	16	25	36	90

Fig. 5 Dilation value corresponds to area of dot

3.2 *Find the Required Amount of Morphological Value*

In this phase, the amount of morphological dilation that have to be applied so the isolated shape of single Mizo script merge into one entity is identified. Incorrect amount of morphological value results in segmentation fault. This identification of morphological values is performed in hit and trail method. Figure 5 gives the corresponding morphological dilation value based on the area of the shape of dot.

3.3 *Train ANN*

In training the network, the areas of circumflex and dot are used as an input and dilation value as a target to the neural network. The input will be of size 16 × 2 matrices, where 2 corresponds to the area of circumflex and dot and 16 corresponds to the number of font size taken into consideration. The target is also composed of 16 classes. Each of these classes corresponds to the amount of morphological dilation values that have to be applied in order to merge the isolated shape into one entity. After training the ANN, the model is saved, which will be used for the actual segmentation. This model likewise will accept input for the shape area of circumflex and dot. The output will be a morphological dilation value.

3.4 *Load Model*

In segmentation, the train model will be loaded and will be used to generate the amount of dilation variable, using this variable a morphological operation of dilation is performed so that the isolated blobs are merged into single entity irrespective of the font size. This makes the segmentation technique adaptive.

Fig. 6 Line segmentation

3.5 Line Segmentation

In this operation, a morphological dilation using the value generated by the ANN is used to merge the character from the background image. After connecting these isolated shapes in the background image, vertical projection profile is performed on the images. This vertical projection generates a histogram. The value of this histogram represents the pixel density distribution. This density distribution only highlights the pixel greater than zero, which in our cases is the character itself, meaning that when there is no character, then the density distribution value will be zero, and this zero projection is known as the valley point. The operation of this line segmentation is depicted as Fig. 6.

3.6 Character Segmentation

Character segmentation is performed using the bounding box of connected component. The output of the segmented line from the previous stage is the process for character segmentation. In this process using connected component analysis, all the connected black pixels are border in a rectangle box known as the bounding box. If the neighboring pixel is not black, then it will be the border by a separate bounding box. In this way, each of the isolated shape will be border by a separate bounding box. In other words, the number of bounding box equals the number of character. This image processing is again performed in the background image. Since, all the isolated shapes are merged from the previous stage, reprocessing is not required. The coordinated location of the bounding box in the background image is used as a reference to draw the same box in foreground image. When a connected region is identified with a bounding box, the coordinates of bounding including the width and height can be determined. Using these determined coordinates, character segmentation is performed in the foreground image.

The character segmentation is depicted in Fig. 7.

Fig. 7 Character segmentation

4 Result Analysis

Mizo character image documents were collected and tested mainly from Mizo daily newspapers. It is observed that line segmentation is done with 100% accuracy. However, character segmentation is accurate as long as the document contains non-touching character. The erosion applied on the touching character is static; therefore, it cannot handle font variation at times. Mizo numerals and special character present in the document do not affect the accuracy in segmentation since by default each numeral is single blobs character. A Mizo character which starts with circumflex and dot as the first character mostly accounts for segmentation fault. This may be due to unbalance area of the shape to the morphological dilation value trained to the ANN. The image document is manipulated for testing by changing its font size within the range of 8–72 as on a document containing an approximate 200 characters shown in below table. The segmentation and recognition accuracy with feed forward back propagation algorithm are given Tables 1, 2 and 3, respectively.

Table 1 Line segmentation accuracy

Font size	Line segmentation		
	Total no. of lines	Correct segmentation	Accuracy
12–72	5600	5600	100%

Table 2 Character segmentation accuracy

Font size	Line segmentation		
	Total no. of characters	Correct segmentation	Accuracy
12–72	224,000	219,744	98.1%

Table 3 Recognition accuracy

FFBP neural network				
Total no. of characters	Correct recognized	Misrecognize character	No. of occurrence	Accuracy
224,000	217,280	!	510	97%
		1	851	
		û	2894	

5 Conclusion

In this paper, we have presented a new technique for Mizo character segmentation for multi-font size, based on the shape and size of the isolated blobs. A hybrid feature extraction [17] is applied for ANN recognition and classification. Line segmentation is performed with 100%. However, character segmentation declines by 1.99% achieving 98.1%. The decline in the accuracy is primarily due to touching character present in some sequences of Mizo word. In spite, morphological erosion operation is applied to disassemble the joint, yet a minimal percentage of error still persists. This error rate may be able to reduce if the erosion value is also trained along with the dilation value to the ANN. Apart from this, the misrecognized character may be improved by using a more robust feature extraction method. These issues will be dealt with in the future for printed and handwritten documents in Mizo script by using various approaches.

References

1. Priyanka, N., Pal, S., Mandal, R.: Line and word segmentation approach for printed documents. IJCA Special Issue on Recent Trends in Image Processing and Pattern Recognition-RTIPPR, pp. 30–36 (2010)
2. Srivastav, A., Sahu, N.: Segmentation of devanagari handwritten characters. Int. J. Comput. Appl. **142**(14) (2016)
3. Upreti, K.K., Bag, S.: Segmentation of unconstrained handwritten hindi words using polygonal approximation. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). IEEE (2016)
4. Saba, Tanzila, Rehman, Amjad, Sulong, Ghazali: Cursive script segmentation with neural confidence. Int. J. Innov. Comput. Inf. Control. (IJICIC) **7**(7), 1–10 (2011)
5. Kumar, M., Jindal, M.K., Sharma, R.K.: Classification of characters and grading writers in offline handwritten Gurmukhi script. In: 2011 International Conference on Image Information Processing. IEEE (2011)
6. Kumar, M., Jindal, M.K., Sharma, R.K.: Segmentation of isolated and touching characters in offline handwritten Gurmukhi script recognition. Int. J. Inf. Technol. Comput. Sci. **6**(2), 58–63 (2014)
7. Siddharth, K.S., et al.: Handwritten Gurmukhi character recognition using statistical and background directional distribution. Int. J. Comput. Sci. Eng. (IJCSE) **3**(06), 2332–2345 (2011)
8. Almuallim, Hussein, Yamaguchi, Shoichiro: A method of recognition of Arabic cursive handwriting. IEEE Trans. Pattern Anal. Mach. Intell. **5**, 715–722 (1987)
9. Erlandson, E.J., Trenkle, J.M., Vogt, R.C.: Word-level recognition of multifont Arabic text using a feature vector matching approach. In: Document Recognition III. vol. 2660. International Society for Optics and Photonics (1996)
10. Amin, A.: Recognition of printed arabic text using machine learning. Document Recognition V. vol. 3305. International Society for Optics and Photonics (1998)
11. Khorsheed, M.S., Clocksin, W.F.: Structural features of cursive arabic script. BMVC (1999)
12. Tse, E., Bigun, J.: A base-line character recognition for syriac-aramaic. In: 2007 IEEE International Conference on Systems, Man and Cybernetics. IEEE (2007)
13. Ramsis, R., El-Dabi, S.S., Kamel, A.: Arabic character recognition system. In: Report KSC027, IBM Kuwait Scientific Center, Kuwait (1988)

14. Broumandnia, A., Shanbehzadeh, J., Nourani, M.: Segmentation of printed Farsi/Arabic words. In: 2007 IEEE/ACS International Conference on Computer Systems and Applications. IEEE (2007)
15. Garain, U., Chaudhuri, B.B.: Segmentation of touching characters in printed Devnagari and Bangla scripts using fuzzy multifactorial analysis. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **32**(4), 449–459 (2002)
16. Chen, Y.K., Wang, J.F.: Segmentation of single-or multiple-touching handwritten numeral string using background and foreground analysis. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1304–1317 (2000)
17. Hussain, J.: A hybrid approach optical character recognition for Mizo using artificial neural network. In: Recent Findings in Intelligent Computing Techniques. Springer, Singapore, pp. 547–553 (2018)

Feasibility Study for a Mini-Hydropower Plant in Dreznica, Bosnia, and Herzegovina

Rodrigo Ramírez-Pisco, Iris Pezic Djukic, Carmen Luisa Vásquez, Amelec Viloria and Noel Varela

Abstract This paper includes an energy, technical, economic, and environmental analysis on the installation of a mini-hydraulic power plant in the Dreznica village in the province of Herzegovina in Bosnia and Herzegovina. The community of Dreznica is relatively isolated since it is completely surrounded by mountains and is composed of about 1500 inhabitants. This region is already provided with an energy system and this study is proposed as an investment in a clean and renewable energy project to contribute in the reduction of greenhouse gas emission.

Keywords Mini hydraulic · Technical analysis · Economic analysis · Environmental analysis · Hydrological study

1 Introduction

The need for alternative and autonomous energy systems in rural communities arises from the difficulties with power transmission lines that are subject to barriers because of isolation (relief) or economic cost [1, 2]. The mini-hydraulic system was selected due to the geography of the Drezanka River (20 km long from its source and with

R. Ramírez-Pisco · I. P. Djukic
UNIBA-Universidad de Barcelona, Barcelona, Spain
e-mail: rramirez@unibarcelona.com

R. Ramírez-Pisco
Universidad Politécnica de Cataluña, Barcelona, Spain

C. L. Vásquez
Universidad Nacional Experimental Politécnica Antonio José de Sucre, Av. Barquisimeto, Bolívar, Venezuela
e-mail: cvasquez@unexpo.edu.ve

A. Viloria (✉) · N. Varela
Universidad de La Costa (CUC), Calle 58 # 5566, Atlántico, Barranquilla, Colombia
e-mail: aviloria7@cuc.edu.co

N. Varela
e-mail: nvarela2@cuc.edu.co

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_23

a height of 485 m). This type of project has been evaluated by private companies in the region but has not been carried out for various reasons. It is also established that the mini-hydraulic power plant will be of the flowing or passing type, that is it will not install a dam or water accumulation system, but the energy will be given by the passage of the current.

After selecting the type of energy solution that best adapts to the region, the technical design of the installation was carried out. The main data are the waterfall and the average annual flow of the water passage in a point of the river, for determining the turbine that best fits. Likewise, the economic analysis of the plant calculated the total cost of the project, including labor and equipment, estimated at approximately € 630,000, with an energy factor of 0.18 €/kWh. Profitability analysis was also carried out, obtaining a value of 5 years.

Finally, an environmental analysis of the project was carried out for evaluating the individual impacts of each component, as well as the consequences and the measures to be taken to alleviate them, developing the environmental impact matrix to evaluate each of the exposed elements. In addition, the environmental impacts of substituting fossil fuel energy sources with renewable energy were considered.

2 Technical and Energy Analysis

Dreznica village is located in the province of Herzegovina in Bosnia and Herzegovina. It is a region of approximately 1500 inhabitants which is completely surrounded by mountains. The partial isolation of the community is an important reason to consider alternative solutions in terms of energy supply. Currently, more alternatives are studied to stop applying the traditional fossil fuels that influence the emission of greenhouse gases, mainly CO_2, which are leading to global warming or climate change. Therefore, this project seeks several objectives for solving different problems. Firstly, supply energy in a rural community and provide cheaper electricity. It is also a technical project that seeks to generate profits for investors and third parties. In addition, no less important, it seeks to collaborate in the transition to decarbonization of the energy model.

The selection of a mini-hydraulic system as a renewable energy source is based on economic evaluation, since this type of plant offers the lowest cost when compared to others such as wind [3] or solar [4], i.e., by investment and energy factor. Likewise, the conditions for the installation of a mini-hydraulic system are favored by the topography of the area and the presence of the Drezanka river, with an appropriate flow and waterfall.

Due to the energy crisis [5, 6] and the technological progress [7], more and more non-conventional energy sources are being built, such as wind turbines, solar energy sources, mini-hydroelectric plants, and biomass. Their construction has essentially a positive, social, economic, and environmental protection character, since they are mainly renewable sources. As a general rule, all countries with a market environment allow the free production of these sources (special regime). In other words, the market is forced to buy all the energy produced at a regulated price, in order to encourage construction. These facts have led to the fast development and construction of renewable sources and all predictions indicate that this trend will continue in the future [8–10].

The electricity system in Bosnia operates on a market basis, where production with renewable sources has priority in meeting the demand. Dreznica village is located in a mountainous area at north of Mostar municipality with the 20 km long river passing through the village. Around the river, there are several small villages. The electricity supply for the area consists of a 20 kV radial distribution power line with several transformation points for the supply. Peak demand is estimated around 3 MW. The problem with this supply is that any failure causes a cut in the whole area. In this sense, a mini-hydropower plant improves the security of supply in this area. The river has a capacity to build several mini power plants with a total capacity of 10 MW. That is, they could provide the total electricity supply of the region with clean energy and also export the surplus energy.

The type of mini-hydraulic power plant to be analyzed is the flow or edge of water type, i.e., it does not consist of a reservoir or accumulation of water. The main elements available to the plant are: water intake, bypass channel, loading chamber, forced piping, building with equipment (turbine, generator, transformer, regulating, and power equipment), and power line.

3 Hydrologic Study

For the calculation of the flow rate, as shown in Fig. 1, water is taken weekly for a one (1) year period. On the other hand, in order to obtain the water cut-off area, two (2) approximations were made (quadratic and triangular). Based on these data of height, the water flow rate is calculated from (1).

$$Q = A^{*} V \tag{1}$$

where:

Q flow rate (m^3/s)
A water cut-off area (m^2)
V water speed (m/s)

For calculating the water speed, the Manning equation is applied (2).

$$V = \big(R(2/3)^{*} S(1/2)\big)/n \tag{2}$$

R profile factor (area/length of contact)
S tilt factor (meters by meters)
n Manning Coefficient: 0,02

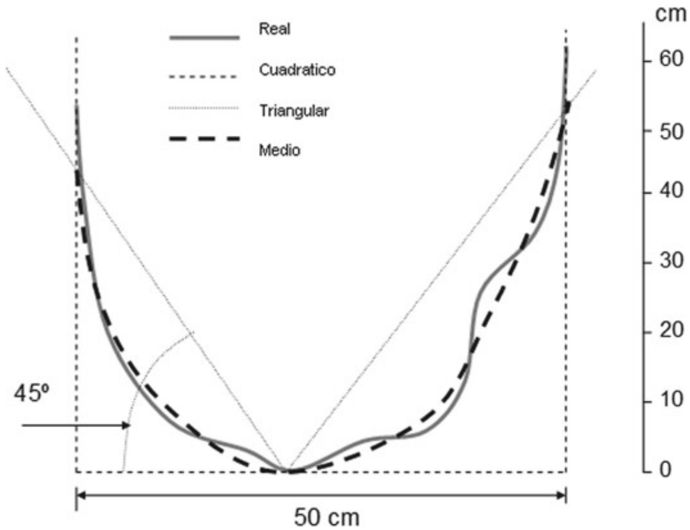


Fig. 1 Chart for the quadratic and triangular analysis of the channel

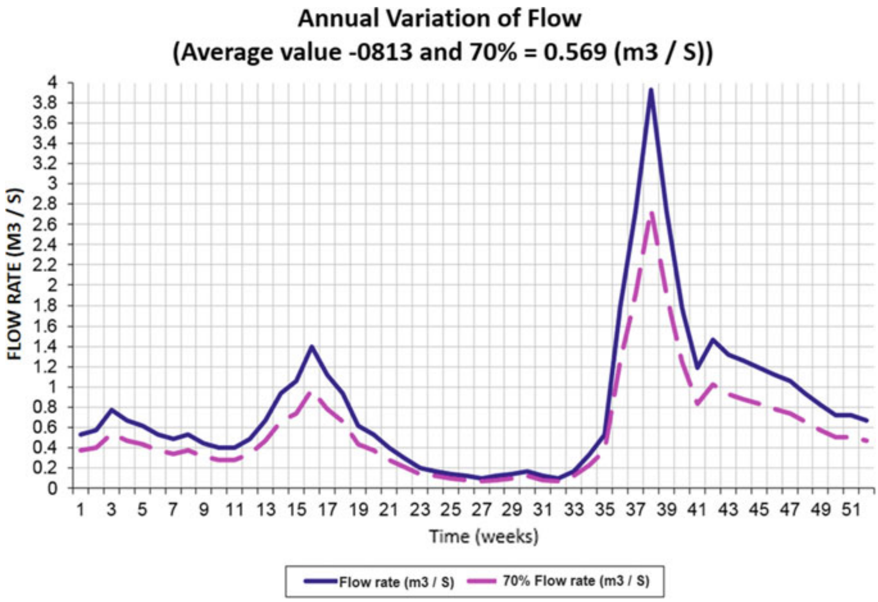


Fig. 2 Annual flow rate of Drezanka River, 2016

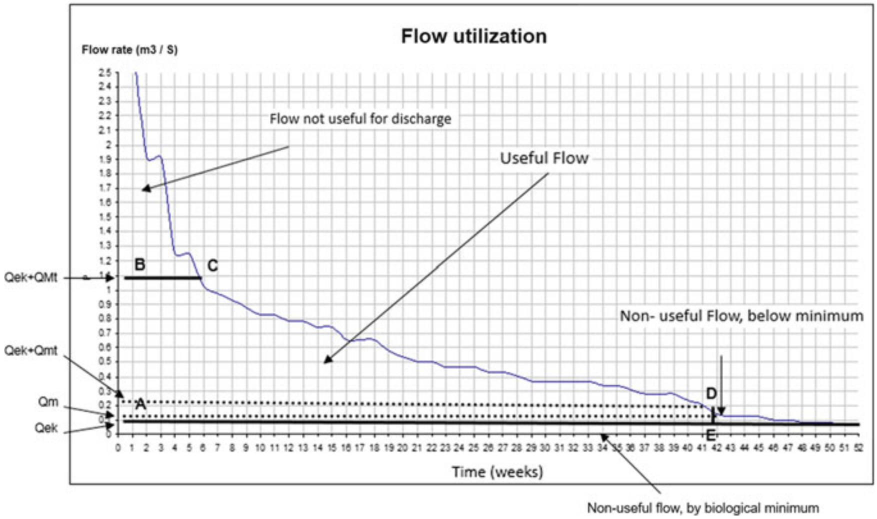


Fig. 3 Useful flow rate for energy production

In order to study the feasibility of the plant operation on the basis of the flow rates obtained, the total and 70% flow rates are represented in Fig. 3. It was also taken into account that there are several additional sources between the water intake and the site where the water height is measured, which is the reason to assert that the flow rate in the water intake will be approximately 30% lower than the measured flow. With the obtained results, the useful flow rate is calculated for the production of electrical power (Figs. 1 and 2).

Where

Q_{m} minimum flow rate
Q_{ek} ecological flow rate (minimum river flow rate)
Q_{Mt} turbine maximum flow rate
Q_{mt} turbine minimum flow rate

For the calculation of Q_{mt}, the following expression is applied (3).

$$Q_{\mathrm{mt}} = K^{*} Q_{\mathrm{Mt}} \tag{3}$$

where K is a factor that depends on the type of turbine (Pelton, $K = 0.1$; Kaplan, $K = 0.25$ Francis, $K = 0.4$, etc.). The criterion for choosing the type of turbine, and therefore QMt, is based on obtaining the maximum area A, B, C, D, E, A as shown in the following graph. In this sense, the maximum flow rate of the turbine, $Q_{\mathrm{Mt}} = 1$ (m^3/s) was chosen. For ten (10) weeks, the plant cannot operate for reasons of ecological flow, i.e., flow below the minimum. On the other hand, there are six (6) weeks a year in which the plant would operate at full load with some unused water (spillage).

For a multi-annual analysis, the analyzed year (2016) is established as a normal year in terms of rainfall. For dry and very dry years, the flow rates decrease by 10 and 20%, respectively. On the other hand, for wet and very wet years, the flow rates increase by 10 and 20%, respectively.

In the operation of a mini-hydraulic power station, the electrical power generated by the power station is directly proportional to the flow and the waterfall, so it is very important to obtain these data as accurately as possible. Likewise, the choice of the appropriate flow is based on the hydrological study based on the criterion of obtaining the maximum production of electric power. Another important factor to consider is to take advantage of the waterfall for investment and environmental reasons. Finally, production equipment must be chosen (turbine + generator) for maximum use of the flow rate and the waterfall.

The installed power is calculated from (4) and a maximum power of 1000 kW is obtained.

4 Economic Analysis

Table 1 shows the costs per concept for civil works, obtaining a cost per unit of power of 627 €/kW.

In order to determine the energy indicator, it is necessary to first calculate the annual output by means of the expression (4).

$$\text{Energy} = P\ \text{installed}^{*} t\,(t - \text{number of hours in a year} = 8760) \tag{4}$$

Given that, according to the hydrological study, it indicates that it is 0.18 €/kWh, as shown in Table 2.

The electricity tariffs in Bosnia (the purchase price of energy from mini-hydroelectric plants) is 50.0 (€/MWh) (100 (km/MWh)), and the usual economic indicators (NPV, IRR, and PAY-BACK) were used to evaluate the profitability of the project.

A. ***VAN (Nominal Actual Value)***: Economic indicator that indicates the value of money earned in the future recalculated in current value. It must be positive for the project to be profitable. A higher value indicates a higher profitability. In our case, it is:

$$\text{VAN} = 1.007.668,16€;\ (\text{Own})\,1.107.167,18€ \tag{5}$$

B. ***TIR (Tax Internal Return)***: Economic indicator for "internal interests." It needs to be positive and higher than current funding interest. A higher value indicates a higher return. In our case:

$$\text{TIR } 21\%;\ (\text{Own})30\% \tag{6}$$

Table 1 Costs for installing the mini-hydroelectric powerplant

Concept	Definition	Amount (€)
Water intake	Concrete azud of 10 m length and 2 m height with grid and gate at 5000 € (includes material and labor)	5.000
Bypass channel	Concrete channel 1 m × 1 m or PVC pipe diameter 60 cm at 50 € per 1 m (includes material and labor). In our case, the length is 1000 m	50.000
Loading chamber	Concrete load dimensions 10 m × 5 m × 3 m, at 10,000€ (includes material and labor)	10.000
Forced piping	Diameter 0.8 m with thickness of 1.0 cm at 200 € per 1 m (includes material and labor). In our case, the length is 400 m	80.000
Building	Power plant building with 1.0 flow (m^3/s) and 120 jump (m)	20.000
Production system	The flow is 1.0 (m^3/s) and the waterfall is 120 (m)	325.000
Transformer	Power of 1000 (kW)	15.000
Electric equipment	Power of 1000 (kW)	30.000
Power line	Line is 10 kV and the difficulty of area is medium	2.500
Access and site preparation	Access roads are 100 € per 1 m (includes material and labor)	10.000
Project and work control	From 6–10% of material costs	23.000
Additional	Estimates between 5 and 15% of total costs	57.050
	Total	627.550

Table 2 Results of the effective hours of work

Description	Symbol	Days	Hours (theoretical)	Hours (effective)
Number of hours in a year	t	365	8760	
Flow below minimum	tmm	91	2184	0
Flow rate above maximum	tMM	35	840	840
Average flow	tsr	239	5736	2575,464
Characteristic time	tc			3415,464
Annual production	Egod (kWh)	3,415,464		
Energy indicator	Ie (€/kWh)	0,18		

Average flow (without min and max.) (0.449/1 = 0.449)
Average flow/maximum flow ratio (0449/1 = 0.449)

C. ***Pay-Back***: Economic indicator for the period of return on investment. A lower value indicates a higher return. In our case, it is:

$$\mathrm{PAY}-\mathrm{BACK}=5\,\text{years};\ (\text{Own})3\,\text{years} \tag{7}$$

5 Environment Analysis

5.1 Inventory of the Environment

As indicated in the section on *Civil Work,* the situation of all the elements of the power station on the ground is shown in Fig. 1. Thus, the entire project contains six elements that could influence the environment: Water intake, bypass channel, loading chamber, forced piping, building, and power line. Ecological aspects of each of these elements were then analyzed.

5.2 Project Analysis

- Water intake: The water intake is of relatively small dimensions and is located in an area where the water passage is narrow with a natural weir. The water intake would only mean increasing this weir by about 1–2 m. The excess water would go over this weir and follow the natural course of the river.
- Bypass channel: In most of the layout, this channel would follow an existing irrigation channel and would be buried. In the sections where it is not possible to bury it, it would adapt to the environment adapting its dimensions and color. However, this element is not relatively large enough to have a great influence on the environment.
- Loading chamber: The loading chamber would be mostly buried. The small visible part would be adapted to the natural environment (similar color to that of the landscape).
- Forced piping: The forced pipe would be completely buried, so that it would not damage the natural environment.
- Building: As stated above, the location would be naturally protected from large waterfalls. The design and color of the building would be adapted to the area, it would not affect the nearby population with noise as it would be far from the towns. It is not located in the route of people or animals.
- Power line: The characteristics of this connection line would be similar to those of the existing distribution network and, therefore, would not present an additional impact on the area.

- Minimum biological flow rate: The water intake would be adapted to keep the minimum biological flow.
- Security: As this is a small accumulation, no special safety measures are required due to rupture or large flows. Practically, the entire weir is used for the evacuation of excess water.

5.3 Identification and Evaluation of Environmental Influences

The identification and assessment of environmental influences can be classified into three periods:

- Construction period: This period is estimated at one year, including preparation works.
- Exploitation period: The concession would be for 30 years. Afterwards, the plant would be managed by the institution that manages the water in this area (this is regulated by law). The useful life of the plant is longer than 30 years, so that the plant would continue to operate.
- Period after the useful life: After the useful life, all the elements of the plant would be dismantled and the environment would be returned to the state prior to the installation of the plant.

During these three periods, there are various influences on the environment. The most commonly used methodology for assessing them is an environmental impact matrix.

5.4 Program for the Protection of the Environment

In the case of mini-hydraulic power stations, the damage that can be caused to the environment can be classified as follows: Loss of arable land; loss of water quality due to its permanence in the pond, and landscape changes due to dam, piping, earthworks, etc.

In our case, there is no loss of arable land. In relation to the change of landscape, it is minimal because all the elements are practically in the river canyon, visible only at a short distance. The water quality is practically the same, there is no pond, and it keeps flowing.

5.5 *Positive Effects*

The following positive effects can be mentioned:

- In general, mini-hydraulic power plants have little effect on the environment compared to other sources of electrical energy.
  - There are no CO_2 emissions and thus no influence on global warming.
  - No influence on polluted precipitations of different types of acids.
  - No special safety measures are necessary.
  - They do not produce poisonous or polluting substances.
- The influence on the environment is practically local and much less than other sources of electrical energy.
- In most of the cases, the consequences are not permanent, after its exploitation the zone can be returned to its initial state.

6 Conclusions

The installation and operation of the mini-hydropower plant on the Drezanka River are presented as a profitable and efficient project.

In relation to the technical design of the equipment, it was possible to find solutions for the required capacity, a maximum flow of 1 m^3/s was obtained for the selected turbine, Ossberger type. With this turbine, an installed power of 1000 Kw was obtained and from the effective working hours of the turbine, an annual production of 3417 MWh/year was obtained. The system can also be fully automated, so it does not require manpower. And the civil work for the channel, the water intake, forced piping, and the building, etc., have resulted in a relatively simple design and construction, due to the selected location.

The economic analysis resulted in an investment of about 600,000 €, including installation and equipment, this amount calculated from the produced energy and the electricity sale rate is estimated to have a payback of about 5 years, knowing that the useful life of the plant would be over 30 years, it indicates that it is a profitable project.

References

1. Domenech, B.: Metodología para el diseño de sustemas de electrificación autónimos para comunidades rurales. UPC, España, Barcelona (2013)
2. Cadena, C.: ¿Electrificación o energización? mediante energías alternativas en zonas rurales. Avances en energías renovables y medio ambiente. **10**, 4–83 (2006)

3. Pérez-Ramírez, R., Pérez, L.A. Martínez-Arremilla, J.: Optimización de sistemas híbridos de energía para la electrificación rural de comunidades índigenas de México. Ingeniería Agrícola y Biosistemas, **3**, 33–39 (2011)
4. Rodríguez-Borges, C., Sarmiento-Sera, A.: Dimensionado mediante simulación de sistemas de energía solar fotovoltáica apliacods a la electrificación rural, » Ingeniería mecánica. **14**(1), 13–21 (2011)
5. López, J.H.: Geopolítica del petróleo y crisis mundial. Dyna. **75**, 156 (2008)
6. Estenssoro, F.: Crisis ambiental y cambio climático en la política global: un tema creciente complejo para América Latina. Universum. **25**(2), 57–77 (2010)
7. Santoyo-Gutiérrez, E., Torres-Alvarado, I.: Escenario futuro de explotación de la energía geotérmica: hacia un desarrollo sustentable. Revista Digital Universitaria. **11**(10), 1–26 (2010)
8. Francés, G.: Entre el mercado y la geopolítica: seguridad de abastacimiento y corredores energéticos en la UE. ICE, Revista de Economía. p. 842 (2008)
9. Manso, E., Miranda, M.: Las reformas liberalizadoras en el sector energético de la Unión Europea. Revista de economíma mundial. **18**, 423–436 (2008)
10. Sanchez, L., Vásquez, C., Viloria, A.: Conglomerates of Latin American countries and public policies for the sustainable development of the electric power generation sector. In: International Conference on Data Mining and Big Data. Springer, Cham, pp. 759–766 (2018)

Feature Selection Using Neighborhood Component Analysis with Support Vector Machine for Classification of Breast Mammograms

N. Kavya, N. Sriraam, N. Usha, D. Sharath, Bharathi Hiremath, M. Menaka and B. Venkatraman

Abstract Recognition of lumps in the breast region is well exploited through mammography. For radiologist, the identification of cancerous tissues is tedious and time-consuming, and many automated computer-aided techniques have been proposed to enhance the clinical diagnosis. This specific research work suggests the application of neighborhood component analysis (NCA) as a feature selection technique for breast mammograms classification. Tamura features (coarseness, contrast and directionality) which provide the characteristics of image surface and objects appearance in images and statistical features (mean, variance, skewness, kurtosis, energy and entropy) were extracted from the breast mammograms, and NCA was applied to identify the best features. Finally, support vector machine classification was performed to classify abnormal condition from normal. Simulation study using the local hospital datasets revealed an overall classification accuracy of 93% by making use of quadratic kernel with SVM classifier.

Keywords Mammography · Tamura features · Statistical features · SVM kernels · Neighborhood component analysis

1 Introduction

The breast cancer incidence is increasing day by day worldwide. In 2017, the American Cancer Society estimated a 2017–2018 years figure as 252,710 new cases were diagnosed among women, and approximately 40,410 women and 460 men died because of breast cancer [1]. Breast cancer has ranked in first place among Indian women with age adjusted rate as high as 25.8 per 100,000 women and mortality 12.7

N. Kavya (✉) · N. Sriraam · N. Usha · D. Sharath
Center for Imaging Technologies, M. S. Ramaiah Institute of Technology, Bangalore 560054, India

B. Hiremath
Department of Surgery, M. S. Ramaiah Medical College and Hospital, Bangalore 560054, India

M. Menaka · B. Venkatraman
Safety, Quality and Resources Management Group, IGCAR, Kalpakkam 603102, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_24

per 100,000 women [2]. The female breasts are usually symmetric in nature. The women's breasts are made up of glandular tissue and fatty tissue. The breast of female adult is composed of 15–20 lobes of branching glands. The size of breast depends on the amount of fat. Each lobe has lobules which produces milk. The breast cancer virtually arises from the glandular tissue. Breast cancer is 100 times more common in women compared to men. The clinical methods for diagnosing breast cancer are self and clinical breast examination, mammography, ultrasound, magnetic resonance imaging, computed tomography, positron emission tomography and biopsy. Among the imaging techniques, mammography is the most acceptable clinical procedure due to its ability to diagnose the cancerous tissue in the breast region. Each mammography imaging yields two images for the right and left regions, respectively. Computer-aided automated diagnostic tool using mammogram image processing gained much attention due to its ability to yield good sensitivity for recognition of normal and abnormal breast mammograms. Several works have been reported in the literature on such automated detection algorithms and performance evaluation parameters like accuracy, sensitivity and specificity were defined [3–8, 14]. It is a well-known fact for pattern recognition problems that appropriate selection of features plays a crucial role in yielding high classification rate. Most of the automated algorithms involves preprocessing, segmentation, feature extraction and classification process. Textural and statistical-based features were generally extracted from the breast region of interest (ROI), and depending on the number of features, appropriate feature selection techniques were employed [9–11].

This specific research work attempts to study the application of neighborhood component analysis (NCA) as a feature selection technique for breast mammogram classification problems. Two views, both mediolateral oblique (MLO) and cranial caudal (CC) of breast mammography were considered, and dataset for the proposed study was collected from the local hospital, Bengaluru, India. After preprocessing, the contrast limited adaptive histogram equalization (CLAHE) method was applied to enhance the images and mathematical operations and connected component labeling were used to extract the region of interest. In the proposed study, the complete breast region was considered as region of interest. In this manuscript, Sect. 2 describes the proposed technique, and Sect. 3 highlights the simulation study results. A brief conclusion was reported in Sect. 4.

2 Methodology

Figure 1 shows proposed technique on NCA-based feature selection in order to develop efficient automated computer-aided quantitative tool for classification of normal and abnormal breast mammograms. CLAHE-based preprocessing procedure was applied followed by tamura and statistical features extraction. The following note provides the short description of the flow process.

(a) Data collection.

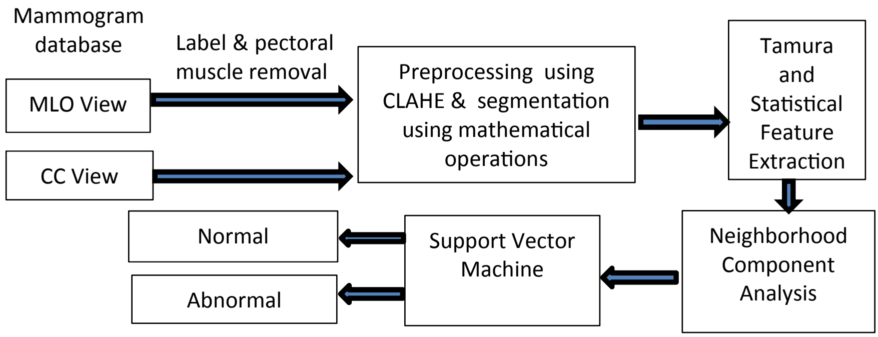


Fig. 1 Proposed technique for mammogram classification

The mammogram images were collected from Ramaiah memorial hospital, Bangalore, India, which included both CC and MLO views. The database has 90 images of CC view and corresponding 90 images of MLO. The actual size of image is 4728*5928.

(b) Image preprocessing, removal of labels and pectoral muscles.

The RGB images were converted into grayscale and resized images. The contrast limited adaptive histogram equalization method was used to enhance the image. Mammogram image had undesirable labels, and it interrupts the feature extraction and proper classification. They were removed using connected component labeling technique. The pectoral muscle was located on the right or left top of MLO view of a mammogram image. Its intensity was high compared to the region of interest. They were removed using manual segmentation.

(c) Tamura and statistical features extraction.

Tamura features give the image characteristics [14], and statistical features are based on the intensity change among pixels [8]. Three tamura features (coarseness, contrast and directionality) and six statistical features were extracted. Totally, nine features were used in the proposed methodology.

- Coarseness measures the granularity of images. It relates to the size of primitive elements forming the texture as shown in Eq. (1) [13].
  - Select the window sizes K and the differences at each scale $E_k\ (x, y)$ are calculated between non-overlapping pixel pairs. Select the proper window size which maximizes $E_k\ (x, y)$ and choose the scale having the largest variations. Then, the coarseness is calculated by averaging the best pixel window size on complete image. For each point (x, y) of the input image, take the averages over the neighborhood, with size $2^l \times 2^l$, where $l \in <0, \ldots, 5>$:

$$A_l(x, y) = \frac{1}{2^{2l}} \sum_{i=x-2^{l-1}}^{x+2^{l-1}-1} \left(\sum_{j=y-2^{l-1}}^{y+2^{l-1}-1} f(i, j) \right) \tag{1}$$

- Contrast stands for quality of images. The gray level varies in an image which was measured using contrast. σ is the standard deviation of the gray-level histogram, α is the kurtosis of the gray-level histogram, and n is set as ¼ as shown in Eq. (2) [12].

$$F_{\mathrm{con}} = \frac{\boldsymbol{\sigma}}{\left(\propto^{4}\right)^{n}} \tag{2}$$

- Directionality was dependent on the presence of orientation. The directionality was the same for two textures if it differs only in orientation. It gives the uniformity between textures in an image, where ΔH and ΔV were horizontal and vertical derivatives, respectively, as shown in Eq. (3) [13].

$$\theta = \tan^{-1}\frac{\Delta V}{\Delta H} + \frac{\pi}{2} \tag{3}$$

(d) Feature selection using neighborhood component analysis technique.

Neighborhood component analysis (NCA) is a non-parametric technique used for feature selection which improves the accuracy of classification algorithm.

The regularization parameter was tuned for the best features. NCA discards irrelevant features which reduces the features dimensionality and improves the algorithm performance. The steps involved in NCA are as follows [11]:

- Data was divided into training and testing sets. Partition of data was done using tenfolds in which the classifier leaves out one division every time and trains on the other nine partitions.
- Lambda value (λ) was tuned, and NCA was trained using each fold in training set for each lambda value.
- The minimum average loss was computed, and best lambda value which corresponds to minimum loss was estimated.
- The features having feature weight greater than the threshold value (T) as in Eq. (4) were extracted, and selected features were trained with SVM classifier and performance of trained classifiers were evaluated.

$$T = \tau \times \max(w) \tag{4}$$

where $\tau = 0.02$ (In proposed method) and $w = \text{weight(feature index)}$

(e) Classification using SVM kernels

The nonlinear classifier, support vector machine has been used for mammogram classification because of its ability to differentiate the binary classes. Kernels convert non-separable data into separable data. Different kernels like quadratic, cubic and medium gaussian were used for analysis. Cross-validation method was used with ten folds.

3 Results

In the proposed method, out of 90 images, 45 images were normal, and 45 images were abnormal in both CC and MLO sets. Different SVM kernels were used to analyze the accuracy. The main aim of the algorithm was to classify the images as normal and abnormal with high accuracy. Both CC and MLO views were considered, and images were labeled based on the annotations given by radiologists.

The tamura features considered were coarseness, contrast and directionality. The statistical features include mean, variance, skewness, kurtosis, energy and entropy. The mean, variance, entropy, coarseness and contrast were high for abnormal images compared to normal images because of the presence of abnormalities and high variance among pixels. The plot for tamura and statistical features of CC and MLO views were shown in Figs. 2 and 3, respectively.

The significant features were selected using NCA. The best lambda value that corresponds to the minimum average loss was plotted as shown in Figs. 4a and 5a for CC and MLO views, respectively. In the proposed method, the regularization parameter (lambda value) was 0.01 and 0.03 for CC and MLO views, respectively. The best features were selected above the threshold value *T*. Out of nine features, six best features were selected in CC view using NCA which are coarseness, directionality, variance, skewness, kurtosis and entropy, and five best features in MLO view which are contrast, mean, variance, skewness and kurtosis were selected as shown in Figs. 4b and 5b for CC and MLO views, respectively.

Tables 1 and 2 show the classification results for CC and MLO, respectively. The tables contain the comparison of results without and with NCA feature selection. As shown in the proposed analysis, the accuracy, sensitivity and specificity improved as the best features selected from the feature set using neighborhood component analysis technique.

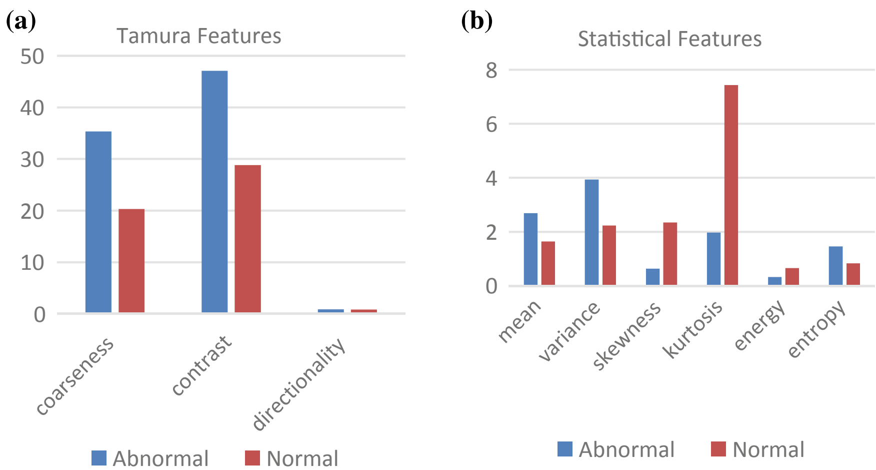


Fig. 2 Feature plots for CC view **a** Tamura features **b** Statistical features

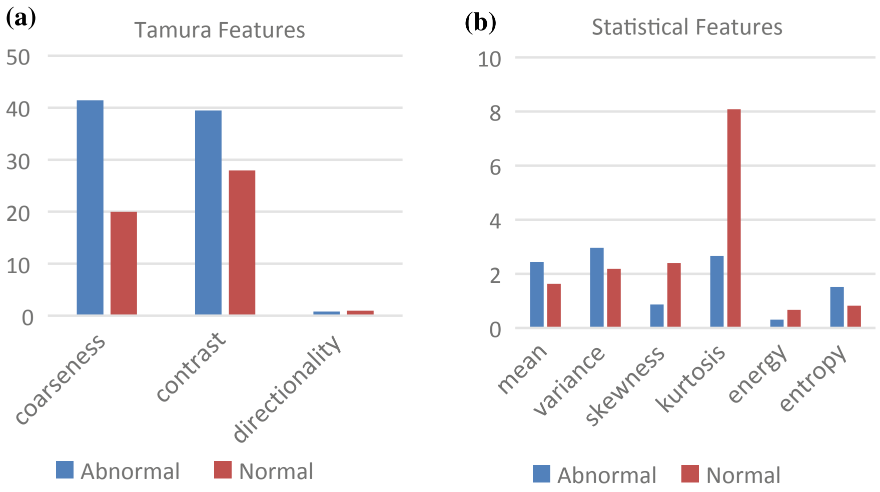


Fig. 3 Feature plots for MLO view **a** Tamura features **b** Statistical features

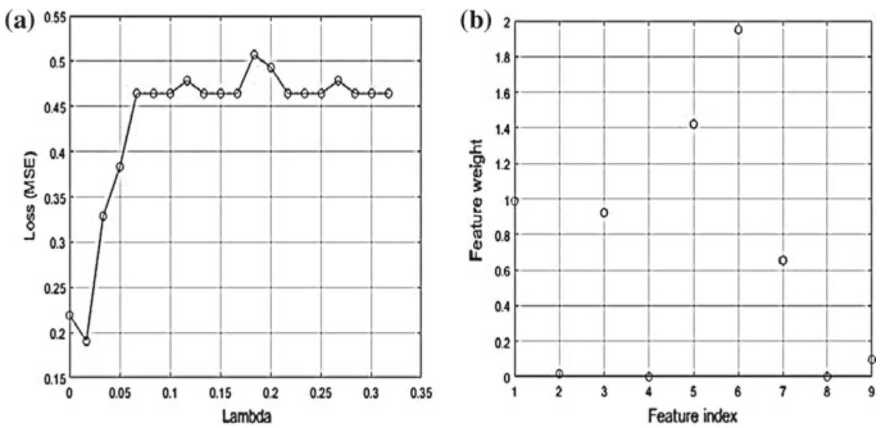


Fig. 4 **a** Plot for best lambda value versus minimum average loss for CC view and **b** Plot for identifying the relevant features of CC view

4 Conclusion

This paper presents the CAD technique based on the texture and statistical features extraction for detection of breast cancer which helps the radiologists and reduces the manual error. The images were collected from the Ramaiah memorial hospital. Totally, nine features are extracted from the mammogram images. The neighborhood component analysis was performed to select the significant features. As a result, six

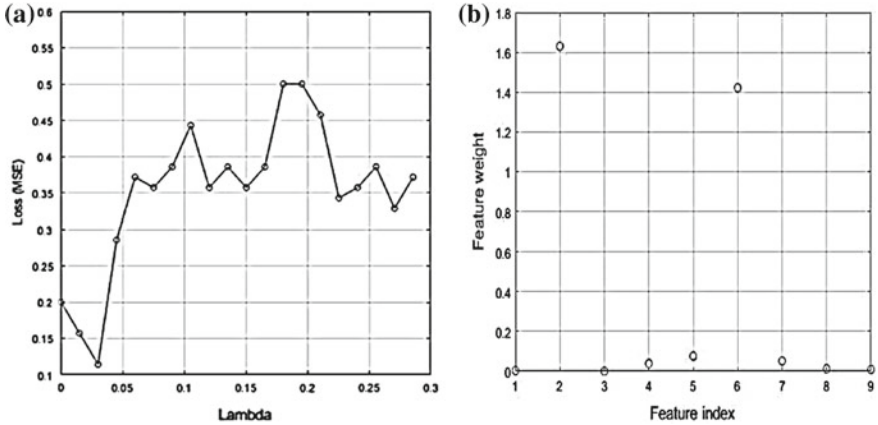


Fig. 5 **a** Plot for best lambda value versus minimum average loss for MLO view and **b** Plot for identifying the relevant features of MLO view

Table 1 Classification results obtained for CC view

SVM kernels	Without NCA feature selection			With NCA feature selection		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Quadratic SVM	90	91.1	89	**92.2**	93.3	92
Cubic SVM	82.2	84.4	80	**84.4**	89	80
Gaussian SVM	85.6	97.7	74	**86.7**	98	76

Bold shows the improvement in the results after applying NCA feature selection technique

Table 2 Classification results obtained for MLO view

SVM kernels	Without NCA feature selection			With NCA feature selection		
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Quadratic SVM	90	84.4	96	**93**	98	88.8
Cubic SVM	87.8	87	88.8	**88.9**	91.1	87
Gaussian SVM	91.1	91.1	91.1	**92.2**	91.1	93.3

Bold shows the improvement in the results after applying NCA feature selection technique features from CC view and five features from MLO view were found to be important. The ten fold cross-validation was used to classify the images, and analysis was done using various kernels. The improvement in the result was obtained by selecting effective features. It can be concluded that the proposed work can be used to classify the normal and abnormal mammogram images using proper feature extraction and selection.

Acknowledgements The authors would like to express their special thanks to IGCAR, Kalpakkam, for funding this project (Project ID: IGC/HSEG/RSD/CP-01/2018) and Department of Surgical and Radiology of Ramaiah Memorial Hospital, Bangalore, for providing the mammogram images for this work. Our special thanks to Mr. Raghu, RIT Bangalore, Dr. Anusha, RMH and all technicians of Radiology department for the annotations and for their consistent support.

References

1. American Cancer society Breast Cancer Facts and Figures 2017–2018. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017–2018.pdf
2. Malvia, S., Bagadi, S.A., Dubey, U.S., Saxena, S.: Epidemiology of breast cancer in Indian women. Asia-Pac. J. Clin. Oncol. **13**(4), 289–295 (2017)
3. Badawy, S.M., Hefnawy, A.A., Zidan, H.E.: Breast cancer detection with mammogram segmentation: a qualitative study. Int. J. Adv. Comput. Sci. Appl. (IJACSA) **8**(10), 117–120 (2017)
4. Goudarzi, Mahsa, Maghooli, Keivan: Extraction of fuzzy rules at different concept levels related to image features of mammography for diagnosis of breast cancer. Biocybern Biomed Eng. **38**, 1004–1014 (2018)
5. Rodriguez-Ruiz, A., et al.: Comparison of breast cancer detection and depiction between planar and rotating synthetic mammography generated from breast tomosynthesis. Eur. J. Radiol. **108**, 78–83 (2018)
6. Shi, Peng, Zhong, Jing, et al.: A hierarchical pipeline for breast boundary segmentation and calcification detection in mammograms. Comput. Biol. Med. **96**, 178–188 (2018)
7. Wang, F., Zhang, S., Henderson, L.M.: Adaptive decision-making of breast cancer mammography screening: a heuristic-based regression model. Omega **76**, 70–84 (2018)
8. Uyun, Shofwatul, Choridah, Lina: Feature selection mammogram based on breast cancer mining. IJECE **8**, 60–69 (2018)
9. Sheba, K.U., Gladston Raj, S., Akhloufi, M.: An approach for automatic lesion detection in mammograms. Cogent Eng. **5**, 1444320 (2018)
10. Gupta, A., Kaushik, B.N.: Feature selection from biological database for breast cancer prediction and detection using machine learning classifier. J. Artif. Intel. **11**(2), 55–64 (2018)
11. Raghu, Shivarudhrappa, Sriraam, Natarajan: Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms. Expert Syst. Appl. **113**, 18–32 (2018)
12. Karmakar, P., Teng, S.W., Zhang, D., Liu, Y., Lu, G.: Improved tamura features for image classification using kernel based descriptors. International Conference on DICTA (2017)
13. Majtner, T., Svoboda, D.: Extension of tamura texture features for 3D fluorescence microscopy. In: 2nd International Conference on 3DIMPVT. pp. 301–307 (2012)
14. Nahid, A.A., Mikaelian, A., Kong, Y.: Histopathological breast-image classification with restricted Boltzmann machine along with backpropagation. Biomed. Res. **29**, 2068–2077 (2018)

Performance Analysis of Implicit Pulsed and Low-Glitch Power-Efficient Double-Edge-Triggered Flip-Flops Using C-Elements

D. Vaithiyanathan, Vikrant Gupta, Santosh Kumar, Alok Kumar Mishra and J. Britto Pari

Abstract In modern electronics, the continuous growth of wearable and portable devices like mobile phones, laptops and smart watches demands low-power consumption. Since flip-flop is a basic storage element of any device, flip-flop designing with low-power consumption is a very critical issue. In this paper, the design of double-edge-triggered (DET) flip-flop belonging to C-element using LECTOR technique is presented. As technology is scaling down continuously so leakage power is an important parameter on which circuit performance mainly depends, in this paper an improvement has been done by introducing effective arrangement of extra transistors in the conventional design. The conventional design and modified circuit are implemented at 45 nm CMOS technology using cadence virtuoso tool at different supply voltage varying from 0.7 to 1.1 V, and reduction in power consumption and improvement in the power delay product (PDP) is achieved as compared with the conventional design.

Keywords Double edge triggered · Low power · C-elements · PDP · Flip-Flop

1 Introduction

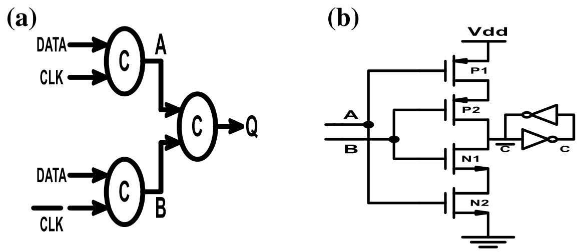
Nowadays, digital electronics technology is mostly working on low-power very large-scale integrated circuits. One of the important elements used in circuit for storage is a flip-flop. In the past, we were using single-edge-triggered (SET) flip-flop (FF) which has more power consumption and propagation delay [15]. So DET flip-flop replaced the SET flip-flop, which affords similar data rate as SET FF at almost half the frequency. So power dissipation of circuit reduces as compared with SET flip-flop

D. Vaithiyanathan (✉) · V. Gupta · S. Kumar · A. K. Mishra
Department of Electronics and Communication Engineering, National Institute of Technology Delhi, Delhi, India
e-mail: dvaithiyanathan@nitdelhi.ac.in

J. Britto Pari
Department of Electronics and Communication Engineering, Presidency University, Bangalore, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_25

Fig. 1 Block diagram and transistor level implementation of C-elements [2]



and delay of the circuit also reduces, and at the same time, DET FF has more glitches in the output [1–14]. DET provides less power dissipation when the glitches in the output get reduced. So C-element-based low-glitch FF was reported in [1, 2, 11], before introducing LG_C DET flip-flop, we must discuss C-element. The C-element is discussed by miller in [5]. C-element has two input and one output in which, when both the inputs are same then it provides the output same as input given. If the input at both the input terminals is not same, then it provides the previous output. Latch-MUX DET flip-flop design is the fundamental design of double-edge-triggered flip-flop, it consists of two latches operated at opposite clock pulse and the output of these latches is connected to multiplexer. C-element (latch) is used to reduce the switching activity in the circuit. If switching transition is less, then power dissipation in transition will be highly reduced, and to remove the problem of power dissipation and delay, the low-glitch LG_C DET flip-flop circuit was introduced in [1] which consists of three C-elements.

Figure 1a, b is the block diagram and circuit implementation of C-element, respectively. Two C-elements are used for the generation of A and B, and output Q depends on A and B in the existing low-glitch circuit, and A is generated when data is given with clock. When the clock is "0", B will generate output otherwise B will maintain the previous output of A. So power dissipation in waiting state for generation of other latch will reduce in low-glitch LG_C_FF circuit. The LG_C_FF is customized to reduce the dynamic power consumption with somewhat increased in power consumption owed for clock transition since at every moment D is either equal to CLK or CLK. So one of the latch will generate either A or B, and latch has no role other than consuming excess power consumption. So in implicit when one latch generates output A and B then other latch will be switched off, so the power consumption gets reduced. As we know at deep submicron technology, leakage power dissipation dominates dynamic power dissipation, so to reduce leakage power dissipation in flip-flop, LECTOR technology is used. In the LECTOR technology, an additional PMOS and NMOS transistor is used between pull-up and pull-down network. Gate of one transistor is coupled to source of other transistor, by this one of the transistors is constantly near the cut-off region, so there is no short circuit path between supply voltage and ground, and hence, the leakage power is highly reduced by using this technique. This paper consists of five sections. Section 1 describes the literature survey and about the C-elements. Section 2 presents the DET FF with C-elements

LG-C-FF and implicit C-FF using LECTOR technique. Section 3 presents the comparison of existing FF with our work. Section 4 discusses the results achieved, and finally, Sect. 5 concludes the paper.

2 Circuit Description

Truth table of C-elements is summarized in Table 1, and in the table, when both inputs *A* and *B* are same then output is same as input applied. At the same time, when the inputs are different, then output maintains the previous value. So in this paper design of various flip-flop circuit is discussed. Which are implemented by using the combination of inverted C-elements, the block and CMOS equivalent diagram of C-elements are presented in Fig. 1a, b.

As we know with the technology scaling, the power consumption is mainly due to leakage power or static power. Glitch problem is common in flip-flop, so in this paper the low-power glitch flip-flop is implemented by using C-elements. With the help of low-glitch LG_C flip-flop, wastage of power is highly reduced. At the same time there is power dissipation due to latches used in the circuit so next circuit implicit pulsed flip-flop is implemented, which is modified version of low-glitch flip-flop shown in Fig. 2. By switching off one of the latch in LG_C_DET_FF at time we can reduce the power consumption. So in this paper, the power consumption is reducing with the help of C-elements, whilst a number of transistors are increasing.

Table 1 Truth table of C-elements

A	*B*	*Q*
0	0	0
0	1	Q_{n-1}
1	0	Q_{n-1}
1	1	1

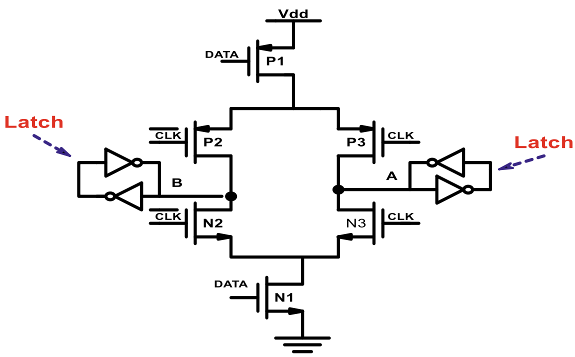


Fig. 2 Circuit Diagram of LG_C_DET_FF [2]

$$Q_{n+1} = AB + (A + B)Q_{n-1} \quad (1)$$

In this paper, we used LECTOR technique, and it is one of the leakage reduction technique discussed for leakage power dissipation reduction in [3]. In the LECTOR technique, two leakage controlled transistors (LCT) which are NMOS and PMOS are placed between pull-down and pull-up network, with the addition of each additional transistor in the gate terminal is connected to the source terminal of other transistor, so gate of each LCT is controlled by source of other LCT transistor. So one of the LCT transistors is constantly close to cut-off region which will provide high impedance path between power supply and ground, and in this way, the leakage current is highly reduced.

The low-glitch power LG_C DET_FF circuit designed by using the C-elements as presented in Fig. 2. Even though inverting and non-inverting C-elements can be used, only inverting topology is used due to the transistor level implementation, being earlier in the inverting configuration. The C-element-based FF has mainly two latches internally and it is named as *A* and *B*. These latches can only switch to CLK and CLK′, when *D* is equal to CLK and CLK′, respectively. So by this way clock transitions of the latch either *A* or *B* will generate [1].

When DATA = 0 and CLK = 0 PMOS P1 is ON, P2 is OFF, P3 is ON, N3 is OFF, N2 is ON; hence, *A* will be logic high.

When DATA = 0 and CLK′ = 0 PMOS P1 is ON, P2 is ON, P3 is OFF, N3 is ON, N2 is OFF; hence, *B* will be logic high.

In low LG_C glitch DET flip-flop, there is no moment when data is neither equal to CLK and CLK′. So latch is either *A* or *B* when one latch is generating, the other latch which is in idle state contributes to power dissipation which is the problem in LG_C DET flip-flop so to overcome this problem we have to switch off the latch which is in idle state and this makes reduction in power consumption. In implicit pulse DET flip-flop shown in Fig. 3, when *A* is generating at that time other latch will be switched off [1].

When DATA = 0 and CLK = 0 then PMOS P4 is ON, P1 is OFF, P2 is ON, P6 is ON, N1 is ON and N4 is OFF, hence *A* will be logic high.

When DATA = 0 and CLK′ = 0 then PMOS P4 is OFF, P1 is ON, P2 is OFF, P6 is OFF, N1 is OFF and N4 is ON; hence, *B* will be logic high.

3 Modified LG-C Flip-Flop

In modified LG_C_FF shown in Fig. 4, we introduced 4 extra transistors (2 NMOS and 2 PMOS) is present in the LG_C DET_FF circuit. In the circuit, the gate of one transistor is coupled to source of other transistor. Therefore, we can say gate of one transistor is controlled by source of other transistor, so one of the transistors is for all time near to cut-off region. This provides high resistance path between supply and ground, so leakage current between supply and ground will be less in LG_C

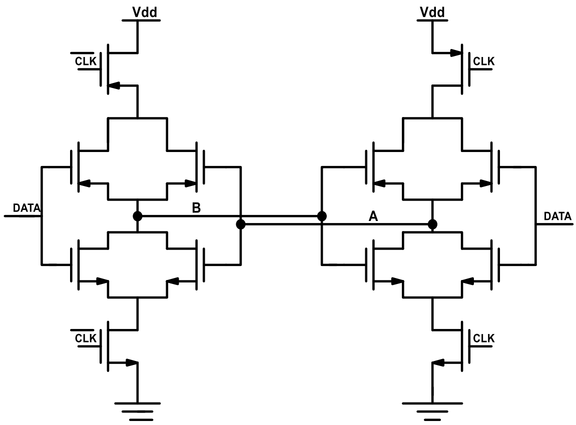


Fig. 3 Circuit of implicit pulse DET flip-flop [2]

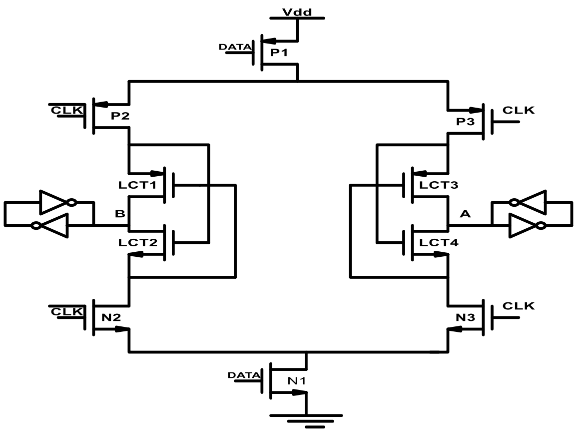


Fig. 4 Circuit of modified LG_C DET_FF

DET_FF. Working of modified circuit is similar to existing LG_C DET flip-flop except reduced the leakage power with the impact of increase in the circuit delay.

3.1 Modified Implicit DET Flip-Flop

In modified implicit DET flip-flop shown in Fig. 5 contains 4 extra transistors (2 NMOS and 2 PMOS) are introduced. With the help of these transistors in which gate of transistor is connected to source of other transistor so we can say gate of one transistor is controlled by source of other transistor so one of the transistor is always near cut-off region which will provide high resistance path between supply and ground so leakage current between supply and ground will be less compared to existing implicit DET_FF.

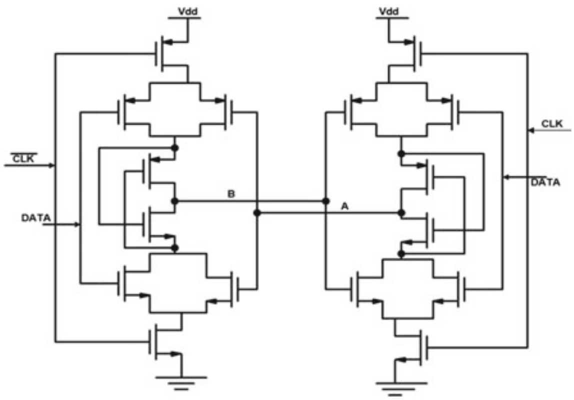


Fig. 5 Circuit of modified implicit DET_FF

4 Results and Discussion

The conventional and modified flip-flop circuits are implemented at 45 nm CMOS technology using cadence virtuoso tool at different supply voltage varying from 0.7 to 1.1 V. It is shown in Figs. 6 and 7, and the output waveform for the LG_C DET_FF and modified LG_C_FF is almost same. The values of *A* and *B* come according to C-elements. When both data and clock are same, then output is same as input but we have inverted C-elements, so output will be in inverted form. From the waveform, we observed that on changing the circuitry from conventional to modified, the operation of circuit is same in both the cases only leakage power and delay is varied.

The comparison of various LG_C DET flip-flop circuit in terms of delay, power and PDP is given in Table 2, and the PDF comparison for all the circuit is plotted and is shown in Fig. 8. This implemented circuit has been calculated at 1 V power

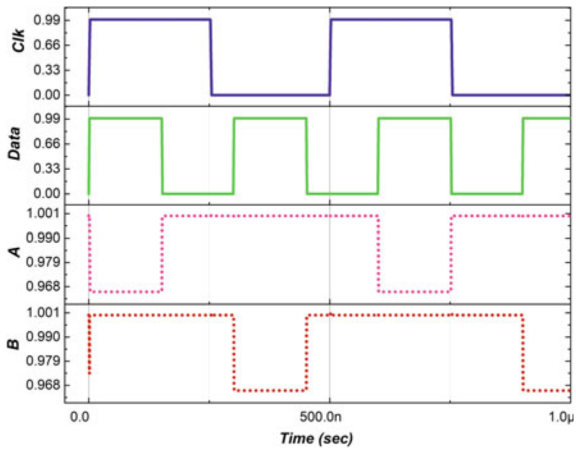


Fig. 6 Waveform of modified LG_C DET flip-flop

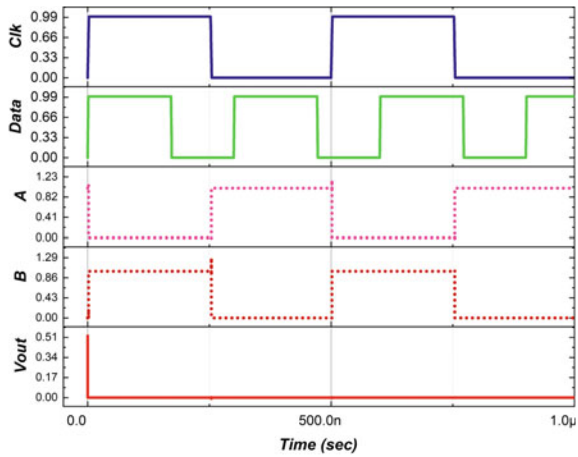


Fig. 7 Waveform of modified implicit DET flip-flop

Table 2 Simulation results for 1 V power supply at 45 nm technology

Flip-flop	Power (µW)	Delay (pSec)	PDP (fJ)
LG_C DET FF [2]	14.27	510.7	7.27
Implicit DET FF [2]	4.51	405.1	1.827
Modified LG_C DET FF	3.40	620.29	2.11
Modified implicit DET FF	0.125	690.1	8.62×10^{-2}

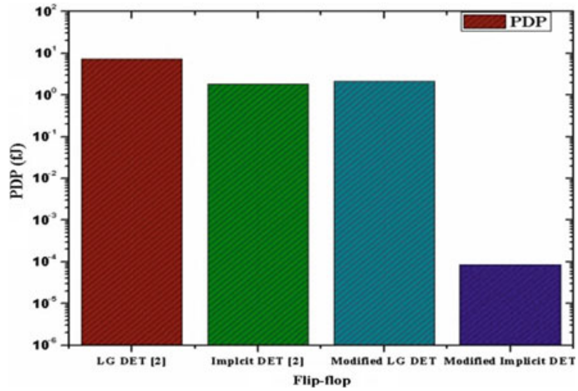


Fig. 8 PDP comparison of different DET flip-flop

supply with room temperature. From Fig. 8, it clearly is shown that the modified DET flip-flop gives better results amongst all other existing circuits [2].

4.1 Temperature Variation

Leakage power is more sensitive to temperature. As the temperature increases, leakage power also increases, and it results in heat dissipation and thermal runway, which can burn the devices. In DET flip-flop, data is mainly split in two parts named as *A* and *B;* therefore, leakage power will different for all the four combinations. Leakage power is the combination of gate leakage current, junction leakage current and sub-threshold leakage current. If input changes, then gate leakage current will change, so we can say changing in input vector leakage will also change. Here, leakage power dissipation is shown in Figs. 9, 10, 11 and 12 for all the four combinations ("00", "01", "10" and "11").

From Figs. 9, 10, 11 and 12, it is observed that for all the input combination other than "00" combination, modified implicit DET flip-flop gives better results amongst all other DET, whether that is implicit DET or LG_C DET FF. For "00" combination alone, modified LG DET flip-flop is better when compared with the other existing techniques.

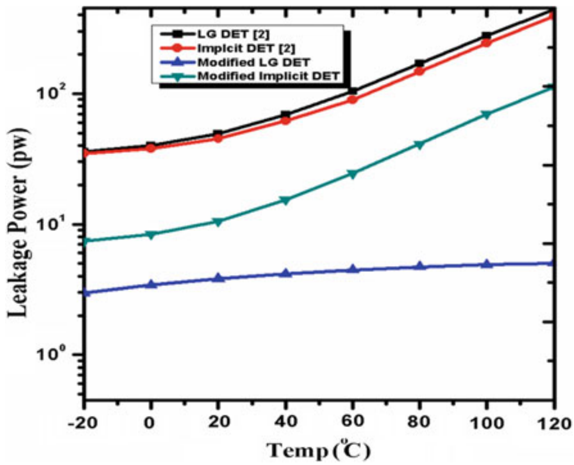


Fig. 9 Power versus temperature for "00" input combination

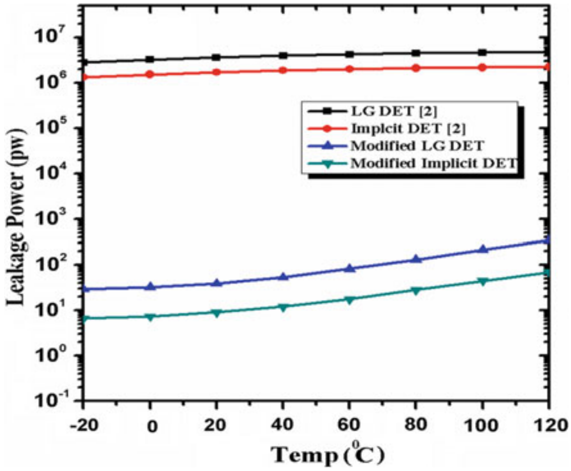


Fig. 10 Power versus temperature for “01” input combination

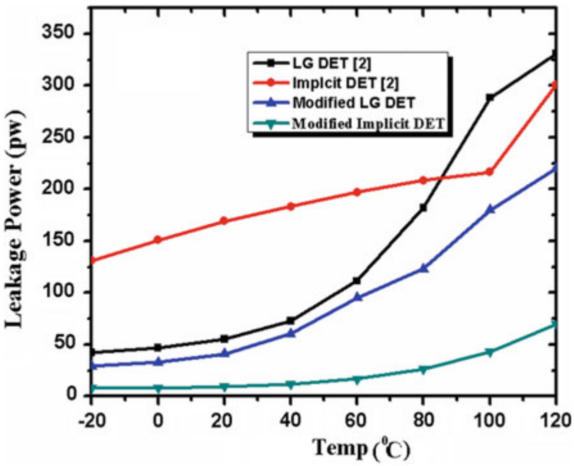


Fig. 11 Power versus temperature for “10” input combination

4.2 Supply Variation

Static power dissipation is nothing but leakage power. We know that as the supply increases, leakage power also increases (Power = $V_{DD}*I_{Leakage}$). We calculated leakage power for different supply voltage for different input combinations (00, 01, 10, 11). Figures 13, 14, 15 and 16 show leakage power for different voltages for different input combinations. From Figs. 13, 14, 15 and 16, it is observed that for almost all the voltages levels and all the input combinations, modified low-glitch flip-flop and modified implicit flip-flop give lower leakage power compared with the existing DET flip-flop.

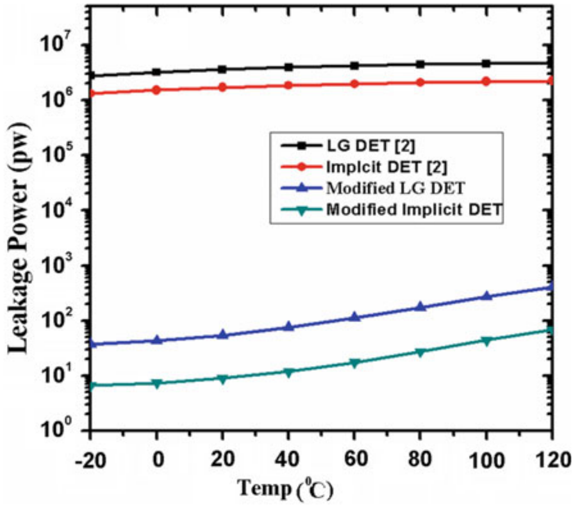


Fig. 12 Power versus temperature for “11” input combination

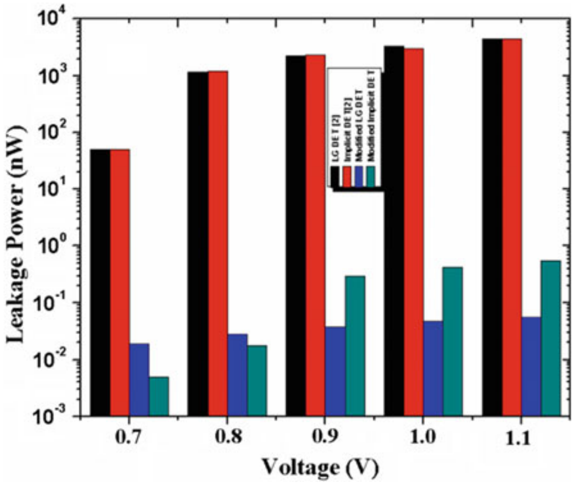


Fig. 13 Power versus voltage for 00 input combination

4.3 Process Variation

The effect of environmental and process variations on transistors can be categorized as: typical (Also called nominal), fast or slow. Process corners are defined when processing variations are combined with the environmental variations. The two complementary types of transistors with independent characteristics are present in CMOS technology, so the speed of each transistor can be distinguished. *A* conventional two-letter designation has been used to describe process corners, where first letter represents an NMOS device and second letter represents a PMOS device. Hence, five classic corners are defined as: tt, ff, ss, fs and sf. The distribution of

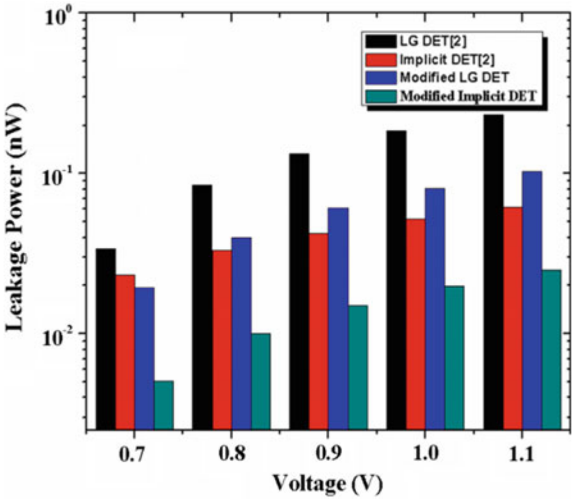


Fig. 14 Power versus voltage for 01 input combination

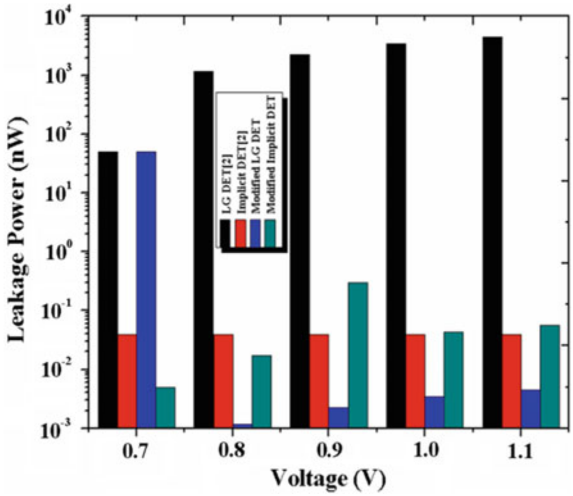


Fig. 15 Power versus voltage for 10 input combination

process variations is provided by manufacturing company. The primary three corners (i.e. tt, ff and ss) are recognized as even corners because at these corners, both devices have been affected evenly. The last two corners (i.e. fs and sf) are known as "skewed" corners because at these corners, one device will switch much faster than the other, resulting in imbalanced switching of the devices. For all the input combinations ("00", "01", "10", "11"), leakage power is shown in Figs. 17, 18, 19 and 20, respectively.

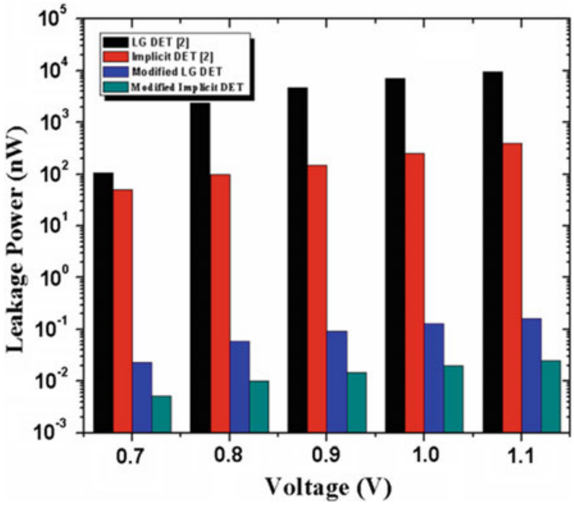


Fig. 16 Power versus voltage for 11 input combination

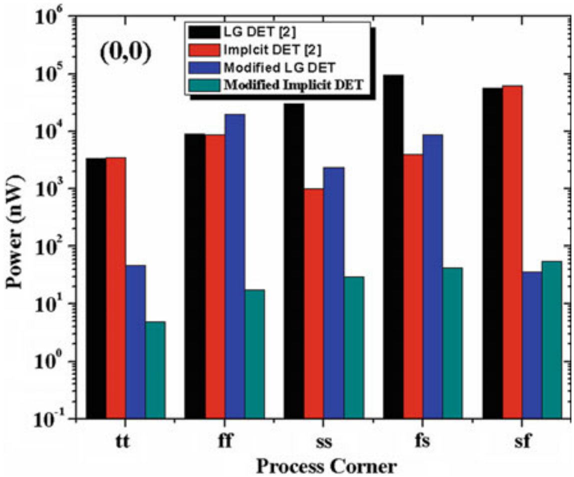


Fig. 17 Process corner analysis for 00 input combination

5 Conclusion

In this paper, performance analysis of low-glitch and power-efficient DET LG_C_FF and modified implicit DET flip-flop was discussed. It is observed from implementation results that the low-glitch DET flip-flop and implicit DET flip-flop CMOS circuit consume lesser power when compared with the existing flip-flop techniques with marginal increase in the delay. Average leakage power dissipation and propagation delay for the modified implicit DET flip-flop are 125nW and 690.1 ps, respectively, for all four possible combinations of the inputs. Modified implicit DET flip-flop has PDP of 70% improvement compared with the conventional DET LG_C flip-flop.

Fig. 18 Process corner analysis for 01 input combination

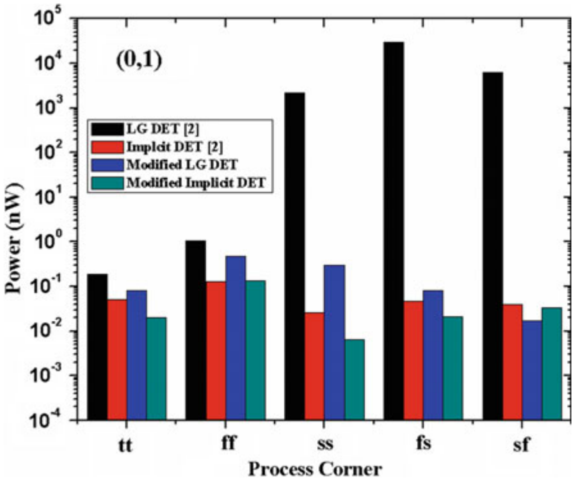
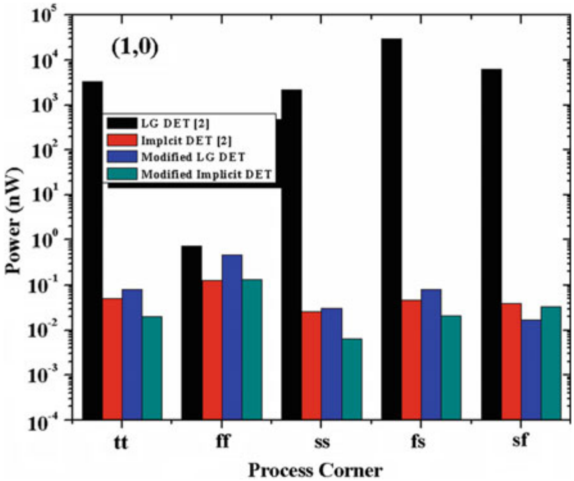


Fig. 19 Process corner analysis for 10 input combination



Due to minimum PDP, we can conclude that modified implicit DET has the best performance amongst the existing DETs and suitable for low-power application such portable and battery operated gadgets.

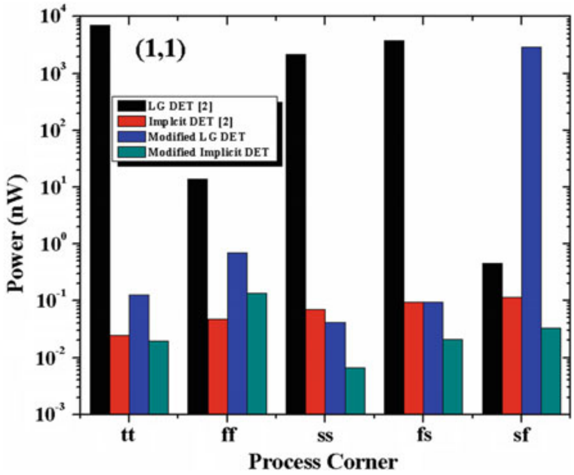


Fig. 20 Process corner analysis for 11 input combination

References

1. Strollo, A.G.M., Napoli, E., Cimino, C.: Analysis of power dissipation in double edge-triggered flip-flops. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **8**(5), 624–629 (2000)
2. Lapshev, S., Hasan, S.M.R.: New low glitch and low power DET flip-flops using multiple C-elements. IEEE Trans. Circuits Syst. I Regul. Paper. **63**(10), 1673–1681 (2016)
3. Rohila, R., Vashishath, M.: CMOS LECTOR technique: a technique for leakage reduction in domino circuits. Int. J. Sci. Res. Dev. **3**(03), 2279–2281 (2015)
4. Tschanz, J., Narendra, S., Zhanping, C., Borkar, S., Sachdev, M., De, V.: Comparative delay and energy of single edge-triggered and dual edge triggered pulsed flip-flops for high-performance microprocessors. In: International Symposium on Low Power Electronics and Design. pp. 147–152 (2001)
5. Muller, D. E.: Theory of asynchronous circuits. Internal Report no. 66, Digital Computer Lab University. Illinois at Urbana-Champaign (1955)
6. Devarapalli, S.V., Zarkesh-Ha, P., Suddarth, S.C.: A robust and low power dual data rate (DDR) flip-flop using C-elements. In: 11th International Symposium on Quality Electronic Design (ISQED). pp. 147–150 (2010)
7. Gago, A., Escano, R., Hidalgo, J.A.: Reduced implementation of D-type DET flip-flops. IEEE J. Solid-State Circuits **28**(3), 400–402 (1993)
8. Bonetti, A., Teman, A., Burg, A.: An overlap-contention free true single-phase clock dual-edge-triggered flip-flop. In: Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1850–1853 (2015)
9. Alioto, M., Consoli, E., Palumbo, G.: Analysis and comparison in the energy-delay-area domain of nanometer CMOS flip-flops: Part I—Methodology and design strategies. IEEE Trans. Very Large Scale Integr. (VLSI) System. **19**(5), 725–736 (2011)
10. Shams, M., Ebergen, J.C., Elmasry, M.I.: Modeling and comparing CMOS implementations of the C-element. IEEE Trans. Very Large Scale Integr. Syst. (VLSI) **6**(4), 563–567 (1998)
11. Alioto, M., Consoli, E., Palumbo, G.: Analysis and comparison of variations in double edge triggered flip-flops. In: 5th European Workshop CMOS Variability (VARI), Palma de Mallorca, Spain, pp. 1–6 (2014)
12. Ji-ren, Y., Karlsson, I., Svensson, C.: A true single-phase-clock dynamic CMOS circuit technique. IEEE J. Solid-State Circuits **22**, 899–901 (1987)

13. Svensson, C., Alvandpour, A.: Low power and low voltage CMOS digital circuit techniques. In: International symposium on low power electronics and design. pp. 7–10 (1998)
14. Renshaw, D., Lau, C.H.: Race-free clocking of CMOS pipelines using a single global clock. IEEE J. Solid-State Circuits **25**, 766–769 (1990)
15. Stojanovic, V., Oklobdzija, V.G.: Comparative analysis of master slave latches and flip flops for high-performance and low-power systems. IEEE J. Solid-State Circuits **34**(4), 536–548 (1999)

Cost-Effective Waste Collection System Based on the Internet of Wasted Things (IoWT)

Chakkrit Snae Namahoot, Michael Brückner, Yoseung Kim and Sopon Pinijkitcharoenkul

Abstract Many municipalities and local organizations are currently planning to improve their services by applying information and communication technologies (ICT), especially Web-based approaches, to serve citizens, business, and the government agencies as well. Besides the quality of service, cost-effectiveness is a major criterion to be met. In this paper, we present a Web-based solution to urban and suburban waste management (Internet of wasted things), which employs smart bins, a wireless sensor network, and a back-end real-time data collection and communication system for managing waste collection in a cost-effective way. Results are gained by setting up a prototype smart waste management system (SWMS) in a local community in the lower north of Thailand. The system may be employed in similar projects, where cost-effectiveness is a major criterion regarding feasibility.

Keywords Smart waste management · Internet of wasted things (IoWT) · Wireless sensor networks · Cost-effectiveness

C. S. Namahoot (✉)
Department of Computer Science and Information Technology, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand
e-mail: chakkrits@nu.ac.th

M. Brückner
Department of Educational Technology and Communication, Faculty of Education, Naresuan University, Phitsanulok 65000, Thailand
e-mail: michaelb@nu.ac.th

Y. Kim
Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University, Phitsanulok 65000, Thailand
e-mail: kingrise@nu.ac.th

S. Pinijkitcharoenkul
Information Technology Center, Pibulsongkram Rajabhat University, Phitsanulok 65000, Thailand
e-mail: sopon_b@psru.ac.th

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_26

1 Introduction

Waste collection and management is a basic service of local government agencies (LGA), and its quality and cost are felt by citizens immediately. Therefore, LGA strive to employ cost-effective and efficient ways for collecting and managing waste in their area of responsibility, even if they outsource services [1]. All of this is in accordance with the Digital and Societal Economy Act adopted by the government of Thailand and accounted for by the Digital Economy Promotion Agency (DEPA) [2].

Cost reduction can be achieved in different ways: (1) reducing hardware cost, (2) simplifying communication and management procedures, and (3) minimizing fuel costs by optimizing collection paths of the trucks. In the following, we outline effective procedures to support these objectives.

(1) Hardware cost reduction: Integrated hardware and software solutions make the most effective use of waste collection, saving time, money, and the environment. The fill level sensor of the clean cap is operated by either a battery or solar energy [3]. In this paper, an ultrasonic sensor is used to monitor fill level, and the NB-IoT network for sending fill level information is battery powered.

    The ultrasonic sensor (interfaced with Raspberry Pi) senses the level of waste in a dustbin. Sensor sends waste level information to Raspberry Pi, and Raspberry Pi sends data to the server continuously [4]. The costs of the NB-IoT modules are lower than the modules for other mobile communication systems on the market such as 3G, 4G, and 5G. The expected value for the NB-IoT module is between 5 and 7 US Dollars in the near future. The use of the modules is becoming more popular in the sensor network industry. The subscription cost, which is one of the most important factors from the user side, will not be more expensive than currently operating mobile communication systems [5]. In our model, users buy NB-IoT modules and sensors at the beginning and pay the subscription fee on a monthly basis.

    Due to the advances in sensor development and decreasing production costs, the use of sensors has increased. The sensor network industry like in a smart city has grown rapidly, and sensors are being applied to new products. These products are becoming more popular at home, at work, and in public facilities such as sports grounds parks and metro stations. People can accrue some advantages from the use of sensors in public areas. Smart cities designed and built in areas where such networks exist. One of the smart city facilities is smart refuse bins which are managed by a local authority. By installing sensor and communication modules, existing bins can become smart bins [6]. The ultrasonic sensor can sense the content levels of the bins, and the information is transmitted to server through NB-IoT network.

    From the perspective of building smart cities, the IoT system could be used for waste management purposes. The IoT system works through an IP address allocated to each object [7]. The Internet Protocol (IP) is the protocol that allows devices to find and communicate with each other through the Internet. A waste management system based on the IoT concept and IP address is therefore necessary in urban or suburban areas for cost-effective waste management.

(2) For simplifying management and communication procedures, different strategies and tools are employed to improve ICT effectiveness in the IoWT sector [3] as outlined in the following.

- Using wireless sensors with data exchange capabilities to measure the distance between waste bin lid and the waste and communicate the data at given times of the day or on demand by service management staff; for cost-effective data communication, a WiFi network can be used
- Maintaining a decision support system (DSS) that works as a recommender system for route planning of the individual trucks collecting the waste; such systems can be supported by a MySQL database (freeware), which holds the data regarding bins, trucks, and other object data
- Shortest path identification can be processed after the target bins, i.e., those bins filled more than a specified threshold, and the positions of the trucks have been identified.
- The respective route information is sent to the trucks, e.g., as a route map displayed via Google Map [8].

(3) Fuel costs and CO_2 reduction: The usual way to collect waste is to empty the bins at certain intervals following a fixed route. The disadvantage of this procedure is that bins are usually filled at different levels: some might be empty and others might be overfilled. Public events have to be taken into account as well since the audience leads to crowded spaces, where waste bins (even the additional ones) are filled quickly. In these cases, they need to be handled faster, otherwise hygiene risks might follow [9]. This means the bins have to be approached quasi "in-time" to make their collection more cost-effective: sometimes earlier than usually scheduled, sometimes later.

Dijkstra's algorithm is an algorithm used to find the shortest path. The algorithm is easy to implement and results in the shortest route between given points (nodes) on a map.

The rest of the paper is structured as follows. Section 2 gives a concentrated overview of the current literature. Section 3 shows the IoWT system model and implementation. Section 4 deals with a use scenario, and Sect. 5 concludes the paper and discusses future work.

2 Related Work

A vast amount of literature exists on the topic of ICT enhanced waste management. Some key words that are often used in this respect are "smart cities," "smart waste collection," "IoT-enabled waste management," and "intelligent waste management" [4].

Guerrero et al. [1] investigate waste collection in developing countries carried out a decade ago. They model stakeholders' role and actions regarding waste collection; finally, they examine the models with limited real data.

Medevedev et al. [10] take up the roles of stakeholders (city and district administration, waste processing companies, truck drivers, recycling factory managers, and traffic police). They focus on the situation of inaccessible waste bins in cities, which are recorded on the smartphones at hand. The annotated data (GPS coordinates, narrative) are sent to the DSS for further processing.

Regarding the different ways of obtaining cost-effectiveness, several papers deal with route optimization. Khan and Gawade [11] report on a procedure, which locks a smart waste bin when it is full and sends an alarm message to the nearest truck to collect the waste from that bin. Unfortunately, this may lead to an erratic itinerary of a given truck through a major city, which will not optimize fuel costs and CO2 emissions significantly.

In their extensive paper on the future of waste management in modern cities, Esmaeilian et al. [12] highlight three areas that need to be covered regarding sustainable development [7]: (1) full transparency of product data throughout their lifecycle, (2) prevention of waste production by supporting new business models, and (3) building of infrastructures for recycling and on-time waste collection. Our work covers area (3) with a focus on cost-effectiveness.

3 Proposed Work

Reducing operating costs by eliminating extra communication fee for each sensor network is the key factor for such long-term projects.

There are some benefits of NB-IoT as follows:

- Low power consumption: ten years battery life depending on configuration
- Extended range: better penetration in indoor
- Less costs: hardware and subscription
- Easy network deployment: AIS NB-IoT the first commercial IoT service in Thailand, expanded to 77 provinces
- Bandwidth: 180 kHz 3GPP licensed (data rate < 100 Kbps).

Based on the low-cost subscribed mobile network, we can provide flexible collection routes and schedules to optimize overall collection tasks, avoiding unnecessary collection.

Then, the system reduces ineffective operational costs when collecting waste and increases the efficiency of collection operations that are currently in place.

3.1 System Design and Implementation

The smart waste management system (SWMS) consists of three main parts: smart bin, smart waste management system, and smart navigation and tracking (Fig. 1).

- Smart bins check the amount of content using ultrasonic sensors to measure the distance between the bin cover and the waste level inside the bin and then send the necessary data to the cloud via NB-IoT. The SWMS collects such information as the bin ID, bin location, and the amount of waste and stores it for further processing.
- The SWMS takes distances in each bin to calculate percentage of waste filling. For example, if the height of the bin is 2 m and the distance for waste is 1 m, then the amount of waste is 50%. If there is more than 80% waste in a bin and the number of such bins exceeds the threshold set up for the SWMS, the system checks the availability of trucks and assigns jobs to staff displaying details on the bin ID, the bin location, the waste level, and the truck ID to get the job order of picking up the waste soon.
- The smart navigation and tracking system calculates optimal routes for the garbage trucks using a modified Dijkstra's algorithm together with a Google map API; then, the truck is navigated by the system to follow the optimal route.

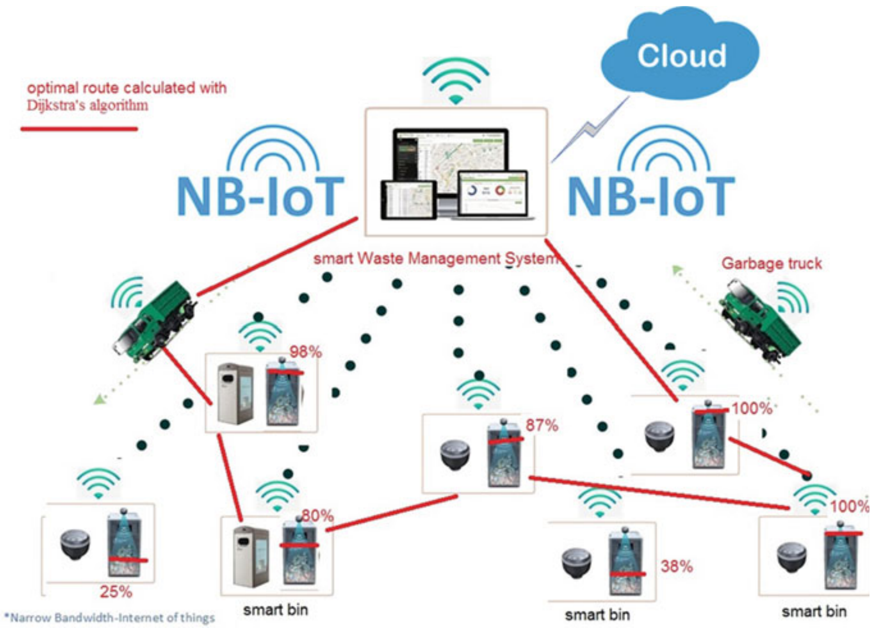


Fig. 1 The smart waste management system

3.2 Shortest Path Implementation

We apply a modified Dijkstra's algorithm [13], which works as follows:

Let the truck start at the initial node B_0, e.g., the location of the agency, and let the nearest waste bin with sufficient filling (>70%) be the target node; bins with sufficient filling are denoted as B_n ($n > 0$) with B_1 as the nearest target node. First, the system extracts all waste bin nodes B_n, their distances and paths between the initial node and the target node from the database. The Dijkstra's algorithm is used in the system, which calculates the shortest path as follows:

1. Set amount of waste 6 ton (AW = 6 ton) for a garbage truck
2. Get total number of bins (TNB) appears on the system (each bin has more than 70 or 80% of waste depending on number of bins)
3. Calculate total amount of waste (TAW)
4. If TAW > 5 ton and ≤ 6 ton, then
5. Set system location node to zero (smart waste management system node) and all other bin nodes to infinity (unvisited nodes)
6. For the system location node, get all distances of unvisited neighbors
7. Choose the shortest distance node as current location.
8. Link current location to unvisited neighbors, calculate, and assign their distances.
9. Do step 3–4 until all the target bin node has been visited.
10. Terminate visited nodes = TNB
11. Return and display the shortest path and send to the truck for picking up waste.

4 Prototype Testing and Result

For prototype tests, we have installed and managed the SWMS and set up smart bins in a location in Tha Pho subdistrict near Naresuan University, where garbage trucks come on a daily schedule. The tested system consisted of three components: the SWMS, the smart navigation and tracking system, and the waste management report builder.

Figure 2 shows the interface of SWMS displaying information regarding the bins, which need to be emptied because of high filling levels. Here, five bins are filled at more than 80%, so these bins are subject to scheduling. The system assigns an available truck for picking up the garbage job (Truck No. 1), and by clicking "Assign job," the operator can start the garbage collection; after that, the system shows information about "waste detail" (Fig. 3).

Figure 3 shows all bins located in the Tha Pho subdistrict. The filling level is represented by red, orange, and green color (80% or more, 50–79% and less than 49%, respectively). The operator clicks "shortest path" to display the shortest route

Fig. 2 The smart waste management system overview window (Tapoe = subdistrict Tha Pho, NU = Naresuan University campus)

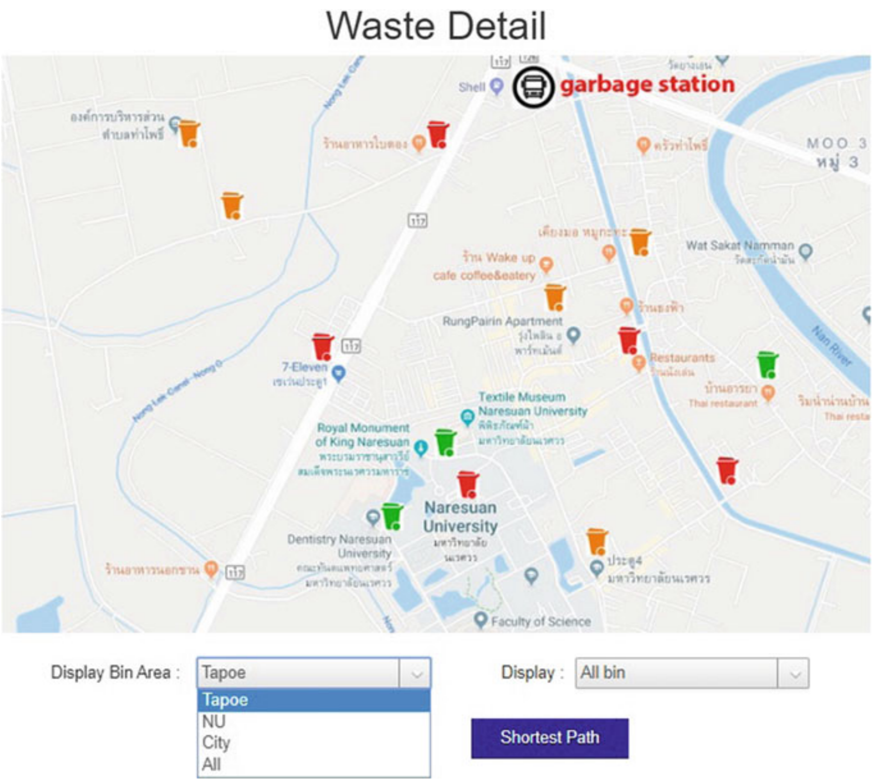


Fig. 3 Garbage collection map with waste details

for picking up the trash with map to navigate the truck into the right direction in Fig. 4.

Figure 4 shows the interface of the smart navigation and tracking system which helps Truck No 1 to find the shortest route to pick up waste from the smart bins efficiently. Truck No 1 gets its job assignment from the system, and then, the truck driver either can click “Get optimal route” or “Routine.” Here, “Get optimal route” has been selected, and the system has displayed the map and locations of smart bins that need to be picked up. The navigation button is used for navigating the truck to the pick up all lusts of bins. During the operation of Truck No. 1, the operator can select “Truck Tracking” to see how many bins have already been emptied so far: bin no. 1, 3, and 4 have been picked up and the bins no. 6 and 7 are waiting (the algorithm

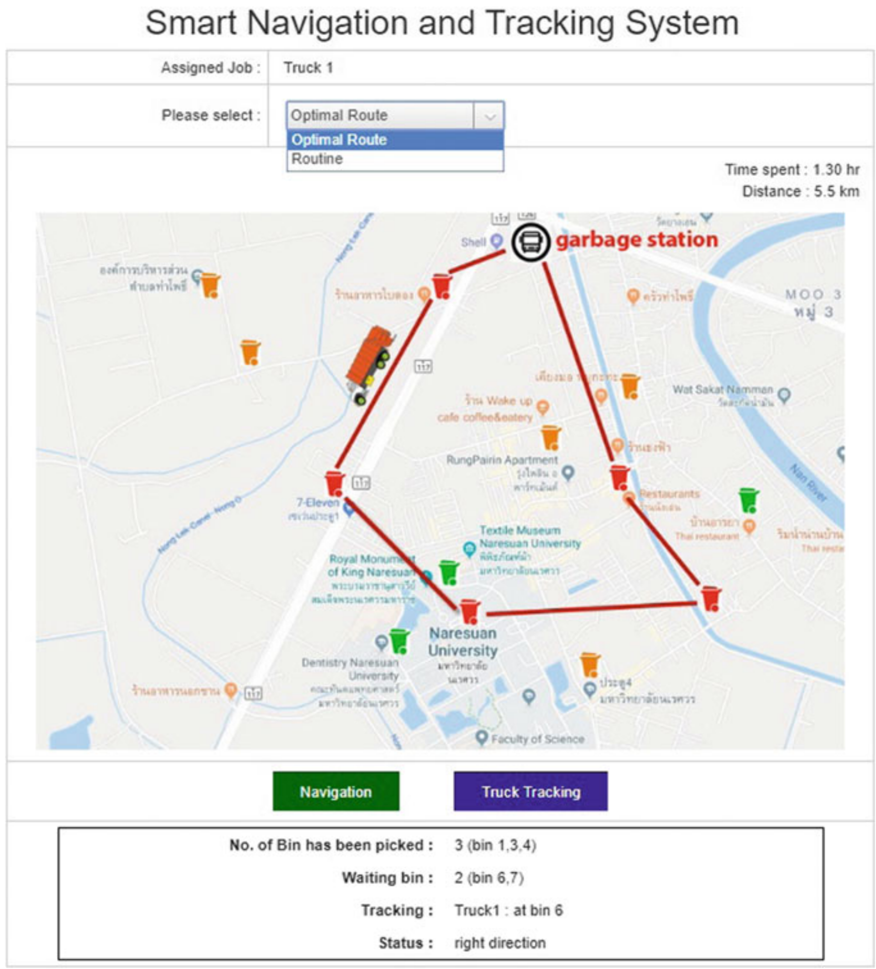


Fig. 4 Smart navigation and tracking system (as operated by the truck driver)

Waste Management Report

Truck no.	Staff	no. of bin	amount of waste	Time spent (hr)	distance (km)	fuel fee (Baht)
Truck 1	Wichit Sopon Arun Andy	10	840%	1.30	5.5 km	150
Truck 2	-	-	-	-	-	-

Fig. 5 Report by the SWMS for each truck

checks the current filling level Fc of each bin and compares it with the filling level Fi before job assignment, and if Fc < Fi, it considers the corresponding bin to be picked up).

After the last bin (bin 7) has been emptied, the status "end job" is displayed on the operator's screen, and the "Waste Management Report" data appear displaying the completed tasks of the job (Fig. 5). The report shows the job report which contains number of trucks and staffs has been on duty and the number of bins, the amount of waste that has been picked up, the time spent and the distance covered by each truck and the total distance together with the fuel fees.

5 Conclusions and Further Work

This paper has reported on the results of a prototype test regarding a smart waste management system (SWMS) based on the IoT, which together builds the Internet of wasted things (IoWT). The objective of the development was to build a cost-effective hardware and software system, which is easy to use and to operate. Therefore, a NB-IoT solution was applied interfaced with Raspberry Pi for communication between the smart bins, the garbage collection station and the trucks. The SWMS lets operators assign jobs to available trucks as deemed necessary according to the current filling levels of the smart bins in the garbage collection area. The SWMS calculates the shortest paths between the smart bins that need to be emptied, which also minimizes fuel costs.

As we have undertaken the test with a prototype SWMS in a small subdistrict in the north of Thailand, the results are limited. Nevertheless, we have shown that the SWMS is feasible and working and may provide a promising starting point for similar projects.

Further work includes (besides extending the waste collection area to an urban one): capability of the smart bins to classify the type of waste, such as bottle, dry or wet waste, recycle waste or trash with value and link with recycling factories nearby. Therefore, the dynamic shortest routing is important to reduce fuel and power consumption [14] and increase the ability to correctly send data from the smart bin to the management system and smart trucks. Also, the multi-path distance vector routing algorithm in multi-homing mechanism-based IoT [15] could be applied for efficient route planning and power consumption. Moreover, the system should provide more detailed data and report them to the municipalities for managing the increasing amount of waste produced in modern cities.

Further work also includes a feasibility study regarding the use of FIWARE in the SWMS, which aims at building a sustainable ecosystem around public software standards [16].

References

1. Guerrero, L.A., Ger, G., William, H.: Solid waste management challenges for cities in developing countries. Waste Manage. **33**(1), 220–232 (2013)
2. Phatcha: Thailand smart city. DEPA brochure (2017)
3. Saha, H.N., Auddy, S.: Waste management using Internet of Thing (IoT). In: 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON). IEEE Press, Bangkok, pp. 359–363 (2017)
4. Datta, D., Alam, I., Mallick, T.C.: Internet of Things (IoT) based waste management system. In: 4th National Conference On Natural Science and Technology 2017 in Asian University for Women (AUW), Chittagong (NCNST'17), (2017)
5. Accent systems, Differences between NB-IoT and LTE-M/What is NB-IoT, https://accent-systems.com/blog/differences-nb-iot-lte-m/. Last Accessed 20 August 2019
6. Samih, H.: Smart cities and internet of things. J. Inf. Technol. Case Appl. Res. **21**(1), 3–12 (2019)
7. Bigirimana, S.: Managing waste through the Internet of Things (IoT): In: ACRID'17 EAI International Conference for Research, Innovation and Development for Africa. Doi: 10.4108/eai.20-6-2017.2270818 (2018)
8. Suresh, A., Udendhran, R.: An efficient framework based on cloud computing integrated with Internet of Things technology for intelligent waste management. In: Raj, P., Koteeswaran, S. (eds.) Novel Practices and Trends in Grid and Cloud Computing. IGI Global (2019)
9. Anagnostopoulos, T., Zaslavsky, A., Kolomvatsos, K., Medvedev, A., Amirian, P., Morley, J., Hadjieftymiades, S.: Challenges and opportunities of waste management in IoT-enabled smart cities: a survey. IEEE Trans. Sustain. Comput. **2**(3), 275–289 (2017)
10. Medvedev, A., Fedchenko, P., Zaslavsky, A., Anagnostopoulos, T., Khoruzhnikov, S.: Waste management as an IoT enabled service in smart cities. In: 8th International Conference on Internet of Things and Smart Spaces, ruSMART (2015)
11. Khan, F.I., Gawade, A.: Dynamic routing for waste management using IoT for cost-efficient service. In: International Conference on Current Trends in Computer, Electrical, Electronics and Communication (ICCTCEEC-2017), pp. 222–230 (2017)
12. Esmaeilian, B., Wang, B., Lewis, K., Duarte, F., Ratti, C., Behdad, S.: The future of waste management in smart and sustainable cities: a review and concept paper. Waste Manage. **81**, 177–195 (2018)
13. Namahoot, C.S., Brückner, M.: SPEARS: Smartphone emergency and accident reporting system using social network service and Dijkstra's algorithm on Android. Lect. Notes Electr. Eng. **310**, 173–182 (2015)
14. Raj, J.S.: QOS optimization of energy efficient routing in IoT wireless sensor. J. ISMAC **1**(01), 12–23 (2019)
15. Krishnaraj, N., Smys, S.: A multihoming ACO-MDV routing for maximum power efficiency in an IoT environment. Wirel. Pers. Commun. 1–14 (2019)
16. FIWARE. FIWARE Community, https://www.fiware.org/about-us/. Last Accessed 15 Sept 2019

Leveraging Artificial Intelligence for Effective Recruitment and Selection Processes

Srirang K. Jha, Shweta Jha and Manoj Kumar Gupta

Abstract In recent years, artificial intelligence (AI) has revolutionized all the management functions across the globe. AI has touched all aspects of management through automation and robotization by substituting human beings at the workplaces with greater efficiency and lower costs. Most of the repetitive and standardized jobs and some of the highly specialized jobs are now being performed by machines powered by AI. Recruitment and selection function, which is quite critical for the success of organizations, has also been significantly disrupted by AI. Now, a good number of small, medium, and large companies are using AI for staffing. Some of the finest AI-driven tools for managing the entire hiring process are available and being adopted rapidly. This research paper is a humble attempt at examining how AI-powered hiring is changing the traditional recruitment and selection processes. Besides, concerns of various stakeholders have also been discussed in the paper.

Keywords Artificial intelligence · Recruitment process · Selection process · Strategic staffing · Workforce diversity · Psychometric testing · Natural language processing

1 Introduction

Artificial intelligence (AI) is the process of activity automation which is associated with human thinking, e.g., decision making, problem solving, and learning [1]. AI is a machines' capacity to read external information accurately, assimilate knowledge from such data, and to employ those inputs to attain particular objectives and tasks via flexible adjustments [2]. AI has an overarching influence on almost all aspects of management systems and processes. Indeed, AI intensifies organizational effectiveness in terms of reduction in errors induced by human discretion and its phenomenal

S. K. Jha (✉) · S. Jha
Apeejay School of Management, New Delhi, India

M. K. Gupta
University School of Information Communication and Technology, Guru Gobind Singh Indraprastha University, New Delhi, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_27

power to analyze big data and make precise predictions to the advantage of business houses. No wonder, several HR systems and processes are also powered by AI these days, resulting in greater efficiency and phenomenal outcomes.

Staffing, being a critical HR function, is one of the major gainers of the unparalleled magic of AI. However, for almost four decades, the processes regarding recruitment and selection have been archaic. Traditionally, companies have recruited potential employees through newspaper advertisements, local employment offices, employee referrals, etc. Likewise, selection processes have also remained somewhat old-fashioned, e.g., screening of resumes manually by matching competencies and job requirements, preliminary telephonic interview followed by personal interview. In some low-profile jobs, written tests or skill tests were also used. At every stage, selection of the employees depended on managerial discretion mixed with inherent biases and subjectivity.

Things have changed drastically with the advent of AI-driven hiring tools. Talent acquisition through advanced AI technologies ensures 'best-fit' thanks to the absence of subjectivity, which cannot be completely neutralized in a traditional recruitment process driven by people at the helm. Using artificial neural network, AI can make precise predictions about candidates who would be most suitable for the jobs vice versa [3] This is also probable because AI-based hiring allows the cross-referencing of previously reported company profile information in order to create a candidate model, as a result, when new vacancies are compared to this model with respect to the tests, experiences, general information in the curriculum vitae in order to verify the similarity with the ideal profile [4].

Use of AI-powered sourcing systems such as Robot Vera, Chabot called Mya, HireVue, by multinational corporations (MNCs) like Amazon, IKEA, L'Oreal, Unilever among others, has helped them augment their candidate sourcing strategies in terms of eliminating unqualified candidates, evaluating potential recruits, and collecting data from body language and facial expressions [5]. AI-based hiring processes have been characterized by speed, accuracy, objectivity, and cost-effectiveness.

The most dynamic factors which have impelled the technology-driven hiring processes in the last decade include (a) changes in the employment conditions during period of economic growth when number of open jobs surpasses the number qualified applicants available, (b) changes in the corporate business practices that force outsourcing or automation of activities that are considered noncore to the business, (c) changes and advancements in the communication, software and hardware technology, and (d) changes and advancements in the data collection and data analysis methodologies to accurately predict a fit between a prospective employee and the organization [6]. It is heartening to note that a number of technology companies have come forward to meet the growing demand for AI-based hiring processes by small, medium, and large enterprises.

2 Related Work

There are majorly three categories of software tools that facilitate AI-based hiring, viz. (a) job aggregator software, (b) candidate assessment software, and (c) application tracking software [7]. In Table 1, we provide a comprehensive view of the AI-based hiring software.

AI-powered algorithms are used generously at all the steps of recruitment and selection processes starting from sourcing to screening of applicants to selection and on-boarding by major corporate houses across the globe. Technology is being used to encourage potential candidates to apply for positions that suit their profiles. Getting the right number of candidates for a few profiles such as nursing, IT, and

Table 1 AI-based hiring software

Job aggregator software	Candidate assessment software	Application tracking software
Functions		
Scanning of jobs posted on various Web sites and collating them on Web site of job aggregator Classifying and listing job openings into a single searchable feed with relevant filters for the potential applicants Attracting potential candidates against job vacancies in sync with their suitability Offer services such as analytics related to candidate profiles to the hiring companies	Scanning of jobs posted on various Web sites and collating them on Web site of job aggregator Classifying and listing job openings into a single searchable feed with relevant filters for the potential applicants Attracting potential candidates against job vacancies in sync with their suitability Offer services such as analytics related to candidate profiles to the hiring companies	Handling recruitment needs of the organization Tracking the candidates Creating and promoting open jobs Collecting resumes Screening and creating a shortlist of suitable job applicants Scheduling interviews of shortlisted job applicants Managing interview process Facilitating on-boarding process for the selected job applicants
Indicative list of companies offering the services		
CareerJet Glassdoor Indeed Jobrapido JobRobot JustJobs LinkedIn Jobs LinkUp Monster SimplyHired StartJobs Trovit	Adaface AmazingHiring Bryq Codility Devskiller eSkill First Choice Hiring Solutions Fluttr HackerRank Harver HireMojo	CEIPAL TalentHire CV Analytics FreshTeam JobItUs Newton Applicant Tracking PeopleApex RecruitPro 360 TalentCube Talentpool Zoho Recruit Zwayam

Source Compiled by the authors

middle management is quite challenging. Mya Systems not only offers a unique AI-driven chatbot for conducting the entire interview process but also helps its clients by actively reaching out to candidates already in their application system, alerting them to new opportunities [8]. Moreover, there is a problem of plenty to complicate the issue of lower number of appropriate candidates.

At times, companies receive huge number of applications, say in millions, which can be handled only by machines. For example, Tesco, one of the largest private sector employers in the UK, got over three million job applications in 2016 [9]. In such scenarios, automated process helps the companies have a look at the profiles of each candidate and sort the most suitable ones which is not possible by deploying human beings at the helm. Besides, the AI-driven algorithm also rules out any bias in the sorting process. AI-based tools for candidate identification use a newly created job posting and go through the company's resume database and bring matching candidates to recruiter's attention and more advanced tools provide further capabilities of visiting the social media sites such as Facebook, LinkedIn, etc., on a predetermined frequency and adding profiles of people that match with certain keywords defined by the recruiters into the applicant tracking system [7].

Interestingly, AI has successfully substituted the psychologists to make precise predictions about the potential candidates. For example, AI-powered HireVue, which specializes in pre-hire assessments, asks potential candidates to answer a few standard questions on camera and makes note of thousands of barely perceptible changes in posture, facial expression, vocal tone, and word choice in order to predict about their competencies and whether they would be 'best fit' for a particular job. HireVue algorithm assigns scores to each video interview based on more than 250,000 data points, including audio, tonality, and speech patterns and makes almost accurate predictions about suitability and 'best fit' of a candidate for a particular job. Besides accuracy in predictions about the competencies of the potential candidates, HireVue algorithm has also succeeded in augmenting workforce diversity with minimal human interface in the entire hiring process, an indication that the HireVue process is absolutely free of any prejudice. Diversity in the talent pool of Unilever increased by 16% after the company partnered with HireVue [8].

AI-based tools, e.g., chatbots are also used significantly to rev up the involvement of the job applicants by using natural language processing (NLP) which demonstrates human-like intelligence while interacting with the job applicants throughout the hiring process [7]. Chatbots have indeed reduced the workload of talent acquisition professionals who can now use their time on more strategic issues rather than focusing on routine but important aspect of hiring that is constantly being in touch with the job applicants so as to avoid any last minute withdrawals by the latter.

3 Proposed Conceptual Model

Artificial intelligence is geared to arouse creativity by escalating the advancement of ground-breaking products, assisting organizations in obtaining additional revenue streams more swiftly, and cutting down superfluous expenses in the process, consequently raising cost-effectiveness [10]. While AI-driven algorithms for recruitment and selection processes have been well received by the industry across the board, job applicants have a number of concerns and apprehensions. First of all, their major concern is the security of their personal data which are handled by third-party organizations which provide services to the recruiting companies. Besides, there might by prejudice in-built in the algorithm itself due to subconscious biases in the minds of software programmers who write codes for the same. In such cases of inadvertent exclusions, the candidates have no way to seek relief.

Besides, there are job seekers who are a bit challenged when it comes to automated screening processes. Testimonies of such candidates indicate that there are rejections even when the applicants have right kind of skills, education, and experiences and they get disillusioned and stop looking for jobs which have AI-driven algorithm for sorting, screening, or pre-hiring assessments [9]. Such unexpected fall out is not good for any of the stakeholders in the process, i.e., the job seekers, the companies looking for people, and the software service providers who thrive on AI-driven algorithm for recruitment and selection processes. Penetration of AI-based hiring processes is quite significant not only among larger corporations but also among small and medium enterprises. Hence, it is imperative to take a break and ponder over genuine concerns of the job seekers' communities across the world. Perhaps a little bit of safeguards to the job seekers may be ensured through the European Union General Data Protection Regulation (GDPR) which mandates the company to disclose whenever a decision that 'significantly affects an individual' was automated with a caveat that the applicant would also be entitled to contest the decision, or request human intervention [9].

Advocates of AI-based hiring processes, however, have a different take on the possibility of any discrimination creeping out during recruitment and selection. All said and done, AI-driven hiring tools are just as biased as the humans who train them [11].

Hence, it is imperative that the organizations pioneering in AI-powered hiring tools are willing to integrate the concerns of potential employees—the stakeholders who are at the receiving ends. While the user organizations are the significant beneficiaries, privacy concerns as well as potential inbuilt discrimination are generally overlooked in AI-powered processes. Here, we propose a conceptual model to reduce the inherent lacunae in the AI-driven staffing processes. The organizations offering AI-driven tools for recruitment and selection may easily follow the proposed of five-stage developmental process so as to continuously improve their offerings without alienating the job seekers who are critical for the success of such automated tools of hiring (Table 2).

4 Conclusion

Dominance of AI in recruitment and selection processes is now an accepted fact. Already hiring function has been outsourced by most of the small, medium, and large

Table 2 Proposed conceptual model of AI-driven staffing processes

AI-driven staffing processes	
Stage 1 development	Stakeholder consultation with the potential candidates to remove any inadvertent prejudice due to subjectivity and cultural biases of the developers and address their privacy concerns
Stage 2 testing	Mock testing on potential candidate to attain reliability and validity of the testing tools (this will boost the confidence of both the developers and the candidates
Stage 3 implementation	Implementing the AI-driven tools in real-life scenarios
Stage 4 feedback	Stakeholder feedback from the candidates taking the tests in terms of their perception about the tools used especially in terms of ethics and privacy
Stage 5 redevelopment	Periodic redevelopment of the AI-driven staffing tools right from stage 1

companies to specialized staffing firms which are using AI in a big way to attain scale, effectiveness, and cost-effectiveness. Hence, the traditional hiring practices may soon become obsolete. It is high time that the fence sitters take a close look at the emerging scenario vis-a-vis AI-powered hiring practices and adopt the new way of filling vacancies in their respective companies with a greater probability of right fit with lower hiring cost. At the same time, the staffing companies also need to consider the concerns of job seekers with regard to exclusion, discrimination, and privacy while leveraging AI-powered tools in the hiring process and find ways to reassure them about the neutrality of the machines and make the interface more user-friendly and comfortable. It is all the more important that issues related to ethics and privacy must be taken very seriously while implementing AI-driven processes because people may have some valid or unfounded apprehensions about the machines taking control over their lives [12].

Moreover, in spite of the importance of the AI-based hiring process, only limited research has been conducted on this topic [13]. It is imperative that research on three fronts should be undertaken simultaneously, i.e., (a) improving the existing AI-based tools/software for recruitment and selection processes, (b) examining the incidence of adoption and overall attitude as well as concerns of the users at both the ends, i.e., job applicants as well as the recruiters, and (c) finding ways to scale up adoption of the AI-based tools by organizations across the board so as to bring down the cost. This requires an interdisciplinary approach to put AI-based hiring systems in proper perspective so that valid concerns of all the stakeholders, such as privacy, prejudice, discrimination, exclusion, joblessness due to process automation, etc., can be addressed to the satisfaction of all.

References

1. Bellman, R.: An Introduction to Artificial Intelligence: Can Computers Think?. Thomson Course Technology, San Francisco (1978)
2. Kaplan, A., Haenlein, M.: Siri, siri in my hand: who's the fairest in the land? On the interpretation, illustrations and implications of artificial intelligence. Bus. Horiz. **62**(1), 15–25 (2019)
3. Rocabert, F.I.: Artificial neural network system applied to human resource management. http://hdl.handle.net/2117/109555
4. Jabota, M., Gutierriz, I., Fernandes, P.O., Teixeira, J.P., Moscon, D.: Artificial Intelligence in the recruitment and selection: innovation and impacts for the human resource management. Economic and Social Development: Book of Proceedings, 96–104 (2019)
5. BasuMallick, C.: 3 B2C Companies using AI to transform their candidate sourcing strategies. https://www.hrtechnologist.com/articles/recruitment-onboarding/3-companies-using-ai-to-transform-their-candidate-sourcing-strategies/
6. Reynolds, D.H., Weiner, J.A.: Online Recruiting and Selection: Innovations in Talent Acquisition. Wiley, Hoboken (2009)
7. Kulkarni, S.B., Che, X.: Intelligent software tools for recruiting. J. Int. Technol. Inf. Manag. **28**(2), 1–16 (2019)
8. Riley, T.: Get Ready, This Year Your Next Job Interview May Be With an AI Robot. https://www.cnbc.com/2018/03/13/ai-job-recruiting-tools-offered-by-hirevue-mya-other-start-ups.html
9. Buranyi, C.: Dehumanizing, impenetrable, frustrating: the grim reality of job hunting in the age of AI. https://www.theguardian.com/inequality/2018/mar/04/dehumanising-impenetrable-frustrating-the-grim-reality-of-job-hunting-in-the-age-of-ai
10. Plastino, E., Purdy, M.: Game changing value from artificial intelligence: eight strategies. Strat. Lead.Ship **46**(1), 16–22 (2018)
11. Economic Times: AI Hiring Tools Can be As Just As Biased as Humans. https://economictimes.indiatimes.com/jobs/ai-hiring-tools-can-be-just-as-biased-as-humans/articleshow/71513678.cms?from=mdr
12. Akerkar, R.: Introduction to artificial intelligence. Artificial Intelligence for Business. SpringerBriefs in Business, pp. 1–18. Springer, Cham (2019)
13. Rodney, H., Valaskova, K., Durana, P.: The artificial intelligence recruitment process: how technological advancements have reshaped job application and selection practices. Psychol. Issues *7*, 42–47 (2019)

Trust Computing Model Based on Meta-Heuristic Approach for Collaborative Cloud Environment

Pooja Pol and V. K. Pachghare

Abstract Collaborative cloud environment required trustworthiness from and to consumer and provides services associated with it. Operations related to consumer's side are able to send task and crucial data to the cloud data center, that operation raises high trust between both the parties. Large flow of user request becomes complicated and heavy to mange in collaborative computation offered by the cloud. So, inter-service bed must provide and tackle large number of consumer request efficiently and provide or allocated trustworthy resources to associate the request. In this context, precision-based manageable service provider platform with respect to trust is emerging as a vital problem. In this paper, we have proposed a precise and collaborative trust computing mechanism by using meta-heuristic approach for trustworthy cloud computing. In initial module, inter-service platform relay response and request for managing and logging of allocated virtual machine. In the second module, resource selection by using nonlinear objective functions with the help of meta-heuristic approach to improve trust evaluation. All those operation depend on quality parameter. Therefore, in the third module, we proposed QoS prediction by appling learning approach, i.e., neural network to understand historical information in collaborative cloud computing environments. Performance analysis and experimental results verified the feasibility and effectiveness of the proposed scheme.

Keywords Meta-heuristic optimization · Evolutionary algorithms · Neural network · QoS cloud parameter · Trust-based system · Cloud computing

P. Pol (✉) · V. K. Pachghare
Department of Computer Engineering and Information Technology, College of Engineering Pune, Pune, India

V. K. Pachghare
e-mail: vkp.comp@coep.ac.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_28

1 Introduction

Cloud computing's latest development is due to its ability to deliver software, infrastructure, and platform services without big investments or expenditures in the management and operation of these services. It includes customers, suppliers, infrastructure service, and so on [1, 6, 7]. They include apps supplied as services, as well as the difficult and softcore technologies providing these services. Collaborative virtualization has drawn the attention of big and small sectors to a big extent. As the Internet becomes an imminent phase of network technology growth, cooperative cloud computing is anticipated to become an approachable era in the economic expansion of cloud computing. Collaborative cloud computing effectively utilizes information systems as a business service and offers highly powerful computing capabilities and huge processing space at reasonable cost to target consumers. Cohesive cloud computing also covers increasing energy or even environmental issues as cooperative cloud technology proponents enhanced resource allocation. Regarding the advantages of cooperative cloud computing, this new frontier still heads several difficulties linked to computing confidence [2, 11], speed of reaction, and instant matching of resources. These difficulties bring fresh integrative designs, techniques for cooperatives, and systems for selection.

2 Motivation

Evolved to address the insufficiency of conventional a process of authorization, the trust computing process is usually regarded as the pillar of cloud infrastructure survival [3, 8, 12]. Building confidence in a cloud setting offers the following main advantages from the user view. Optimize and Precise security [10, 12].

In between consumer and cloud environment, most important mechanism is trustworthiness. Thus, trust block takes care of transaction between cloud and consumers to optimize whatever least negotiation.

3 Related Work on Cloud Monitoring and Trustworthy Cloud Service [5, 9, 12]

In cloud computing, asset distribution performs an important role to play in less time span. Meta-heuristic techniques have been used to provide fast ideal performance for the complicated allocation process within the necessary timeframe. Paper suggested cloud-based module. On cloud computing-based popular meta-heuristic techniques, and neural network for service selection as well as nonlinear objective function for trust match making.

A. **Collaborative Cloud Integration [6]**

To quickly identify the best service in a massive cloud environment with big network entities is a challenging task. The technology of tracking agents offers extensible processes that can be the best solution for efficient cloud surveillance. In this job, we use a compact design based on dispersed surveillance agents to track big-scale VM service conduct through its modular and disseminated design. This structure may rapidly interpret VM's service conduct with big network entities in the cloud setting. Customers can use a chosen cloud agent to acquire the service. Offering readily agent's primary job is to provide reliable and safe facilities. In addition, this design is a cloud service agent that is enriched by security. The outcome of trust computing, such as identity management, authorization, can be used for promise-based security measures and the matching of resources. As illustrated in Fig. 1, the secure, reliable cloud service realtor consists of three key modules.

The first is the module for communication and agent handling, which has two fundamental tasks: link and alignment to the cloud service, and perception of object-based information. To retrieve and inflation rate all resource data from various suppliers, the cloud service link and adaption sub-module are used. This sub-module encompasses various tool chains provided by various cloud service providers or cloud pages. So that the certain trusted service surveillance design modules need to know even one set of interfaces. The publishing agents and sub-module interpreting records track the true-time service data of the assigned resources to ensure user SLA. In the immersive system, this sub-module interactively controls the service actions of VMs by disseminating operatives on remote cloud sites and is accountable for interpreting data on runtime service behavior, including information on conduct related to safety and QoS. The controlled information is stored on an evidential basis that the surveillance system maintains. This sub-module also releases the recent version of tracking units immediately to all remote cloud site broker executives when a fresh version of a broker is being created. To avoid the hacking of faith agents by fraudulent customers, we can implement a monitoring process around agents and the monitor to reinforce the trust system's safety and remove the problem of information manipulation. In order to promote a verifying system, two network entities, several other efficient techniques can be implemented.

B. **Trust Evaluation [4, 9]**

To evaluate trust, taking into account a sector-wide cooperative cloud service setting with a host handling thousands of service suppliers, real-time surveillance of each VM is a difficult job from the customer demands standpoint. In the system suggested, we have embraced a disseminated and monolithic strategy to managing agents. The agent-based cloud surveillance system can rapidly interpret wide-scale VM service conduct in the cloud setting with big network organizations through such a dispersed and patriarchal architecture. Consumers can use a chosen cloud broker to acquire the service. The fundamental job of broker is to provide quick, reliable, and safe service. In a cooperative cloud service setting across the business sector, it is necessary to

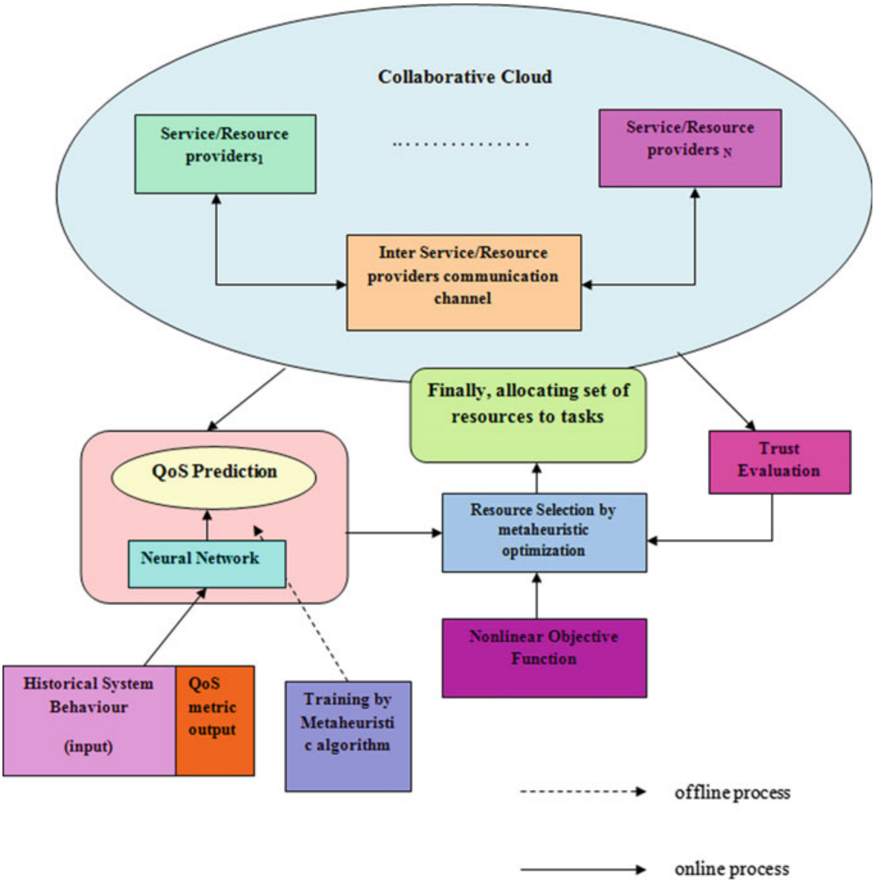
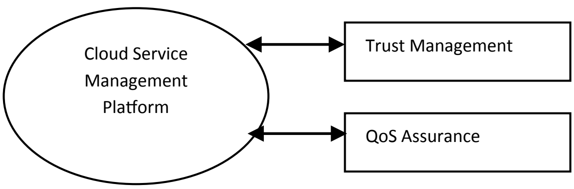


Fig. 1 Base proposed model

deploy various cloud brokers. Such brokers are frequently attempted by a well-known provider such as intermediaries. As illustrated in Fig. 2, the cloud surveillance scheme predicated on the agent consists of three groups (Fig 3).

B.1. **Reputation Measurement for trust management**

Fig. 2 Cloud service behavior



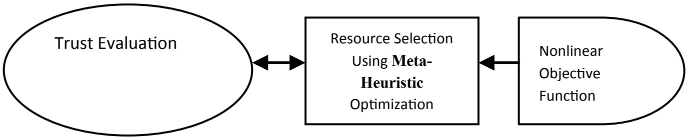


Fig. 3 Trust evaluation

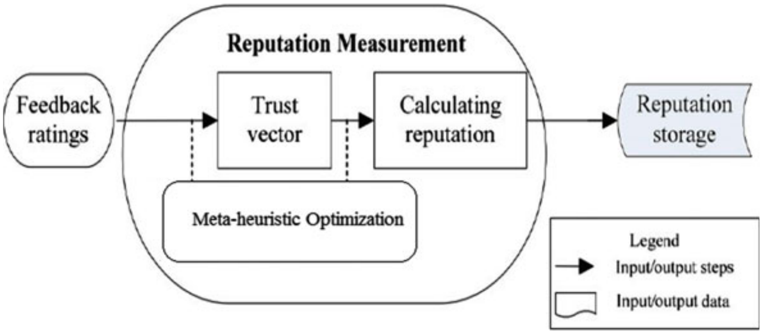


Fig. 4 Reputation measurement

We suggested a design of confidence called the Favorability Trust Model for accessibility feedback. Cloud services suggested credibility calibration strategy includes two stages. The first stage is Trust Vector, where we use the cloud model to analyze the amount of uncertainty of response ratings. The second phase is Credibility Calculation, where we use heuristics to measure each cloud service's credibility rating. The legacy results that are significant yardstick for cloud service applications are ultimately collected (Fig. 4).

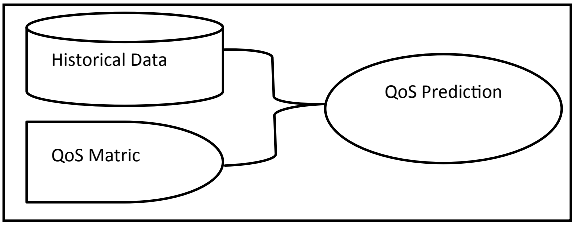
Trust Significance is the calculation of the trust values of provides recorded by agent. The measured trust values against either the corresponding registered providers are then presented in agent. From the leftmost column, trust factor can be measured.

$$\begin{aligned}&\text{Trust(Significant)} = \text{Trust}\\&\text{prev(Significant)} + \text{Trust}\\&\text{prov(Significant)} + \text{Trust}\\&\text{cons(Significant)}\end{aligned} \tag{1}$$

where Trustprev(Significance) is Broker's prior pi trust value, Trustprov(Significance) is the average Significance validation value allocated by other trustworthy suppliers. The average Significance validation value allocated by other reliable service is Trustcon (Significance)

$$\text{Trust prov(Significant)} = \sum fb(Pi, Pj)nj = 1, \quad j \neq in \tag{2}$$

Fig. 5 QoS trust behavior



where Function(*Pi*, *Pj*) is the significant provider's response value allocated by the *Pj* supplier. Fb is the estimate of the complete elevations provided to each query in the reviews that are between 0 and 5.

$$\text{Trust cons}(pi) = \sum \text{function}(Pi, Cj)nj = 1, \quad j \neq in \tag{3}$$

C. **Security and QoS-based Trust Behavior**

As a supplementary safety features, trust solves the problem of offering appropriate authentication based on assessing client satisfaction and makes conventional safety facilities extremely reliable and secure by maintaining that all communication devices are secure during authentication, authorization, or the stewardship of the key. We concentrate on two convince characteristics of cloud service conduct from the view of safety improvement and QoS certainty, notably safety-related and QoS-related behavior (Fig. 5).

4 Trust-Based Pricing Includes Two Main Problems that Need to Be Addressed

1. Trying to calculate the client's confidence. The trust between the customer and the supplier relies on many variables: the specific trust, the track record reported by all colleagues, the trust strength node, and the weights assigned to immediate trust and credibility by a specific client. Precise trust and credibility can be numerically estimated, but all straps are entirely personal specifications depending on the client's needs.
2. Identifying a feature for valuations. The impact of confidence in the tendered rates must be determined by each supplier. There is a lack of information about the customers' desires to the suppliers' confidence. Our past research has shown that evolutionary algorithms are appropriate for offering prices in concealed data economies even though they evolve the availability feature quickly to an evolving unknown setting nevertheless for the reasons of rationalization.

5 Analysis and Discussions

In this section, the GA without trust, GA with trust, meta-heurastic without trust, meta-heurastic with trust methods are used. The average schedule length is shown in Table 1 and Fig. 6.

From Fig. 6, it can be observed that the HEFT without trust has higher average schedule length by 6.75, 9.29 and 12.44% for 300 number of jobs, by 5.84, 8.8 and 10.87% for 600 number of jobs, by 9.35, 13.31 and 17.2% for 900 number of jobs, by 12.92, 15.75 and 17.94% for 1200 number of jobs and by 6.45, 9.39 and 12.02% for 1500 number of jobs when compared with HEFT with trust, GA without trust, and GA with trust.

Table 1 Average schedule length

Number of jobs	GA without trust	GA with trust	MH without trust	MH with trust
300	811	758	739	716
600	1707	1610	1563	1531
900	2651	2414	2320	2231
1200	3582	3991	3875	3774
1500	4257	3991	3875	3774

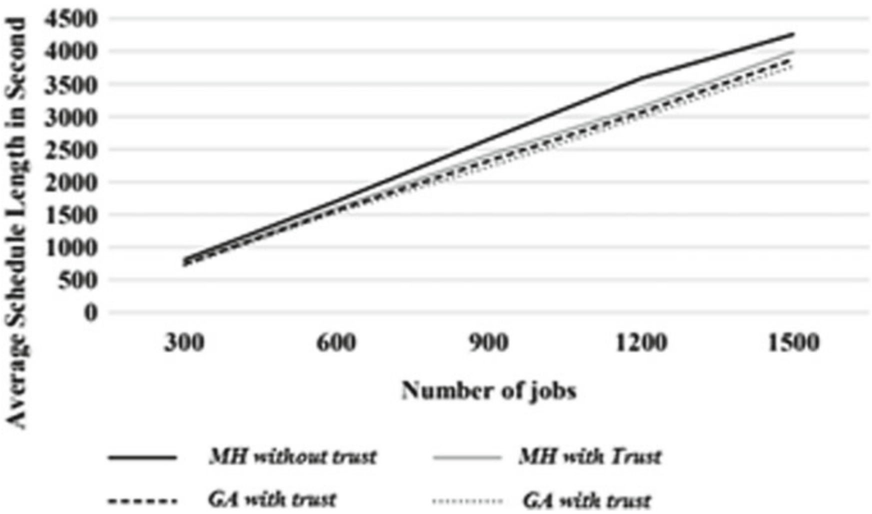


Fig. 6 Average schedule length

6 Conclusion and Feature Scope

As a supplementary design with a conventional safety mechanism, trust avoids the problem of offering appropriate identity management by assessing service conduct and makes conventional safety services extremely reliable and accurate by maintaining that all connecting devices are trusted during verification, permission, or key management.

In this job, we suggested an unconventional system for perceiving and quarrying security and QoS-related convince conduct by depending on dispersed and smart agents. Based on propose architecture design, the suggested system can effectively interpret the service conduct of significant-scale VMs and rapidly perform the integrity calculation of service assets on the basis of the information collected in real time.

To our greatest understanding, in future, for reliable cloud services with a versatile and adjacent confidence computing, system based on large data analysis, assessment of quality and trust can be calculated with different trust and QoS parameters.

References

1. Ivona, B., Schahram, D.: Grid vs cloud—a technology comparison. Vienna University of Technology, 4 (2011). https://doi.org/10.1524/itit.2011.0640
2. Wenjuan, L., Lingdi, P., Qinlong, Q., Qifei, Z.: Research on trust management strategies in cloud computing environment. J. Comput. Inf. Syst. **8**(4), 1757–1763 (2012)
3. Parameswari, R., Priya, C., Prabakaran, N.: A trust, privacy and security infrastructure for the inter-cloud. Int. J. Comput. Technol. Appl. **3**(2), 691–695 (2012)
4. Talal, N., Quan, S., Sherali, Z., Jian, Y.: Trust management of services in cloud environments: obstacles and solutions. ACM Computing Surveys (2013)
5. Zhi-Yong, Y., Heng-Wei, Z., Kan, N., Jin-Dong, W.: Research on service trust evaluation approach under cloud computing environment. IET (2015)
6. Xianrong, Z.: QoS representation, negotiation and assurance in cloud services. A thesis submitted to the school of computing (2014)
7. Jinbo, X., Ximeng, L., Zhiqiang, Y., Jianfeng, M., Qi, L., Kui, G., Patrick, C.: A secure data self-destructing scheme in cloud computing. IEEE Trans. Cloud. Comput. (2014)
8. Priya, C., Prabakaraan, N.: A research on trusted cloud computing with secure resources for multiple clouds using data coloring. IJIRCCE **3**(4) (2015)
9. Salman, M., Ahmedand, T., Neeraj, S.: Trust validation of cloud IaaS: a customer centric approach. IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. (2016)
10. Jingjing, G., Jianfeng, M., Xinghua, L., Junwei, Z., Tao, Z.: An attribute-based trust negotiation protocol for D2D communication in smart city balancing trust and privacy. J. Inf. Sci. Eng. **32** (2016)
11. Bahador, S., Amir, M., Nooruldeen, Q.: Cloud computing service negotiation: a systematic review. Comput. Stand. Interfaces (Jan 2018). https://doi.org/10.1016/j.csi.2017.08.006
12. Xiaoyong, L., Yuan, J., Huadong, M., Yao, W.: Fast and parallel trust computing scheme based on big data analysis for collaborative cloud service. IEEE Trans. Inf. Forensics Secur. (Feb 2018). https://doi.org/10.1109//tfis.2018.2806925

Integration of Generation Y Academician Attributions with Transformational Leadership Style: Association Rules Technique on Minimizing Turnover Intention

Charles Ramendran SPR, Anbuselvan Sangodiah, Vimala Kadiresan, Ramesh Kumar Moona Haji Mohamed and Che Supian Mohamad Nor

Abstract Private universities in Malaysia are facing difficulties in retaining Generation Y academics due to divergence in attributions with leadership styles (head of department). In this study, 150 academics were participated to determine patterns of attributions most suited transformational leadership style by using association rules technique. Results showed all the attributions have met the minimum support level which is 30% except for four attributions (seek new ways to solve problems, creative and rethinker); however according to association rules technique, these attributes have shown a unique finding whereby it found strongly significant relationship with intellectual stimulation leadership style. As a result, apart from 150, 76 Generation Y academics have chosen intellectual stimulation leadership style as their choice.

Keywords Transformational leadership style · Association rules technique · Intellectual stimulation · Generation Y · Transformational leadership style

Charles Ramendran SPR (✉) · R. K. M. H. Mohamed
Faculty of Business and Finance, Universiti Tunku Abdul Rahman, 31900 Kampar, Perak, Malaysia
e-mail: charlesr@utar.edu.my

R. K. M. H. Mohamed
e-mail: rameshk@utar.edu.my

A. Sangodiah
Faculty of Information Communication and Technology, Universiti Tunku Abdul Rahman, 31900 Kampar, Perak, Malaysia
e-mail: anbuselvan@utar.edu.my

V. Kadiresan
Faculty of Business, Economics and Accounting, HELP University, 40150 Shah Alam, Selangor, Malaysia
e-mail: vimala.k@help.edu.my

C. S. M. Nor
School of Distance Education, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia
e-mail: supian@usm.edu.my

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_29

1 Introduction

Transformational leadership was first introduced by Downton in 1973, followed by James McGregor Burns popularized transformational leadership in 1978 by mentioning on morality and motivation of subordinates rely on leadership; however, it was widely spoken and further elaborated by Bass in the 1980s. Bass [1] categorized transformational leadership into four; the components are idealized influence, individual consideration, intellectual stimulation and inspirational motivation. Each component of transformational leadership has its attributions in which subordinate's praise and see as their role model known as idealized influence, leader who gives personal attention on subordinate's growth known as individual consideration, subordinates are nurtured on creativity and innovative part as well as diminishing traditional way of performing and thinking known as intellectual stimulation [2] and lastly, inspirational motivation known as leaders who able to inspire and motivate subordinates.

Leithwood et al. [3] argued that transformational leadership style is one construct that needs to give attention for affecting and enhancing the successful education institution learning process. It is further supported by [4] indicated leadership attributions crucial in impacting on project performance in academic field. Similarly, transformational leadership is more appropriate to apply during multifaceted project or state [5]. Transformational leaders can run the institution toward effectiveness, and subordinate's needs are met and encourage working harder [6]. According to [7], Generation Y are entering the workforce and by 2020, they are the dominant workforce in global. Gen Y'ers have high expectations on their leader whereby it must be knowledgeable and assign challenging task for them [8].

Gwavuya [9] stated leadership style of a leader has influenced employee's turnover intention. Cohen et al. [10] highlighted that turnover intention is the predictor of actual turnover, which means if an employee displays strong intention to leave he or she has high tendency to quit. Harhara et al. [11] quoted in order to address employees leaving organization, turnover intentions used as the measurement. Misunderstanding and frustration with superior can create unsupportive working environment consequently causing turnover intentions [12].

Various researches have been done about characteristics of Generation Y and leadership and its attributes, however a few described leadership attributes and styles that are preferred by Gen Y employees. McCrindle [13] posits specific leadership attributes required need to be examined with [14] added there is a need to integrate leadership style that preferred by Gen Y employees at workplace. Apparently, current and past research work uses the common statistical methods and techniques to investigate the relationships between Generation Y and leadership styles [15]. A case in point is the work by [16] where the researcher has used basic statistical analysis and multiple regression analysis to find the association between compensation package and the retention of Generation Y employees in Pakistan. In another work [17], identifying the factors that influence the purchase intentions of Generation Y students toward the fast-food industry in Nilai, Malaysia, relies on descriptive statistics, correlation analysis and multiple regression analysis. The association of factors such as food quality, service quality, restaurant environment and price with purchase intentions used correlation analysis. Apart from those aforementioned works, the researcher in [18] used basic statistical methods to study the extent to which generations of B and Y and gender differences respond to flexibility at work.

The fact is any study that relates to finding relationships or associations between two factors or variables, and it is very common for researchers to use statistical techniques. However, this study explores an alternative to statistical technique to study the relationships between two variables that are attributes of Generation Y and leadership styles. Association rules can also be used to find the relationships or associations between two or more than two variables. This non-statistical technique has been used and applied widely in the areas such as business and educational [19]. The usage of association rules mining in other areas can also be found in [20–22]. Generally, association rules technique is one of the most popular data mining methods [23]. In view of this, this study explores using association rules technique which is novel in the context of Generation Y academician attributions with transformational leadership style.

2 Literature Review

2.1 Related Work

Only two levels of headings should be numbered. Lower-level headings remain unnumbered; they are formatted as run-in headings. Gen Y'ers are different from other characteristics in terms of work-related attributes and attitudes and leadership styles. This illustrates that the managers need to understand and consider the best approach of leading people in order to retain their generation of workforce [24, 25]. In line with that, performance in workplace will operate at the optimal level when the managers are able to capitalize on the leadership and followership that exist within the Gen Y [26]. Differences of work attributions between generational groups should be identified by leaders and implemented in various work environments to boost employee motivation, commitment and performance [20, 21] as each generation will be nurtured with their respective time zone's values as they were raised with these values which they were brought up [16].

Sheu and Chu [22] explained that Generation Y had a positive attitude toward work where they are hopeful, talented, patriotic, collaborative, inclusive, civic minded and they lived with strong social stressors. The findings by [17] espoused several key attributions to Generation Y, namely competent, as the most important characteristics followed by determination, self-controlled, honesty and seek for problem solving. The Generation Y known for self-absorbed and self-reliant, and has a strong sense of independence and autonomy that emphasize on inner satisfaction, work–life balance and freedom and ambitious, optimistic and tech-savvy [23–25]. Besides, optimism, civic duty, confidence, achievement, a preference to a polite relationship with authority are several characteristics preferred by Gen Y [26]. Employees as such prefer to be supervised by leaders who get the consensus of all when decision making is concerned and have the drive to make changes happen.

However, they will opt to move jobs from one place to another in order to take up a new challenge [27, 28]. Similar findings by [29] explained that employees often will move on to find new horizons of opportunities if the intended needs are not met through opportunities to develop their inborn skills. Such employees are focused entirely on goals and achievement as loyalty is not the priority to pursue a career while personal life is [30] as job security is not a motivator because they are less interested in long-term employment [25, 28]. This explains that Gen Y will project better career when platform to express their abilities is created and push the employees to exploit these abilities fully by leaders [31] as they are extraordinarily confident of their abilities [32].

Generation Y demands for freedom and overall flexibility to be more creative in thoughts and innovation in action [33]. Gen Y'ers' relationships between their supervisors differ from the older generations [34]. This is apparent when Gen Y employees place a great value on friendships, coworkers and managers' relationships, social collaborations, and they thrive on working together to achieve goals, better knowledge creation, commitment and performance [23, 28, 29, 35, 36].

Gen Y tends to demonstrate an inclusive management style where they expect immediate feedbacks for their efforts rendered and emphasized [31]. At the same time, leaders must understand that younger generations want clear direction and constructive feedback on their performance on a constant pace [16].

Moreover, Generation Y expects organization to adapt to changes based on their needs and experience in a rapidly changing environment which expose to heavy reliance on technology [37]. They use technological platforms effectively, and they are technological savvy as they grew up with technology and perceived as most technically literate [22, 29, 31, 38–42]. These are through actively integrating technology into their lives by communicating using Internet, mobile phones and electronic gadgets which allow them to involve in receiving information and are connected to the peers in real time [33, 43]. Gen Y expects these as a support in working environment in order to prevent time waste [35].

The findings of [44] proposed that there could be a possible connection between Gen Y and their leader's leadership style. This is due to the fact that the respondents that were picked concluded that there are three main elements; Competency, determination and honesty were the most preferred attributes and influence the expected leadership style. These traits match with the high preference for idealized influence under the transformational leadership style. Moreover, transformational leadership has a better score compared to transactional leadership across all the factors in Generation Y. Besides, it is mentioned that older aged employees are more likely to practice transactional leadership [29], and on the other hand, younger managers favoured using individualized consideration and intellectual stimulation [45] which can be perceived that Gen Y most spontaneously appears to adopt transformational leadership. This is supported by [46] where Gen Y'ers expect more from their leader than the older ones, and they challenge the leader when preferring transformational leadership over the other styles. Transformational leadership studies have brought a lot of attentions to leadership styles which can influence leaders to apply and be involved in that style. Thus, it can be said that Generation Y has a connection with transformational leadership style. Therefore, the associations that exist between Generation Ys with transformational leadership style that were not previously studied would be investigated using association rules technique in this study.

3 Methodology

3.1 Data Sets

In this study, a total of 150 responses have been obtained from academic staff of Generation Y from several local universities in Malaysia which intends to find relationships between attributes of Generation Y and types of leadership styles of superior. The attributes of Generation Y that have been of interest in this study are technology savvy, rethinker, optimistic, ambitious, creative, innovative, self-absorbed, autonomy, independence and seek new ways to solve problems, while the types of leadership styles are individualized consideration, idealized influence, inspirational motivation and intellectual stimulation.

These ten attributes of Generation Y and four types of leadership styles have been used in this study to identify the relevant attributes that are matching with the specific type of leadership style. One of the data mining techniques which is association rules has been used in this study to find the relationships and inherent patterns between attributes and leadership styles. By using the technique, it is possible to derive patterns of combination or co-occurrence of several attributes matching or associating well with a specific type of leadership. Deriving these hidden and interesting patterns in this manner emanates from the concept of market basket analysis using association rules.

According to [47], market basket analysis (MBA) is a data mining technique to discover association rules in large data sets. It is very popular in the retailer industry where it [47] can be used to determine the products which are bought together by customers. In the context of MBA, association rules find interesting association among large set of data items where it is derived from frequent item sets using support and confidence as threshold levels. According to the author in [48], the support and confidence measures are defined as follows:

In this study, the minimum support is capped at 30% while minimum confidence at 70%, respectively. The main reason to set such minimum support is to avoid generating trivial frequent item sets when the minimum support is lower than 30% or missing the interesting patterns and associations between attributes of Generation Y and leadership styles when minimum support is higher than 30%. As the main focus

Table 1 Data sets by type of leadership style

Type of leadership style	Individualized consideration	Idealized influence	Inspirational motivation	Intellectual stimulation
Size of data set	24	22	28	76

of this study is to discover interesting relationships between attributes of Generation Y and leadership styles, the support measure and confidence measure are given more emphasis than other measures in the context of association rules in the results and discussion section later. We recognize left-hand side rule (X) to represent the attribute of Generation Y while the right-hand side rule (Y) to be another attribute of Generation Y, so the data set consisting of 150 responses is sliced into 4 data sets by type of leadership style. And association rules have been applied separately on individual data set as shown in Table 1 to generate frequent item sets and interesting rules of attributes of Generation Y for a particular type of leadership style. The possible frequent item sets that can be generated range from 1, 2, 3 and more. In practice, most of the research works related to association rules have indicated that the appropriate the number of frequent item sets that are useful in decision making are not beyond 5. If frequent item set is 1, it means that there exists only one attribute while frequent item set 2 means there will be two attributes, and in order for the confidence measure to be used in association rules, the minimum frequent item set is 2.

Frequent item set

1 frequent item set {self-absorbed}
2 frequent item set {self-absorbed, autonomy}
3 frequent item set {seek new ways to solve problems, creative, rethinker}.

For instance, if the support level is 50% for item set {ambitious, innovative} in the data set of intellectual simulation, it can be interpreted that 50% of total responses/transactions (76) have the occurrence of ambitious and innovative attributes of Generation Y in respect of leadership style of intellectual stimulation. While for the confidence level at 60%, using the same item set, it can be interpreted that an academic staff has 60% of ambitious characteristic, and they are also likely to possess innovative characteristic in association with the intellectual simulation leadership style. In this case, left-hand side (LHS) rule is ambitious attribute and right-hand side (RHS) rule is innovative attribute.

Although the association rules can also be used as a predictive modeling to predict type of leadership style based on the attributes of Generation Y, it is not the main focus as the study is more concerned about discovering patterns of co-occurrence of attributes of Generation Y for a particular type of leadership style. Future work may involve collecting more data for training and testing the accuracy of prediction model based on association rules.

3.2 Association Rules Algorithms

There are several association rules algorithms such as FP Growth, Apriori, Eclat, RElim and others that can be used to carry out market basket analysis (MBA). However, in this study, FP Growth has been chosen as it is generally a better algorithm in terms of performance in generating association rules. The work by [49] highlights better performance of FP Growth compared to other algorithms. Rapid Miner Studio tool has been used in this study as it is widely used in academic research [8].

4 Results and Discussion

Based on Table 2, the maximum number of item set (item set size) that has been generated in this study is 2 with the minimum support of 30% and the minimum confidence level of 70%, respectively. However, despite the confidence level for some association rules of 2 item sets belonging to intellectual simulation of leadership style hovering around 70%, the rules are interesting as the co-occurrence of the attributes is so distinct and unique which sets it apart from other co-occurrence attributes of Generation Y for various leadership styles. It has been found that the attributes such as seek new ways to solve problem, creative, rethinker, innovative and technology savvy are very significant to the leadership style intellectual simulation as those attributes were not found in influencing other leadership styles.

Although rules number 27 and 28 are below the minimum support level, it is rather interesting as it is 3 item sets and the co-occurrence of attributes is closely related to 2 item sets in the same leadership style. Strong association rules are also observed for leadership styles of individual consideration and inspirational motivation where the confidence level surpasses 70% exceeding the minimum support of 30%. It can be concluded that when academic staff has the characteristic of self-absorbed more than 70% likely they will also possess the characteristic of independence or autonomy in respect of individual consideration leadership style. The same applies for inspirational motivation leadership style where one being optimistic, more than 70% likely the person will also have the characteristic of ambitious or autonomy.

Generally, the co-occurrence of attributes of Generation Y of 2 item sets or 3 item sets for each type of leadership style is somewhat unique and different from those in other leadership styles. In this study, the use of association rules technique is appropriate as it unfolds favorable outcome where patterns of co-occurrence attributes are visible and distinct which can be used to determine a particular leadership style. When academic staff of Generation Y exhibit some attributes related to the Generation Y, one will be able to determine if a particular academic staff is suitable to the current leadership style being practiced in an organization. With this knowledge in place, minimizing turnover intention will be possible.

Table 2 Frequent item sets and association rules by type of leadership style

Rule no.	Leadership style	Items	Item set size	Support (min support level—30%)	Confidence (min confidence level—70%)
1	Individual consideration	Self-absorbed	1	0.840	NA
2		Autonomy	1	0.440	
3		Independence	1	0.400	
4		Self-absorbed, autonomy	2	0.375	0.818
5		Self-absorbed, independence	2	0.333	0.800
6	Inspirational motivation	Optimistic	1	0.750	NA
7		Ambitious	1	0.714	
8		Autonomy	1	0.393	
9		Optimistic, ambitious	2	0.500	0.700
10		Optimistic, autonomy	2	0.321	0.818
11	Idealized influence	Ambitious	1	0.727	NA
12		Independence	1	0.545	
13		Autonomy	1	0.545	
14		Optimistic	1	0.318	
15		Ambitious, independence	2	0.409	0.750
16		Ambitious, autonomy	2	0.318	0.583
17	Intellectual simulation	Seek new ways to solve problems	1	0.763	NA
18		Creative	1	0.645	
19		Rethinker	1	0.513	
20		Innovative	1	0.513	
21		Technology savvy	1	0.395	
22		Seek new ways to solve problems, creative	2	0.487	0.638
23		Seek new ways to solve problems, rethinker	2	0.395	0.517

(continued)

Table 2 (continued)

Rule no.	Leadership style	Items	Item set size	Support (min support level—30%)	Confidence (min confidence level—70%)
24		Creative, rethinker	2	0.316	0.615
25		Creative, innovative	2	0.303	0.590
26		Innovative, technology savvy	2	0.342	0.667
27		Seek new ways to solve problems, creative, rethinker	3	0.211	0.533
28		Seek new ways to solve problems, innovative, technology savvy	3	0.211	0.727

5 Conclusion

This work observes the potential use of association rules technique on integrating Generation Y academics attributes with transformational leadership component. The results obtained using association rules are promising and prove that it can be used as an alternative to commonly used statistical technique in finding hidden and interesting relationships between variables in the form of patterns. Based on the analysis done, the findings led to a direction for human resource department to pay attention to Malaysian private universities' faculty head leadership style as one of the factors in order to minimize turnover intention. Future work may consider building a prediction model using association rules, and this requires large data set for training and test data. The created model then can be integrated into any existing HR system being practiced in universities. Apart from that, comparison of results between statistical technique and association rules would be worth looking into in the future study.

References

1. Bass, B.M.: Leadership and performance beyond expectations. Free Press, Collier Macmillan (1985)
2. Ahmed, A.D., Bach, Ch.: Major traits/qualities of leadership. Innov. Space Sci. Res. J. **3**(1), 47–53 (2014)
3. Leithwood, K., Leonard, L., Sharratt, L.: Conditions fostering organizational learning in schools. Educ. Adm. Q. **34**(2), 243–276 (1998)
4. Keller, R.T.: Transformational leadership and the performance of research and development project groups. J. Manag. **18**(3), 489–501 (1992)
5. Dulewicz, S.V., Higgs, M.J.: Design of a new instrument to assess leadership dimensions and styles. Sel. Dev. Rev. **20**(2), 7–12 (2004)
6. Sadeghi, A., Pihie, Z.A.L.: The role of transformational leadership style in enhancing lecturers' job satisfaction. Int. J. Bus. Soc. Sci. **4**(8) (2013)
7. Raman, G., Ramendran, C., Beleya, P., Nodeson, S., Arokiasamy, L.: Generation Y in institution of higher learning. Int. J. Econ. Bus. Model. **2**(2), 142 (2011)
8. Meier, J., Crocker, M.: Generation Y in the workforce: managerial challenges. J. Hum. Resour. Adult Learn. (1), 68 (2010)
9. Gwavuya, F.: Leadership influences on turnover intentions of academic staff in institutions in Zimbabwe. Acad. Leadersh. J. **9**(1), 1–15 (2011)
10. Cohen, G., Blake, R.S., Goodman, D.: Does turnover intention matter? Evaluating the usefulness of turnover intention rate as a predictor of actual turnover rate. Rev. Public Pers. Adm. 0734371X15581850 (2015)
11. Harhara, A.S., Singh, S.K., Hussain, M.: Correlates of employee turnover intentions in oil and gas industry in the UAE. Int. J. Organ. Anal. **23**(3), 493–504 (2015)
12. Jawahar, I.M.: A model of organizational justice and workplace aggression. J. Manag. **28**(6), 811–834 (2002)
13. McCrindle, M.: New generations at work: attracting, recruiting, retaining and training Generation Y. The ABC of XYZ (2006)
14. Sujansky, J.: Leading a multi-generational workforce. Occup. Health Saf. (Waco, Tex.), **73**(4), 16–18 (2004)
15. Stevenson, H.: Teacher leadership as intellectual leadership: creating spaces for alternative voices in the English school system. Prof. Dev. Educ. **38**(2), 345–360 (2012)
16. Ameen, M.M.: The retention of generation Y employees in pakistan **13**(2), 21–40 (2018)
17. Xiao, A., Yang, S., Iqbal, Q.: Factors affecting purchase intentions in generation Y: an empirical evidence from fast food industry in Malaysia. Adm. Sci. **9**(1), 4 (2018)
18. Ciarniene, R., Vienazindiene, M.: Flexible work arrangements from generation and gender perspectives: evidence from Lithuania. Eng. Econ. **29**(1), 84–92 (2018)
19. Sathye, M.: Leadership in higher education: a qualitative study. In: Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, vol. 5, no. 3 (2004)
20. Li, K., et al.: Impact factors analysis on the probability characterized effects of time of use demand response tariffs using association rule mining method. Energy Convers. Manag. **197**, 111891 (May 2019)
21. Zeng, L., Wang, B., Fan, L., Wu, J.: Analyzing sustainability of Chinese mining cities using an association rule mining approach. Resource. Policy **49**, 394–404 (2016)
22. Sheu, J.J., Chu, K.T.: Mining association rules between positive word-of-mouth on social network sites and consumer acceptance: a study for derivative product of animations, comics, and games. Telemat. Informatics **34**(4), 22–33 (2017)
23. Land, S., Fischer, S.: RapidMiner in academic use (2012)
24. Rai, S.: Strategic HRM perspectives: study of Indian gen Y management professionals. J. Strat. Hum. Resour. Manag. **3**(3) (2014)
25. Williams, K.C., Page, R.A.: Marketing to the generations. J. Behav. Stud. Bus. **3**(1), 37–53 (2011)

26. Baruch, Y.: Transforming careers: from linear to multidirectional career paths: organizational and individual perspectives. Career Dev. Int. **9**(1), 58–73 (2004)
27. Zemke, R., Raines, C., Filipczak, B.: Generations at work: managing the clash of veterans, boomers, Xers, and Nexters. Amacom, Your Workplace. New York (2000)
28. Kerslake, P.: Words from the Ys. Management 44–46 (May 2005)
29. Broadbridge, A.M., Maxwell, G.A., Ogden, S.M.: Experiences, perceptions and expectations of retail employment for Generation Y. Career Dev. Int. **12**(6), 523–544 (2007). http://dx.doi.org/10.1108/13620430710822001
30. Martin, C.: From high maintenance to high productivity: what managers need to know about generation Y. Ind. Commer. Train. **37**(1), 39–44 (2005). http://dx.doi.org/10.1108/00197850510699965
31. Yeaton, K.: Recruiting and managing the 'why?' Generation: gen Y. CPA J. **78**(4), 68 (2008)
32. Lowe, D., Levitt, K.J., Wilson, T.: Solutions for retaining generation Y employees in the workplace. Bus. Renaiss. Q. **3**(3), 43–57 (2008)
33. Harris-Boundy, J., Flatt, S.J.: Cooperative performance of millennials in teams. Rev. Bus. Res. **10**, 30–46 (2010)
34. Holt, S., Marques, J., Way, D.: Bracing for the millennial workforce: looking for ways to inspire generation Y. J. Leadsh. Account. Ethics **9**(6), 81 (2012)
35. Myers, K.M., Sadaghiani, K.: Millenials in the workplace: a communication perspective on millenials' organizational relationships and performance. J. Bus. Psychol. **25**(2), 225–238 (2010)
36. Solnet, D., Kralj, A., Kandampully, J.: Generation Y employees: an examination of work attitude differences. J. Appl. Manag. Entrep. **17**(3), 36 (2012)
37. Ng, E., Schweitzer, L., Lyons, S.: New generation, great expectations: a field study of the millennial generation. J. Bus. Psychol. **25**, 281–292 (2010)
38. Hershatter, A., Epstein, M.: Millennials and the world of work: an organization and management perspective. J. Bus. Psychol. **25**, 211–223 (2010)
39. Tulgan, B.: Common misconceptions about generation X. Cornell Hotel. Restaur. Adm. Q. **37**(6), 46–54 (1996)
40. Gorman, P., Nelson, T., Glassman, A.: The millennial generation: a strategic opportunity. Organ. Anal. **12**(3), 255–270 (2004)
41. Welsh, M.J., Brazina, P.R.: Gen Y anatomy lesson: they're not alien, just different. Pa. CPA J. **81**(3), 1–5 (2010)
42. Immordino-Yang, M.H., Christodoulou, J.A., Singh, V.: Rest is not idleness: implications of the brain's default mode for human development and education. Perspect. Psychol. Sci. **7**(4), 352–364 (2012)
43. Bolton, R.N., Parasuraman, A., Hoefnagels, A., Migchels, N., Kabadayi, S., Gruber, T., Solnet, D.: Understanding generation Y and their use of social media: a review and research agenda. J. Serv. Manag. **24**(3), 245–267 (2013)
44. Goman, C.K.: Communicating for a new age. Strat. Commun. Manag. **10**(5), 8–9 (2006)
45. Horeczy, A., Lalani, A., Mendes, G., Miller, M., Samsa, L., &Scongack, T.: Leadership preferences of generation Y (2012)
46. Oshagbemi, T.: Age influences on the leadership styles and behaviour of managers. Empl. Relat. **26**(1), 14–29 (2004)
47. Kultalahti, S., Edinger, P., Brandt, T.: Expectations for leadership-generation Y and innovativeness in the limelight. In: Proceedings of the European Conference on Management, Leadership and Governance, pp. 152–160 (2014)
48. Annie, L., Mc, C., Kumar, A.D.: Market basket analysis for a supermarket based on frequent itemset mining. Int. J. Comput. Sci. Issues (IJCSI) **9**(5), 257–264 (2012)
49. Lai, K., Cerpa, N.: Support vs confidence in association rule algorithms. In: Proceedings of the OPTIMA Conference. Curicó, pp. 1–14 (2001)

An Improved Self-tuning Control Mechanism for BLDC Motor Using Grey Wolf Optimization Algorithm

Murali Muniraj, R. Arulmozhiyal and D. Kesavan

Abstract Brushless DC motor employed wide role actuator plays a significant role in many real-time applications. This paper investigates modelling and simulation of BLDC motor with an optimization algorithm for self-tuning parameters in an unknown alleyway. Grey wolf algorithm (GWA), an intelligent control algorithm is developed with the behaviour of a wolf while hunting the pathways. Further the algorithm also reduces noise cancellation that accurately reduces the impact on load during unknown alleyways. GWA is employed to acquire the gain values and self-tune parameters for the inverter to drive BLDC motor, and constant term is introduced to reduce overreach of motor speed and position of shaft. A comparative analysis is conducted among the feedback controller for BLDC motor optimization such as neuro-ANN and fuzzy-PID through simulation. Results suggest the proposed GWA algorithm holds better performance and reduce error in comparison with the other two optimization methods.

Keywords BLDC motor · Wolf · Optimization · Optimization · Alleyway

1 Introduction

Brushless DC motor recommended high and low-power applications provide advantages of high stability, high torque and speed ratio, silent and inverter quadrant operation and low thermal losses [1]. A BLDC motor has a silent operation, compact form, high torque–speed characteristics reliability and low maintenance [2]. Stator with three-phase winding arranged in trapezoidal nature excited with permanent magnets on the rotor. BLDC motor adds advantage of brushless in commutator arrangement and an electronic-based commutation of hall-based position sensors used as a feedback signals [3]. Limitations in BLDC motor variable torque/speed operation failed to use in commercial applications. Over decades continuous development in technology of power electronics and parallel processing of processors. Industrial process drivers and control schemes applied with recent power electronic

M. Muniraj (✉) · R. Arulmozhiyal · D. Kesavan
Department of Electrical and Electronics Engineering, Sona College of Technology, Salem, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_30

technology will enable better working drives with efficient and existing BLDC motor [4]. Adaptative neural network shapes more attention towards self-learning algorithm and parallel working. Many distinct research papers considered as controller with ANN are solution for nonlinear pathways [5–8]. In ANN methods design of training algorithm is complex and not suitable to nonlinear loads and pathways [9–11]. GWA designed for all variables geometries and over come drawbacks of Fuzzy-PID and ANN [5, 12, 13]. Further GWA provides low ripples in inverter current also suitable for low speed and high torque applications [14]. A real-time software MATLAB is used to simulate GWA to obtain satisfactory with in-built position commands to controller and validates GWA.

2 Design of BLDC Motor

Design of BLDC motor comprising voltage source inverter (VSI), encoder and hall sensor. BLDC motor is verified with predefined parameters listed in Table 1 and simulated in MATLAB.

2.1 Modelling of BLDC Motor

BLDC motor consists of stator side winding and permanent magnet rotor winding side, and rotor winding-induced currents are neglected. A three-phase winding connected as input and converter with the BLDC motor excited to induce a rotating electric field are shown in Fig. 1. Damper winding is excluded, and BLDC motor of stator and rotor is modelled as per circuit equation phase variables.

$$\text{Van} = R\left(i_r - i_y\right) + L\frac{\mathrm{d}}{\mathrm{d}t}\left(i_r - i_y\right) + \text{ea} - \text{eb} \tag{1}$$

$$\text{Vbn} = R\left(i_y - i_b\right) + L\frac{\mathrm{d}}{\mathrm{d}t}\left(i_y - i_b\right) + \text{eb} - \text{ec} \tag{2}$$

Table 1 BLDC motor

BLDC motor	Values	Units
Nominal voltage—Vn	32	Volts
Terminal resistance—R	1.16	Ohms
Output power—P_{2max}	101	Watts
EMF gain constant—KE	2107	mV/rpm
Stand still torque—MH	301	NM

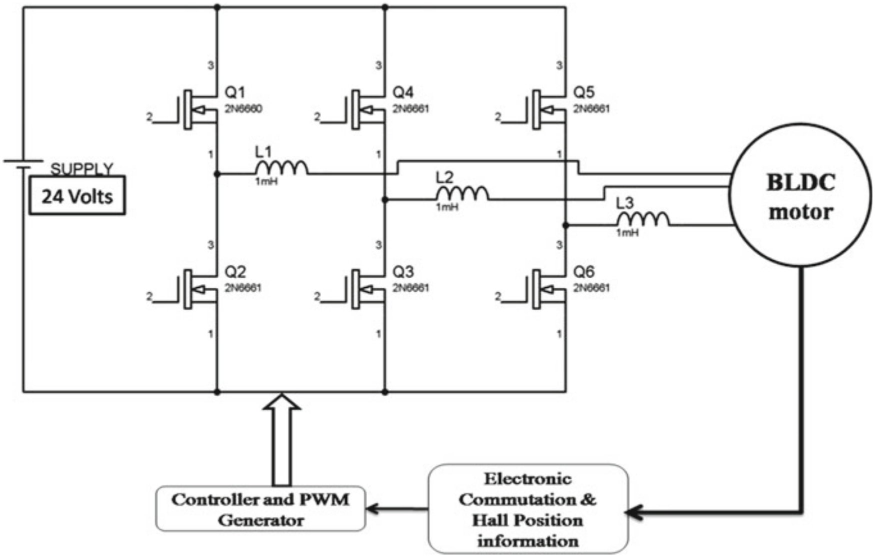


Fig. 1 Block diagram feedback control of BLDC motor

$$\mathrm{Vcn} = R(i_b - i_r) + L\frac{\mathrm{d}}{\mathrm{d}t}(i_b - i_r) + \mathrm{ec} - \mathrm{ea} \tag{3}$$

where Van, Vbn and Vcn are input phase voltages; R—winding resistance of per phase; i_r, i_y and i_b are stator phase currents; L is the inductance;

Motor parameters are assumed to change in the rotor as per reluctance and angle as per salient operation rotor.

Output torque of BLDC motor is given as.

$$T_e = \frac{k_t}{2}\left[F(\theta_e)i_r + F\left(\theta_e - \frac{2\pi}{3}\right)i_y + F\left(\theta_e - \frac{4\pi}{3}\right)i_b\right] \tag{4}$$

where K_t is the torque constant, and F(θ_e) is function back EMF (Fig. 1).

2.2 Modelling of Encoder

Inverter drive control unit of three-phase winding BLDC motor with feedback encoder pulse signals (Ha, Hb, Hc) and decoder to control algorithm with the subsystem model are shown in Fig. 2. Subsystem comprises the logic to drive motor in forward and reverse direction with PWM changing fixed frequency as per control logic. Shaft quadrature encoder with 2200 pulse per revolution (ppr) simulated in

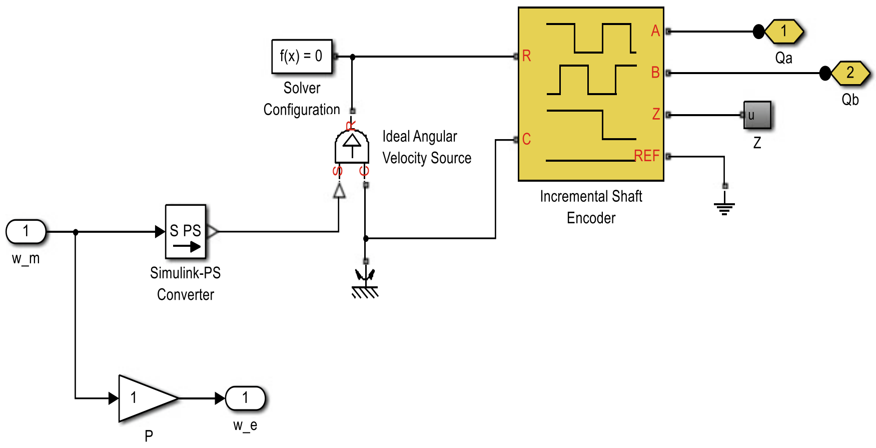


Fig. 2 Encoder model for BLDC motor subsystem

MATLAB Simscape block sets, solver block connected with physical transmission block sets from Simpower tools.

$$N_a = \left(\frac{60}{2000}\right) \times f_{Qa} \tag{5}$$

N Rotation Per Minute
Qa Puse Frequency in Hz.

Pulse signals from encoder fed from feedback unit are measured in frequency and converted to equal speed value. Output pulse multiplexed responds to speed and direction as per Eq. 5. Encoder pulses of rotation as Qa and Qb with phase angle difference of three-phase winding correspond to the feedback given to the controller to change rotation in BLDC motor either in forward or reverse direction.

3 Design of Grey Wolf Algorithm

Generally, grey wolves live together and hunt in numbers. Initially, wolves follow the attack in pattern which forms a encircle attack desired target. Similarly, an algorithm is proposed by Sun et al. [5]. In general, a model of GWA designed with the fittest solution is termed with symbol alpha (α), best second fittest termed beta (β) and last solution to fittest problem as delta (δ). Auxiliary solution of the fittest problem is termed as omega (ω) which guides all wolves for optimizing target. The function of GEA optimization to reduce of BLDC motor tuned with function Eq. 6 and design for attaining less shoot values through flowchart Fig. 3.

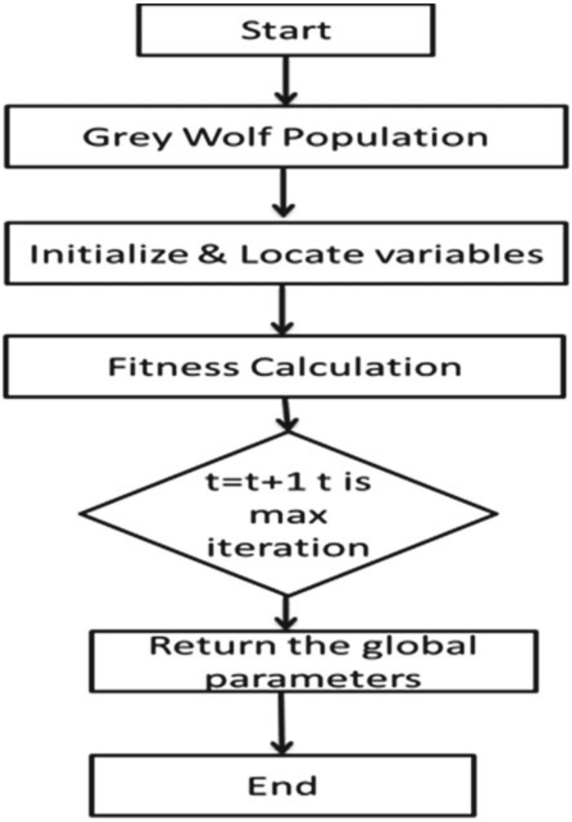


Fig. 3 Flowchart of GEA optimization

$$F = \frac{\sum_{n=0}^{N} [\Delta e\omega(n)]}{nTs} + \left(\frac{\Delta eid(n)}{nTs}\right) \quad (6)$$

where ω is speed reference and id reference value of current reduction.

Design procedure of controller:

Step 1: To generate weighting matrices M and N from test solution.
Step 2: MATLAB model of BLDC motor with feedback mechanism obtained to tune gain parameters of inverter Kp and Kd.
Step 3: Gain matrix for algorithm modelled in Simulink and run the MATLAB model of motor drive with algorithm gain drive values with objective function evaluation.
Step 4: Derive the values of fitness function.

4 Results and Discussion

To validate the proposed GEA optimization for BLDC motor, simulation is carried out in MATLAB model shown in Fig. 4. The performance parameters of BLDC motor are obtained from Table 1. Wolf population need to select from reference values which depend on accuracy of optimization. More test data will influence time delay in population estimation. Performance comparison of fuzzy-PID, ANN and GWA and tuning process of GWA and fuzzy-PID is shown in Fig. 5 through the MATLAB model.

Simulation carried load torque value is 10 nm for three different controller are set with set pulse counts of 225 pulse counts as step response. From Fig. 6, a fuzzy-PID controller holds less overshoot values and could not achieve desired response 0.12 time. As in ANN network, control response is poor due much overshoot and produces oscillatory response (Table 2).

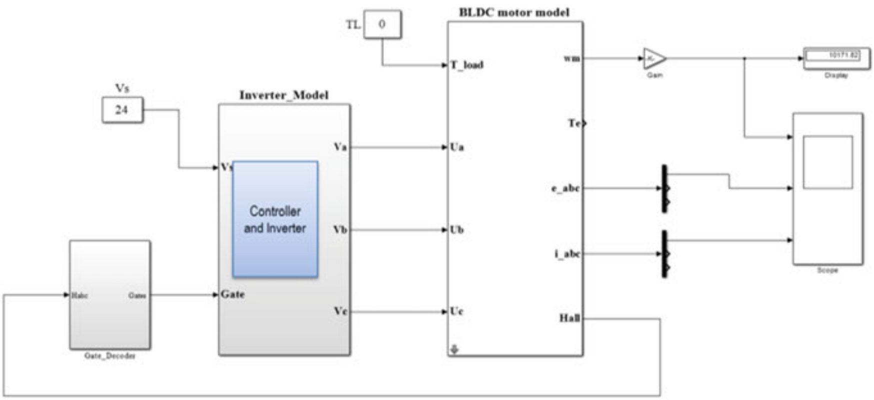


Fig. 4 MATLAB simulation model of BLDC motor

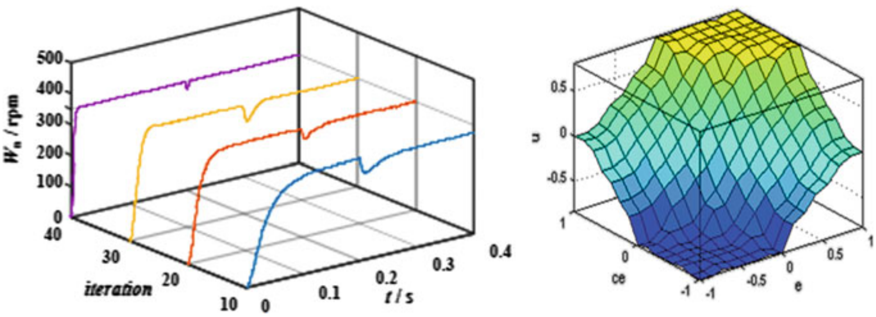


Fig. 5 Auto-tuning procedure of GWA and fuzzy-PID algorithm

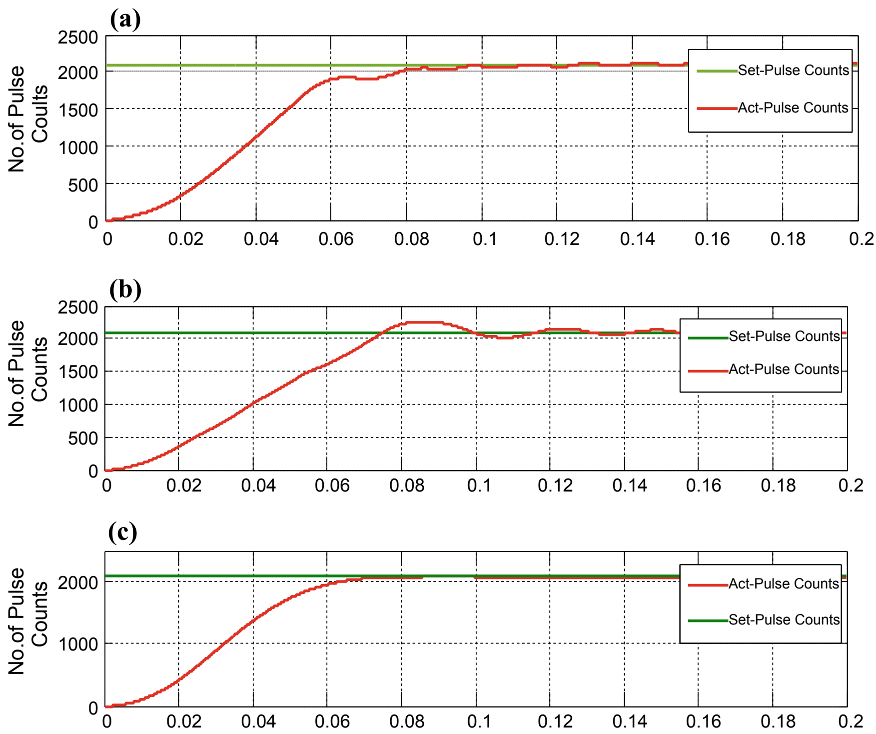


Fig. 6 Simulation response of BLDC motor for **a** fuzzy-PID **b** ANN **c** GWA

Table 2 Response of control algorithms at no load and constant load

Controller	Load condition	Overshoot	Final time(S)
Fuzzy-PID	No load	−7.6	0.09
	Load 10 Nm	16.5	0.12
ANN	No load	11.6	0.01
	Load 10 Nm	19.8	0.14
GWA	No load	0.01	0.056
	Load 10 Nm	0.02	0.085

Hence, the proposed GWA control algorithm with tuning gain parameters of inverter achieves steady state response, in corresponding, even in loaded condition. Further, the proposed model analysed for different values as tabulated that holds better performance for all the loads and all responses proved superiority of using GWA control mechanism of BLDC motor.

5 Conclusion and Future Scope

A high-performance GWA is proposed in this paper as a controller for BLDC motor. GWA algorithm ensures maximum dynamic performance from motor load during constant and variable load conditions. Feed back controllers designed with encoder pulse acquire real-time data of BLDC motor and obtain pulse through GWA control algorithm. To suppress the overcome shoot, an index is produced which in function optimizes the current and speed gain values. In order to validate the GWA is better than other controllers, a Simulink model is developed in MATLAb and results suggest GWA control offers better response for BLDC motor. This algorithm further implemented to harness renewable energy source and obtain better efficiency.

References

1. Djerioui, A., Houari, A., Ait-Ahmed, M., Benkhoris, M., Chouder, A., Machmoum, M.: Grey wolf based control for speed ripple reduction at low speed operation of PMSM drives. ISA Trans. **74**, 111–119 (2018)
2. Bianchini, C., Immovilli, F., Cocconcelli, M., Rubini, R., Bellini, A.: Fault detection of linear bearings in brushless AC linear motors by vibration analysis. IEEE Trans. Ind. Electron. **58**(5), 1684–1694 (2011)
3. Gao, Z.M., Zhao, J.: An improved grey wolf optimization algorithm with variable weights. Comput. Intell. Neurosci. **2019**(Article ID 2981282), 13 (2019)
4. Jordehi, A.R.: Optimal scheduling of home appliances in home energy management systems using grey wolf optimisation (Gwo) algorithm. In: 2019 IEEE Milan Power Tech. Milan, Italy, pp. 1–6 (2019)
5. Sun, X., Hu, C., Lei, G., Guo, Y., Zhu, J.: State feedback control for a PM hub motor based on grey wolf optimization algorithm. IEEE Trans. Power Electron. https://doi.org/10.1109/tpel.2019.2923726
6. Jordehi, A.R.: Optimal scheduling of home appliances in home energy management systems using grey wolf optimisation (Gwo) algorithm. In: 2019 IEEE Milan Power Tech. Milan, Italy, pp. 1–6 (2019). https://doi.org/10.1109/ptc.2019.8810406
7. Al Saaideh, M.I., Mazideh, B.B., Abu-Al-Nadi, D.I.: Grey wolf optimizer for optimal design of digital HR filter. In: 2019 10th International Conference on Information and Communication Systems (ICICS). Irbid, Jordan, pp. 256–261 (2019)
8. Murali, M., Arulmozhiyal, R.: Modeling, simulation of control actuation system with fuzzy-PID logic controlled brushless motor drives for missiles glider applications. Sci. World J. Hindawi Publishing Corporation **2015**, ISSN: 1537–744X, 1–15 (2015)
9. Murali, M., Arulmozhiyal, R.: Intelligent optimum control of brushless DC motor. Int. J. Control. Theory Appl. Int. Sci. Press. ISSN: 09745572, **10**(8), 49–58 (2017)
10. Sun, C., He, W., Ge, W., Chang, C.: Adaptive neural network control of biped robots. IEEE Trans. Syst. Man, Cybern. Syst. **47**(2), 315–326 (2017)
11. Ma, H., et al.: Neural-network-based distributed adaptive robust control for a class of nonlinear multi-agent systems with time delays and external noises. IEEE Trans. Syst. Man, Cybern. Syst. **46**(6), 750–758 (2016)
12. Ramya, A., Balaji, M., Kamaraj, V.: Adaptive MF tuned fuzzy logic speed controller for BLDC motor drive using ANN and PSO technique. J. Eng. **2019**(17), 3947–3950 (2019)

13. Verma, V., Chauhan, S.: Adaptive PID-fuzzy logic controller for brushless DC motor. In: 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA). Coimbatore, India, pp. 445–449 (2019)
14. Peng, X., Jia, M., He, L., Yu, X., Lv, Y.: Fuzzy sliding mode control based on longitudinal force estimation for electro-mechanical braking systems using BLDC motor. CES Trans. Electr. Mach. Syst. **2**(1), 142–151 (2018)

A Wearable Wrist-Based Pulse Oximetry for Monitoring Cardiac Activities—A Pilot Study

Ramya Shekar, N. Sriraam, Prabhu Ravikala Vittal and Uma Arun

Abstract Pulse oximetry is a method of determining the oxygen saturation in the arterial blood by placing the sensor on the finger, toe, wrist or ear lobe of the subject and acts as an indicator of overall health. This research study attempts to propose a wearable wristwatch-based pulse oximeter for monitoring cardiac activities. Typical physiological parameters such as heart rate and oxygen saturation (SpO_2) were estimated through laboratory-based first-level experimental settings. The pulse oximeter sensor with the real-time processor, ATmega328p was embedded into a wristwatch, and through Bluetooth mechanism, one can see the result of activities on a mobile phone or on a laptop. The use of MAXREFDES117# sensor provides an added advantage compared to the available sensors in the market with an integrated, level translator and power converter. The proposed study confirms the suitability for real-time monitoring of cardiac activities and resulted in a low-cost, user-friendly and low complex device design.

Keywords Pulse oximetry · Oxygen saturation (SpO_2) · Heart rate (HR) · Bluetooth HC-05 · Android-based smartphone · Serial terminal app

1 Introduction

Medical devices have become very common in today's developing world. Wireless technology is the most emerging field which helps in monitoring the health conditions of human beings during regular daily activities. The health monitoring appliances found in hospitals, like the electrocardiogram (ECG), require the electrodes to be attached to the human body and human intervention to check the working of the machines [1]. The inventions of a wearable sensor for health monitoring are more advantageous and user-friendly. Pulse oximeter is one of such non-invasive device which can be used to continuously monitor the oxygen saturation (SpO_2) and heart rate (HR). The proposed device is a wrist-based reflectance mode pulse oximeter,

R. Shekar (✉) · N. Sriraam · P. R. Vittal · U. Arun
Center for Medical Electronics and Computing, M.S Ramaiah Institute of Technology, M.S.R Nagar, MSRIT Post, Bangalore 560054, India

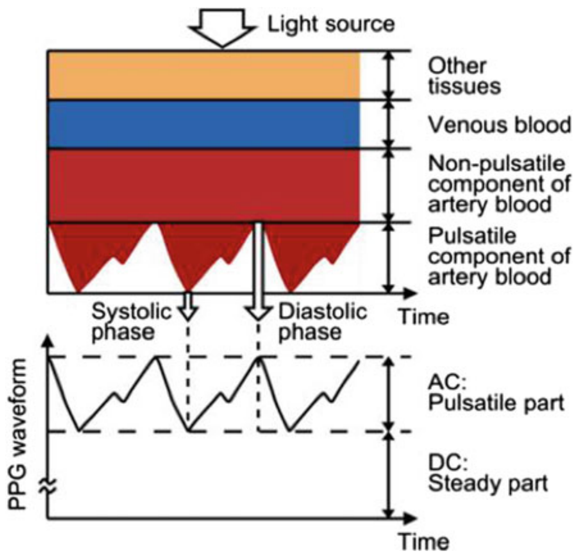
V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_31

since the wrist is considered to be more a ideal site for long-term measurement without much motion artifacts. The developed device is a low-cost, low-complexity and non-invasive with real-time analysis.

1.1 Background of PPG

Photoplethysmography as the name suggests is an optical technique used to detect the blood volume changes in blood in the peripheral circulation [2]. The pulse oximeter uses red and IR LEDs of low intensity which is illuminated onto the arterial points in the human body, and the light travels through the tissues. This light is absorbed by the bones, skin pigments and venous and arterial blood. Since the blood absorbs most of the light compared to the surrounding tissues, a change in the intensity is observed which is detected by the photodetector. There are two types of pulse oximeter attachment based on the placement of the source and detector, which is reflectance and transmittance mode. In reflectance mode, both the sources consisting of the two LEDs and the detector which consists of photodiodes are placed on the same side of the finger. In transmittance mode, the source and detector are placed on opposite sides of the finger. The detected signal consists of varying AC and constant DC component. The AC component corresponds to the blood volume changes which are in synchronization with the heart beat, and the DC component corresponds to the reflection or transmission of the optical signals by tissue and blood volume changes. Figure 1 shows the pulse oximeter waveform and its characteristic parameters [3].

Fig. 1 Pulse oximeter waveform



1.2 Related Work

Pulse oximeter technology is being used in monitoring the SpO_2 levels, and many researches have been conducted in this domain.

Nur et al. conducted a prototype design on wrist-based wireless PPG data acquisition using Bluetooth module for heart-rate measurement and wireless transmission to the nearby connected smartphone. However, the processing was done in MATLAB, and GUI was displayed using LabVIEW [4]. Ghamari et al. proposed a wristband-type wireless PPG device and SD card to save data, and processing was done using MATLAB and Kubios HRV software [5]. Reyes et al. conducted research on wireless radio frequency-based system. The data was collected using MATLAB, and LabView was used to display the waveform and data post-processing for HRV analysis [6]. Mukherjee et al. proposed e-health monitoring architecture where multiple sensors were connected to the processing device, and the post-processing results were sent to the cloud [7]. However, the incorporation of many sensors resulted in the high-power consumption. Simi et al. proposed a design of wrist-based biowatch for BP measurement; however, the accuracy of the device could not be achieved due to calibration [8]. Zhe et al. acquired the PPG signals under different types of daily life motion artifact and applied various noise cancelation algorithms [9]. Lin et al. proposed a contactless heart-rate detection along with motion artifact monitoring by incorporating additional LEDs on the camera, but since it is based on imaging resolution which impacts the data analysis [10]. Saquib et al. proposed a device which uses two controllers one at patient side and the other at the receiver side to measure the heart rate and viewed via RS232 communication; however, the use of two controllers increases the device complexity and power consumption [11]. However, most of post-processing in the above literature was performed offline. It can be inferred from the related background that there is a huge need to develop a low-cost, user-friendly module which can provide the real-time data analysis of HR and SpO_2 and transmission of data to the nearby doctors during emergency conditions in countries where tertiary healthcare facilities are yet to be established.

2 Methodology

2.1 Data Collection

Two hundred healthy subjects 18–50 years age were included in the study with the informed consent. Ethical clearance was taken, and the recordings were performed in the room temperature (25 °C). For the initial pilot study, only the resting condition of the subjects is considered. A 10 min recording of the data is performed and displayed on the smartphone or PC through Bluetooth which can be logged and saved for further analysis.

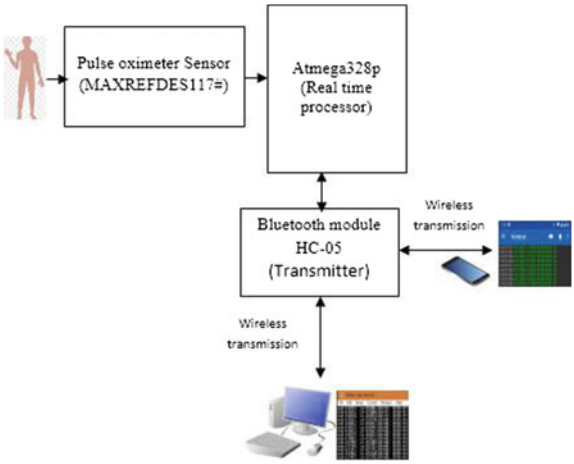


Fig. 2 Block diagram of the proposed design

The proposed design with the sensor interfaced with the microcontroller along with the Bluetooth is shown in Fig. 2.

The following note highlights the features of each module being used in this proposed study.

(1) ATmega328P: It is an 8-bit open-source microcontroller board which belongs to AVR microcontroller family. The device can be programmed with arduino integrated development environment (IDE) via type B USB cable. The operating voltage range is 7–20 V.
(2) Pulse oximeter sensor(MAXREFDES117#): MAXREFDES117# is a I2C (Inter-integrated circuit) interface-based heart rate and SpO_2 monitoring subsystem purchased from maxim integrated. The chip consists of MAX30102 sensor integrated with red and IR LEDs. It also consists of a step-down converter MAX1921 which is used to convert the supply input to 1.8 V which is suitable for the heart-rate sensor. It also consists of a level translator MAX14595 which provides communication between the sensor and the microcontroller operating at different voltage range. The integration of the step-down converter and level translator provides an advantage along with ultra-low power consumption compared to the other sensors available at the market. The operating range of the sensor is 2–5.5 V. The chip is tiny with board size of 12.7 mm × 12.7 mm which makes it easy to be embedded onto the Velcro and used on finger or wrist-based applications.
(3) HC-05 module: HC-05 is a Bluetooth module, which is mainly used for the wireless data transmission. It is based on the universal asynchronous transmitter and receiver (UART) interface. It consumes low power which is a great advantage and operates on voltage ranging from 1.8 to 3.6 V. In this project, HC-05 is used to transfer the data wirelessly to the smartphone via Serial Terminal app or computer via Tera Term App.

(4) Serial Bluetooth Terminal App and Tera Term App: They are line-oriented and Emulator Terminal app used by microcontrollers, arduinos and other devices with a serial and UART interface connected with a Bluetooth to serially convert the data to the android device and PC. It has logging and sharing capabilities which can be used to share the real-time data to known doctors during emergency conditions over Whatsapp or emails or any other applications.

(5) Pulse sensor and HC-05 calibration: The baud rate of the HC-05 module and the pulse sensor have to be set same for wireless transmission to be achieved. This is done using the AT commands by writing a simple code in the IDE and the baud rate is set to 115,200 bps (bits per second) which is same for the pulse sensor.

2.2 Hardware Development

All the electronic components are controlled by the ATmega328p. The pulse sensor (MAXREFDES117#) is used as input from which the raw IR and red values are extracted along with displaying the heart-rate and SpO_2 values. The entire setup is powered by a rechargeable 9 V battery.

The experimental setup as shown in Fig. 3 cannot work without suitable coding. The complete coding is done on the IDE platform which uses embedded C. The maxim provides few inbuilt functions that are used to obtain the raw IR and red values with suitable delay. The mean of the raw values of IR and red LED is taken which corresponds to DC component and subtracted from the signal to get signal consisting of only AC component which is used for further processing. A four-point moving average filter is applied to the signal, and a threshold is applied to allow the signal varying from 0.5 to 5 Hz range corresponding to 30–300 BPM (beats per minute). Using the maxim peak detection function, the systolic peaks are detected. The signals are normalized, and minimum distance and minimum height are set to

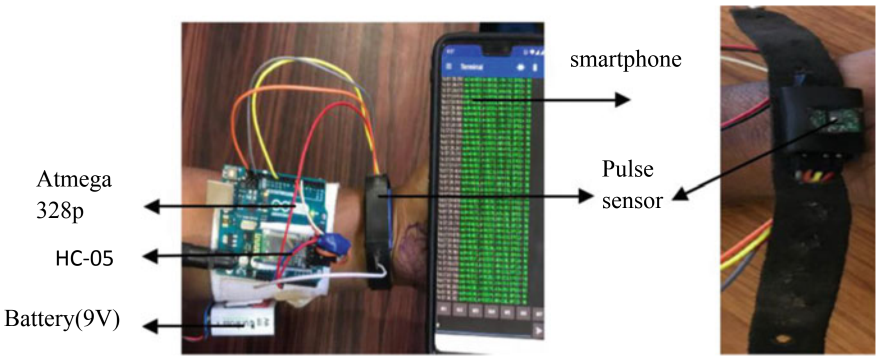


Fig. 3 Experimental setup

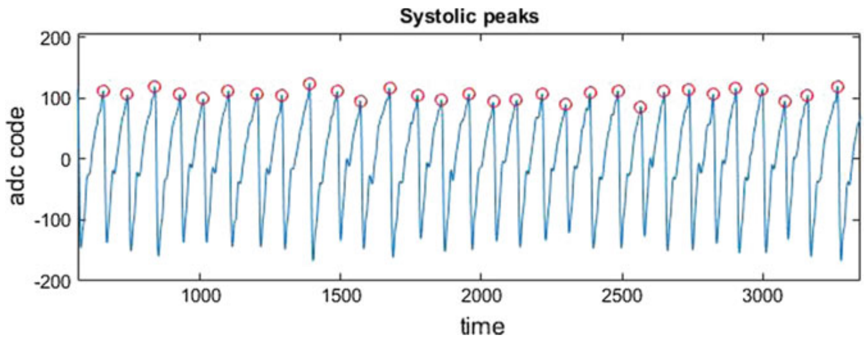


Fig. 4 Systolic peak detection

detect a selected point as peak. This way the error that occurs due to motion artifacts is reduced in the signals.

An infinite while loop runs displaying the raw ir, red, heart-rate and SpO_2 values on the smartphone or computer. The complete algorithm is shown in Fig. 6.

Once the peaks are detected, the heart rate and SpO_2 are calculated using Eqs. 1–3 [12].

$$\text{HR} = \frac{\text{total number of peaks}}{\text{Duration of thesignal}} * 60 \tag{1}$$

Equation 1 represents the heart rate calculated as the ratio of total number of peaks as shown in Fig. 4 to total duration of the signal represented by *x*-axis (axes).

Equation 2 represents the SpO_2 calculation in percentage where 'K' represents the proportionality constant which is based on calibration results, and 'R' represents the absorption ratio. In this experiment, $K = (108\text{-}25)$ is used.

$$\text{SpO}_2\% = K * R \tag{2}$$

$$R = \frac{\frac{\text{AC}_{\text{red}}}{\text{DC}_{\text{red}}}}{\frac{\text{AC}_{\text{IR}}}{\text{DC}_{\text{IR}}}} \tag{3}$$

Equation 3 represents the absorption ratio which is the ratios of the AC and DC components of red LED to AC and DC components of the IR LED.

3 Results

Table 1 shows the mean heart rate and SpO_2 values of ten selected subjects out of 200 recordings to represent the reliability of the system. Since the heart rate and SpO_2 are displayed for every second by running a while loop, the mean of the values of the

Table 1 Mean heart rate and SpO_2 values obtained from the device

Subject id	Age	Height (cm)	Weight (Kg)	Mean HR	Mean SpO_2 (%)	HR and SpO_2 valid
1	23	160	80	68	95	1
2	22	153	47	78	99	1
3	22	150	52	83	96	1
4	21	165	54	75	99	1
5	21	174	72	93	98	1
6	22	162	80	88	98	1
7	22	160	42	93	100	1
8	21	155	53	78	99	1
9	25	165	63	60	95	1
10	26	160	55	78	99	1

1-Normal, 0-Abnormal

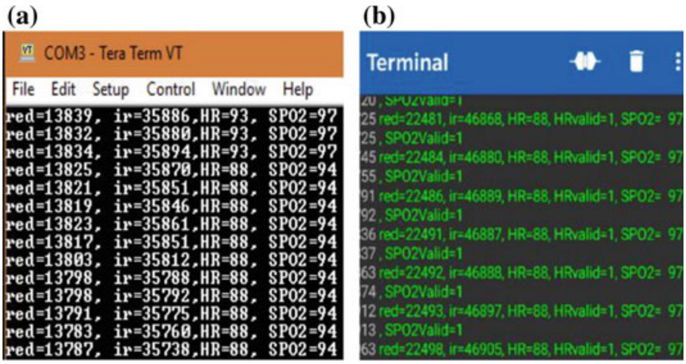


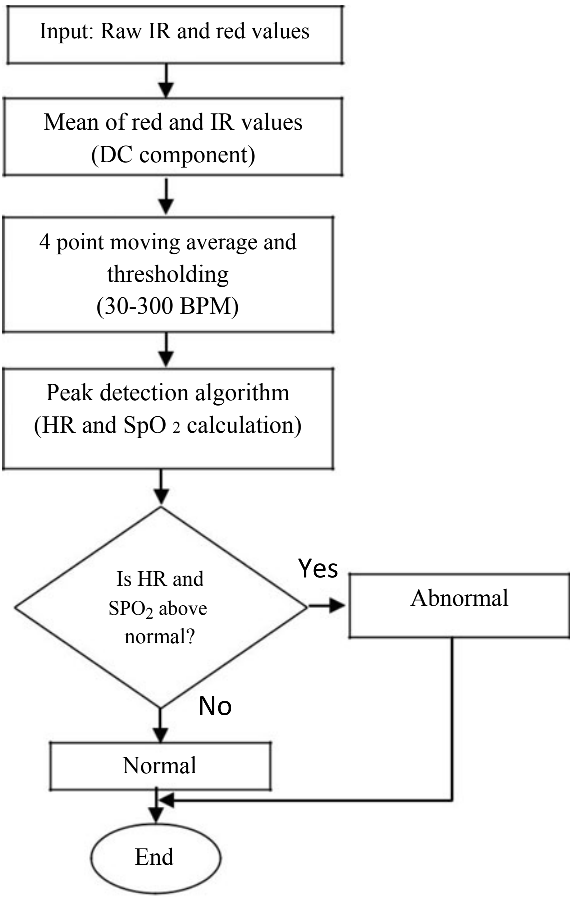
Fig. 5 Output displayed wirelessly on **a** PC using Tera Term application, **b** smartphone using Bluetooth Serial Terminal application

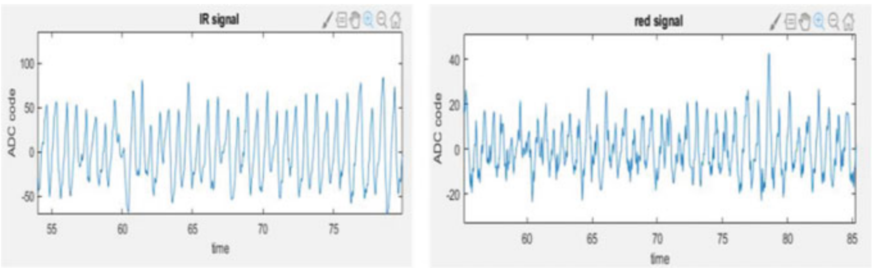
respective parameters is tabulated. The result shows that the heart rate (60–100 BPM) and SpO_2 (95–100%) are in normal range for healthy volunteers. If the values exceed the normal range, it is indicated on the display by a ‘0’ and ‘1’ for normal range and red values obtained by the wireless transmission on the smartphone (Figs. 5, 6 and 7).

4 Conclusion

The first-level prototype on the wrist-based pulse oximetry for cardiac activities monitoring is implemented and validated. In this study, the complete experimental

Fig. 6 Flowchart representing pre and pro-processing steps





a) IR signal b) Red signal

Fig. 7 Graph plotted using MATLAB 2018b by logging the raw IR and red data setup which includes the sensor, microcontroller and the wireless transmission of the collected data are discussed. The real-time data acquisition of the signals is established and transferred to the nearby smartphone or computer via Bluetooth HC-05 module. The necessary preprocessing steps are discussed to get a clean PPG signals and plotted using MATLAB 2018b. The device is validated by collecting the recordings from 200 healthy volunteers. The reliability of the device can be further extended by incorporating the accelerometer and studying the effects of motion artifacts on subjects during physical exercises and collecting more data based on gender and age variations.

Acknowledgements This research work was supported by the Department of Biotechnology (DBT), Govt of India (Project Ref NO: BT/PR14751/MED/32/422/2015).

References

1. Nitzan, M., Taitelbaum, H.: Of oxygen saturation in arterial and venous blood, pp. 9–15 (2008)
2. Kim, I., Lai, P.H., Lobo, R., Gluckman, B.J.: Challenges in wearable personal health monitoring systems. In: International conference IEEE engineering, vol. 2014, pp. 5264–5267 (2014)
3. Tamura, T., Maeda, Y., Sekine, M., Yoshida, M.: Wearable photoplethysmographic sensors—past and present. Electronics (2014)
4. Zainal, N.I., Rodzi, M.Z.M., Khan, S., Habaebi, M.H., Gunawan, T.S.: Design and development of wireless PPG data acquisition for health monitoring application using bluetooth module. IEEE (2016)
5. Paper, C., Paso, E.: Design and prototyping of a wristband-type wireless photoplethymographic device for heart rate variability signal analysis design (Aug 2016)
6. Reyes, I., Nazeran,H., Franco,M., Haltiwanger, E.: Wireless photoplethys-mographic device for heart rate variability signal azquisition and analysis, pp. 2092–2095 (2012)
7. Mukherjee, S., Dolui, K., Datta,S.K.: Patient Health management system using e-health monitoring architecture, pp. 400–405 (2014)
8. Thomas, S.S., Nathan, V., Zong, C., et.al.: Biowatch-a wrist based signal acquisition system for physiological signals including blood pressure (2015)
9. Lin, Z., Zhang, J., Chen, Y., Zhang Q.: Heart rate estimation using wrist-acquired photoplethysmography under different types of daily life motion artifact (2015)
10. Lin, Y.C., Chou, N.K., Lin, G.Y., Li, M.H., Lin, Y.H.: A real-time contactless pulse rate and motion status monitoring system based on complexion tracking. Sensors (2017)
11. Saquib, N., Papon, M.T.I., Ahmad, I., Rahman, A.: Measurement of heart rate using photoplethysmography. IEEE(2015)
12. Bagha, S., Shaw, L.: A real time analysis of PPG signal for measurement of SpO_2 and pulse rate. Int. J. Comput. Appl. **36**(11), 0975–8887 (2011)

Data Sciences and Teaching Methods—Learning

Jesus Silva, Rafael Portillo, Ana Emilia Hernandez, Noel Varela, Hugo Martinez Caraballo, Hugo Hernández Palma, Osman Redondo Bilbao and Nadia Leon Castro

Abstract Data Science (DS) is an interdisciplinary field responsible for extracting knowledge from the data. This discipline is particularly complex in the face of Big Data: large volumes of data make it difficult to store, process and analyze with standard computer science technologies. The new revolution in Data Science is already changing the way we do business, healthcare, politics, education and innovation. This article describes three different teaching and learning models for Data Science, inspired by the experiential learning paradigm.

Keywords Data Science · Deep learning · Experiential learning · Gamification

J. Silva (✉)
Universidad Peruana de Ciencias Aplicadas, Lima, Peru

R. Portillo · A. E. Hernandez · N. Varela
Universidad de la Costa (CUC), Calle 58 # 5566, Atlántico, Barranquilla, Colombia
e-mail: rportill3@cuc.edu.co

A. E. Hernandez
e-mail: ahernand48@cuc.edu.co

N. Varela
e-mail: nvarela2@cuc.edu.co

H. M. Caraballo
Universidad Simón Bolivar, Barranquilla, Colombia
e-mail: hugo.martinez@unisimonbolivar.edu.co

H. H. Palma · O. R. Bilbao · N. L. Castro
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

O. R. Bilbao
e-mail: oredondo@ul.edu.co

N. L. Castro
e-mail: nleon@ul.edu.co

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_32

1 Introduction

The wide variety of applications and the growing demand for subject matter experts has made DS courses, books and manuals proliferate. The standard pedagogical method that we can see in these courses consists of: (1) the explanation of the different branches of machine learning (supervised, unsupervised, by reinforcement); (2) the detail of some learning paradigm such as decision trees or networks of artificial neurons; and (3) illustration using toy datasets such as Weather or Iris [1, 2]. This method is used in some highly popular DS courses in the community such as Machine Learning from Stanford University or the Johns Hopkins University DS specialization [3, 4].

Few fields allow the student to take the lead in profiles as diverse and interesting as DS: economists, business managers, doctors, biologists, Web site administrators, etc. Similarly, few disciplines can offer rewards as attractive to experiential learning as the three million dollars the contest winner earned to predict patients who were admitted to a US hospital in the following year; or the million dollars with which the Netflix company awarded the best predictor of movie ratings [5, 6].

In this work, three models of teaching and learning of DS are presented that, inspired by experiential learning [7], seek that the student learns through reflection on experimentation, rather than being passive containers of content undisputed. In addition, these models have been used in a master's subject at the University of Mumbai in India, which is focused on one of DS's most leading and complex fields, deep learning or deep learning (DL) [8].

2 Description of Methods

This section describes the three proposed models for teaching and learning in DS including the activities and resources used.

2.1 Artificial Neural Networks (ANN)

In this teaching and learning model, neural architectures are first introduced, which are the cornerstone of the DL field. Many of the great advances in twenty-first century artificial intelligence are based on the use of this machine learning paradigm. NARs are neuroscience-inspired computational models capable of predicting an output from tagged data (supervised learning), as well as finding underlying and hidden structures in untagged data (unsupervised learning) [9]. Second, practical troubleshooting tips are given with ANN, and two DS working environments are presented: Weka and H20.ai [10]. Finally, an experience is proposed to students: training and evaluating neural architectures for a specific problem, providing a set of data that allows to infer an intelligent prediction model, such as those available in the UCI repository.

The teaching and learning model described and follows the classic flow of a course in DS, albeit with a marked practical and experiential component, where it is asked to address a concrete and realistic problem of great social interest.

2.2 Artificial Vision (CV)

This second teaching and learning model addresses the problem of artificial vision, i.e., the ability of computers to gain high-level understanding from images or videos [11]. One of the fundamental aspects of this model is that students are sought to verify that the lessons learned in ANN do not allow acceptable results for artificial vision. Consequently, after an introduction to the field, students are proposed to predict the object containing an image. To do this, the CIFAR-108 dataset, which contains 60k color images in 32×32 size, is provided with objects belonging to ten different categories (classes in supervised learning terminologies) [12]. In addition, this task is proposed in the competition modality to include a gamification scheme: the ranking of the best predictions is shown in the virtual platform of the subject. Thanks to this, students can reflect on the problems that lead NNAs to have low percentages of correct predictions. No student exceeded 60% success.

After this contest, the networks of convolutional neurons (ConvNets) and; in a hands-on approach, a working environment for the design, training and use of these networks is explained called Caffe [13]. Artificial vision requires specific models that are capable of, among others, taking advantage of spatial information from images and videos: a pixel close to another will probably belong to the same object. ConvNets have achieved the best CV results since they were used in the ImageNet competition in 2012 to reduce the prediction error from 26 to 15% [14].

After introducing ConvNets, the same experience is re-ensued in a new contest. This time some students reached 82% correct predictions. Finally, there is a third experience in which training data is significantly reduced, but it is allowed to use ConvNets already trained to fit the new data, i.e., learning by transfer in DS terminology.

The key to this second teaching and learning model for resumes is that students see their previously acquired knowledge challenged in the face of a new problem. You have time to try to apply known methods to new situations; reflect on the results; conceptualize new knowledge; and experimentally test the new ideas in the next contest. This covers the four phases of the Kolb cycle for experiential learning [15]. In addition, the competitions allow experiential learning guided by games as proposed [16] for the field of information technologies. However, in this model, the methods and solutions used by students are guided and not free to propose innovative solutions or research methods other than those treated during the course.

2.3 Natural Language Processing (NLP)

This model addresses another problem of great complexity in DS: the processing and understanding of natural language by computers. The NLP, together with the CV, discussed in the previous module, are two of the most difficult problems of artificial intelligence [16]. We call these problems AI-complete, or AI-hard, as they seek to make computers as smart as people. The roots of the NLP lie in the work "Computing Machinery and Intelligence" where Alan Turing speculated on the requirements that a machine should meet in order to claim that it is really smart [17]. Since 2014, DL methods have led the results in NLP for a variety of tasks such as answering questions, automatically translating into different languages, summarizing large documents or generating texts automatically [18].

For this teaching and learning model, the order of the previous models is reversed: first a realistic experience is given to students and then some alternative solutions are offered. More specifically, it addresses the problem of text classification: it is looking for a predictive model that is capable of cataloging a text in a specific class. Text classification is a generalization of problems such as SPAM filters, which exist in email accounts, the analysis of social media sentiments by companies and governments to detect trends in public opinion, or the prediction of the subject of a document. This problem is very challenging with previous knowledge of the subject since, unlike ANN or CV, obtaining a vector of input characteristics of equal length for different texts is not trivial [19].

In this experience, students form groups of up to five people, as research groups, in a laboratory equipped with one job per person. Its task is to make a proposal for a competitive call for a research project in a large company. The research problem is to predict the relevance of an article's title to the body of the article, based on a dataset with three attributes: title, body, and class (relevant/irrelevant). In addition, to facilitate the exploration of various innovative solutions, instead of requesting a code implementation as in the previous models, workflow diagrams are requested for the proposals as well as the working environments to be used and bibliographic references.

Students, then, present their proposals which, while they may include flaws in terms of the interfaces of the technologies used, were definitely very creative and made good use of many of the concepts already seen in DL and other subjects of the master. In this line, some students considered the use of ConvNets, studied in CV, for text analysis. Other proposals included state-of-the-art NLP technologies such as the use of recurrent neural networks and dense forms of word representation such as Word2Vec and GloVe.

Students should then review an excerpt of selected readings (about five pages) with an introduction to the NLP field, the problem of text classification, and DL development platforms. Finally, a master class is taught where various solutions to the problem raised are discussed. These solutions are also contextualized with the different proposals made by students.

As in CV, this teaching and learning model is part of a challenging and unresolved problem with prior knowledge. The text requires a very elaborate preprocess before it can feed a neural network with it. In addition, the request for a design rather than an implementation is more complicated for the students of the master's degree, given their technical profile. It is also key in this model for give students freedom and confidence to explore different alternatives and innovative solutions. Finally, the content of the master class should be flexible and address the main concepts used in student proposals, rather than simply presenting a series of non-debatable solutions.

3 Results

The objective of the study is to analyze the evolution of student satisfaction for each of the three teaching and learning models implemented. As described in the previous section, the models correspond to the three parts in which the deep learning subject is divided: artificial neuron networks (ANRs), artificial vision (CV), and natural language processing (NLP). At the end of the course, a questionnaire was provided to the 321 students present out of a total of 330 enrolled. For each of the models, three statements were made: $A1$ "The content of the course meets my training needs," $A2$ "What I have learned will apply to my work," $A3$ "Applied methodology, technical resources, and teaching materials were appropriate." Participants were asked to rate their agreement or disagreement on a five-point Likert scale. These questions are similar to those asked by [20]. In this work, however, a scale of five points has been chosen instead of six, as it is more widespread and the possibility of giving an equidistant response between the maximum and the minimum was considered.

Questions were also raised based on a pure experiential learning model, where there is no material supplied or any guidance when it came to addressing a realistic experience. "Consider a fourth model in which only a realistic experience is proposed at the beginning of the subject, deep learning, for example. In this case, students, all together as a group, would be free to implement and investigate the best methods to complete this task." This model is hypothetical, in the sense that it was not followed during the subject, but we consider that the opinion of the students is valuable for new editions of the master's degree. For this model of learning, three other statements were made: $A4$ "I would like at least one subject to follow this scheme," $A5$ "I would like at least one subject to follow this scheme, but with smaller groups," and $A6$ "I would like the whole master's degree to follow this scheme." Again, they were asked to rate their agreement or disagreement on a five-point Likert.

Table 1 shows the descriptive statistics for student satisfaction, with the first column being the teaching and learning model employed (ANN, CV, NLP, or pure experiential learning), and the above-mentioned statements, numbered from $A1$ to $A6$.

With regard to the first three models, it can be seen in all cases that the average of the responses given to the three statements is above half the scale (3). It also highlights a very low standard deviation, less than one point, except in one case.

Table 1 Evaluation of teaching and learning models

	Mean	Std. dev.
Networks of artificial neurons		
*A*1	3.40	1.07
*A*2	3.62	0.79
*A*3	3.12	0.95
Artificial vision		
*A*1	4.65	0.60
*A*2	4.48	0.54
*A*3	4.01	0.89
Natural language processing		
*A*1	4.25	0.70
*A*2	4.09	0.98
*A*3	4.54	0.99
Pure experiment		
*A*4	3.95	1.21
*A*5	3.99	1.25
*A*6	3.01	1.35

This indicates a small expected variation from the arithmetic mean and without the need for an ANOVA test, that the means within the different models evaluated are similar. In terms of comparison between the different models, CV is the best valued in the $A1$ and $A2$ claims, while NLP is the best valued in $A3$, which includes the methodological aspects. In this sense, students seem to have appreciated the advantages of experiential learning and the freedom to research and apply unguided solutions to realistic situations. The teaching and learning model used at ANN, which is closest to traditional DS teaching, is the worst valued, albeit exceeding half the scale. There is almost a point of difference between ANN and the maximum reached in the question by the other models (CV or NLP).

It is also highlighted, from the results obtained in Table 1, that there is no strong support for the pure experiential learning model, at least in the terms set out in the $A4$–$A6$ claims. Results are greater than 3 for $A4$ and $A5$ and neutral for $A6$. The professional nature of the master's degree could justify the student's suspicion of subjects where no support material is provided even if, as in the case of NLP, this material is provided afterward, after a realistic experience without guides and adjusting the material to some of the solutions previously proposed by students.

In order to delve deeper into the comparison between the different models, students have been asked a vote where for each of the models used in the subject the following statements are proposed regarding the methodology used: $A7$ "My favorite methodology is used in," $A8$ "The methodology that allows me to learn more is," $A9$ "The methodology that (I believe) that allows me more lasting learning is," $A10$ "The methodology that I consider closest to the daily work of a data scientist is."

Table 2 Voting between different models

	ANN	CV	NLP
$A7$	41	218	62
$A8$	03	245	73
$A9$	15	247	59
$A10$	47	231	43

Answers for each of these statements can be ANN, CV, or NLP. It was decided to omit the “does not know/do not answer” column to force an unambiguous answer, considering that there should always be a favorite model. Similarly, the possibility of voting several options simultaneously was omitted. However, some students decided to hold multiple votes and these responses were counted as extra votes for the chosen models.

Table 2 shows the results of the vote. As you can see, the machine vision (CV) model is superior in all four statements, reaching an absolute majority in all categories, i.e., more than half of the total votes. Consequently, students seem to have appreciated the innovative methodological aspects of this model, such as: experimentation with prior knowledge in a new and challenging task; and gamification through competitions to face students in a competition with objective assessment criteria. It should also be noted that the NLP model gets a number of votes very close to CV in the $A10$ statement. This NLP model is close to pure experiential learning, in the sense that students are given the freedom to seek their own solutions in the resolution of the task in question. Therefore, students seem to have identified this out-of-life model, along with CV, as something very close to a data scientist’s day-to-day.

4 Conclusions

This article has described three teaching and learning models that enable students to gain new knowledge by iterating on realistic experiences, reflection on them, conceptualization of new knowledge and experimentation. The models are designed in such a way that the responsibility of the students in the construction of their own knowledge is increased, giving themselves total freedom when proposing solutions in the latest proposal.

Although all three models have received assessments above half the scale employed (3) by students, they show a clear favorite: the methodology used in artificial vision. In this model, three realistic experiences are proposed under the modality of competition that is interspersed with master classes. Both anonymous surveys conducted and material used in the experiences proposed in this work is available to the interested reader by soliciting the authors.

References

1. Alonso, F., López, G., Font, J.M., Manrique, D.: Learner satisfaction when applying an instructional model in e-learning: an experimental study. In: Proceedings of the 2nd International Conference on Computer Supported Education, vol. 1. CSEDU, pp. 141–146 (2010)
2. Jacobson, M., Ruddy, M.: Open to outcome: a practical guide for facilitating and teaching experiential reflection. Wood 'N' Barnes (2004)
3. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT Press (2016)
4. Hahsler, M., Karpienko, R.: Visualizing association rules in hierarchical groups. J. Bus. Econ. **87**, 317–335 (2017)
5. Silverstein, C., Brin, S., Motwani, R., Ullman, J.: Scalable techniques for mining causal structures. Data Min. Knowl. Discov. **4**(2–3), 163–192 (2000)
6. Viloria, A., Lis-Gutiérrez, J.P., Gaitán-Angulo, M., Godoy, A.R.M., Moreno, G.C., Kamatkar, S.J.: Methodology for the design of a student pattern recognition tool to facilitate the teaching-learning process through knowledge data discovery (big data). In: International Conference on Data Mining and Big Data, pp. 670–679. Springer, Cham (2018)
7. Viloria, A., Lezama, O.B.P.: Improvements for determining the number of clusters in k-means for innovation databases in SMEs. Procedia Comput. Sci. **151**, 1201–1206 (2019)
8. Viloria, A., et al.: Integration of data mining techniques to PostgreSQL database manager system. Procedia Comput. Sci. **155**, 575–580 (2019)
9. Lanzarini, L., Villa-Monte, A., Ronchetti, F.: SOM + PSO. A novel method to obtain classification rules. J. Comput. Sci. Technol. (JCS&T) **15**(1), 15–22 (2015)
10. Lanzarini, L., Villa Monte, A., Aquino, G., De Giusti, A.: Obtaining classification rules using lvqPSO advances in swarm and computational intelligence. In: Lecture Notes in Computer Science, vol. 6433, pp. 183–193. Springer-Verlag, Heidelberg, Berlin (2015)
11. Zurada, J., Levitan, A.S., Guan, J.: A comparison of regression and artificial intelligence methods in a mass appraisal context. J. R. Estate Res. **33**(3), 349–387 (2011)
12. Mayer-Schonberger, V., Cukier, K.: Big data: a revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt (2013)
13. Viloria, A., Acuna, N., Mejia, H., Galofre, M.: Determination of the Influence of thermal comfort in care and concentration of media education students: case Colombia. Indian J. Sci. Technol. **9**(46), 1–7 (2016)
14. Witten, I., Frank, E., Hall, M.: Data mining: practical machine learning tools and techniques. Elsevier Science (2011)
15. Shiralkar, S.: IT through experiential learning: learn. Apress, Deploy and Adopt IT through Gamification (2016)
16. Venugopal, K.R., Srinivasa, K.G., Patnaik, L.M.: Soft computing for data mining applications, p. 354. Springer, Springer-Verlag, Berlin, Heidelberg. ISBN 978-3-642-00192-5 (2009)
17. Lock, D., Nettleton, D.: Using random forests to estimate win probability before each play of an NFL game. J. Quant. Anal. Sport. **10**(2), 197–205 (2014). https://doi.org/10.1515/jqas-2013-0100
18. Choudhury, A., Jones, J.: Crop yield prediction using time series models. J. Econ. Econ. Educ. Res. **15**, 53–68 (2014)
19. Wu, J., Chen, L., Zheng, Z., Lyu, M.R., Wu, Z.: Clustering web services to facilitate service discovery. Knowl. Inf. Syst. **38**(1), 207–229 (2014)
20. Zhu, J., Fang, X., et al.: IBM cloud computing powering a smarter planet. Libro Cloud Computing, vol. 599.51/2009, pp. 621–625. Páginas (2009)

Data Security in Cloud Computing Using Three-Factor Authentication

Sunanda Nalajala, B. Moukthika, M. Kaivalya, K. Samyuktha and N. L. Pratap

Abstract Cloud storage is tremendously increasing its services, huge range of storage and communication of massive data over network. This practically has an adverse effect on the way of upholding this data, when it especially comes to the privacy of the user-secured and highly confidential data. We first view you with a system that is vulnerable to this authentication protocol with its misuse of biometrics and incorrect password generates no user to lost the mobile device. We even went along with this scheme and gave out a major issue to overcome with impersonation attack. However, this scheme makes the way easy to attack for offline password guessing attack. We included a three-factor authentication which carries the smart card into card reader, gets the password and identity and conveys the user details' requesting time. We then came up with a scheme to overcome these security flaws of this prescribed authentication scheme combining passwords, mobile devices and biometrics. The proposed system is robust three-factor authentication with the help of password, biometrics and mobile device which provides reliable security strength to the user's data and makes counterattack to existing attack, giving with more benefits compared to the previous scheme. This scheme will not only encounters with the security issues, but also provides with most enhanced security functionalities.

Keywords Three-factor authentication · Counter attacks · Privacy · Biometrics · Smart cards

S. Nalajala · B. Moukthika · M. Kaivalya · K. Samyuktha
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
e-mail: sunanda@kluniversity.in

N. L. Pratap (✉)
Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
e-mail: lakshmanpratap@kluniversity.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_33

1 Introduction

The revolutionary of data and information over the communicating channel has widely changed the file storage accessories to organized and clustered cloud environment, which makes a provision for flexibility, data security, recovery and collaboration. Cloud storage has provided a massive amount data including health, insurance, banking, business and sensor data. The most important in the sense of many business holding is the cost advantage.

Moving the data to the cloud takes away the cost of hardware and maintenance with the increased productivity and cloud allowances. This kind of storage gives access not only PC operator along with mobile clients. It can access computational results, applications, services that are developed in the cloud. Sometimes this confidential data can be exposed by some unknowns without our permissions like without valid authentications. Since this cloud storage mainly relies on the network which is public service with the wide extension of services and policies.

This type of storage is suspect able to a consequence of attacks such as denial of service attack, insider threats, insecure API which means an API is the initial entry point for the attackers and spectra and meltdown in which the attackers exploit the meltdown to view data on the user's servers and spectra is harder to exploit and fix it. Moreover, attacks on identity such as masquerade of identity and tracing and account hijacking have become the most routine in the network. It takes a way out to advantage which includes scalability, resilience, flexibility, efficiency and outsourcing. The issues we have encountered in this scheme suffer from biometric misuse, wrong password and fingerprint login and no user revocations mechanism. It faces a weal assault on the service providers' impersonation.

The main aim of this paper is that we want to improve the security of the data in the cloud by implementing the three-factor authentication protocol using EEC to preserve the biometric template (user biometrics) privacy by having the concept of the fuzzy extractor. The reason that we have chosen the three-factor authentication, which is the user is using three different authenticated factors to authenticate himself in the cloud environment ensures the user's privacy and also help in increasing the security level in the cloud and there is a possible overcome of the misuse of the biometrics. For that, we have adopted bio-cryptosystem such as fuzzy logic in field of computational intelligence instead of direct ways to inculcate hash functions.

The remaining paper is structured as follows. Section 2—Related Work Sect. 3—Proposed Work, Sect. 4—Result, Sect. 5—Conclusion and Sect. 6—Future Scope.

2 Related Work

Our core concept outlines the safety variables that need to be closely monitored, and most transactions need to be monitored by key consumers and service providers [1]. In the authentication phase, when the user wants the access to the service provider, the user shares the passwords and the fingerprints with the smart card which later computes them with an encryption algorithm, and then the user and the service provider start authenticating each other [2]. In the recent years, cloud computing and the mobile cloud computing are ruling over many industries mainly when we consider of storing data or to sharing data [1].

In the authentication phase, when the user wants the access to the service provider, the user shares the passwords and the fingerprints with the smart card which later computes them with an encryption algorithm, and then the user and the service provider starts authenticating each other [2]. Apart from these, we can also make use of cloud computing during the retrieving of data and to host the applications over the Internet [1].

Privacy and security level, in any aspects of the production, utilization and launching or updating of the applications, the service providers must always take care of the privacy and security to avoid various attacks [1]. In order to maintain the protection, we need to ensure general cloud safety in which we will host the request and also the crucial part of security as the end-user level [1]. We should also keep the client's credentials, details and data private and secured from any of the attacks [1].

In the authentication phase, when the user wants the access to the service provider, the user shares the passwords and the fingerprints with the smart card which later computes them with an encryption algorithm, and then the user and the service provider starts authenticating each other [2]. Next, the client sends a service request to the service provider there, and with the aid of a random number, the service provider determines the identity and responds to the end of the user [2]. With this type, series of sharing the authentication phase according to algorithm was proposed by this scheme [2].

Here, we will explain how to establish a model using user's identity, and this relates to the encryption and signature systems, which in turn helped to present an identity authentication scheme [3]. Through simulation test, the lightweight process is being applied to the user side even though it is on the massive scale cloud [3]. Identity-based signature is also required to build a complete identity-based authentication for the cloud computing to be done [3]. In the verification phase, the signature equation will be equated with the encrypted equation generated earlier and will be validated [3].

In current solutions to mobile cloud computing, the main concerns relate to operational level, end-user level, [2] service-level issues as well as data management, context awareness and security [5]. In many other fields, the above problems also relate not only to mobile cloud computing. These similarities can also be applied for other related fields, not only in cloud computing area. These issues can highlight a few unique sets of challenges [1].

At the service-level module production, we need to make sure that we have the correct and relating applications which are able to fill the criteria for which they are made [3]. Along with these applications, performance is also to be considered because the user must sense the quality of the service that is provided for them, and the maintenance of the application plays a crucial role at the level of execution of the applications and also cloud API; the applications that are created by the service providers must be available over the Internet for the usage [4].

Privacy and security level, in any aspects of the production, utilization and launching or updating of the applications, the service providers must always take care of the privacy and security to avoid various attacks. In order to maintain the protection, [7] we need to ensure general cloud safety in which we will host the request and also the crucial part of security as the end-user level We should also keep the client's credentials, details and data private and secured from any of the attacks [6].

Data management is major security concern for the users as one of the main reasons for the usage of the networks is to store, update, share and utilize data from any of the places just over the Internet [4]. While storing data, the issues majorly come from the personal data storage on mobile cloud under which entire media of the mobile even the location history and chat export can also be done [5]. There are many applications available in the market for the storage and the access of the data over the Internet.

One among these issues is data access issues like unavailability of data, error pages and updating problem or data can be stolen via Internet by the man-in-the-middle attack or any of such attacks. In general, eavesdropping technique can also be done [7]. To avoid such attacks, we should make sure our network is secured and we must make sure that source entity, i.e., cloud, or any other source stations must be secured [7].

Everyday, programmers are coming up with new social networking sites. Such platforms are making enormous profit by providing users with a forum. Since the social network users are created a few thousands per day, it is difficult to store such huge amounts of data. Cloud provides a platform for this tiny amount of data to be processed as a solution. To some extent, the exposed information is personal data that has resulted in the possibility of disclosure of individuals ‘ confidential information. Hence, the safety of cloud-based social networking data is priority of highest concern. The code encrypts the information before it is placed in the cloud, and only the private key decrypts the data that makes data to be more secured in cloud data storage [8].

The cloud computing is continuously evolving and showing a exponential growth in the field of computing, and it is also applied in various fields such as big data analytics, Internet of things, e-health clouds and various new technologies, but still the security issues and the threads that are associated with it are unwieldy. This paper analyses and implements various methods that are involved in conserving the privacy such as writing the access rights, permissions and policies, anonymizing the data, fragmenting and reconstructing the data. All these approaches would protect the privacy of the authorized information. It concentrates on the privacy of the cloud-based IaaS, SaaS and PaaS [9].

3 Proposed Work

The objective of this paper is to make sure the data security enhanced in cloud computing using three-factor authentication using elliptical cryptography to preserve user privacy.

3.1 *Elliptical Curve Cryptosystem*

The elliptic curve when is discussed is represented on a from F_p. The elliptic curve equation is to be represented with the alphabet $E_{p,}$ and the equation is $\boldsymbol{E_p(a, b)\!:\!y^2 = x^3 + ax + b(\mathrm{mod}p)}$ **on infinite field** $\boldsymbol{fp}$, which is said to be a form and makes sure that the elliptic curve equation satisfies the condition of the line equation. Perfect elliptic system curve is symmetric on both sides of the quadrants, and the line equation can only satisfy the elliptic curves. The ECC's security depends on the following computationally implausible problems.

3.2 *Fuzzy Extractor*

This also allows standard cryptographic inputs and the cryptographic techniques for the security purpose. The extraction of the input must be uniform and has the tolerance for the noise. Make sure the input changes, if made, are still related to the string that is reconstructed. If any other biometric model, $B_{1,}$ relatively similar to B, with the aid of an alternative string P, the derived random string R remains constant. A fuzzy extractor contains of a *Generate* probabilistic process of generation and a *Rep* deterministic system of reproduction.

3.3 *Three-Factor Authentication*

Three-factor authentications are generally a process we follow in many of the account registration process. In our procedure, we are mainly oriented about the three factors' usage such as user credentials, fingerprints and passwords where we held the login phase. User must have the control of his possession like nounce and templates on server side. There is a possibility to implement one-time password with the smart phone.

3.3.1 Initialization Phase

Server (S) chooses a point (P) from a finite-circular additional group (G) (F_p) based on an input area (*fp*) on an elliptical curve. Server chooses X as its private key and computes public key X_p. Finally, the server saves X and restores E(*fp*), g and p.

User (U_i) takes the following sequences with S undergoes following step to verify him as valid person: User (U_i) has to choose an identity (ID), password (P), nonce and along with these users have to provide the biometric imprints (b) and obtain.

$$(R, P) = \text{Generate}(b)$$

Then, U_i submits identity (ID) along with the hashed values of password and biometrics, i.e.,

$$h = \text{hash}(P_i \| R) \oplus \text{nonce}$$

After getting request from the user to the server, S verifies whether ID of the user matches with any stored values in the server. If the user identity is same, then S decline the request. Otherwise, S selects a random integer t and calculates.

$$b^* = \text{hash}(\text{ID}\|x\|t) \oplus h$$

Then, S updates its values with the new look of $\{\text{ID}, t\}$ and sends b^* and P.

User (U_i) computes $B_1 = b^* \oplus$ nonce and $B_2 = $ hash(hash(ID||P||nonce)mod sp), where sp is a random range integer, $2^8 \leq \text{sp} \leq 2^{16}$, which determines the capacity of the pool of the <ID, P> pair against offline password guessing attack. U_i then stores B_1, B_2, P and P to mobiles.

3.3.2 Login Phase

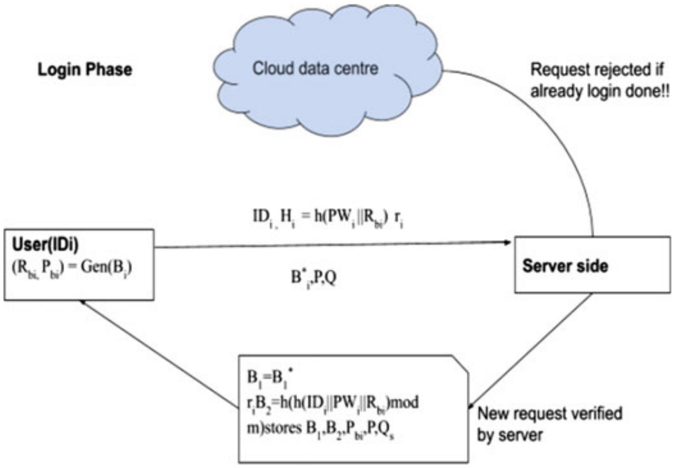
In this phase, U_i inputs ID, P and b^*on the gateway. The mobile device chooses a integer $z \in Z * n$ and u_i, and calculates $R_{bi} = \text{Rep}(b^*, \text{nounce})$.

$$B_2^* = \text{hash}(\text{hash}(\text{ID}\|P\|\text{nounce}) \bmod m).$$

If the equation $B_2^*? = B_2$ does not state, the mobile device declines the request. Otherwise, it continues to validate.

$$c_1 = zP, c_2 = B_1 \oplus \text{hash}(P\|\text{nounce})$$

$c_3 = zQ_s$, $c_4 = \text{hash}(c_1\|c_2\|c_3)$ and $c_5 = Ec_3x(\text{ID}\|c_4)$, where c_3x represents x-coordinate of the ECC point c_3. Then, U_i sends the message $m_1 = \{c_1, c_5\}$ to S.



3.3.3 Authentication Phase

In this area, U_i and S are verified properly when U_i needs request time and then proceed with following step:

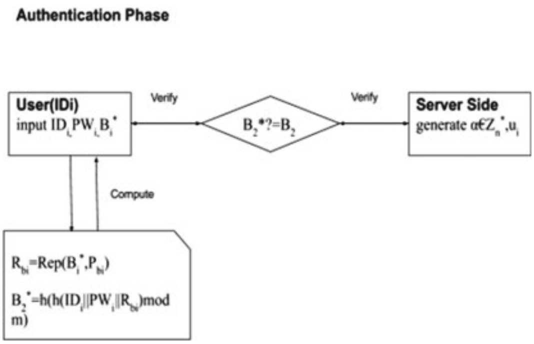
Upon receiving m_1, S calculates $c_3 = xc_1$ and de-encrypts c_5 to get ID and c_4. Then it calculates $c_2 = \text{hash}(\text{ID}\|x\|t)$ and verifies.

$$c_4? = \text{hash}(c_1\|c_2\|c_3).$$

If it is not accepted, S declines the request. If three illegal times message m_1 with the same identity ID appear in a phase, ID will be stopped. S randomly takes a number spk $\in Z * n$ and computes.

$$\begin{aligned} c_6 &= (\text{spk})P, \text{sk} = \text{hash}(c_1\|c_6\|(\text{spk})c_1), \\ c_7 &= \text{hash}(\text{ID}\|\text{DS}\|c_2\|\text{sk}), \text{ and} \\ c_8 &= Ec_3x(c_7). \end{aligned}$$

Then, S sends $m_2 = \{c_6, c_8\}$to U_i.



3.3.4 Biometric Phase

In this phase, U_i updates the password and biometrics through the following steps.

U_i when done with login phase, and gives the new password and if possible biometrics also request to *S*. After that, *S* has done the first step of the authentication phase. If it is verified, *S* computes.

$$\mathrm{Sa} = \mathrm{hash}(c_3 \| c_2 \| c_1),$$

and Sa accepts U_i. Otherwise, *S* declines the request. Upon getting Sa and the acceptance, U_i sees checks *S* is verified $=$ hash(c_3||c_2||c_1). If it accepted, U_i inputs extra password *P*, new *i* and new biometric information.

S. no	Variable	Meaning
1.	x	Server private key
2.	P	Server chooses the finite-field cyclic additional group generated by *A* point *P*
3.	Q_s	Server public key
4.	ID user	User's ID
5.	Password user	User's password
6.	Nounce	One-time password
7.	Biometric user	Biometric of user
8.	hash	Hash function
9.	sku, skd	The session key between user and server
10.	$\mathrm{R_{bi}}$	By giving biometrics, the extracted string
11.	$\mathrm{P_{bi}}$	By giving biometrics, the auxiliary string
12.	m	Medium integer which determines the capacity of the pool of <ID_i, PW_i>
13.	$\vert$	The concatenation operation
14.	$\oplus$	The bit-wise XOR operation

4 Result

We have implemented the proposed algorithm in Python in which initially the user is welcomed to the server and asked to give the three factors that are user biometrics and user identification and password and the nounce (session key), and the user is given the options of register, login, authenticate, biometric exchange and exit.

```
Welcome to our server
Entering user biometrics
Enter the user id11
Enter the user password12
Enter the nounce20
Enter your option for the process
1.Register
2.Login and Authenticate
3.Biometric change
4.Exit
1
Enter the user id: 11
checking your userid:


Enter the user password: 12
Creating a user id in server


Registration successful
```

```
Registration successful


Enter your option for the process
1.Register
2.Login and Authenticate
3.Biometric change
4.Exit
2
Enter the user id: 11
Enter the user password: 12
Login successful


Do you want to authenticate??Yes
You have logged in successful


User is sending m1={C1,C5} to Server
Server is decryptying and obtaining user id and C4


Sending m2={c6,c8} to the User
Decrypting c8 and c7


You are authenticated


Enter your option for the process
1.Register
2.Login and Authenticate
3.Biometric change
4.Exit
```

```
Enter your option for the process
1.Register
2.Login and Authenticate
3.Biometric change
4.Exit
3
Enter the user id: 11
Enter the user password: 12
Login successful
Do you what to change the Biometrics??Yes
Sending the server request to biometric change...


Verfying the request....
Request has been granted
Previous Biometric: 232bc709310220054613b4ec493021e1


Biometric has been changed
New Biometric: 232be709310225054613b4ec493721d1


Enter your option for the process
1.Register
2.Login and Authenticate
3.Biometric change
4.Exit
4
Thank you
```

5 Conclusion

Authentication is most required to specify the security and privacy of highly confidential information. So, as a purpose, we have implemented using three-factor authentication, namely password, phones and biometrics, of user surely to match the necessity of highly formatted data. As of that, we can make the data to be more secured than the traditional way of storing data. In our referred paper, we have suggested a protocol and explained the challenges in designing identity-based protocol. We have demonstrated that the base process used is defective towards impersonating the user attack and revocation and also password offline guessing attack technique.

6 Future Scope

Cloud data security during the process of communication makes the way to be little complicated as seen with other practical solutions. Management of the central data and the user application development improves the same operator in the direct ways within the same central location. As of the point, here in this case we have inculcated the architecture of a single cloud server, whereas to achieve more data security, we can uphold for multi-server architecture as solution to get a clear-cut strategy. Make sure any inside user who uses a computer as the cloud server to access a total network has secured access, regardless of where they make the access that network from. Data classification can be made to provide the level of security based on data prioritization.

Acknowledgements This work was supported by the KL University Centre for Computer Science Students—an industry research collaborated laboratory under the guidance of N. Sunanda, Asst. Professor, KL University.

References

1. Dinh, H.T., et al.: A survey of mobile cloud computing: architecture, applications, and approaches. Wirel. Commun. Mob. Comput. **13**(18), 1587–1611 (2013)
2. Jiang, Q., Ma, J., Wei, F.: On the security of a privacy-aware authentication scheme for distributed mobile cloud computing services. IEEE Syst. J. (2018)
3. Jiang, Q., Khan, M.K., Lu, X., Ma, J., He, D.: A privacy preserving three-factor authentication protocol for e-health clouds. J. Supercomput. (2016)
4. Cybernetics approaches in intelligent systems. Springer Science and Business Media LLC (2018)
5. He, D., Zeadally, S., Kumar, N., Lee, J.H.: Anonymous authentication for wireless body area networks with provable security. IEEE Syst. J. (2016)
6. Jiang, Q., Kumar, N., Ma, J., Shen, J., He, D., Chilamkurti, N.: A privacy-aware two-factor authentication protocol based on elliptic curve cryptography for wireless sensor networks. Int. J. Netw. Manag. (2017)
7. Vinod, A.R.: Hindering data theft attack through fog computing. Int. J. Res. Eng. Technol. (2014)
8. Praveena, A., Smys, S.: Ensuring data security in cloud based social networks. In: 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), vol. 2, pp. 289–295. IEEE, 2017
9. Karthiban, K., Smys, S.: Privacy preserving approaches in cloud computing. In: 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 462–467. IEEE, 2018

An Effective Machine Learning-Based File Malware Detection—A Survey

Ashwin A. Kumar, G. P. Anoosh, M. S. Abhishek and C. Shraddha

Abstract The objective of this paper is to enable computers to learn on their own, identify malicious activities, increase scanner efficiency and sensitivity. The machine learning algorithm enables the identification of patterns in observed data, the development of models that explains the world and the prediction of things without explicitly preprogrammed rules and models. There have been huge research interests in the cybersecurity industry as well as in universities in the subjects of how to effectively block malicious documentation without a sign of slowing down. The main aim of the paper is to investigate the efficiency of large files and increase sensitivity in malware detection.

Keywords Malware · Machine learning · Scanner · Vulnerabilities

1 Introduction

Machine learning (ML) plays a key role in a wide range of serious applications, such as data mining, the processing of natural languages, image recognition, and skilled systems. ML provides likely solutions in all these and more domains and is set to support our future development.

Cybersecurity is a set of technology and approaches designed to save attack, unofficial access, change or destruction of computer systems, networks, applications, and facts. Network security systems and computer (host) security structures encompass cyber protection systems. At a minimum, everyone has a firewall, antivirus software and intrusion detection system (IDS). IDS helps to become aware of, manage and decide the unauthorized use, duplication, change, and demolition of the statistics machine [1]. Violations of security consist of external invasions (assaults from outside the company) and inner invasions (attacks inside the agency). However, the surprisingly complex nature of many real global issues often way that it's miles

A. A. Kumar (✉) · G. P. Anoosh · M. S. Abhishek · C. Shraddha
Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, Karnataka, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_34

unreasonable, if not possible, to find out particular algorithms to be able to resolve them results easily at all times.

It is possible to use a weak document in order to carry out malware payload, which can be embedded or downloaded from a document file, to spread malware across documents. In most cases, JavaScript supports the document to enable entire weaknesses and then execute code of choice for the attacker. This includes ambiguity and storage management techniques such as buffer overflow, ROP, and encrypted heap shellcode [2].

Many methods for chunk-based documented attacks ranging from passive detection, for example, to dynamic analyzes by means of sandboxing technologies have been developed. One of the advantages of signature-based detection is that the sensing of known malware is fairly low. On the other hand, since signature-based detection uses byte commands to match a particular malware, zero-day attacks or malware deviations are not dynamic. This presents major challenges for AV scanners that rely heavily on signature detection. Another solution is to perform analysis and behavior-based malware detection.

This utilizes sandbox software to add an additional identification layer. This monitors enforcement shown in a text folder when it is accessed in a determined context rather than using byte instructions, and if a certain comportement is detected, a warning action will be made. Even with highly complicated content like JavaScript to attack, this increases the detection rate. It should be remembered that the sandbox-based technology operates only if an identified file conducts malicious actions in a real world. They also regularly find ways to avoid sandboxing systems with so-called anti-sandbox approaches. If an atmosphere for sandboxes is observed, for instance, then good behavior and sleep mode can be seen. There are also other boundaries. Some sandboxing tools address only detailed PDF types of attacks, for example, MDScan for JavaScript [3], Nozzle for heap spray [4], or record the lively compliance of a system and require manual analysis to make an unveiling decision, as in the case of CWSandbox [5].

Printers are part of every corporate and personal network these days, so that every network has a good survival chance. In many networks, network printers and multifunction printers (MFP), throughout general, have arisen as cyber-attackers because they are meant to serve several (wired and wireless) interfaces and direct many protocols in order to support a huge base of domestic workers and ad hoc traffic. The viability of cooperation with all manner of printers [6–10], which could be supplied by the relevant section of the study, has been accentuated previous research and guides. Protocol assaults include: Denial-of-service assaults, privilege escalation, the leakage of print jobs, or system filings and even code execution on a printer itself [13]. Printing protocol abuses include select malicious motions an attacker might perform on a target printer.

2 Literature Review

In Zhang [11], the MLPDF model uses a back-propagating algorithm with a stochastic descent search to update a model to effectively PDF-based malware detection using a machine learning-based approach, the neural network model of multi-layer perceptron (MLP) to pinpoint PDF-based malware, called MLPDF. The datasets used are brain and malicious PDF papers. The findings of the study show an impressive MLPDF approach that is well above all eight well-known commercial antivirus scanners evaluated.

Chen et al. [12] have used the technology of IoT, WSN, and cloud technology to give information to the farmers in their phone. They have introduced nodes, which are a set of different sensors, combined to form a single unit to measure various physical and environmental factors. The details are then stored in the cloud and compiled. The datasets are Drebin and MaMaDroid (5879 malware samples) and are analyzed using the R programming and the Cloud MongoDB. The result is the rate of detection of malware decreased in Mama Droid from 96 to 1% and in Drebin from 97 to 1%, with only a little distortion caused by our method of manipulation for example.

Hecht and Sagi [13] included behavioral systems used in infogain, gain ratio and correlation (Pearson) methods to analyze and detect network-printer attacks to achieve the best results in print-protocol traffic detection. Base data is mild and deceptive in terms of experimental observations, whereby the proposed architecture identifies printable protocol attacks efficiently, offering a marginal fall-positive rate of 99.9% accuracy.

Liu et al. [14] describe the detection of adversarial examples based on steganalysis where the author presents adversarial examples that can be detected effectively with the steganalysis detector. Attack methods are based on network gradient calculations like fast gradient sign method (FGSM), fast gradient value (FGV), and Jacobian-based saliency map attack (JSMA). In comparison, some techniques, such as L-BFGS, Deepfool, and Carlini and Wagner (C&W) assault, are focused on solving optimization problems. Dataset is an ImageNet10-class.

Clements and Lao [15] the fast gradient sign method (FGSM) generate in the direction of the sign of the cost functions gradient to produce an adversarial input with very slight perturbation. The Jacobian-based saliency maps attack (JSMA) algorithm uses the gradients of the learned function, rather than the cost function, to produce a saliency map of the input. Datasets are MNIST and CIFAR10. Experimental results show that the proposed algorithms achieve 100% stealthiness for both datasets under all adversarial scenarios.

Sohi et al. [16] have used network intrusion detection systems focus on signature-based intrusion detection methods exhibiting a lower level of false positives, compared to the anomaly based detector. Synthetic datasets generated using overlay methodology, where four different scenarios are taken into account. The number of alarms raised by the Bro running against the same pool of mixed data can indicate how much improvement can be achieved by applying our method.

Alkasassbeh and Almseidin [17] describe machine learning methods for network intrusion detection of the use of the KDD99 dataset in which the detection price became 88 percent of widespread attacks, whether recognized or unknown. The basic benefit of this research is the minimum amount of academic information that wants to produce top traffic category outcomes.

Zhang and Su [18] blanketed Machine Learning Attack and Defense on Voltage Over-scaling Lightweight Authentication results show that ANN, RNN, and CMA-ES can clone the mission reaction behavior of VOS-based fully authentication with predictive accuracy of up to 99.65%, while predictive accuracy is much lower than 51.2% after deploying our proposed ML resilient method.

Cai [19] proposed a preliminary study on Android Malware Detection's sustainability. Datasets include the first collection of 1221 harmless applications (oldBen) in each device class by installing the top 50 popular apps. By uploading the top 100 popular apps in each device class, the latest benevolent dataset (newBen) was collected. DroidSpan achieved F1 accuracy of 91% (versus MamaDroid's 75%). This procedure has shown that not only does DroidSpan effectively spot malware, but it also maintains high accuracy of detection for four years (93% F1 measurement) (81% F1 five years).

3 Discussion

Comparing how PDF and other commercial scanners do with larger information will be useful as part of the future research, particularly including more recent PDF documents to the dataset. To address such mechanisms, it will raise awareness of protective mechanisms against such assaults and attack changes. No previous research has focused on detecting attacks by learning and testing supervised ML classifiers on traditional (non-3-d) printer protocols.

Detector cannot have very good performance when it is not trained and tested on the same adversarial method. So we will try to explore methods for training one detector against different kinds of adversarial attacks. These techniques such as detection using side-channel information suffer from reduced sensitivity toward small Trojans. Attempts to improve the ability of NIDS systems to defend against them by extending their signature databases and generating a more realistic and close to the real-world ground truth to test a NID. The model has a limited amount of time to examine whether the input image is natural or not.

Individuals put up a digital or tough replica image in the passive face recognition to register their identity in a destiny identification gadget. This approach is smaller in magnitudes than non-malicious ones for properly modeling and forecasting the destiny values of time malicious activities. This is a common problem in the detection of anomalies. The MLP classifier has the lowest end result for the Brute Force attack, which implies that MLP can not interpret Brute Force assault details among all the different information. KDD database has 41 attributes and all of them have been registered, but as part of destiny research, additional classifiers and the role selection to see the most relevant characteristics could be investigated.

4 Conclusion

Machine learning has been developed as a new computer system capability. Machine learning will make our future stronger than any other innovation this century. Rapid progress in information storage and the strength of computer processing have dramatically changed the game over the last few years. The facts are very large, the time taken to calculate is improved, and this is where machine learning takes place to help people with large information in a minimum of time. In this paper, we examined the technique of malware detection based entirely on the behavior of documents that distributed primitive access. Our findings show that files are regularly distinguished for use beyond a few years. This location explains the negative aspects of the survey papers and also the high degree of accuracy and resilience to various obstruction systems.

References

1. Mukkamala, S., Sung, A., Abraham, A.: Cyber security challenges: designing efficient intrusion detection systems and antivirus tools. Enhancing Comput. Secur. Smart Technol. 125–163 (2005)
2. Zhang, J., Rabaiotti, J.: The PDF exploit: same crime, different face. https://www.symantec.com/connect/blogs/pdf-exploit-same-crime-different-face/. Last accessed 18 March 2018
3. Tzermias, Z., Sykiotakis, G., Polychronakis, M., Markatos, E.P.: Combining static and dynamic analysis for the detection of malicious documents. In: Proceedings of the Fourth European Workshop on System Security. EUROSEC'11, pp. 1–6 (2011)
4. Ratanaworabhan, P., Livshits, B., Zorn, B.: NOZZLE: a defense against heap spraying code injection attacks. In: Proceedings of the 18th Conference on USENIX Security Symposium. SSYM'09. Berkeley, CA USA (2009)
5. Willems, G., Holz, T., Freiling, F.: Toward automated dynamic malware analysis using CWSandbox. IEEE Secur. Priv. **5**, 32–39 (2007)
6. Kim, F.X.: Phenoelit: attacking networked embedded devices. Presented in Black Hat USA 2002. http://www.blackhat.com/presentations/bh-usa-02/bh-us02-phenoelit-network.pdf. Last accessed 18 March 2018
7. Adrian, C.: Hacking network printers. http://www.irongeek.com/i.php?page=security/networkprinterhacking. Last accessed 18 March 2018 (2017)
8. Sibert, W.O.: Malicious data and computer security. In: Proceedings of the 19th National Information Systems Security Conference (1996)
9. Muller, J., Mladenov, V., Somorovsky, J., Schwenk, J.: SoK: exploiting network printers. In: 2017 IEEE Symposium on Security and Privacy, pp. 213–230 (2017)
10. Cui, A., Costello, M., Stolfo, S.J.: When firmware modifications attack: a case study of embedded exploitation. Ndss (2013)
11. Zhang, J.: MLPdf: an effective machine learning based approach for PDF malware detection, pp. 1–6 (2018)

12. Chen, X., Li, C., Wang, D., Wen, S., Zhang, J., Nepal, S., Xiang, Y., Ren, K.: Android HIV: a study of repackaging malware for evading machine-learning detection (2018)
13. Hecht, A., Sagi, A.: PIDS: a behavioral framework for analysis and detection of network printer attacks, pp. 1–20 (2018)
14. Liu, J., Zhang, W., Zhang, Y., Hou, D., Liu, Y., Zha, H., Yu, N.: Detection based defense against adversarial examples from the steganalysis point to view (2018)
15. Clements, J., Lao, Y.: Hardware trojan attacks on neural networks (2018)
16. Sohi, S.M., Ganji, F., Seifert, J.-P.: Recurrent neural networks for enhancement of signature-based network intrusion detection systems (2018)
17. Alkasassbeh, M., Almseidin, M.: Machine learning methods for network intrusion detection and intrusion prevention systems. Pro. Quest Diss. Theses. **106** (2018)
18. Zhang, J., Su, H.: Machine learning attack and defense on voltage over-scaling-based lightweight authentication, pp. 1–12 (2018)
19. Cai, H.: a preliminary study on the sustainability of android malware detection. arXiv Comput. Sci. (2018)

Data Mining and Neural Networks to Determine the Financial Market Prediction

Jesus Silva, Jesús García Guliany, Lissette Hernandez, Rafael Portillo, Noel Varela, Hugo Hernández Palma, Osman Redondo Bilbao and Lesbia Valero

Abstract Predicting stock market movements has been a complex task for years by gaining the increasing interest of researchers and investors present all around the world. These have tried to get ahead of the way in order to know the levels of return and thus reduce the risk they face in investments [1]. Capital markets are areas of fundamental importance for the development of economies and their good management that favors the transition from savings to investment through the purchase and sale of shares [2]. These actions are so important that they are influenced by economic, social, political, and cultural variables. Therefore, it is reasonable to consider the value of an action in an instant not as a deterministic variable but as a random variable, considering its temporal trajectory as a stochastic process.

J. Silva (✉)
Universidad Peruana de Ciencias Aplicadas, Lima, Peru

J. G. Guliany
Universidad Simón Bolivar, Barranquilla, Colombia
e-mail: jesus.garcia@unisimonbolivar.edu.co

L. Hernandez
Universidad del Atlántico, Puerto Colombia, Colombia
e-mail: lissettehernandez@mail.uniatlantico.edu.co

R. Portillo · N. Varela
Universidad de La Costa (CUC), Calle 58 # 5566, Atlántico, Barranquilla, Colombia
e-mail: rportill3@cuc.edu.co

N. Varela
e-mail: nvarela2@cuc.edu.co

H. H. Palma · O. R. Bilbao · L. Valero
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

O. R. Bilbao
e-mail: oredondo@ul.edu.co

L. Valero
e-mail: lvalero@ul.edu.co

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_35

Keywords ACP · Market prediction · Mexican stock exchange · Stock market · RNA

1 Introduction

The challenge of predicting the behavior of stock exchange assets has led to multiple studies being published in recent decades; the first models used for the prediction of these time series were the linear ones for their easy interpretation. In the early 1980s, the most commonly used to model shareholder behavior was the ARIMA. However, after the decade, stock volatility is given more importance than in the average data and ARCH and GARCH became popular [3]. In general, ARIMA models such as the one presented by [4], multiple regression such as that developed by [5] genetic algorithms such as the one shown by [6], and in the last decade, more frequently applications with artificial neural networks (RNAs) such as that published by [7] have been used. The most accurate in the prediction have been those that have used computational methodologies such as RNs [8]. In this sense, [9] he used the HDRs for the forecast of the South American action of Inversions SA, compared the results with econometric models, obtaining greater accuracy in daily predictions with RNA. Arrieta concluded that the skill of statistical techniques is reduced as time series become complex while ARNs with sufficient data and proper architecture are much more efficient.

In general, neural networks have been used in the prediction of stocks, stock indices, futures, commodities, and currencies. Studies have confirmed that the use of backpropagation (BP) networks is ideal in short-term forecasts [10]. An RNA, a set of interconnected layers consisting of one or more artificial neurons, has the ability to learn, generalize, find hidden patterns, and process information efficiently [11]. The most popular architecture in financial prediction problems has been with three layers (input, hide, and output), and it has been shown that increasing the number of layers does not result in significant improvements in predictions [12].

Other popular networks have been multilayer, radial-based, and self-organizing networks. In the prediction of the stock market, the most common are the multilayer FF networks that use the BP algorithm; in these networks, the information is moved in a single direction from the input layer to the hidden layer and then to the output layer. After comparing the network output with a target or meta-value, the error is propagated backward to adjust the network parameters; this type of learning is known as supervised learning [13].

The research work presented in this document develops an application for the prediction of returns presented by the stock of a group of companies listed on the Colombian Stock Exchange with a short-term (daily) approach. Multilayer FF neural networks are used with the BP algorithm and are considered a hybrid analysis that integrates both variables of fundamental analysis and technical analysis, based on the reasoning that both analyses are considered by investors in their speculations.

2 Method

This methodology is an adaptation of the proposal by [14] that results from a compilation of the guidelines that various researchers have provided to guide predictive experiments with RNA.

- Data selection: For this activity, data series are taken into account that relate to variations in the action under study and are potential input patterns to the RNA. The series initially selected were 837; these included: currencies, stock exchange indices, commodities, futures, and other fundamental analysis data as financial indicators of the selected companies. The Macroeconomic indicators, are show in the Table 1. The selection of the action of the sample of companies met the criteria of: availability of information, relevance within the stock exchange, diversified industrial sector, impact on the Colombian economy, and relationship with indices and indicators. In addition, it is part of the Price and Quotes Index (CPI), the main stock index of the exchange.
- Data preprocessing: Ensuring the seasonality of time series and eliminating trends are fundamental requirements in the parametric and prediction models, for this reason simple return series were used instead of price series, as they used price series, as they have better statistical properties [15], for which equation was applied 1.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{1}$$

where:

R_t Simple return.
P_t Price of the day t.
$P_{t\text{-}1}$ Price of the day before t.

Table 1 Input data series (potential)

Category	Data series	No. of time series
Exchange rates	Closing price, opening, maximum, minimum	251
Indexes	Closing price, opening, maximum, minimum, volume	325
Raw materials	Closing price, opening, maximum, minimum, volume	47
Future	Closing price, opening, maximum, minimum, volume	124
Macroeconomics	Closing price, opening, maximum, minimum	32
Financial ratios	Interest coverage, debt reduction, gross margin income, net margin, operating margin, acid test, return on funds, operating profitability, earnings per share, sales, cash flow	58
Total		837

Likewise, the data were standardized so that they entered the active domain of the tansig transfer function, that is, to oscillate in a range of -1 and 1.

- Selection of input variables: Of the 837 series, it was necessary to debug those that had high input–input correlation because they would provide little or no additional information to the network and make it difficult to learn [16], for this reason, a correlation analysis between the series and removed those with correlations greater than 0.85.
- Build the model and select the best architecture: The data are divided into three groups: training, validation, and testing. During these stages, it was necessary to perform an experimentation process supported by MATLAB software, and its toolbox, in this, the number of inputs, and neurons in the hidden layer were combined. Once the experiments were carried out, the architecture that presented the best prediction performance, measured with the MSE statistics and the success rate in all three stages, was selected.
- Results assessment: The final phase consisted of expressing conclusions on the performance measures (MSE and success rate) obtained and with better results, analyzing the usefulness of predictions and establishing considerations for future research.

3 Results

The implementation of the ACP, in order to reduce the problem of dimensionality to speed up network learning, was carried out with the Kaiser criterion, which indicates that factors whose own values are greater than one must be preserved because they contribute to one unit of variance to total variance. The result of the GPA was the extraction of 18 components that explain 84.3% of the total variance of the model; Table 2 presents the result of the ACP expressed in the own values of each component.

After performing the linear combinations of the components, to extract the 18 factors that served as inputs to the network, training and experimentation are proceeded. A multilayer FF network was used, trained under the BP algorithm, and consisted of three layers (input, hidden and output); the choice of this architecture is due to the use of its is widely supported works such as [17–20]. Supervised learning was employed, which means that an expected output (objective) was provided to the network so that its parameters (synaptic weights and bias) were adjusted until the outputs approached the target. The expected output was the daily action returns presented in the study period.

The optimization algorithm used was that of decreasing gradient; its use is justified in its ability to predict economic or financial problems and its ability to generalize nonlinear time series models [4]. From the total of the experiments, it was necessary to extract those networks that in their training, validation and test phases obtain the same results or upcoming results because not necessarily the network that presents the lowest MSE in the training stage, is the one that performs best in the forecast.

Table 2 Own values and total variance explained

No.	Own values		
	Total	% Variance	% Accumulated
1	18.254	20.3514	19.987
2	11.253	15.2147	34.254
3	7.254	8.2587	40.254
…	…	…	…
15	2.0143	1.2354	84.257
18	1.00	1.0025	84.257
22	0.9458	0.933	84.365
…	…	…	…
94	−1.12E-12	−1.12E-14	102
96	−1.33E-12	−1.31E-02	103

The reason is that there may have been an overfitting, meaning that the network learned well from the training dataset, but is unable to generalize the model [21]. The best architectures for their proximity to training, validation, and testing values are presented in Table 3.

The network that presented the best performance in predicting daily action returns was the 451 network. Figure 1 shows the predictions of the performances made by this network and contrasts with the actual values of the period. Some significant deviations are observed; however, the graph somehow manages to simulate the behavior of the yields without reaching the accuracy. Another aspect that was analyzed was if the network managed to predict the direction the action took, that is, if the stock went up or down priced, the network managed to do it in 77.5% of the opportunities, being an important result for such problems.

Table 3 Architectures with better forecasting performances

No. RNA	MSE training	MSE tests	% Successful direction	Architecture	
				Neurons in hidden layer	Days (inputs)
451	0.00016547	0.00019521	77.5	5	57
487	0.00018754	0.00019874	70.2	6	20
587	0.00019854	0.00019635	68.25	8	12
802	0.00020147	0.00020214	68.29	10	27

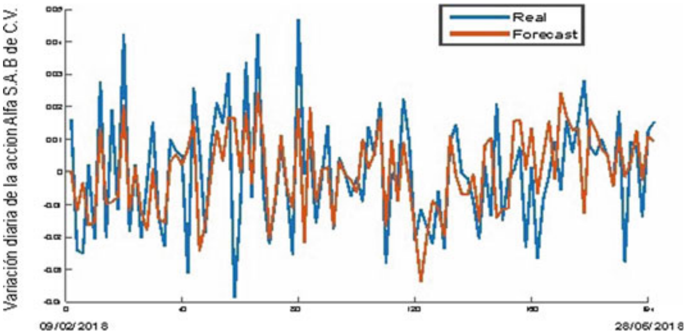


Fig. 1 Projections of action yields (forecasts)

4 Conclusions

The results show that it is possible to predict the direction of the stock with 75.5% success; this is an important result, considering that it would provide the investor with some degree of certainty in their investments. It was found that the accuracy of the predictions reduces after 30 days so it is ideal to re-train the neural network regularly to maintain forecast performance. As for the predictions of company performance, the SSM presented a value of 0.00018 for the best network. At this point, it is important to note that variations in the action under study range from −0.03 to 0.04, so this MSE value is not a reliable result for predicting yields.

It remains a challenge to predict the magnitude of the gain or loss an investor could get. By converting the predicted series of returns to a price series, it is possible to note that the network successfully generalized the trend of action, although not accurately predicted, shows that if it is possible to approach the results in short time period. It can be inferred that, with some other configuration of variables and architectures, or even with another type of network, more accurate results could be obtained. Finally, the results of this research add evidence to the global debate focused on the possibility or not of predicting financial markets in the short term. From the researchers' point of view, the results reflect a possibility of approaching market behaviors through the use of data analysis and artificial intelligence techniques.

The success of the results will depend on the availability and quality of the related information. Future research aims to improve the accuracy of predictions through the design of new prediction models, further delving into variables that could affect specific action. Some of the selected variables may have hindered the learning process and did not provide relevant information. Also, it is considered to use a larger sample size, other networks, and different learning algorithms to measure their impact on the results.

References

1. Zaharia, M., Xin, R.S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M.J., Ghodsi, A., Gonzalez, J., Shenker, S., Stoica, I.: Apache spark: a unified engine for big data processing. Commun. ACM **59**(11), 56–65 (2016)
2. Charris-Fontanilla, A., Parody-Camargo, E., Rodríguez, E.C.: Aplicación Del Modelo Estocástico Winer Gaus Para la predicción del precio de acciones del Mercado Bursátil Colombiano. In: XXI Congreso Internacional de Contaduría, Administración e Informática, Ciudad De México: Scielo, vol. 20 (2016)
3. Alonso, J.C., García, J.C.: ¿Qué tan buenos son los patrones del IGBC para predecir su comportamiento? Estudios Gerenciales **25**(112), 13–36 (2009)
4. Bechara, A., Enrique, J., Cruz, J.C.T., Ceballos, Hermilson Velásquez: Predicciones de Modelos Econométricos y Redes Neuronales: El Caso de La Acción de Suraminv. Semestre Económico **12**(25), 95–109 (2009)
5. Hahsler, M., Karpienko, R.: Visualizing association rules in hierarchical groups. J. Bus. Econ. **87**, 317–335 (2017)
6. Yuan, M., Ouyang, Y., Xiong, Z., Sheng, H.: Sentiment classification of web review using association rules. In: Ozok, A.A., Zaphiris, P. (eds.) Online Communities and Social Computing. OCSC 2013. Lecture Notes in Computer Science, vol. 8029. Springer, Berlin, Heidelberg (2013)
7. Silverstein, C., Brin, S., Motwani, R., Ullman, J.: Scalable techniques for mining causal structures. Data Min. Knowl. Discov. **4**(2–3), 163–192 (2000)
8. Amelec, V., Carmen, V.: Relationship between variables of performance social and financial of microfinance institutions. Adv. Sci. Lett. **21**(6), 1931–1934 (2015)
9. Viloria, A., Lezama, O.B.P.: Improvements for determining the number of clusters in k-means for innovation databases in SMEs. Procedia Comput. Sci. **151**, 1201–1206 (2019)
10. Kamatkar, S.J., Kamble, A., Viloria, A., Hernández-Fernandez, L., Cali, E.G.: Database performance tuning and query optimization. In: International Conference on Data Mining and Big Data, pp. 3–11. Springer, Cham (2018)
11. Cagliero, L., Fiori, A.: Analyzing twitter user behaviors and topic trends by exploiting dynamic rules. Behavior Computing: Modeling, Analysis, Mining and Decision, pp. 267–287. Springer (2012)
12. Gupta, A.: A survey on stock market prediction using various algorithms. Int. J. Comput. Technol. Appl. **5**, 530–533 (2014)
13. Moghaddam, A., Moghaddam, M.H., Esfandyari, M.: Predicción Del Índice Del Mercado Bursátil Utilizando Una Red Neuronal Artificial. J. Econ. Financ. Adm. Sci. **21**(41), 89–93 (2016)
14. Parody-Camargo, E., Charris-Fontanilla, A., García-Luna, R.: Modelo log-normal para predicción del precio de las acciones del sector bancario. Dimensión empresarial **14**, 137–149 (2016)
15. Vui, C.S., et al.: A review of stock market prediction with artificial neural network (ANN). IEEE International Conference on Control System, Computing and Engineering. Penang, IEEE (2013)
16. Vui, C.S., et al.: A review of stock market prediction with artificial neural network (ANN). International Conference on Control System, Computing and Engineering, pp. 477–482. Penang, IEEE (2013)
17. Viloria, A., et al.: Integration of data mining techniques to PostgreSQL database manager system. Procedia Comput. Sci. **155**, 575–580 (2019)
18. Lanzarini, L., Villa-Monte, A., Ronchetti, F.: SOM + PSO. A novel method to obtain classification rules. J. Comput. Sci. Technol. (JCS&T) **15**(1), 15–22 (2015)
19. Lanzarini, L., Villa Monte, A., Aquino, G., De Giusti, A.: Obtaining classification rules using lvqPSO Advances in swarm and computational intelligence. Lecture notes in computer science, vol. 6433, pp. 183–193. Springer-Verlag, Heidelberg, Berlin (2015)

20. Nanda, A., Shaked, M.: The hazard rate and the reversed hazard rate orders, with applications to order statistics, pp. 853–864 (2001)
21. Amelec, V., Carmen, V.: Validation of a model for productivity evaluation for microfinance institutions. Adv. Sci. Lett. **21**(5), 1610–1614 (2015)

Research on the Cloud Archiving Process and Its Technical Framework of Government Website Pages

Xinping Huang

Abstract This paper proposes the connotation of the government Web sites' Web page cloud archive and clarifies the archiving process of the government Web site webpage from the aspects of data collection, management, storage, utilization, and protection. On this basis, it builds the government Web site Web cloud archiving technology framework which contains the data acquisition layer, data management layer, data storage layer, data utilization layer, and application presentation layer.

Keywords Government website · Web page archiving · Cloud archiving · Technical framework

1 Introduction

At present, the amount of information resources held by the Chinese government departments has accounted for more than 80%. The resources are the true record of government network activities and have important preservation values. In the document issued by the National Archives Bureau for the Filing of Electronic Documents for Government Affairs, it has clearly stated that "The website serves is an important window for the information dissemination and external services of various units, it has formed a large number of webpage contents with credentials and value for examination, which need to be archived and secure in time with save conditions" [1]. Therefore, the archiving and preservation of government Web page information has risen to the national level, and it is of great significance to study the archiving process and technical issues of government Web pages.

Cloud computing is an emerging green IT service mode, which allows users to conveniently obtain shared and configurable computing resource pool on demand. IT can achieve rapid allocation and distribution of a large number of computing resources while reducing costs and is one of the most important means to save government IT expenditure. Cloud storage technology has the advantages of easy

X. Huang (✉)
School of Public Policy and Management, Tsinghua University, Beijing 100084, China
e-mail: hxp0730@163.com

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_36

storage of the architecture, mass storage of unstructured data, high performance, stability, etc., and also the secure management of the stored data through cloud security protection technology [2], which is a long-term storage of the government Web site pages as the novel solution to the problem.

2 Connotation of Government Web site Pages' Cloud Archive

2.1 Conceptual Definition

As early as the 1990s, relevant scholars have successively carried out theoretical research and practical exploration on long-term preservation of Web information resources (Web archive, referred to as WA). Web archive is a collection and the storage activity for valuable network information resources under the premise of ensuring complete, authentic, and readable data. The purpose is to achieve long-term preservation and flexible access of network information resources [3]. Cloud archiving is a remote data storage acquisition model. Different from the traditional archiving method, cloud storage service providers use cloud computing and cloud storage technologies to provide long-term data storage optimization services for saving data, thus ensuring the data security and flexibility for usage [4].

"Cloud archiving of government web pages" is a process of the long-term data preservation and timely feedback of visiting requests, which is based on the clear demand of long-term preservation of government Web pages resources, using cloud computing method to carry out a series of the archiving and preservation of valuable government Web pages.

2.2 The Basic Characteristics

The Web pages of government Web sites have the particularity of resources, so the cloud archiving of the government Web sites is different from the general archiving method, and it is with characteristics of original record, credential value, mass storage, authority, and easy expansibility. In the process of cloud archiving, it is necessary to fully consider the basic characteristics of the government Web site pages cloud archiving, so as to adapt the technical method and the archiving process. The following are some of the primary procedures.

Original Recordability The government Web site pages are digital cultural heritage that is easy to disappear. It records the government's various government affairs activities, as well as the information on the changes of government internal organization personnel and the time and space information of relevant government affairs. In the process of archiving and storing government Web pages in the form of cloud storage, it is necessary to ensure that original record of Web information remains unchanged, that is, the Web information will not have any deviation in information content due to the change of storage mode, and storage space can be expanded to ensure the value and quality of archives.

Voucher Value From the content point of view, the government Web site Web pages is directly converted from the file and is not prepared separately for the usage after use; from the form, the government Web site Webpage is recorded by the electronic file's metadata, so it cannot be changed. In particular, the archived Web pages of government Web sites are different from those of general archives. They have relatively few restrictions on utilization and do not have the characteristics of less usability. However, its voucher value is also limited by the time and regional characteristics of the published content.

Mass Storage Cloud archiving method utilizes cloud storage technology to the dynamically expand storage space as needed. Not only can it flexibly and efficiently deal with the archival storage problem of massive government Web site pages, but also provide the long-term storage optimization services for the mass storage data, which further protects data storage security while saving costs.

Authoritative Feature When using cloud storage technology for archival storage, in order to maintain the authority of government Web page information, on the one hand, it is necessary to identify whether the source of information meets the authoritative requirements, and on the other hand, to ensure that the storage content is not subject to change to ensure its authority. It is not subject to changes in form of storage, and attention is paid to authority of the government Web page information to ensure that its archival record is authoritative and stable.

Easy Extensibility Web pages of government Web sites use cloud storage technology to realize cloud archiving. The easy expansibility of cloud archiving of government Web sites mainly reflects two aspects. One is the storage scale effect realized by cloud storage, and the other is the elastic expansion level of storage space. The cloud storage architecture adopts the parallel capacity expansion mode, which is simpler than previous capacity expansion link operation without restrictions. This feature helps to reduce operating costs and effectively reduce resource waste.

3 Government Web site Pages Cloud Archiving Processes

Government Web site pages cloud archiving is the process of archiving and saving government Web site Web resources by using various IT technologies and resources in the cloud environment. The government Web page file belongs to an electronic file. According to the electronic file life cycle theory, the generation, distribution,

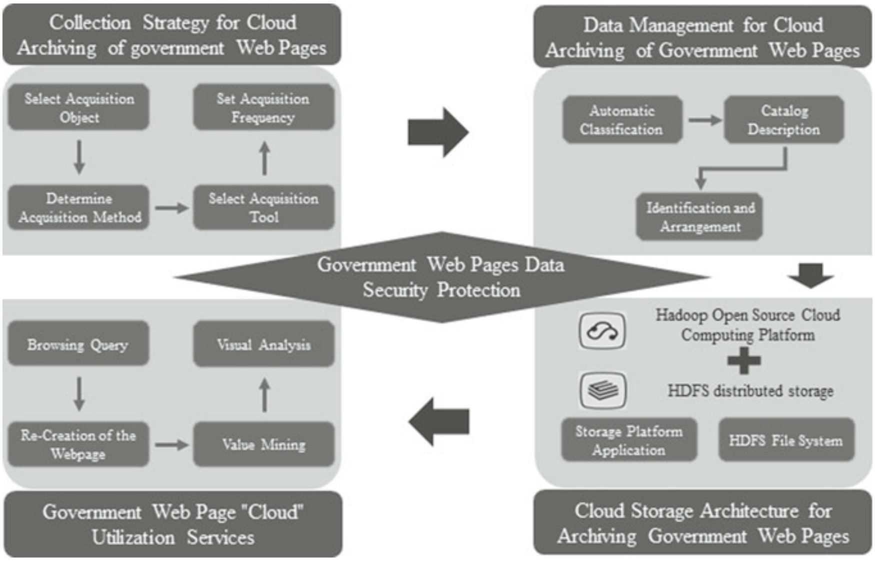


Fig. 1 Cloud archiving procedures of government Web pages

capture, archiving, and utilization of the government Web site Web pages file is the complete life cycle. This paper analyzes and also elaborates the cloud archiving process of government Web pages along with the idea of collection, management, storage, and utilization, and runs data protection throughout the archiving process to ensure continuity and standardization of information resources of government Web pages from collection to the long-term preservation. The proposed process is shown in the Fig. 1.

3.1 Collection Strategy for Cloud Archiving of Government Web Pages

3.1.1 Select Acquisition Object

The information collected by government Web page is collected by the government website with "gov.cn" in the domain name. The collection follows the principle of goal orientation, and from the actual needs of file management, the online data resources in the government Web site Web page are collected in real-time manner and dynamically. The collected content includes various forms of the multimedia, such as text, pictures, and video. When selecting the collection object, it is necessary to determine its storage value, authority, integrity, and systematicness. At the same time, it is necessary to determine different collection objects according to different business requirements, and comprehensively consider timeliness of the collection content and whether there are copyright issues. Different types of information release have the different aging value. Classifying the collection according to the aging characteristics of the collection objects helps to ensure storage or utilization value of the collection results. At the same time, only after solving the copyright problem of acquisition can we determine the acquisition object and implement the acquisition behavior. Otherwise, it should not be included in the collection scope.

3.1.2 Determine Acquisition Method

There are two common collection methods for the network information resources, namely selective acquisition and also integrity collection. As a kind of network information resource, the government Web site Web page can be integrated and then expanded based on the two collection methods to ensure the effectiveness of the collection behavior. In the collection, the integrity collection is the main means, supplemented by selective acquisition. That is, while all the information in the target government Web site is automatically collected, the intelligence of domain experts is used to manually identify the government Web site information, and the government Web site information with the higher value is divided into separate categories. Information depth is frequently collected, so that it is targeted. The hybrid acquisition method helps to collect and save the information resources of the national government Web site in a comprehensive and in-depth manner.

3.1.3 Select Acquisition Tool

At present, the common open-source tools for government website information collection include heritrix, httrack, nutch, smart crawler, etc., under the premise of fully considering specific requirements of integrity collection and selective acquisition, combining the characteristics and cost of existing acquisition tools, selecting two open-source tools, Heritrix and HTTrack, and improving and expanding functions based on original to achieve the collection of government web page information [5]. HTTrack has the strong link analysis function. Therefore, for small-scale important government Web sites, HTTrack can be used to obtain the deep structure of such Web sites from the server, so as to complete the in-depth collection of the important government Web page information.

3.1.4 Set Acquisition Frequency

This article refers to the "Website Macro Evaluation Score Sheet" [6] developed by the US "North Carolina Government Website Archives" project, and evaluates the collected government Web sites, and finally sets the frequency of the collection according to the score. Among them, for Web sites with scores of "7.0–7.99," the frequency of collection is once a year; the scores are "8.0–10.99," the frequency of collection is four times a year; the score is on the Web site of “11.0–21.0,” the frequency of collection for once a month.

3.2 Data Management for Cloud Archiving of Government Web Pages

3.2.1 Automatic Classification

The classification of government Web site pages needs to consider two points: one is the classification standard, and the other is the classification method. The classification standards should be demand-oriented and also convenient for the subsequent storage and utilization. According to different classification standards, corresponding classification schemes should be formed to guide implementation of subsequent classification. The selection of classification methods should take into account the frequent updates of government Web sites and the huge volume of data. Therefore, automatic classification is an inevitable means to ensure the rapid and orderly classification. At present, the methods of automatic text categorization mainly include knowledge engineering-based classification method and machine learning-based classification method [7]. Machine learning-based classification method is the mainstream choice for the automatic text classification. This paper applies the guided machine learning method to realize the automatic classification of large-scale government Web pages. The basic idea of automatic classification based on machine learning is to first collect the collected government Web sites. The text of the Web page is represented by the VSM vector space model. After assigning the text features of the document, the classifier is selected to realize the automatic classification of text [8]. In the specific implementation, it is necessary to pay the attention to the quantity and quality of the government Web site pages selected as the training set. Firstly, the experts manually classify, then analyze and mine the links between the keywords and the classes, and then train the classifiers through guided machine learning model. Finally, the classifiers are used to realize automatic classification of government web pages.

3.2.2 Catalog Description

In order to realize the long-term preservation and sustainable use of government Web page information, it is also necessary to catalog the collected Web pages. Specifically, it is necessary to analyze, organize and also record the acquired Web pages, accurately describe content, structure, management process, management system, and some other description objects of the Web site, so as to realize the cataloging description of government Web pages. The purpose is to establish the relationship between archived Web pages and their metadata to give identification to the archived government Web pages, and to lay a foundation for the flexible access and utilization of massive government Web pages. Metadata can be used to discover and then retrieve government information resources, and more and more government agencies around the world have started to establish and promote the use of metadata standards. Based on the features of DC metadata, such as ease of operation and scalability, and integrating current wide range of DC applications [9], this paper is relatively mature DC metadata standard as the data description format for government Web page information cataloging.

3.2.3 Identification and Arrangement

The government Web page file belongs to a special electronic document. It draws on the electronic document identification method. The identification and collation of government Web page documents includes two aspects: document identification and usability judgment. Usability judgment of Web page has a certain complexity, which is different from the definition of usability judgment in the user experience (UX) domain. The main body of the usefulness judgment of government Web site Web page includes internal and external features of government Web site Web page information, including the information source, information content, release date and release form of webpage information, etc. [10]. Through the analysis of these characteristics, the preservation value and content quality of government Web files are tested. When making usability judgment, it is necessary to avoid information deviation caused by the subjectivity and the variability of individual judgment and adopt objective and unified criteria for judgment.

3.3 Cloud Storage Architecture for Archiving Government Web Pages

The dynamic and linked nature of the Web site and the demand for its access in the future make the preservation of government Web page information more dependent on the long-term preservation strategy and the storage architecture of the dynamic expansion. The cloud storage technology has the advantages of easy expansion, high performance, and stability of the storage architecture. In addition, it provides a storage service platform. At the same time, technologies such as the cloud data isolation, cloud data integrity verification, and data availability protection provided by the server provide data security management guarantee [11]. HDFS distributed storage technology is usually adopted to realize the cloud storage architecture of government Web pages. The cloud storage architecture based on this technology scheme includes the virtual storage platform application and file entity storage environment. Among them, the storage platform application part by calling the Web application server running on HDFS API application to complete the Hadoop storage cluster of government Web page stored in the data in a relational database to read and write, delete, backup operation, and based on the browser/server mode user support government Web sites Web archiving cloud storage service. At same time, the paper refers to common media format in MIME as the proposed storage format, adopts the maintenance method of periodic detection and backup, migrates the Web pages of the different format types as needed, and solves the long-term preservation problem of archiving government Web pages.

3.4 Government Web Page "Cloud" Utilization Services and Safety

First of all, the browsing query is a basic service in the "cloud" utilization service provided by the government Web site Web page and is a key interface for realizing other services. It can respond to users' specific information needs dynamically and in a timely manner, and realize fast retrieval and flexible access to the archived massive government Web site pages, which requires the professional and perfect retrieval system support. Secondly, the recreation of the Web page is to present the content of the archived Web page to the user in its original appearance, so that the user feels like accessing the original Web page. Then, value mining is to use the Web-based data mining technology to analyze the archived government Web page information, and to mine the new and useful tacit knowledge, which will make the tacit knowledge explicit and further deductive reasoning knowledge of the value. Finally, visual analysis is the use of the data visualization technology to display government Web page data stored in the form of large data sets in the form of the graphics or images such as interactive maps, timelines, tag clouds, photo galleries, etc., making the stored data more intuitive with vivid display [12].

Under cloud computing service mode, archiving government Web page data is stored in the cloud, but the transparency of the cloud services and the uncertainty of service providers make the security protection of data become the primary consideration for cloud archiving [13]. Key technologies such as trusted cloud computing, data isolation and data protection can be adopted to ensure the confidentiality, authenticity, integrity, and reliability of archiving government Web page data in the cloud computing environment.

4 Technical Framework for Cloud Archiving of Government Web Pages

The author proposes the technical framework of the government Web site Web page cloud archiving, as shown in Fig. 2. The technical framework is a five-layer architecture, from bottom to top for the data acquisition layer, data management layer, data storage layer, data utilization layer, and application presentation layer.

Data Acquisition Layer The data acquisition layer implementation functions mainly include Web page evaluation and Web crawling. Combining the cloud computing architecture with general Hadoop-based distributed computing framework, MapReduce technology, and HDFS distributed file system, the open-source Web crawler Heritrix system and HTTrack system are used to achieve parallel collection and integrity collection to improve the efficiency of Web page collection with the effect. At the same time, according to the results of Web page evaluation, the collection strategy is formulated, and the collection of large-scale archived Web pages is achieved in strict accordance with the collection strategy.

Data Management Layer The functions implemented by the data management layer include classification, recording, and identification. In the cloud archive classification, machine learning algorithms such as the support vector machine and BP neural network and Bayesian text classification algorithm can be used to automatically build the feature lexicon, and the automatic classification of collected government Web pages can be realized through feature collocation.

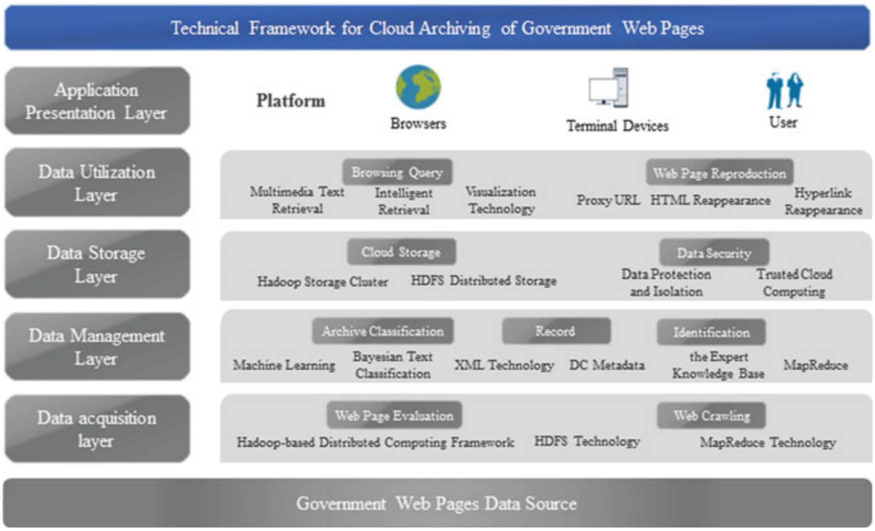


Fig. 2 Technical framework for cloud archiving of government Web pages

In the aspect of the cloud archiving description, DC metadata standard is used to describe the collected government Web sites in XML format. DC metadata has the characteristics of good expansibility and easy to understand and use. As an extensible markup language, XML has unique advantages in the integration with DC metadata [14]. XML allows user-defined tags and document structures, so the core element names of DC metadata can be directly applied to create XML format tags for the resource description. In the identification part of the cloud archive, MapReduce technology mode under the cloud environment can be adopted to build the intelligent identification system of cloud archive based on the expert knowledge base to realize the automatic identification of Web page information. The authentication of government Web pages includes content authentication and technical authentication. Among them, content identification is the identification of documents, which can realize the distinction between government Web files and non-governmental Web files. Technical appraisal is to judge authenticity, validity, and availability of electronic documents.

Data Storage Layer This layer is based on the distributed data storage management system of cloud storage, and the application data access component provides the data access service of the corresponding database for the realization of the business functions such as the collection, management, storage, and also utilization of the government Web site. In addition, data security management is a key issue to be solved at this layer, mainly based on data security protection technology to ensure the originality and credibility of the file [15].

Data Utilization Layer The utilization of the cloud archive involves multimedia information retrieval technology, intelligent retrieval technology, and Web page reproduction technology. The cross-reference index method is used to realize the multi-mode retrieval mechanism of multimedia text in government Web pages to improve the accuracy and the efficiency of multimedia text retrieval in government Web pages. Aiming at the organization strategy and access method of archiving the distributed storage data of government Web pages on the platform, the intelligent semantic retrieval of the distributed storage data of government Web pages is realized by using the relevant technologies and methods of intelligent agent and semantic analysis [16].

Application Presentation Layer The application presentation layer is the external interface to realize the cloud archiving function. The main body of the application presentation layer includes users, terminal devices, and browsers. With the help of terminal devices and browsers, users can archive and utilize the Web pages of government Web sites through portals. The portal Web site can simplify the complex process and technology of the cloud archiving to the greatest extent and reduce the difficulty of users' operation to make the interactive experience better, improve user satisfaction, and also realize the efficient utilization of the archives resources of government Web site Web page.

5 Conclusion

The government Web site Web page has high research value and intelligence value, and it is of great significance to archive it for the long-term preservation and the utilization. This article is mainly from the cloud archive of the overall process and technical implementation of the government Web site Web page cloud archive in detail. On the basis of defining the content of government Web pages, this paper expounds the archiving process of government Web pages information from the aspects of data collection, management, storage, utilization, and protection. To a certain extent, it has enriched the relevant research on the government's Web site's original digital government information and also provided theoretical support and practical reference for the long-term accessibility of government information in the cloud environment.

Acknowledgements This research is supported by the National Social Science Youth Fund Project "Online archiving and development and utilization of web pages based on the cloud computing" (Project No.: 18CTQ040).

References

1. Notice of the national archives office on the consultation of six archival industry standard projects such as the "Regulations on the Filing of Electronic Documents for Government Affairs" (EB\OL) 09 March 2019. http://www.saac.gov.cn/daj/tzgg/201903/5e26a638415c490a92f6ce50636dba2b.shtml?tdsourcetag=s_pcqq_aiomsg
2. Tao, S.: Analysis and research on cloud backup strategy of archives digital resources. Arch. Commun. (04), 12–16 (2012)
3. Guangyuan, Y.: A review of web archive research at home and abroad. Libr. J. **282**(10), 88–94 (2014)
4. Xinrong, H., Wenqi, P., Xiaojie, W.: Cloud archive—a new archive method in cloud environment. Arch. Constr. **04**, 4–7 (2014)
5. Huang, X., Wang, P.: Research and application progress of web archive technology in recent years at home and abroad. Libr. Sci. Res. (18), 30–35 + 19 (2016)
6. He, H.: Research on government website information resource preservation system. Wuhan University (2010)
7. PANG, G., JIANG, S.: A review of text automatic classification technology. Inf. Theory Pract. **35**(02), 123–128 (2012)
8. Prajapati, B., Garg, S., Chauhan, N.C.: Some investigations on machine learning techniques for automated text categorization. Int. J. Comput. Appl. **71**(71), 32–36 (2013)
9. Tambouris, E., Tarabanis, K.A.: An overview of DC-based e-government metadata standards and initiatives. In: Electronic Government: Third International Conference, EGOV 2004. Zaragoza, Spain. Proceedings Springer, Berlin, Heidelberg. August 30–September 3 (2004)
10. Wei, W.: Some thoughts on the file filing of websites. China Arch. **10**, 68–69 (2017)
11. Chaosheng, F., Zhiguang, Q., Ding, Y.: Cloud data security storage technology. Chin. J. Comput. **38**(01), 150–160 (2015)
12. Ping, W., Xinping, H., Nanxue, Z.: The ways and trends of the development and utilization of foreign web archive resources. Res. Libr. Sci. (23), 43–49 (2015)

13. Mcleod, J., Gormly, B.: Using the cloud for records storage: issues of trust. Arch. Sci. **17**(02), 1–22 (2017)
14. Bosch, T., Mathiak, B.: How to accelerate the process of designing domain ontologies based on XML schemas. Int. J. Metadata Semant. Ontol. **8**(3), 254 (2013)
15. Askhoj, J., Sugimoto, S., Nagamori, M.: Preserving records in the cloud. Rec. Manag. J. **21**(03), 175–187 (2011)
16. Xiang Jing, W., Zhenxin, S.Z.: Research on webpage re-creation method and application based on web archive. Digit. Libr. Forum **07**, 17–21 (2009)

Determination of Contents Based on Learning Styles Through Artificial Intelligence

Jesus Silva, Lissette Hernandez, Jenny Romero, Noel Varela, Hugo Hernández Palma, Nataly Orellano Llinás, Yasmin Florez and Carlos Vargas Mercado

Abstract The study presents the development of a platform for structuring adaptive courses based on active, reflexive, theoretical and pragmatic learning styles using artificial intelligence techniques. To this end, the following phases were followed: search, analysis and classification of information about the process of generating content for courses; analysis and coding of the software component for generating content according to learning styles; and application of tests for validation and acceptance. The main contribution of the paper is the development of a model using neural networks and its integration in an application server to determine the contents that correspond to the active, reflexive, theoretical and pragmatic learning styles.

Keywords Artificial intelligence · Neural networks · Learning styles

J. Silva (✉)
Universidad Peruana de Ciencias Aplicadas, Lima, Peru

L. Hernandez
Universidad del Atlántico, Puerto Colombia, Colombia
e-mail: lissettehernandez@mail.uniatlantico.edu.co

J. Romero · N. Varela
Universidad de la Costa (CUC), Calle 58 # 5566, Atlántico, Barranquilla, Colombia
e-mail: jromero58@cuc.edu.co

N. Varela
e-mail: nvarela2@cuc.edu.co

H. H. Palma · C. V. Mercado
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

C. V. Mercado
e-mail: cvargas@ul.edu.co

N. O. Llinás · Y. Florez
Corporación Universitaria Minuto de Dios—UNIMINUTO, Barranquilla, Colombia
e-mail: nataly.orellano@uniminuto.edu

Y. Florez
e-mail: yasmin.flores@uniminuto.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_37

1 Introduction

Among the characteristics of the desirable education is that all students can learn in spite of their different learning styles, which are defined as the set of strategies designed to address their preferences to certain academic resources, such as videos, text and images, among others [1]. In order to comply with the above, it is important to offer adaptive courses based on the learning styles of each student.

Under these references, the study focuses on a platform for structuring adaptive courses based on active, theoretical, reflexive and pragmatic learning styles using artificial intelligence techniques, in order to collaborate with teachers and students at the university level.

The development of the proposal began with an analysis of the methods and tests for determining the students' learning styles [2]. Once the study was carried out, the Honey-Alonso test [3] was selected since it was the most suitable tool according to the research carried out with 10,325 students from different faculties of the University of Mumbai in India. An additional comparison of the results of this test with strategies linked to the different learning styles was proposed for improving the post-evaluation results [4].

Next, a study was carried out on the various artificial intelligence techniques, showing that neural networks are a feasible solution to this type of problem [5, 6]. After choosing the artificial intelligence technique, the research developed a base of knowledge based on keywords and identifying each of the learning styles.

Then the requirements and use cases were collected regarding the administration of students and courses in which the contents will be included. Once this was done, the researchers proceeded to design the model for the prediction of adaptive courses according to the learning styles in the platform [7].

Finally, a service was implemented in the application server of the platform for the Weka framework and thus achieves the identification of each of the contents of a specialized course for each student. In order to monitor and validate its operation, the corresponding tests were applied to third-cycle students from the Systems Engineering career at the University of Mumbai in India.

2 Method

The implementation of the neuronal network model for the creation of adaptive courses according to the learning style (active, reflective, theoretical and pragmatic) of each student was done through the programming environment in Java, JavaScript, Java Server Faces and MySQL, integrating all the operation in a block so that it can be executed on a web platform based on Java [8]. The method is based on three phases described below [9].

2.1 Search, Analysis and Classification of Information About the Content Generation Process for the Courses

As initial phase of the research, it consisted of a bibliographic search of studies with the same nature in terms of intelligent generation of contents for the courses. Once the search was completed with respect to the generation of contents according to learning styles, the researchers selected the one that offers the best results. Finally, a test was selected to determine the learning style of each student, leaving Honey-Alonso test as such. In the post-evaluation of the test results, an improvement was made by inserting an additional comparison of the results of this test with the strategies linked to the different learning styles [10].

2.2 Analysis and Coding of the Software Component for Content Generation According to Learning Styles

At this stage, the functional requirements of this component were analyzed and determined based on the form of qualification and the necessary components suggested by the Chaea-Alonso test [11]. It was modeled and programmed; additionally, the improvement in the results obtained by this test was integrated through the comparison of the preferences in terms of individual strategies of each student. For the design of courses, a subject of a Systems Engineering course was selected, delimiting its content [12].

In addition, the requirements and cases of use were collected regarding the administration of students and courses in which such content will be included. Once completed, a neuronal network was used to design the model for the prediction of adaptive courses according to the learning styles in the platform [13]. Finally, all the components were integrated into a web platform, having as its execution environment an application server that provides data source services, web services, support for the implementation of the neural network, development of administrative environments for teachers and administrators of the platform, as well as an area for courses that are adaptive to the learning style of students [14].

2.3 Validation and Acceptance Tests

In this phase, a test plan was developed to validate the platform with the previous results obtained in the surveys. In other words, the results obtained on the platform were consistent with those obtained in the aforementioned surveys [15].

With regard to software quality assurance, the tests were carried out with 300 students and 10 professors from the Systems Engineering career at the University of Mumbai in India, obtaining the acceptance and approval of the platform.

Table 1 Functional requirements

Code	Description	Complexity
TH011	To the administrator of the platform to administer teachers	Medium
TH022	To the administrator to manage roles	Medium
TH033	To the teacher to administer students	Medium
TH044	To the teacher to administer his courses	Medium
TH055	To the teacher to manage the contents	Medium
TH066	The teacher will have an intelligent software component that allows him to generate the contents for a course according to the style of each student	Medium
TH077	The teacher determines a student's learning style using the combination of the Chaea-Alonso test and the preferences of learning strategies	Medium
TH088	The student will be evaluated only once autonomously by the platform to determine the learning style through the Chaea-Alonso test and selection of preferred strategies	Medium
TH0110	The student will be shown the contents of the course of according to the previously defined learning style	Medium

3 Experimental Design

This section provides details of the developed activities, separated into phases that allowed the design and implementation of the platform for structuring the adaptive courses based on active, reflexive, theoretical and pragmatic learning styles using artificial intelligence techniques [16].

3.1 *Analysis and Determination of Functional Requirements, Roles and Actors*

The following functional requirements were obtained for the e-learning system, see Tables 1 and 2.

3.2 *Design of the System Knowledge Base Model*

For the implementation of the intelligent part of the system, the multilayer perceptron algorithm from the Weka library [17] was applied. The .arff files (file format used by Weka) were created for the approximate determination of the learning styles of the contents entered into the platform.

Table 2 Roles and actors of the platform

Actor	Roles
Administrator	• Manage teachers • Manage students • Manage courses • Assign teacher to course
Teacher	• Manage students • Manage contents • Add students to course • Manage evaluations
Student	• Display contents • Take evaluations • Get learning style by Chaea-Alonso test and selection of preferred strategies

The parts of an .arff file are described below:

- Relationship Name: Declared as @RELATION name.
- Declaration of attributes next to their data type: Declared as @ATTRIBUTE name_of_attribute type. The type of data used in the attributes was NUMERIC. The last declared attribute corresponds to the class, that is, the possible answers that the algorithm will return. In this case, the possible results are: "Yes" or "No."
- Training data header: Declared with @DATA.
- Instances: After the header line of the training data, each instance is specified in a single line. The instances detail the values of each attribute, separated by "," (comma) and the last element of the instance corresponds to the target/prediction value.

Table 3 shows the content of an .arff file specifying each of its parts by columns (e.g., in the column "Training data (Header + Instances)" only 10 of the 106 instances contained in the original .arff file used in this study were considered).

Once the knowledge base model was designed, a service was implemented in the platform application server for the Weka framework. For this, the knowledge base

Table 3 .arff file content (knowledge base)

Relation	Attributes	Training data (Header + Instances)
@RELATION act	@ATTRIBUTE act NUMERIC @ATTRIBUTE prag NUMERIC @ATTRIBUTE teo NUMERIC @ATTRIBUTE refl NUMERIC @ ATTRIBUTE **result** {yes, no}	@data 0.2, 0.35, 0.75, 0.3, **no** 0.9, 0.5, 0.4, 0.65, **no** 0.9, 0.4, 0.066, 0.0, **yes** 0.66, 0.75, 0.7, 0.4, **no** 0.8, 0.33, 0.0, 0.5, **yes** 0.0, 0.4, 0.5, 0.16, **no** 0.88, 0.89, 0.06, 0.7, **yes** 0.44, 0.04, 0.2, 0.2, **yes** 0.8, 0.58, 0.35, 0.7, **yes** 0.4, 0.55, 0.9, 0.98, **no**

was placed in the public access folder of the server (generally/bing) and the service provided by the application server in the lib folder of the server with the content entry form. Finally, the neural network module becomes lemmatized terms [18]. The relationship of learning terms-styles is analyzed, obtaining an input for the neural network.

4 Results

4.1 Knowledge Base

After designing the knowledge base, the .arff files were used to train the neural network in Weka and thus generate the prediction model. Weka is a tool with a graphical interface that allows the processing of any set of data through the application, analysis and evaluation of different techniques, mainly those coming from automatic learning [19]. The neural network model (multilayer perceptron algorithm) for the application generated by Weka is the following Fig. 1.

By using the multilayer perceptron algorithm for the approximate determination of the learning styles of the contents entered into the platform, 96.33% accuracy was obtained in the predictions.

Cross-validation with k groups [20], with $k = 12$, was used as a technique to measure the performance of the algorithm. This technique is based on dividing training data into equal or nearly equal groups. Then k executions of the algorithm are performed. In each execution, a different group is taken from among the k groups to use it as a test set in order to evaluate the classifier, and the remaining $k - 1$

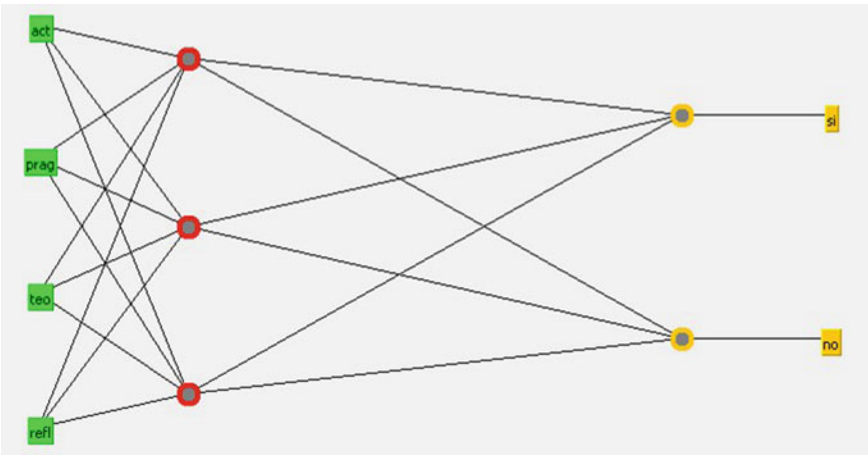


Fig. 1 Neural network model generated by Weka

groups are used to train the neural network. In the end, the precision percentage of the multilayer neural network was obtained by calculating the mean of the precision obtained in the *k* runs.

5 Conclusions

Once the study on the topic about the platform for structuring adaptive courses based on active, reflective, theoretical and pragmatic learning styles using artificial intelligence techniques is concluded, the following conclusions and recommendations for future studies are established:

- The students' learning style was determined by means of the Honey-Alonso test and strategies linked to the different learning styles.
- The various artificial intelligence techniques studied represent a viable solution for predicting the structuring of contents according to the student's learning style. However, the one that prevailed among all of them was the neural network, since, through this technique, one could estimate that content corresponds to a learning style using a vocabulary with themed terms.
- By using neural networks and their multilayer perceptron algorithm, an efficient uncertainty model was designed to estimate the structure of the content and thus be able to offer a course to each student depending on their learning style.
- The neural network model was implemented on a web platform that generates estimated results on the way students learn, and this was validated in a real educational setting through an experimental group consisting of students and teachers of the Systems Engineering career at the University of Mumbai in India.

References

1. Refaeilzadeh, P., Tang, L., Liu, H.: Cross-validation. In: Encyclopedia of Database Systems, pp. 532–538. Recuperado el 26 de diciembre de 2014, de http://leiting.net/papers/ency-cross-validation.pdf (2009)
2. Jamae, J., Johnson P.: JBOSS in Action. Manning Publications Co. Recuperado el 26 de diciembre de 2014, de http://testa.roberta.free.fr/My%20Books/Computer%20programming/Java/Manning%20%20JBoss%20in%20Action%20Configuring%20the%20JBoss%20Application%20Server.pdf (2009)
3. Jendrock, E., Cervera-Navarro, R., Evans, I., Gollapudi, D., Haase, K., Markito, W., Srivathsa, C.: The Java JEE 6 Tutorial. Recuperado el 26 de diciembre de 2014, de http://docs.oracle.com/javaee/6/tutorial/doc (2013)
4. Hahsler, M., Karpienko, R.: Visualizing association rules in hierarchical groups. J. Bus. Econ. **87**, 317–335 (2017)
5. Yuan, M., Ouyang, Y., Xiong, Z., Sheng, H.: Sentiment classification of web review using association rules. In: Ozok, A.A., Zaphiris, P. (eds.) Online Communities and Social Computing. OCSC 2013. Lecture Notes in Computer Science, vol. 8029. Springer, Berlin, Heidelberg (2013)

6. Silverstein, C., Brin, S., Motwani, R., Ullman, J.: Scalable techniques for mining causal structures. Data Min. Knowl. Discov. **4**(2–3), 163–192 (2000)
7. Amelec, Viloria, Carmen, Vasquez: Relationship between variables of performance social and financial of microfinance institutions. Adv. Sci. Lett. **21**(6), 1931–1934 (2015)
8. Viloria, A., Lezama, O.B.P.: Improvements for determining the number of clusters in *k*-means for innovation databases in SMEs. Proc. Comput. Sci. **151**, 1201–1206 (2019)
9. Kamatkar, S.J., Kamble, A., Viloria, A., Hernández-Fernández, L., Cali, E.G.: . Database performance tuning and query optimization. In: International Conference on Data Mining and Big Data, pp. 3–11. Springer, Cham (2018)
10. Cagliero, L., Fiori, A.: Analyzing Twitter user behaviors and topic trends by exploiting dynamic rules. In: Behavior Computing: Modeling, Analysis, Mining and Decision, pp. 267–287. Springer (2012)
11. Erlandsson, F., Bródka, P., Borg, A., Johnson, H.: Finding influential users in social media using association rule learning. Entropy **18**, 164 (2016)
12. Meduru, M., Mahimkar, A., Subramanian, K., Padiya, P.Y., Gunjgur, P.N.: Opinion mining using Twitter feeds for political analysis. Int. J. Comput. (IJC) **25**(1), 116–123 (2017)
13. Abascal-Mena R., López-Ornelas, E., Zepeda-Hernández, J.S.: User generated content: an analysis of user behavior by mining political tweets. In: Ozok, A.A., Zaphiris, P. (eds.) Online Communities and Social Computing. OCSC 2013. Lecture Notes in Computer Science, vol. 8029. Springer, Berlin, Heidelberg (2013)
14. Dehkharghani, R., Mercan, H., Javeed, A., Saygin, Y.: Sentimental causal rule discovery from Twitter. Expert. Syst. Appl. **41**(10), 4950–4958 (2014)
15. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by Gibbs sampling. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005), pp. 363–370 (2005)
16. Viloria, A., et al.: Integration of data mining techniques to PostgreSQL database manager system. Proc. Comput. Sci. **155**, 575–580 (2019)
17. Torres Samuel, M., Vásquez, C., Viloria, A., Hernández Fernandez, L., Portillo Medina, R.: Analysis of patterns in the university Word Rankings Webometrics, Shangai, QS and SIRScimago: case Latin American. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligent and Lecture Notes in Bioinformatics (2018)
18. Lanzarini, L., Villa-Monte, A., Ronchetti, F.: SOM + PSO. A novel method to obtain classification rules. J. Comput. Sci. Technol. (JCS&T) **15**(1), 15–22 (2015)
19. Lanzarini, L., Villa Monte, A., Aquino, G., De Giusti, A.: Obtaining classification rules using lvqPSO advances in swarm and computational intelligence. In: Lecture Notes in Computer Science, vol. 6433, pp. 183–193. Springer, Berlin, Heidelberg (2015)
20. Martinez, P.: Investigación y análisis de los estilos de aprendizaje del profesorado y alumnos del primer ciclo de educación secundaria obligatoria en el ámbito del C. P. R. de LAREDO. CANTABRIA. I Congreso Internacional de estilos de aprendizaje. Recuperado el 9 de junio de 2014, de http://www.estilosdeaprendizaje.es/PMGeijo.pdf (2004) (Consulta: 9 de junio del 2014)

Evaluation Computing of Cultural Tourism Resources Potential Based on SVM Intelligent Data Analysis and IoT

Jun Chen and Mang Lu

Abstract Evaluation computing of the cultural tourism resources potential based on SVM intelligent data analysis and IoT is analyzed in this paper. The research highlights are as follows: (1) In order to effectively improve classification performance of symbol data, a symbolic space representation method is defined by deepening the spatial structure relationship between different attribute values and labels of symbol data. (2) Some parallel operators and parameters are adjusted to optimize the performance of the new algorithm. (3) In the security analysis of the actual protocols, the proof method of reduction is often adopted to reduce the security proof of the protocols to some recognized difficult problems. The simulation results prove the effectiveness of the proposed method.

Keywords Evaluation computing · Resources potential · SVM · Intelligent data · Internet of Things (IoT)

1 Introduction

Complex event processing technology captures and also transforms raw data into a dynamic continuous event stream for the real-time processing. It has become a key technology for real-time data analysis in many application fields such as logistics and manufacturing. In this situation, the current variety of the application software systems, including IoT applications, is evolving from a closed single system to an open system. The characteristics of this kind of development have also attracted the attention of the software industry. The research on various methods and also theories for this feature has also received extensive attention. Web services are the

J. Chen (✉)
Department of Tourism Management, Ningbo City College of Vocational Technology, Ningbo 315100, China
e-mail: juluanyan074@163.com

M. Lu
Department of Art and Design, Ningbo City College of Vocational Technology, Ningbo 315100, China

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_38

set of modular APIs that can be invoked over network to perform remote system request services, enabling inter-machine and platform interoperability and the data interaction between machines. In order to realize real-time control of IoT devices across networks and platforms, it is necessary to implement their Web services [1–3].

Academia is conducting active research on various issues in the management of IoT devices. A number of relevant standards organizations at home and abroad have started related work, including discussion and formulation of IoT equipment management standards [4–6]. As the sample, listed framework is recommended.

(1) CEP registers the event mode in the processing engine runtime and only performs a single pass scan of passed event stream within the specified time window to find the expected result which is also not achieved by the traditional request response query mode [7–9].
(2) CEP may regarding come from in the distributional many data pool simple event to carry on the synthesis inference, the recursion has new high rank level complex event and also may judge between event time order each kind of complex relations that can fully information content which contains using the atomic event carry on the semantic inference.
(3) The service of IoT is oriented to a large number of heterogeneous and distributed data sources and adopts EDA-SOA architecture to coordinate the business between various service components through events. CEP takes events as the query object and can handle the event flow in the EDA with characteristics of persistence, concurrency, uncertainty, and mixing of related and irrelevant information [10–13].

The framework has hierarchical relationships between the models. The event representation model is the component of the event flow model, while the time model is an important feature of the event representation model and the event flow model. The core detection model is responsible for calculating and processing the input event stream. The model and optimization model stipulate the content and the method of event pattern detection. Figure 1 gives the presentation [14].

To complete the efficient model, SVM should be integrated. The parallelization of SVM algorithm has made some progress, but still has room for improvement. For example, SMO algorithm after parallel optimization has a limited processing scale; while the latter level of cascade SVM algorithm consumes a lot of time, but the support vector is excluded as contribution is not very large, the MapReduce calculation model requires multiple accesses to disk during iterative processing, which affects the training speed. In order to effectively improve the classification performance of symbol data, a symbolic space representation method is defined by deepening the spatial structure relationship between different attribute values and labels of symbol data, combined with mutual information and conditional entropy information measurement methods. The method can not only retain the original information of the symbol data, but also reflect correlation between the attribute value and the label, and can also effectively measure the difference between the different attribute values [15–17]. Figure 2 gives the sample.

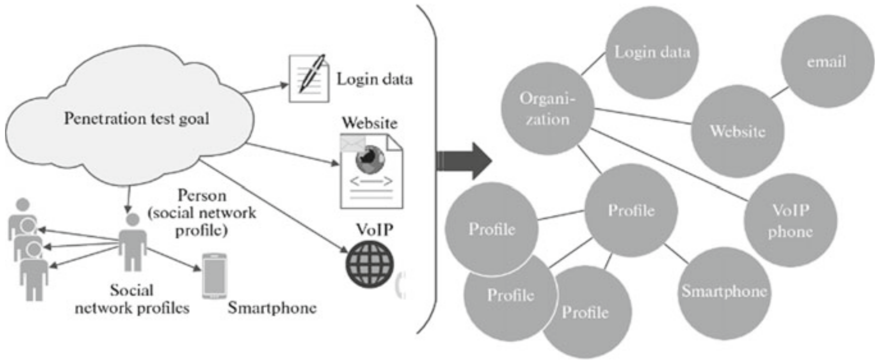


Fig. 1 IoT testing paradigm demonstration

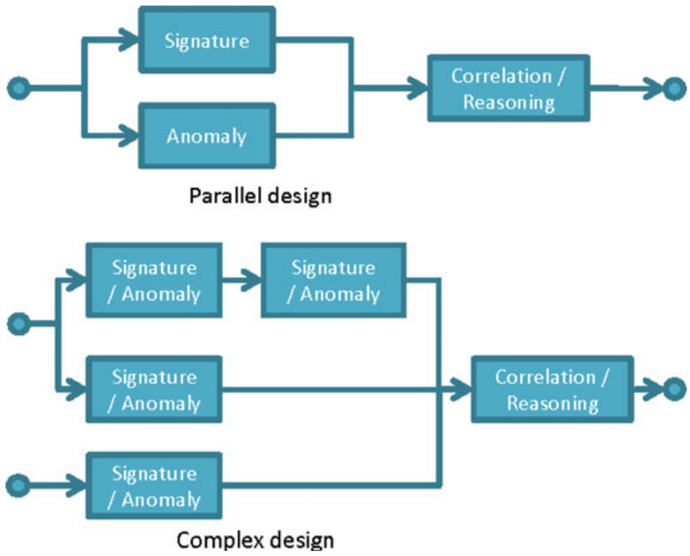


Fig. 2 SVM parallel model demonstration

Inspired by the mentioned methodologies, this paper proposes the evaluation computing of cultural tourism resources potential based on the SVM intelligent data analysis and IoT. In the following sections, proposed model is analyzed.

2 The Proposed Methodology

2.1 Resources Potential Model

Market potential refers to the largest period of market sales of ICT standards in a given period of time and under certain conditions. Due to the early development of the urban tourism market, the development of the urban tourism market has reached a certain level. According to the principle of the demand consumption, as people continue to consume, the consumption efficiency of products will continue to decline, so the development space of the tourism market will become smaller and smaller. The difficulty will be bigger and bigger. Compared with the urban tourism market, the rural tourism market is still in its infancy when various facilities are not perfect and tourism enterprises pay insufficient attention to the market. However, with the continuous improvement of the income of the rural residents, while the continuous improvement of people's lives, the continuous improvement of rural economic conditions and various facilities, the potential of the rural tourism market will be directly revealed and the blank. These gaps are market opportunities. The market potential of the BASS model is parabolic, and the market potential of the EPHP model is flat and also straight, while the actual number of users of the GSM technology standard is flat and linear, and is located in the EPHP model. Below the market potential line, the trends of the two are consistent. Based on this perspective, Fig. 3 gives the circle analysis [18–22].

Fig. 3 Resources potential circle demonstration

2.2 SVM Data Analytic Framework

The kernel method is an important technique for solving nonlinear problems in machine learning. It uses the kernel function to describe the inner product between the feature space vectors, avoiding the explicit expression of the nonlinear map, which not only ensures the generalization performance of the learner to avoid the dimensional disasters. In essence, the kernel can also calculate the similarity of inner product measurement data after mapping two samples to a high-dimensional space, but the commonly used linear kernel, the Gaussian kernel, and polynomial kernel cannot directly measure the similarity of interval type data.

The basic idea of SVM is to find a classified hyperplane so that the two types of samples in the training sample can be separated and as far away from the plane as possible. SVM has the algorithm simply, the anti-chirp nature strong, the study sample small, the efficiency high with the promotion good and so on the merits, can the traditional method big sample request restraint, the final policy-making function only not determine by the minority support vector, it may "the rejection" the massive redundant samples to then solve the high spectrum dimension disaster problem which speeds up the operation efficiency [23, 24].

In our proposed model, the ideas have the following novelties. (1) Improve the training structure and merging strategy of the Cascade SVM, and implement the algorithm by using the Spark parallel framework. (2) Some parallel operators and parameters are adjusted to optimize the performance of the new algorithm.

In general, for the Gaussian kernel function, it simply selects an optimal single kernel parameter. In fact, the dimensional features of the target feature vector play different roles in the recognition, that is, the importance of the features is different. The M-SVM algorithm uses the interval median to measure the similarity of the interval data, only considers the internal of the interval data, and loses the key information of the interval size; the IBV-SVM algorithm uses the left and right boundary values to measure the similarity of the interval data. Features, although considering the size of the interval, the internal distribution is not involved; the GIK-SVM algorithm combines the similarity between the interval median and the half-width to measure the similarity of interval data, and combines the information of these two aspects. Therefore, combined with the above time complexity, the GIK-SVM algorithm is relatively more comprehensive and feasible. Figure 4 gives the SVM data simulation results.

These precise nuclear parameters can reflect the nature of the corresponding features of the target, making the separability between different categories of the targets better. In addition, the selection of the different kernel parameters for each feature also produces additional information. Analysis of these different nuclear parameters is another way of character selection. This article unifies the genetic algorithm the superiority, proposed non-balanced data target identification SVM Gaussian nuclear function multi-parameter parallel optimization choice method, and will be able to reflect the comprehensively the sorter performance the F measure function will take the nuclear function multi-parameter optimization criterion.

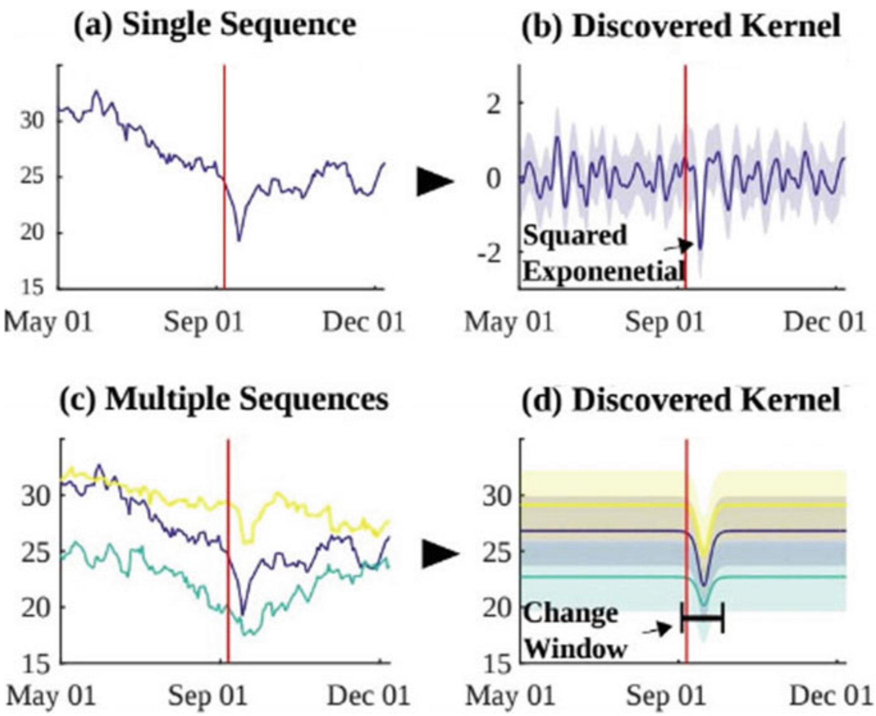


Fig. 4 SVM data simulation results

2.3 *IoT Framework for the Measurement*

The condition monitoring system is required to be versatile and also scalable. The Internet of Things is widely used in monitoring of complex equipment conditions in the fields of transportation, power generation, engineering machinery, aviation, etc. The types of equipment are diverse, the number is large, and the requirements for monitoring and analysis applications are complex. Logic control module took the software logic control center, through transfers in the factory class function in logic control module example first floor various modules control function class to use control function class transfer control function realization to equipment logical control. Logic control module simultaneously has the demonstration function, as during realization concrete operation also may the equipment real-time condition demonstration in the operation contact surface.

It should be noted that in the RFID system, since the background server and the reader usually have strong computing power and can also implement a relatively secure encryption algorithm, we assume that there is a core secure communication channel between the background servers, and the background server is the same. There is also a secure communication channel between the readers connected to it. However, due to cost constraints, tags generally have limited storage space and limited computing power, so we assume that there is an unsecure communication channel between the reader and the tag.

Accordingly, in Fig. 5, we denote pipeline. The computational complexity method judges the security of the protocol by evaluating the computational cost and success probability of the adversary. That is, as long as the time and resources required for a protocol to be compromised exceed capabilities of the adversary, the protocol is considered to be computationally safe. In the security analysis of the actual protocols, the proof method of reduction is often adopted to reduce the security proof of protocols to some recognized difficult problems that cannot be solved in probabilistic polynomial time, such as the difficulty in solving discrete logarithm, the difficulty in factorizing large prime numbers, the existence of the unidirectional functions, etc. Our model will be based on these regulations.

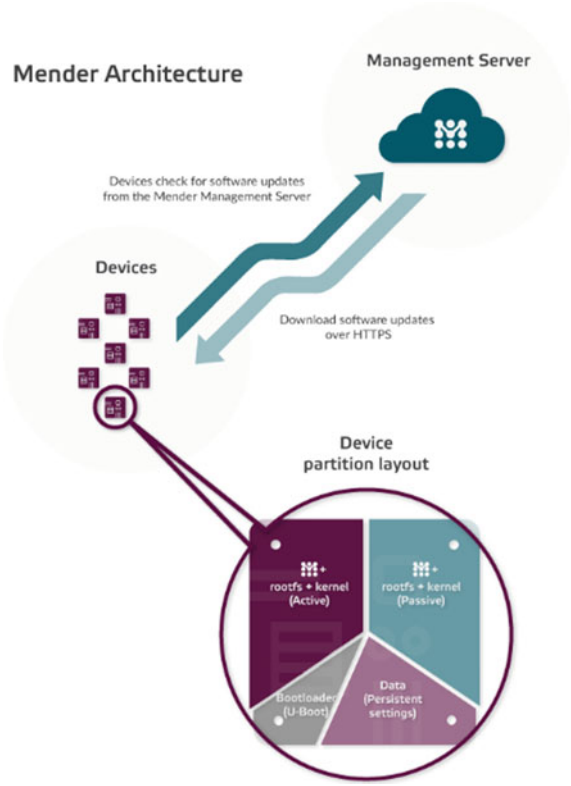


Fig. 5 IoT framework for measurement

2.4 Evaluation Computing of Cultural Tourism Resources Potential

The development of tourism resources is based on the investigation and research of tourism resources. The investigation and research of tourism resources involve the acquisition of resource information, the summary management of some related information and the preliminary analysis and mapping. For the further analysis, we should avoid the following aspects. (1) Wide regionalism is a major characteristic of the spatial distribution of tourism resources. This characteristic determines the coverage of tourism resources investigation and research must be wide, only in this way can we have a comprehensive and accurate understanding of the widely distributed tourism resources. Due to the restriction of human, material, and natural resources, the traditional methods of tourism resources investigation and research are often insufficient in coverage and low in accuracy. (2) Comprehensiveness is a prominent feature of tourism resources in terms of content. All kinds of tourism resources are not simple and independent elements in content, but a combination of the general mutual inclusion, mutual penetration, mutual complementation, and complementation. Dynamic is the outstanding feature of tourism resources in time. It is mainly reflected in the seasonal changes of tourism resources and their own life cycle changes. However, traditional survey and research methods of tourism resources cannot provide dynamic information of tourism resources, and the speed of information update is far behind the speed of resource changes.

The value of tourism resources and the conditions for tourism development are the main subjects of evaluation. Under the guidance of market demand, tourism resource evaluation pays more attention to the "attractive value" of resources to the market, and the attraction value is significantly affected by the development conditions such as general transportation location conditions and geographical environment. The improvement of development conditions prompts the viewing, culture, and section of tourism resources. The value of examinations and the rest periods has gradually become prominent and realized. At the same time, tourism development conditions have been continuously improved due to the exploration of the value of tourism resources. Therefore, listed components should be seen.

(1) Landscape combination. Involved in landscape space, type combinations, etc. If the landscape combination is not ideal, the scenery area will be monotonous; the landscape combination is good, and cultural, religious, artistic, and the ethnic customs landscapes of periods are cross-integrated, which can simultaneously reflect natural beauty and cultural rhyme.
(2) Ecological capacity. The nature and degree of the ecological impact of tourism activities are related to the ecological bearing capacity of tourism destinations. The evaluation of the natural ecological bearing capacity involves the bearing capacity, safety, and natural disaster risk of natural resources and environment on tourism development.
(3) The position and the market condition, the accessibility and the spatial distribution affect the resources attraction the important attribute, the neighboring

two place tourist resources form mutually generation of or the supplementary relations according to the different size that has the big market development potential from source of tourists near resources.

According to China's own situation, we should rationally develop the tourism resources according to local conditions, and strive for greater tourism space and a more suitable tourism environment. Traditionally, tourism resources are abundant, such as beautiful scenery, numerous historical sites, profound cultural symbols, unique architectural features, and typical folk customs. Throughout this class can be clearly seen development of tourism, natural, and cultural resources occupies an important share in market competition, and at the same time, infrastructure, reception conditions as well as the improvement of human environment, etc. have also played a role, but to make the tourism sustainable development bottom go to, we must put an end to the occurrence of excessive or unreasonable development; otherwise, it will bring irreparable damage tourism development. Green tourism is a term used to describe best environmental practices in the tourism sector, including topics such as operational efficiency, environmental management, waste, transportation, and social responsibility and biodiversity. By agreeing to a code of conduct for the green tourism enterprise program and an independent audit of their actions and companies are committed to reducing their impact on the environment. Based on the proposed model, the evaluation will be finalized.

3 Conclusion

Evaluation computing of the cultural tourism resources potential based on SVM intelligent data analysis and IoT is analyzed in this paper. The framework has hierarchical relationships between the models. The event representation model is the component of the event flow model, while the time model is an important feature of the event representation model and the event flow model. With the proper combination of the mentioned models, the finalized framework is done. The simulation results prove the effectiveness of the methodology. In the future, we will test the more scenarios to validate the robustness.

References

1. Huang, S., Cai, N., Pacheco, P.P., Narrandes, S., Wang, Y., Xu, W.: Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genom. Proteom. **15**(1), 41–51 (2018)
2. Plageras, A.P., Psannis, K.E., Stergiou, C., Wang, H., Gupta, B.B.: Efficient IoT-based sensor BIG data collection-processing and analysis in smart buildings. Future Gener. Comput. Syst. **82**, 349–357 (2018)
3. Agarap, A.F.M.: A neural network architecture combining gated recurrent unit (GRU) and support vector machine (SVM) for intrusion detection in network traffic data. In: Proceedings

of the 2018 10th International Conference on Machine Learning and Computing, pp. 26–30. ACM (2018)

4. Muhammad, K., Hamza, R., Ahmad, J., Lloret, J., Wang, H., Baik, S.W.: Secure surveillance framework for IoT systems using probabilistic image encryption. IEEE Trans. Ind. Inf. **14**(8), 3679–3689 (2018)
5. Aljarah, I., Ala'M, A.Z., Faris, H., Hassonah, M.A., Mirjalili, S., Saadeh, H.: Simultaneous feature selection and support vector machine optimization using the grasshopper optimization algorithm. Cogn. Comput. **10**(3), 478–495 (2018)
6. Misra, A., Vojinovic, Z., Ramakrishnan, B., Luijendijk, A., Ranasinghe, R.: Shallow water bathymetry mapping using Support Vector Machine (SVM) technique and multispectral imagery. Int. J. Remote Sens. **39**(13), 4431–4450 (2018)
7. Zhang, S., Wang, H., Huang, W., You, Z.: Plant diseased leaf segmentation and recognition by fusion of superpixel, *K*-means and PHOG. Optik **157**, 866–872 (2018)
8. Zareapoor, M., Shamsolmoali, P., Jain, D.K., Wang, H., Yang, J.: Kernelized support vector machine with deep learning: an efficient approach for extreme multiclass dataset. Pattern Recogn. Lett. **115**, 4–13 (2018)
9. Chen, W., Pourghasemi, H.R., Naghibi, S.A.: A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. Bull. Eng. Geol. Env. **77**(2), 647–664 (2018)
10. Anguraj, D.K., Smys, S.: Trust-based intrusion detection and clustering approach for wireless body area networks. Wireless Pers. Commun. **104**(1), 1–20 (2019)
11. Kumar, R.P., Smys, S.: A novel report on architecture, protocols and applications in Internet of Things (IoT). In: 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 1156–1161. IEEE (2018)
12. Sartakhti, J.S., Afrabandpey, H., Ghadiri, N.: Fuzzy least squares twin support vector machines. Eng. Appl. Artif. Intell. **85**, 402–409 (2019)
13. Al-Dabagh, M.Z.N., Alhabib, M.H.M., Al-Mukhtar, F.H.: Face recognition system based on kernel discriminant analysis, *K*-nearest neighbor and support vector machine. Int. J. Res. Eng. **5**(3), 335–338 (2018)
14. Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., Gupta, B.: Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. J. Comput. Sci. **27**, 386–393 (2018)
15. Dohare, A.K., Kumar, V., Kumar, R.: Detection of myocardial infarction in 12 lead ECG using support vector machine. Appl. Soft Comput. **64**, 138–147 (2018)
16. Eseye, A.T., Zhang, J., Zheng, D.: Short-term photovoltaic solar power forecasting using a hybrid wavelet-PSO-SVM model based on SCADA and meteorological information. Renew. Energy **118**, 357–367 (2018)
17. Singh, A., Chatterjee, K.: Cloud security issues and challenges: a survey. J. Netw. Comput. Appl. **79**, 88–115 (2017)
18. Jaramillo, F., Orchard, M., Muñoz, C., Antileo, C., Sáez, D., Espinoza, P.: On-line estimation of the aerobic phase length for partial nitrification processes in SBR based on features extraction and SVM classification. Chem. Eng. J. **331**, 114–123 (2018)
19. Gutiérrez, G., Ponce, J., Ochoa, A., Álvarez, M.: March. Analyzing students reviews of teacher performance using support vector machines by a proposed model. In: International Symposium on Intelligent Computing Systems, pp. 113–122. Springer, Cham (2018)
20. Liu, H., Cocea, M., Ding, W.: Multi-task learning for intelligent data processing in granular computing context. Granular Comput. **3**(3), 257–273 (2018)
21. Tabares, Z.E.M., Campos, A.C., Silva, E.V., Milanés, R.A.I.: Intelligent data analysis to calculate the operational reliability coefficient. In: International Workshop on Artificial Intelligence and Pattern Recognition, pp. 68–76. Springer, Cham (2018)
22. Zhu, S., Qiu, X., Yin, Y., Fang, M., Liu, X., Zhao, X. and Shi, Y.: Two-step-hybrid model based on data preprocessing and intelligent optimization algorithms (CS and GWO) for NO_2 and SO_2 forecasting. In: Atmospheric Pollution Research (2019)

23. Neuberger, A., Ahmed, Z., Dandekar, T.: March. IntelliEppi: intelligent reaction monitoring and holistic data management system for the molecular biology lab. In: Future of Information and Communication Conference, pp. 392–407. Springer, Cham (2019)
24. Zhou, G., Lv, M., Bao, T., Mao, L., Huang, K.: Design of intelligent carpooling program based on big data analysis and multi-information perception. Cluster Comput. **22**(1), 521–532 (2019)

Data Mining and Social Network Analysis on Twitter

Jesus Silva, Noel Varela, David Ovallos-Gazabon, Hugo Hernández Palma, Ana Cazallo-Antunez, Osman Redondo Bilbao, Nataly Orellano Llinás and Omar Bonerge Pineda Lezama

Abstract The emergence of a networked social structure in the last decade of twentieth century is accelerated by the evolution of information technologies and, in particular, the Internet has given rise to the full emergence of what has been called the Information Age [1] or the Information Society [2]. Social media is yet another example of people's extraordinary ability to generate, disseminate and exchange meanings in collective interpersonal communication with a massive, real-time networked system where everything tends to be connected. The analysis of the climate of opinion on Twitter is presented around the Common Core State Standards (CCSS), one of the most ambitious educational reforms of the last 50 years in USA.

J. Silva (✉)
Universidad Peruana de Ciencias Aplicadas, Lima, Peru

N. Varela
Universidad de la Costa (CUC), Calle 58 # 55-66, Atlántico, Barranquilla, Colombia
e-mail: nvarela2@cuc.edu.co

D. Ovallos-Gazabon · A. Cazallo-Antunez
Universidad Simón Bolívar, Barranquilla, Colombia
e-mail: david.ovallos@unisimonbolivar.edu.co

A. Cazallo-Antunez
e-mail: ana.cazallo@unisimonbolivar.edu.co

H. H. Palma
Corporación Universitaria Latinoamericana, Barranquilla, Colombia
e-mail: hhernandez@ul.edu.co

O. R. Bilbao
Corporación Politécnico de la Costa Atlántico, Barranquilla, Colombia

N. O. Llinás
Corporación Universitaria Minuto de Dios—UNIMINUTO, Barranquilla, Colombia
e-mail: nataly.orellano@uniminuto.edu

O. B. Pineda Lezama
Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras
e-mail: omarpineda@unitec.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_39

Keywords Social media mining · Social media · Twitter · Social network analysis · SNA · Common core state standards

1 Introduction

Twitter is a microblogging social network that is widely used today. Collecting and protecting data has been a constant in order to recover, analyze and generate information and knowledge. Social media present the same need with a significant difference [3, 4]: the speed at which new data is produced, which the sociologist Paul Virilio anticipated when stated that "real time prevails over real space and the geosphere [5]. The supremacy of real time and immediacy, over space and surface is a fait accompli and has an inaugural value (heralds a new era)" [6], so that "the emergence of hyper-immediate social media highlights the need for new forms of real-time research" [7].

Social Media Mining (SMM) [8] can be defined as the process of extracting, storing, representing, visualizing and analyzing user-generated data with the aim of discovering significant patterns (dissemination of information or rumors, influence, homophilia, social or consumer behavior, prediction, etc.) from social interactions in Internet social media. To deepen the understanding, real and potential, both of the patterns of interactions and of the identification of those nodes with a disproportionate influence or superhubs within the studied network and their effects have become an increasingly significant issue for more fields of knowledge [9].

The objective of the research is to propose the identification of superhubs in large networks extracted from Twitter in relation to their most prominent explicit relationships.

2 Data

For presenting a new method linking social network analysis (SNA), SMM and Twitter, it was necessary to choose an object of study that fulfilled two conditions: (1) to be as significant as possible in its social impact, without necessarily having a global scale and (2) the climate of opinion must keep active over time [10, 11].

The fieldwork consisted of tracking, capturing, storing, representing and analyzing the information generated on Twitter about US President Donald Trump.

3 Method

The capture and extraction of the data presented in this research were carried out directly from the application programming interface (API) of Twitter during six

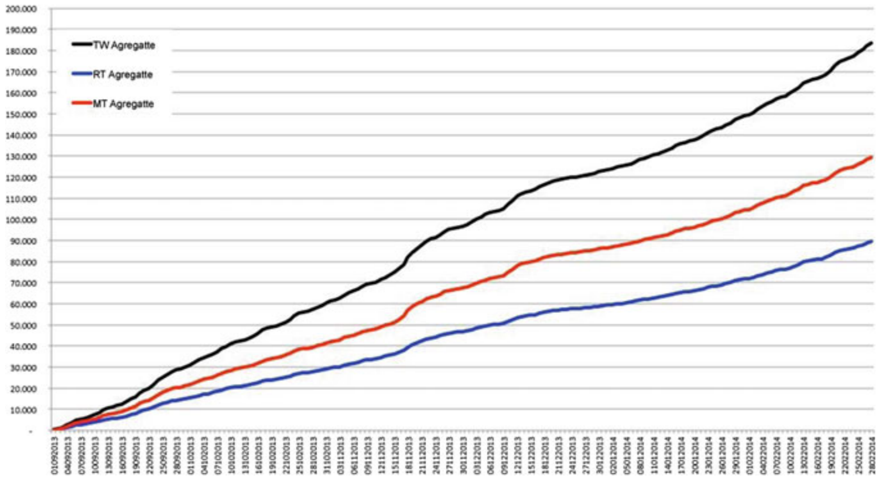


Fig. 1 Cumulative evolution of tweets, MTs and RTs

months, from September 2, 2018 to February 28, 2019. After defining the study object by means of the keyword, with or without hashtag, “DonaldTrump” or “Trump”, 250,236 tweets were captured, of which 98,258 were identified as retweets (RTs) and 151,978 mentions or answers (MTs). See Fig. 1 for details.

The .gexf format was chosen to generate the complete network. Then, the file was imported and exploited with two types of ARS software (first in Gephi11 and then exporting it to .dl files for analysis with UCINET12 of the 1 and 0.25% networks) for representation, visualization and analytical exploitation. Finally, nicknames or user names were anonymized with a unique ID to guarantee anonymity.

4 Results

With the data from the complete network, the modularity, in-degree and out-degree metrics were calculated and filtered until the identification of the network nodes of 1 and 0.25% by in-degree and out-degree [12, 13].

- The resulting Complete Network is an addressed network of 85,324 nodes and 154,258 relationships that were filtered to obtain the subsequent networks.
- The giant component (GC) with 54,235 nodes and 125,365 relationships gives an idea of the significant relational density within the climate of opinion.
- Application of the modularity metrics to the GC (Fig. 2) which shows the existence, represented in colors, of three large structural communities.
- The 1% network by in-degree: 754 nodes; 4520 relationships (Fig. 3). Applying the key player procedure to this network the study obtains that 93.5% of the network

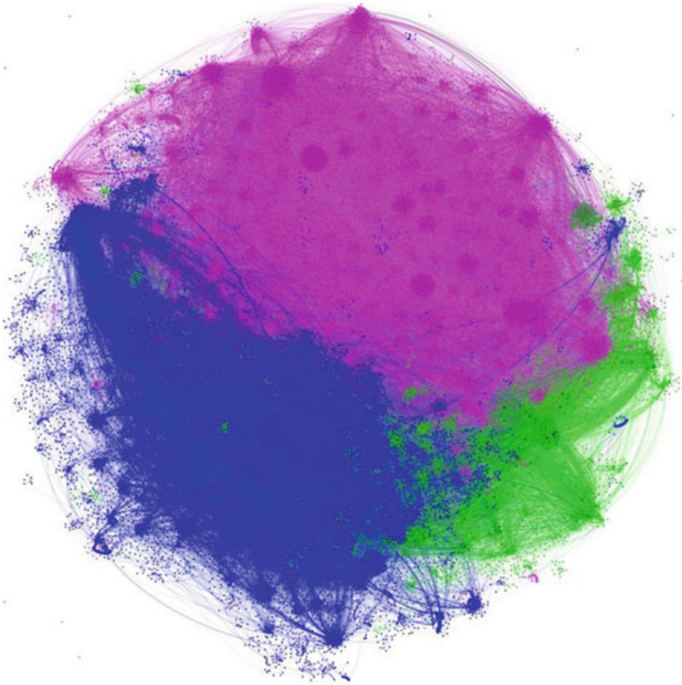


Fig. 2 Modularity of the giant component

would be reached with only 18 superhubs nodes (3.59%) which is a very small fraction to achieve site percolation [14].

- Network of 0.25% per in-degree: 251 nodes; 1023 relationships (0.7% of the total). Starting from the hypothesis that vaccinating/infecting a small fraction of nodes—in this case, superhubs identified by third parties as influential—and their relationships in a network, site percolation can have very significant effects on it when it comes to slowing down/expanding information, ideas or perceptions and provoking a knock-on or chain effect in the network in which the benefit of slowing down/expanding this reduced number of nodes will provoke significant changes not only in the structure of the whole network but also in the behavior of the individuals that form it. Applying the key player procedure to this network the results show that 96.3% of the network would be reached with only 9 superhubs nodes to get the site percolation.
- Network of 1% per out-degree: 789 nodes; 4250 relationships. Applying the key player procedure to this network, the results show that 95.3% of the network would be reached with only 19 superhubs nodes, which is a very small fraction to get the site percolation.
- Network of 0.25% per out-degree: 250 nodes; 920 relationships. Starting from the same hypothesis that in the network of 0.32% by in-degree and applying the

Fig. 3 Network of 1% per in-degree

procedure of the key player to this network, the results show that 94.2% of the network would be reached with only 5 superhubs nodes to get the site percolation.

- Geolocation by in-degree. With those individuals who make public the information of their geolocation in the USA in their profile, a representation by their in-degree was performed (Fig. 4), which shows the activity of the two large structural communities identified throughout the USA and a certain balance in the distribution of influence of viral information and the debate about Donald Trump.
- Geolocation by out-degree. With those individuals who make public the information of their geolocation in the USA in their profile, the representation was obtained by their out-degree, which shows the main focuses of emission, with very few nodes exercising in practice as main broadcasters. It also reflects the geographical origin of the activity of the two large structural communities on the two coasts and in the center of the country on the debate about Donald Trump.

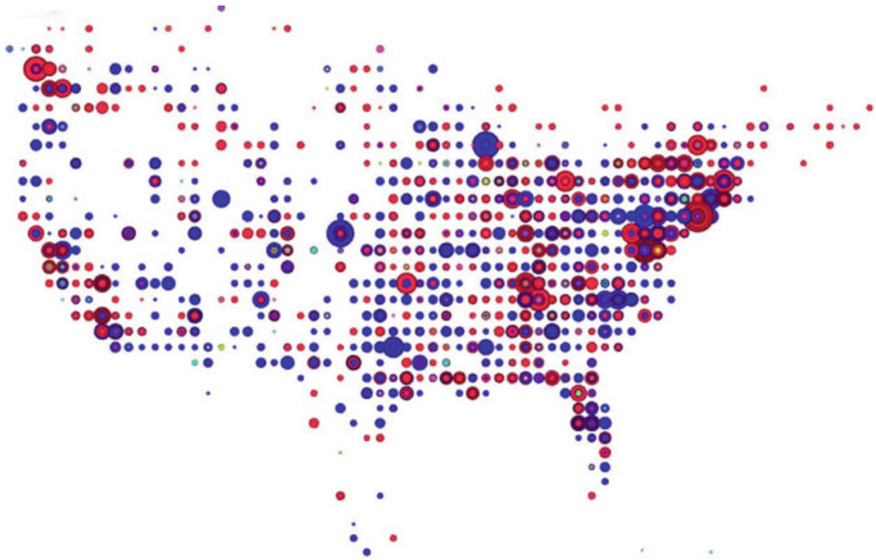


Fig. 4 Complete network. Geolocation by in-degree

5 Conclusions

The intersection between collective interpersonal communication on Twitter, SMM techniques and ARS presents four key characteristics for this study:

(1) A structural intuition of social relations is assumed, (2) relational empirical data are systematically captured, collected, represented and analyzed, (3) mathematical models are used for analysis along with technology and (4) visualizations of relations and patterns of interaction are created and shared, allowing the generation of meaningful structural ideas and their communication to others, which fully coincides with [14] on the development of ARS as a social discipline.

The objective to study networks in this research is the identification of the main actors or superhubs in the social, cultural and political debate about Donald Trump on Twitter, from his most significant relationships, using those data generated by the user's communication. Twitter offers, as a singular feature, the fact that it is playing, in practice, the function of intersecting medium of the rest of the media. This is a sort of spinal column or central nervous system through which the contents of the collective interpersonal communication, facilitated by the Internet architecture, can be identified, captured, analyzed and represented.

There is evidence that shows clear connections between the contagion/diffusion/development of infectious diseases and the diffusion of information, since both are propagated from person to person through networks, of influence or homophilia, which shows a great structural similarity [15]. This feature has led to the diffusion of ideas being conceptualized as social contagion [16] and is applicable to Internet social media and, notably, to Twitter.

This type of data presents new opportunities and challenges to researchers, where the most interesting resides, not only in the amount of data, but also in what they can do with these large amounts of data that cannot be done with small amounts. At least two major challenges for researchers already exist:

(1) The challenge of complexity or how to capture and add multidimensional data in a consistent way, not very homogeneous, not very structured and massive that are produced endlessly at any time or place and that have heterogeneous and unstable sources (that can appear and disappear) keeping the search and identification of significant patterns as a central objective.
(2) The challenge of $N =$ everything or how to develop methodologies that allow to work with the totality of the data that is produced, that is, how to investigate with complete universes. If sampling is a technique developed for times of information scarcity, it is a reality that was collectively abandoned already. However, the era of Big Data or the Petabyte Era predicts "a world in which vast amounts of data and applied mathematics replace any other instrument" implying that "the volume of data will obviate the need for theory, and even scientific method" [17]. In the meantime, researchers must keep the focus on analysis processes and correct decision making by identifying significant patterns and being aware that "the implicit promise of Big Data is that the solution to information overload passes through greater amounts of data" [18].

An obvious limitation of this type of research would lie in the lack of access to the meanings that circulate through the networks. To this end, there are emerging methodologies such as netnography [19], a proposal between sociology and anthropology that responds to the need for interdisciplinary approaches to unite the study of structure and access to meanings.

References

1. Hahsler, M., Karpienko, R.: Visualizing association rules in hierarchical groups. J. Bus. Econ. **87**, 317–335 (2017)
2. Yuan, M., Ouyang, Y., Xiong, Z., Sheng, H.: Sentiment classification of web review using association rules. In: Ozok, A.A., Zaphiris P. (eds.) Online Communities and Social Computing. OCSC 2013. Lecture Notes in Computer Science, vol. 8029. Springer, Berlin, Heidelberg (2013)
3. Silverstein, C., Brin, S., Motwani, R., Ullman, J.: Scalable techniques for mining causal structures. Data Mining Knowl. Discov. **4**(2–3), 163–192 (2000)
4. Amelec, V., Carmen, V.: Relationship between variables of performance social and financial of microfinance institutions. Adv. Sci. Lett. **21**(6), 1931–1934 (2015)
5. Java, A., Finin, T., Song, X., Tseng, B.: Why we Twitter: understanding microblogging usage and communities. In: WebKDD/SNA-KDD'07 Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (2007)
6. Naaman, M., Boase, J., Lai, C.-H.: Is it really about me? Message content in social awareness streams. In: CSCW'10 Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (2010)

7. Zafarani, R., Ali Abbasi, M., Liu, H.: Social Media Mining. Cambridge University Press. http://dx.doi.org/10.1017/cbo9781139088510 (2014)
8. Morris, M.R., Teevan, J., Panovich, K.: What do people ask their social networks, and why? A survey study of status message Q&A behavior. In: CHI'10 Proceedings of the 28th International Conference on Human Factors in Computing Systems (2010)
9. Viloria, A., Lezama, O.B.P.: Improvements for determining the number of clusters in *k*-means for innovation databases in SMEs. Proc. Comput. Sci. **151**, 1201–1206 (2019)
10. Kamatkar, S.J., Kamble, A., Viloria, A., Hernández-Fernandez, L., Cali, E.G.: Database performance tuning and query optimization. In: International Conference on Data Mining and Big Data, pp. 3–11. Springer, Cham (2018)
11. Mayer-Schönberger, V., Cukier, K.: Big data: a revolution that will transform how we live, work, and think. Mariner Books. http://dx.doi.org/10.2501/ija-33-1-181-183 (2014)
12. Cagliero, L., Fiori, A.: Analyzing Twitter user behaviors and topic trends by exploiting dynamic rules. In: Behavior Computing: Modeling, Analysis, Mining and Decision, pp. 267–287. Springer (2012)
13. Erlandsson, F., Brodka, P., Borg, A., Johnson, H.: Finding influential users in social media using association rule learning. Entropy **18**, 164 (2016)
14. Meduru, M., Mahimkar, A., Subramanian, K., Padiya, P.Y., Gunjgur, P.N.: Opinion mining using twitter feeds for political analysis. Int. J. Comput. (IJC) **25**(1), 116–123 (2017)
15. Abascal-Mena, R., López-Ornelas, E., Zepeda-Hernández, J.S.: User generated content: an analysis of user behavior by mining political tweets. In: Ozok, A.A., Zaphiris, P. (eds.) Online Communities and Social Computing. OCSC 2013. Lecture Notes in Computer Science, vol. 8029. Springer, Berlin, Heidelberg
16. Dehkharghani, R., Mercan, H., Javeed, A., Saygin, Y.: Sentimental causal rule discovery from Twitter. Expert Syst. Appl. **41**(10), 4950–4958 (2014)
17. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005), pp. 363–370 (2005)
18. Viloria, A., et al.: Integration of data mining techniques to PostgreSQL database manager system. Proc. Comput. Sci. **155**, 575–580 (2019)
19. Guandong, X., Lin, L., (eds.). (2013). Social media mining and social network analysis; emerging research. Reference and Research Book News **28**(3). Portland: Ringgold Inc. http://dx.doi.org/10.4018/978-1-4666-2806-9

Numerical Modeling and Simulation of High-Efficiency Thin Cu(In,Ga)Se Photovoltaic by WxAMPS

M. J. Alam, Sheikh Shahparan Mahtab, A. A. Mamun, A. Monsur, Sabiha Sunjida Ahmed, Md. Israq Azız, Ashraful Islam and Mahedi Hasan

Abstract CIGS (C = Copper, I = Indium, G = Gallium, S = Selenium) or CIS is a thin-film photovoltaic which is used for converting light into electricity. It is produced by deposition on glass or on plastic body, and there is two electrodes act as back and front contact, whose function is to collect charges. It has a high absorption co-efficiency, and for that, it can easily absorb sunlight [1]. A I–III–VI alloying semiconductor compound Cu(In,Ga)Se better known by CIGS is a composition made by copper, indium, gallium, and selenium. It is one of the most promising ingredients for CIGS solar cells in the photovoltaic industry which is better known as thin-film technology. A zinc oxide: aluminum (ZnO:Al)/zinc oxide (ZnO)/cadmium sulfide (CdS)/Cu(In,Ga)Se (CIGS)/molybdenum (Mo)/substrate-based photovoltaic has

M. J. Alam · S. S. Mahtab (✉) · A. Monsur · A. Islam
Department of EEE, Feni University, Trunk Road, Feni 3900, Bangladesh
e-mail: mahtabshahzad@gmail.com

M. J. Alam
e-mail: alameee1993@gmail.com

A. Monsur
e-mail: Monsurdpt23@gmail.com

A. Islam
e-mail: ashrafulislam@gmail.com

A. A. Mamun
PWD, Dhaka, Bangladesh
e-mail: al_mamun_cuet@yahoo.com

S. S. Ahmed
Department of CSE, Feni University, Trunk Road, Feni 3900, Bangladesh
e-mail: ahmedsunjida@gmail.com

Md. I. Azız
Department of EEE, Dhaka University, Dhaka, Bangladesh
e-mail: israqaziz26@yahoo.com

M. Hasan
Department of BASc (Marine Engineering), University of Tasmania, Hobart, TAS, Australia
e-mail: mahedih@utas.edu.au

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_40

been designed in this paper, and besides an analytical numerical simulation is also performed by WxAMPS. 18.7% power conversion efficiency has been found under 1.5 spectrum at the formation of ZnO:Al/zinc oxide/CdS/CIGS/molybdenum/substrate where the applied temperature is 300 K. In the apparatus, a quasi-electrical field conducted on the part of the back contact is persuaded through the absorber, Ec, Ev exacerbated proximate to the back contact, in this manner the energy gap remains constant throughout the depth. For this reason, the conversion power efficiency is higher in CIGS thin-film solar cell. Moreover, thermal stability has been inspected in thin-film CIGS solar cell. These results will influence the future experimental work in technological optimization of CIGS solar cell.

Keywords CIGS · WxAMPS · Solar cell · Photovoltaic · Power conversion efficiency · Quasi-Ohmic contact

1 Introduction

With the development of civilization, the energy demand of the world is increasing day by day continually. Researchers have been looking for low-cost high-efficient novel materials in comparing to other existing materials. The Cu(In,Ga)Se_2 (CIGS) quaternary chalcopyrite semiconductor debasing is one of the most promising materials in photovoltaic solar cell which produces high efficiency. On the other hand, the use of CIGS instead of silicon technology has been enhanced due to its good materials properties and low cost. Furthermore, the availability of fossil fuels has been greatly depleted. The rest of fossil fuels reserve will go for a maximum of 100 years. In addition, fossil fuels have an adverse effect on the environment due to its large emission of carbon dioxide, which is the prime cause of global warming. To save the environment, no other source is better than solar power. Chapin, Fuller, and Pearson are the persons who first devised the photovoltaic structure, and they got 6% efficiency [1].

After that, due to hard work of researchers, the efficiency gets improved and many kinds of thin-film solar cell have been discovered as like organic inorganic, CIGS, perovskite, CdT, tandem solar cell, etc. In the beginning, photovoltaic solar cell had low efficiency which was limited within 4–6%. But the low efficiency did not last for long. With the hard work of researchers, solar cell's efficiency improved and eventually gained popularity because of green energy, low cost, and clean source of energy. The another advantage of CIGS semiconductor is the ability of band gap tuning to absorb different length of photon but Gallium employment which varies from 1.02 to 1.68 eV. Different researchers have found excellent conversion efficiency, over 20%. In crystalline silicon cells, almost 25% efficiency is illustrated [2]. Researcher is committed to overcome the basic properties and challenges which keep down the electrical parameters. In several papers, it is narrated that CIGS solar cell is presenting literal inhomogeneities on micrometer scale, so the main parameter is affected. This effect brings some complexity on compulsory parameter on quasi-Fermi (QF) level [3]. Alleviating or eliminating these complexities is more important. Finally, it should be referred that complex structure materials can give highly interesting and expecting results. Originally, CIGS comprises materials with alloying copper indium gallium selenium. There might be anode and cathode, where ZnO is traditionally used as anode. Cadmium sulfide acts as buffer layer. The layers can be substrated on soda–lime glass. Molybdenum or aluminum can be used as back contact.

2 Simulation and Modeling

2.1 Simulations Details

Copper indium gallium selenide (CIGS) photovoltaics' main component is its absorber layer, which is a semiconductor material (an alloy) of group 1-3-6. It is very much possible to tune its band gap by the equation given below [4] and theoretically it ranges from 1.01 to 1.67 eV which can be obtained by this equation:

$$\mathrm{Eg} = 1.011 + 0.421x + 0.244x2$$

Ohmic metal to semiconductor surface contact (SC) and a contact which is rectifying can be obtained with this to formula

$$\Phi m > \mathrm{Eg} + \chi$$
$$\Phi m < \mathrm{Eg} + \chi$$

From the SC interface, majority hole carriers sees Φ_B which is used as barrier height, as they travel from the semiconductor to metallic contact. But it is not seen in SC interface. The height of contact barrier for majority carrier holes can be obtained with this formula:

$$\Phi B = \mathrm{Eg} + \chi - \Phi m$$

From the study, it has been found that CIGS needs a large number of various layers. Therefore, the numerical solution cannot be achieved directly by using theoretical knowledge. It is needed to perform this analysis via simulator. With this, it is very much possible to find out relationship between performance parameters which may have effect on performance or behavior varies in different state.

For this study, WxAMPS [5] was used which was originated by S. Fonash and improved by A. Rockett at the Illinois University. It is one of the most used and reliable simulators in photovoltaics field. Here, it is very easy to find different characteristics data and graph, like *I–V* curve, Voc, Jsc, FF, efficiency, band diagram, G&Rn, electric field, QE etc.

2.2 Reference Cell

From the basis of M. Gloecler theorem [6], it can be said that simulation which is numerical subject can be performed with input parameters and using partial knowledge obtained from theoretical knowledge. The reference cell of CIGS solar cell can be seen from Fig. 1.

The reference cell is widely used in the CIGS photovoltaics world which is composed of aluminum-zinc oxide and acts as transparent conductive oxide. n-doped zinc oxide was used as electron transport material, and CdS was used as buffer layer. After that, the absorber layer CIGS was employed, then molybdenum was employed as back contact, and finally soda–lime glass was used. The input parameters can be seen from Table 1 for this different layer (Table 2).

C. WxAMPS simulation process

Generally, WxAMPS is the updated version of AMPS for better performance. In this system, firstly it looks like following (Fig. 2).

Here, in ambient part, we have to give the general parameters needed. The ambient window is showed in Fig. 3. We have set the input parameters like temperature (we can check our model in different temperatures). Then, the light spectrum file needed to be on.

We have taken 1.5 light spectrum for this work. Then, the quantum efficiency file for CIGS has been loaded. Quantum efficiency calculates how many photons

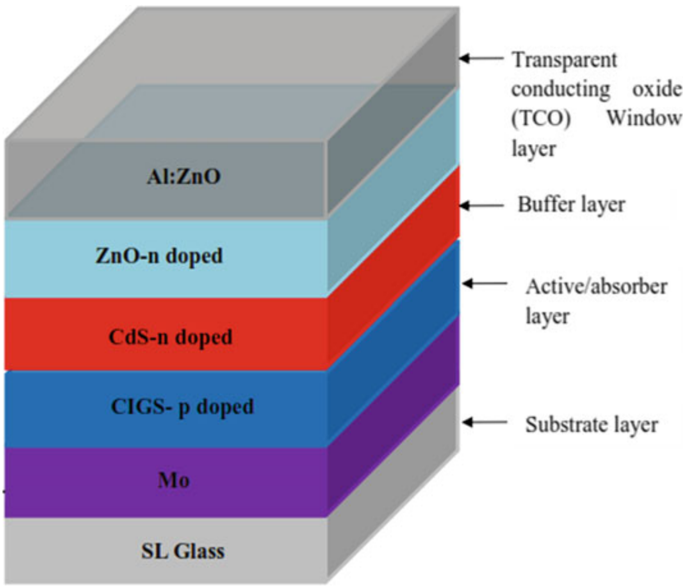


Fig. 1 Reference cell of CIGS photovoltaics

Table 1 Input parameters of different layer

Layer properties	ZnO–Al	ZnO	CdS	CIGS
Dielectric constant	9	9	10	13.6
Electron mobility (cm^2/Vs)	100	100	100	100
Hole mobility (cm^2/Vs)	25	25	25	25
Carrier density (cm^{-3})	10^{18}	10^{18}	10^{17}	2×10^6
Band gap (eV)	3.3	3.3	2.4	1.12
Effective density (cm^{-3}) N_c	2.2×10^{18}	2.2×10^{18}	2.2×10^{18}	2.2×10^{18}
Effective density (cm^{-3}) N_v	1.8×10^{19}	1.8×10^{19}	1.8×10^{19}	1.8×10^{19}
Electron affinity (eV)	4.0	4.0	3.8	4.1
Thickness (nm)	200	200	50	3000

Table 2 Simulation data Gaussian defect (GD) for the reference cell

Parameter	ZnO:Al	ZnO	CdS	CIGS
Defect type	Donor	Donor	Acceptor	Donor
Energy level (ev)	1.65	1.65	1.2	0.6
Deviation (eV)	0.1	0.1	0.1	0.1
σ_n	1×10^{-12}	1×10^{-12}	1×10^{-17}	1×10^{-13}
σ_p	1×10^{-15}	1×10^{-15}	1×10^{-15}	1×10^{-15}
N_t	1×10^{17}	1×10^{17}	1×10^{18}	1×10^{14}

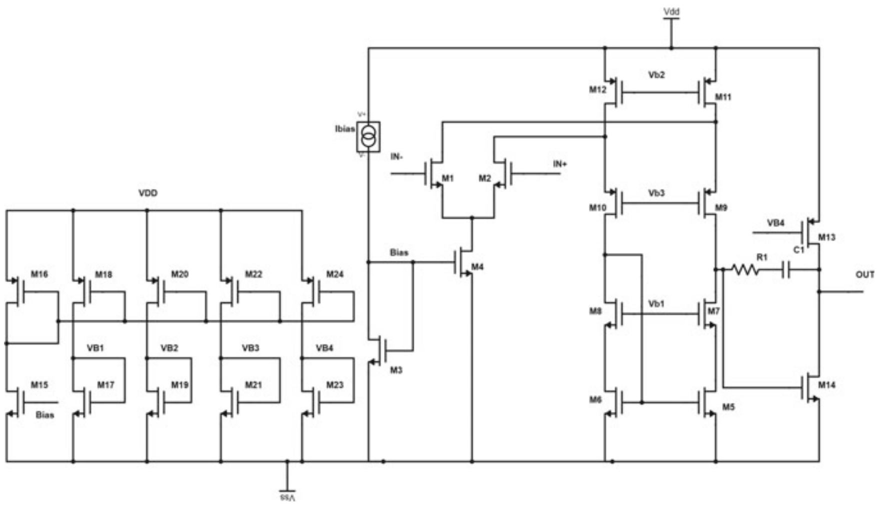


Fig. 2 Starting window of WxAMPS

Fig. 3 Ambient input parameters

absorbed by the absorber layer. Then, the front contact and back contact metals input should be inputted. Finally, after inputting bias voltage file, the work is done for this window (the values have been given in Table 3).

Now, we need to input the layers parameter showed in Table 1. The parameters organized as our reference cell (Fig. 1). In this window, the optical data and defect data were also inputted.

Table 3 Contact layers data used in the simulator

Parameter	Back contact	Front contact
Φ_B (eV)	Variable	0
S_n (cm/s)	2×10^7	1×10^7
S_p (cm/s)	2×10^7	1×10^7
Reflectivity	0	0.01

3 Results and Discussions

There may be a question of how the simulator simulates the reference cell. Here, before declaring the result, we will discuss about it.

Based on the data given in chapter Two, the simulation work was done with WxAMPS. The output characteristics have been shown below in subchapter. The solution algorithm of WXAMPS has been redesigned and modified to solve tunnel equation and that has extended the model for drift/diffusion model. Here, in this improved system, the rate of recombination is changed and differed from usual SRH (Shockly Read Hall) solution system model, and took it into a new dimension. The tunneling layer calculation is very important as it makes drastic change in after inputting the layers parameters where each of the cell layers has been separated and categorized in four different groups. And that is accessible by various tabs. After completion of the input, we can achieve the result from third window showed in Fig. 2.

The output data canbe seen at this time. From there, we can get the images or CSV file needed for calculation is shown in Fig. 4 (Fig. 5).

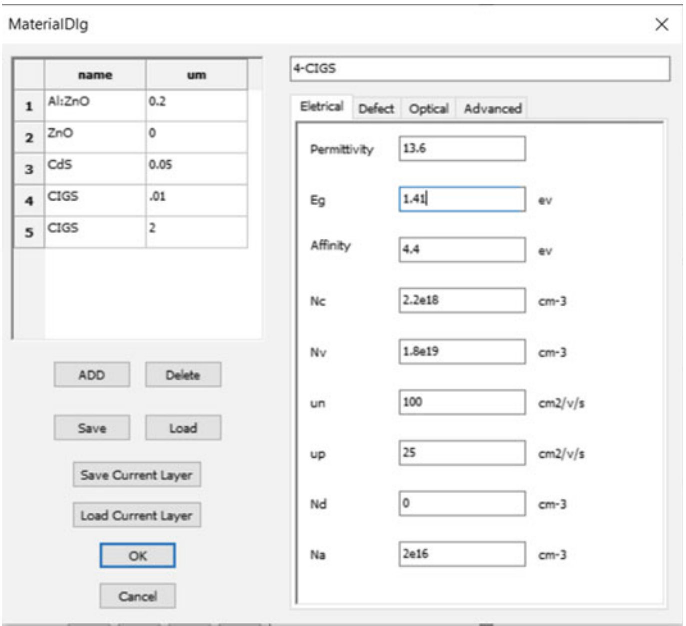


Fig. 4 Layers input parameters

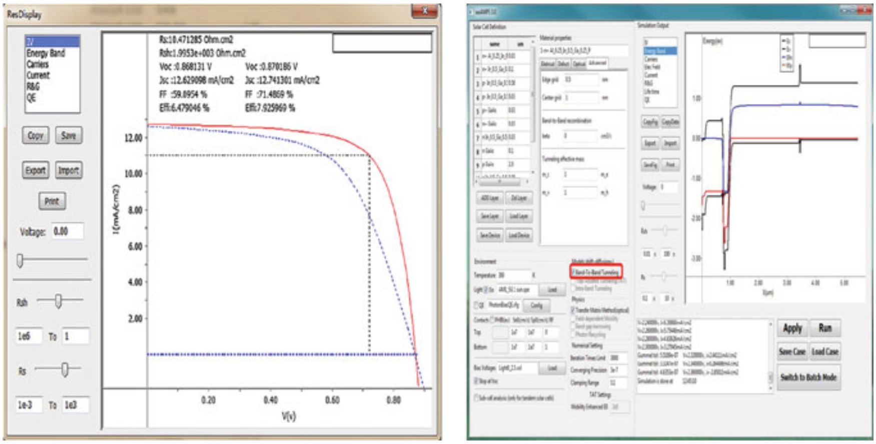


Fig. 5 Sample output of WxAMPS

3.1 I–V Characteristics of Single Junction CIGS Solar Cell at 300 K

I–*V* characteristics are the most important output figures from where we can calculate the efficiency of a solar cell. Right panel (Fig. 2) shows the active layer of the proposed structure. The values of SC current density Jsc and Voc have been found 36.68 mA/cm^2, and 0.64 V that produces fill factor (FF) is of 79.42%. Here, we can see that, in this cell, the value of fill factor is quite higher. The values of shunt resistance (Rsh) and series resistance are found to be 701.8 Ω-cm^2 and 0.03588 Ω-cm^2 respectively. The calculated value of cell efficiency (η) is found 18.7%. Here, Sc means short circuit current and V_{oc} means open circuit voltage (Figs. 6 and 7).

Fig. 6 *I*–*V* characteristics of the CIGS solar cell

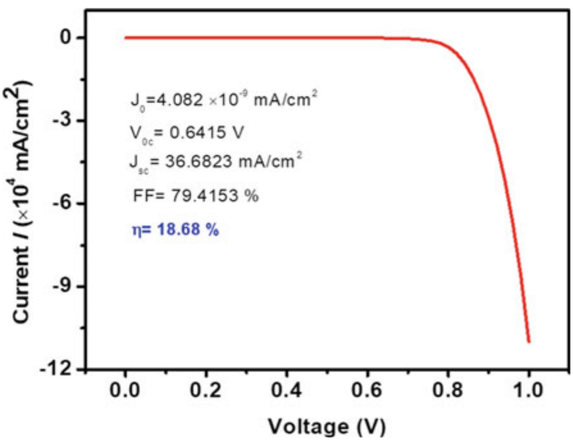
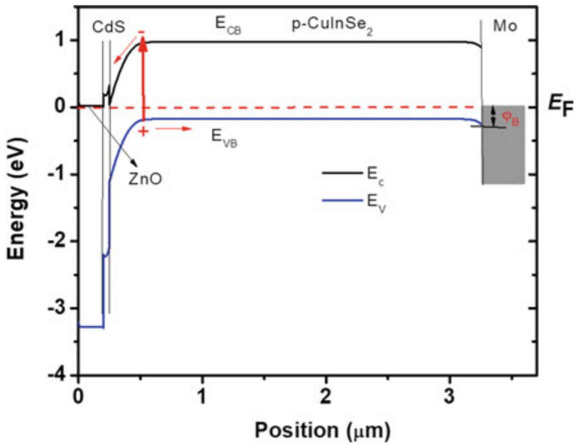


Fig. 7 Band diagram of CIGS solar cell



The energy band diagram of zinc oxide: aluminum (ZnO:Al)/zinc oxide (ZnO)/cadmium sulfide (CdS)/Cu(In,Ga)Se (CIGS)/molybdenum (Mo)/substrate at 300 K under 1.5 AM spectrum has been illustrated. Here, we have found that CIGS photovoltaic form a recombination in back surface. Because there created a spike in conduction band (CB) at 0.2 um which can be observed from Fig. 8. This spike also create quasi-ohmic contact (QOC) is bended in the valance band near the BC region. This recombination at back surface which is low makes the energy conversion more efficient. And QOC also helps this energy conversion more effective for the simulated structure of CIGS photovoltaic.

Free electron and hole concentration in the cell structure are represented in Fig. 4. It is noteworthy that free electrons are available only in the window layer ZnO

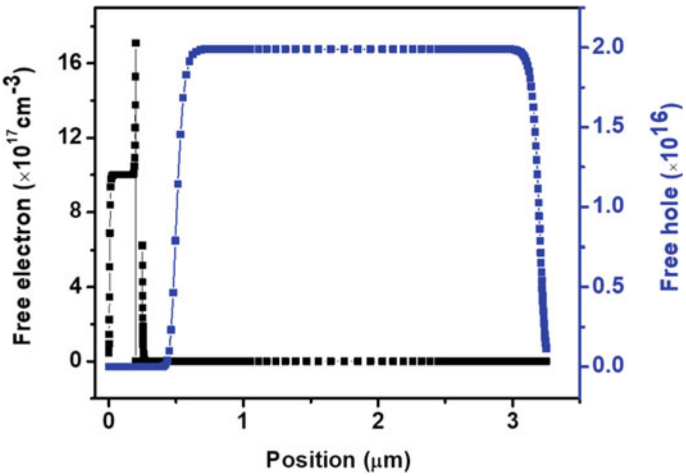


Fig. 8 Free electron and hole

and buffer layer CdS region, however free holes exist in the p-type $CuInSe_2$ region (Fig. 9).

It is evident that the recombination between electron and hole is negligible throughout the device that consequently raises efficiency of the cell.

Charge acceptor like and charge donor like in the device have been calculated and illustrated. Charge donors like carriers in the device are much higher and maintain wide range almost constant up to end of the device. Therefore, more electrons accumulated in the conduction band and contributed to current flow in the device which results increased efficiency of the cell.

Calculated electric field at the CdS to CIGS junction is represented in Fig. 10. Electric field appears only in the CdS to CIGS junction which is due to the band of the

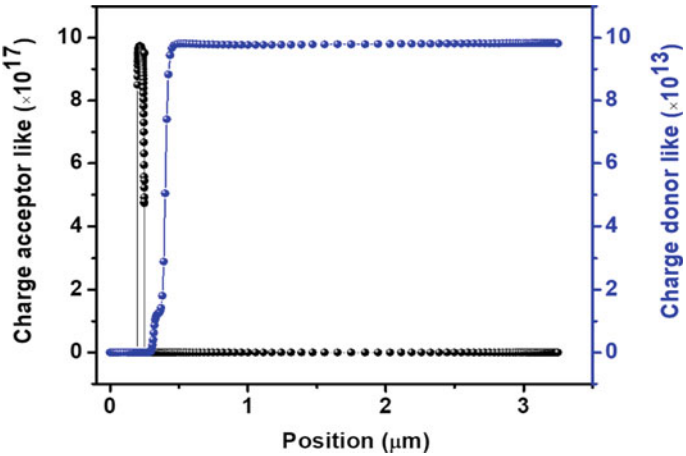
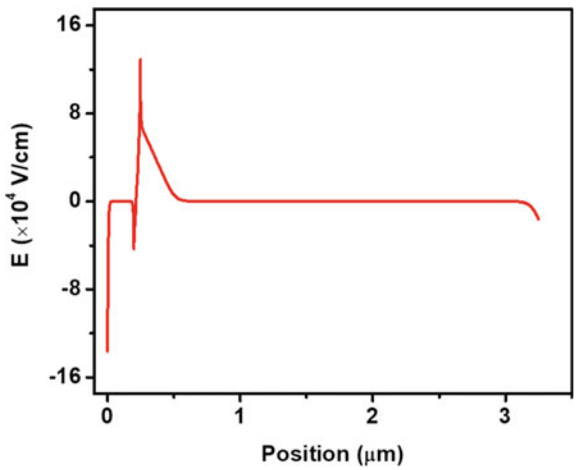


Fig. 9 Charge donor and acceptor generation in optimal solar cell

Fig. 10 Electric field of the CdS to CIGS interface in the proposed CIGS solar photovoltaic

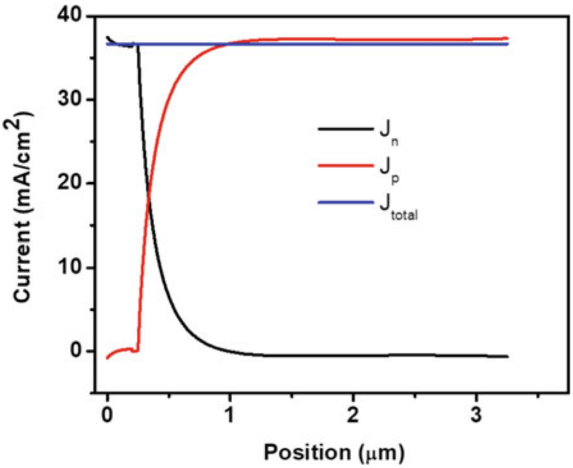


p-type $CuInSe_2$ semiconductor to align the Fermi level (EF). The electric filed does not exist in other region within the device which is good for carriers' movements to the electrodes and efficiency as well.

From Fig. 8 it can be observed that the J_n rapidly decreases with increasing thickness of the device while the J_p increases gradually and remain constant. The J_{total} is kept almost constant and always maintain a particular value (Fig. 11).

Because of the quasi-electrical (QF) field, the electrons which are minority current carriers have changed their direction just before the back contact (BC) in reference cell. So that the backsurface region which is absorber/back contact turned into dominant interjection (Fig. 12).

Fig. 11 Electron, hole and total current density of CIGS solar cell structure



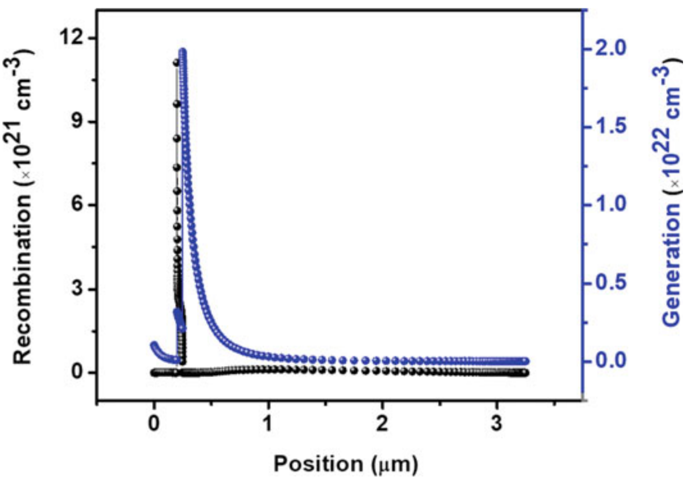
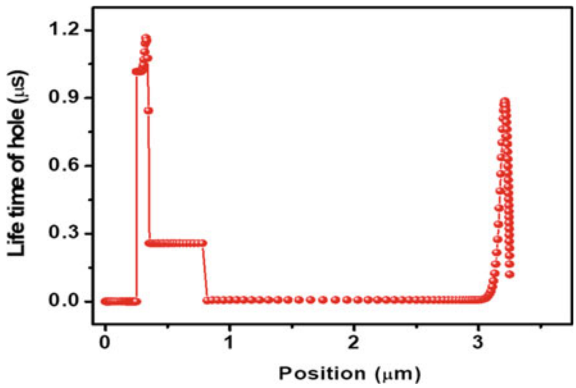


Fig. 12 R&G in the CIGS photovoltaic

Fig. 13 Lifetime of minority carriers in the CIGS solar cell structure



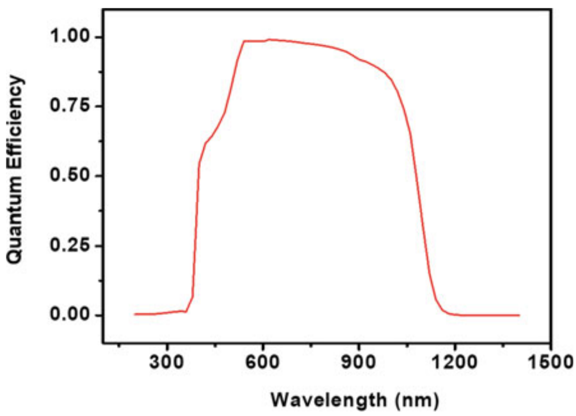
The carrier generation is one-order higher than that of recombination in a very narrow region, and on the other hand, generation is comparatively in wide region and in effective region of $CuInSe_2$ (Fig. 13).

The minority carriers' lifetime is shorter what is good for device efficiency. It is seen that lifetime of minority holes in the CdS is found to be 1.1 µs that is reasonable in the range and it is in very short region. The lifetime of minority carrier electrons in the p-$CuInSe_2$ region is found to be 0.25 µs and after 0.78 µm becomes zero.

The calculated quantum efficiency is shown in Fig. 14. It is found that front surface recombination and low diffusion length are dominant in our device. The maximum QE achieves with 99% at 548 nm in the device (Fig. 15).

Increasing temperature decreases the band gap (Eg) of the semiconductor, and as a result power conversion efficiency decreases. This phenomenon can be observed from Fig. 12 which shows high efficiency at low temperature but with the increasing temperature the efficiency falls down slowly in CIGS solar cell. At room temperature, the obtained efficiency was 18.56%.

Fig. 14 Quantum efficiency (QE) of reference cell



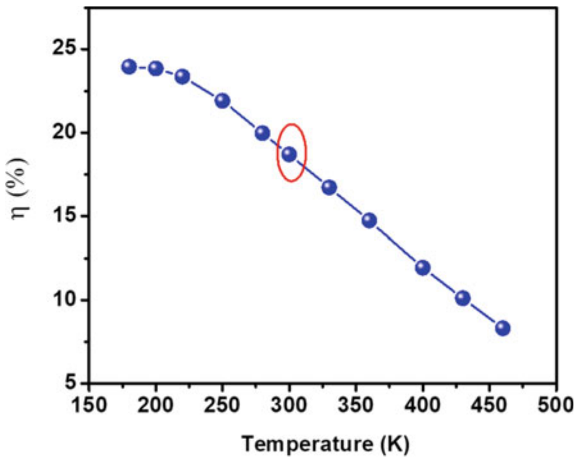


Fig. 15 Temperature effect on efficiency of CIGS cell

4 Conclusion

The design and performance analysis of CIGS solar cell have been investigated using analytical numerical simulator WxAMPS. The conversion efficiency of the cell ZnO: Al/ZnO/CdS/CIGS/Mo/substrate at 300 K under 1.5 AM spectrum has been obtained 18.7%. In the device, a quasi-electrical field conducted on the part of the back contact is persuaded through the absorber, Ec, Ev exacerbated proximate to the back contact. In this manner, the energy gap remains constant throughout the depth. For this reason, the power conversion efficiency is higher in CIGS thin-film solar cell. Moreover, the thermal stability has also been inspected in submitted thin-film CIGS solar cell. By the way, future experimental work for technological optimization will be influenced by these results.

References

1. Chapin, D.M., Fuller, C.S., Pearson, G.L.: A new silicon P–N junction potential for converting solar radiation into electrical power. J. Appl. Phys. **25**, 676 (1954)
2. Saji, V.S., Choi, I.-H., Lee, C.-W.: Progress in electrodeposited absorber layer for $CuIn_{(1-x)}Ga_xSe_2$ (CIGS) solar cells. Sol. Energy **85**(11), 2666–2678 (2011)
3. Solar Frontier Press Release: Solar frontier achieves world record thin-film solar cell efficiency: 22.3%, 8 Dec 2015
4. Spies, J.A.: Thesis on Inorganic Thin-Film Solar Cells. Oregon State University, 1 Nov 2007
5. Liu, Y., Heinzel, D., Rockett, A.: A new solar cell simulator: WxAMPS. In: Photovoltaic Specialists Conference (PVSC), 37th 2011
6. Gloeckler, Y.Y.M., Sites, J.R.: Potential of submicrometer thickness $Cu(In,Ga)Se_2$ solar cells. J. Appl. Phys. **98**. Article ID 103703 (2005)

Design of a Two-Stage Folded Cascode Amplifier Using SCL 180 nm CMOS Technology

Vanitha Soman and Sudhakar S. Mande

Abstract This paper presents design of a two-stage folded cascode amplifier with CMOS Technology. Maximum DC gain is the important required factor for analog and mixed signal circuits.The proposed circuit is designed to achieve more than 100 db and the obtained DC gain is 107.615 db. Phase margin is measured as 62.65°. The power is measured as 1.97 mW. The proposed circuit is developed and performed with the technology of SCL 180 nm cadence tool. The designed values are compared with the system performance.

Keywords Cascode · CMRR · PSRR · UGB

1 Introduction

Op-amp is an essential building block for ADC which is used in wireless and wireline communication system. This circuit is required to obtain the high DC gain with fully differential input and with the single-ended output. Two-stage folded cascode amplifier is presented.

Current mirror is the major block of op-amp and to improve the power consumption and gain Folded cascode op-amp is needed [1, 2]. Input common-mode voltage is compared with the reference voltage to get the right value [3] so common-mode feedback is used in many circuits. Low power with maximum gain is the requirement for ADC. Gain-boosted cascode op-amp is developed [4]. CFC architecture is used to develop the cascode amplifier [5].

V. Soman (✉)
Terna Engineering College, Navi Mumbai, India

S. S. Mande
Don Bosco Institute of Technology, Mumbai, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_41

This paper presents the design of a folded cascode topology using 180 nm technology. This paper is organized into four sections. Section 2 deals with the architecture and design specification of the circuit with estimated parameters. Section 3, discusses the design procedures and calculations of cascode amplifier, Sect. 4 shows the simulation results and its measurement. Section 5 concludes the paper.

2 Architecture of the Two-Stage Amplifier Circuit

2.1 Proposed Circuit of Folded Cascode Amplifier

Figure 1 shows the proposed architecture of the folded cascode amplifier used for ADC. Systematic design of cascode circuit for ADC in 180 nm technology with a 1.8 V supply voltage. The main modules of the proposed folded cascode amplifier circuits are biasing circuit and the two-stage folded cascode amplifier. These modules are designed and simulated.

2.2 Methodology

In this paper, the circuit is designed using SCL 180 nm CMOS technology, with a process parameter as shown in Table 2 and the simulations were performed using the Cadence software.

3 Folded Cascode Amplifier Design

Figure 2 shows the circuit diagram of the proposed folded cascode amplifier. Table 1 shows the input specification of the proposed cascode amplifier which needs the maximum DC gain of more than 100 db. Input common voltage is 0.9 V with the Vdd of 1.8 V. The required power consumption is 2 mW or less than that (Table 2).

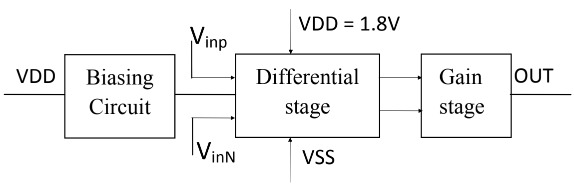


Fig. 1 Folded cascode amplifier circuit

Table 1 Design specification

Parameters	Value
DC gain	>100 db
Phase margin	60°
Vdd	1.8 V
I/P common-mode voltage	0.9 V
Unity gain frequency	90 MHz
Slew rate	200 v/μs
Output voltage swing	1 V p–p
Load capacitance	1 pF
Power consumption	2 mW
Input referred noise	$20\,\text{nV}/\sqrt{\text{Hz}}$

Table 2 Process parameter—180 nm technology

Parameters	Value
μ_n	$300\,\text{cm}^2/\text{Vs}$
μ_p	$70\,\text{cm}^2/\text{Vs}$
Cox	$8.78\,\text{fF}\,\mu\text{m}^{-2}$
Tox	4 nm
μnCox	$263\,\mu\text{A}/\text{V}^2$
μpCox	$62\,\mu\text{A}/\text{V}^2$
For L = 180 nm	V_{TP} = 400 mV
	V_{TN} = 480 mV
	$\lambda_n = 0.48\,\text{V}^{-1}$
	$\lambda_p = 0.328\,\text{V}^{-1}$

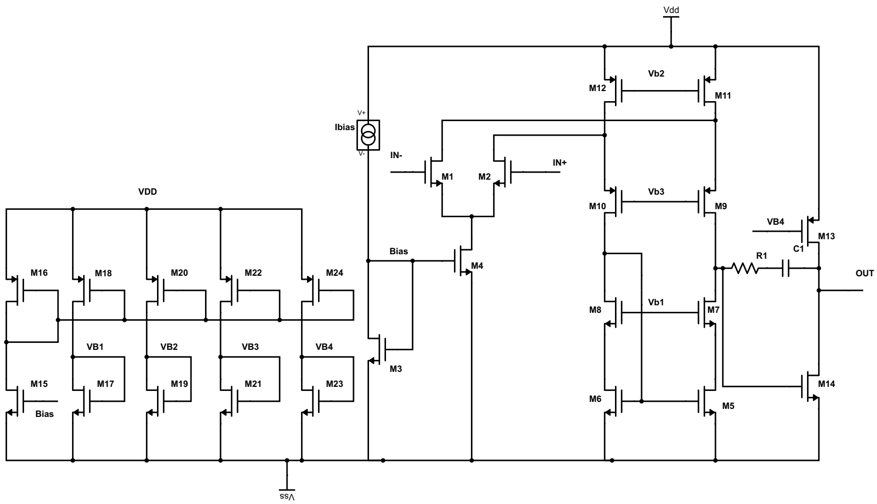


Fig. 2 Folded cascode amplifier circuit

The differential pair of NMOS receives the differential input signal. M3 will act as a current sink. Bias voltage is generated from the bias circuit. The circuit has two stage like first gain stage coupled with the output stage (second gain stage).The bias voltages like VB1, VB2, VB3, and VB4 are applied to gate of PMOS and NMOS pairs.

The circuit is modified with the current mirror connected between gate of M5, M6, and drain of M10. The drain saturation voltage of output stage M13 is equal to M14. The first gain stage M9 drain saturation voltage is equal to the M11.

3.1 Design Procedure

1. $\left(V_{ineq}\right)^2 = 2\frac{4KT\gamma}{gm1}\left[1 + \frac{gm11}{gm1} + \frac{gm5}{gm1}\right]$

   Let $\frac{gm11}{gm1}$ and $\frac{gm5}{gm1}$ be approximately 0.5 and designing for a lower noise value, say $10\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$, $gm1 = 441.6\,\mathrm{A}$

2. Unity gain frequency $\frac{gm11}{Cc2\pi}Cc = 0.78\,\mathrm{pF}$

3. Slewrate $= \frac{Iss}{Cc} = 200\,\mathrm{V/s}$, $Iss = 160\,\mathrm{A}$

   Now Idss $= \left[1 + \frac{C_L}{C_C} + I_{SS}\right] = 360\,\mathrm{A}$, $gm1 = \sqrt{\frac{2I_{SS}K_P W_1}{2L_1}}$ $but\ gm1 = gm2$

   $hence\ \left(\frac{W}{L}\right)_1 = \left(\frac{W}{L}\right)_2 = 4.634$

4. Output voltage swing $= V_{omax} - V_{omin}$, $V_{omax} = V_{DSAT14}$, $V_{omin} = V_{DD} - V_{DSAT13}$, $V_{OP-P} = V_{DD} - V_{DSAT13} - V_{DSAT14}$, $V_{DSAT13} = V_{DSAT14} = 0.4\,\mathrm{V}$

   $V_{DSAT13} = \sqrt{\frac{2I_{DSS}}{K_P\left(\frac{W}{L}\right)_{13}}} = 0.4\,\mathrm{V}$, $\left(\frac{W}{L}\right)_{13} = 72.58$

   $V_{DSAT14} = \sqrt{\frac{2I_{DSS}}{K_n\left(\frac{W}{L}\right)_{14}}} = 0.35\,\mathrm{V}$, $\left(\frac{W}{L}\right)_{14} = 22.338$

5. For sytematic zero offset error $V_{GS5} = V_{GS14}$ and

   $V_{DSAT5} + V_T = V_{DSAT14} + V_T$, $V_{DSAT14} = V_{DSAT5}$ hence $\left(\frac{W}{L}\right)_{5,6} = 4.96$

6. DC gain $A_V = A_{V1} * A_{V2}$,

$$A_{V1} = \frac{gm1}{\frac{gds8gds6}{gm8} + \frac{gds10(gds12+gds2)}{gm10}} \text{ and } A_{V2} = \frac{gm13}{gds13 + gds14}$$

$gds_{13} = \lambda_P, I_{DSS} = 21.25s, gds_{14} = \lambda_n, I_{DSS} = 31.04s,$ so $A_{V2} = 34.3784$

The first stage gain should be $A_{V1} = 2941.47$ so let $\lambda_P = \lambda_n$ for M_1, M_{12}

In this design $I_{11}, I_{12} = Iss$. We have $I_{1,2} = \frac{Iss}{2}$,

then we get $I_{5,6,7,8,9,10} = \frac{Iss}{2}$, $(gm)_{7,8} = 333.3S$ and $(gm)_{9,10} = 333.3S$.

By solving the first stage gain equation $\lambda_P = \lambda_n = 0.04356$

To get the lengths, $\lambda_1 L_1 = \lambda_2 L_2$ is used in this design.

$L_n = 1.98347$ m and $L_{1,2,5,6,7,8} = 1.9843$ m.

When $L_n = 1.355$ m and $L_{9,10,11,12} = 1.355\,\mu$m.

$V_{DSS} = 0.35,\ V_{GS5} = 0.83,\ V_{D7} = 0.83,\ V_{DS7} = 0.48,\ V_{S7} = 0.35,\ V_{DSAT7} = 0.48\,\text{V}$

from this values $\left(\frac{W}{L}\right)_{7,8} = 2.64$. $V_{DSAT9,11} = 0.48\,\text{V}$

using this $\left(\frac{W}{L}\right)_{9108} = 10.97$ and $\left(\frac{W}{L}\right)_{11,12} = 21.9419$

7. Input common-mode voltage $V_{incm} = V_{DSAT4} + V_{DSAT1} = 0.9\,\text{V}$

$V_{DSAT1} = \sqrt{\frac{2I_{ss}}{2K_n\left(\frac{W}{L}\right)_1}} = 0.19\,\text{V}$, $V_{DSAT4} = 0.29\,\text{V}$, $\left(\frac{W}{L}\right)_4 = 14.467$

in this design $I_3 = 16I_{bias} = \left(\frac{W}{L}\right)_3 = \frac{\left(\frac{W}{L}\right)_4}{16} = 0.904.$

See Table 3.

Bias Voltages

The calculated values for bias voltages are
For M_{11} and M_{12}, $V_{B2} = 0.915\,\text{V}$,
For M_9 and M_{10}, $V_{B3} = 0.45\,\text{V}$,
For M_7 and M_8, $V_{B1} = 0.96\,\text{V}$,
nd For M_{13}, $V_{B4} = 0.8\,\text{V}$.

Table 3 Aspect ratio

W/L	Value	L	W
$\left(\frac{W}{L}\right)$ 1, 2	4.634	9.1	1.98
$\left(\frac{W}{L}\right)$ 3	0.904	1.6	1.8
$\left(\frac{W}{L}\right)$ 4	14.467	26	1.8
$\left(\frac{W}{L}\right)$ 5, 6	4.96	9.8	1.98
$\left(\frac{W}{L}\right)$ 7,8	2.64	5.236	1.98
$\left(\frac{W}{L}\right)$ 9, 10	10.97	1.355	14.9
$\left(\frac{W}{L}\right)$ 11, 12	21.9419	1.355	29.7
$\left(\frac{W}{L}\right)$ 13	72.6	1	72.6
$\left(\frac{W}{L}\right)$ 14	22.348	1	22.348

4 Simulation results

See Tables 4 and 5 and Figs. 3, 4, and 5.

Table 4 Stability analysis

Cornner	SS		TT	FF	
Temperature (°C)	0	100	27	0	100
DC gain	116.235	−293.043	107.615	−180.41	−128.258
Unity gain frequency	117.59 MHz	NAN	146.05	NAN	NAN
Phase margin	49.79	NAN	62	NAN	NAN

Table 5 Simulation results

Parameters	Value
DC gain	107.615 dB
Unity gain frequency	146.05 MHz
Phase margin	62.65
Slew rate	241.81 V/µs
Output swing (Vpp)	1.65 V
PSRR plus	−63.611 dB
PSRR minus	−70.872 dB
ICMR (input common-mode range)	1.04 V
Power	1.979226 mW
RMS thermal noise voltage	15.18 nV/$\sqrt{\text{Hz}}$
CMRR	121.57

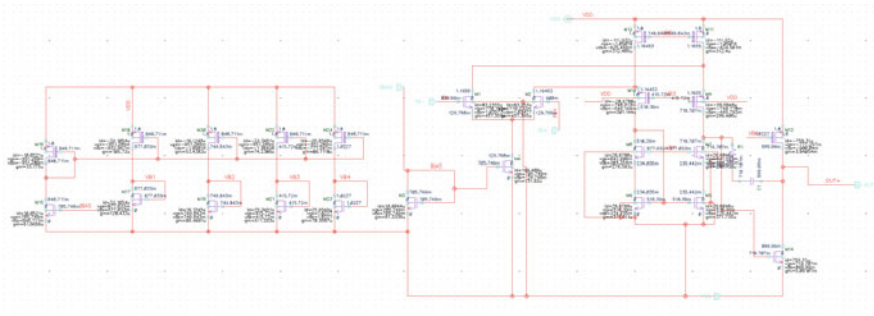


Fig. 3 DC operating point analysis

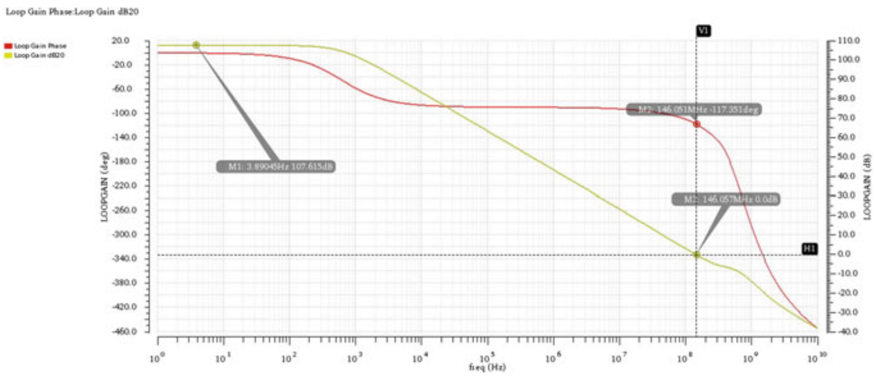


Fig. 4 STB analysis

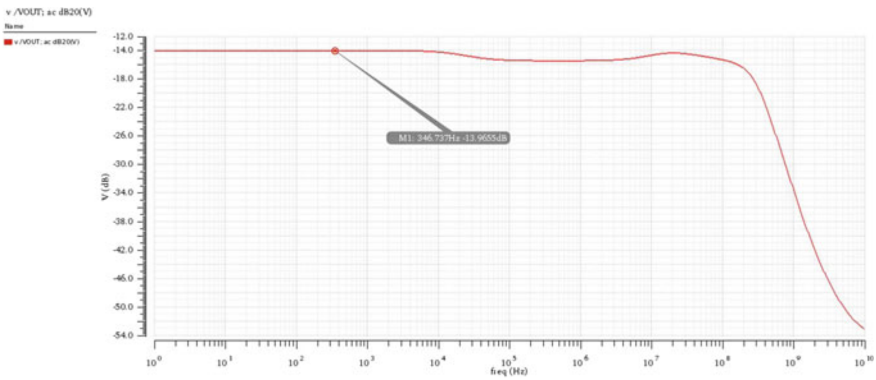


Fig. 5 CMRR

5 Conclusion

This paper presented the design of two-stage folded cascode amplifier. DC gain of 107 dB is achieved from the design parameters. The circuit was simulated and the performance of was tabulated. From the calculation of the transistor sizing, the two-stage amplifier has been designed and achieved for the given specification. The maximum DC gain is the bottleneck of ADC work. This achieved DC gain makes it suitable for low-power high-speed application. From the results, this circuit can be suitable for hybrid ADC which is the next phase of our research work.

References

1. Snathanalakshmai, M., Aha, A.: An improved folded cascode amplifier with low power current mirror circuits. Int. J. Pure Appl. Math. **118**(10), 45–50 (2018)
2. Rajni, E.: Design of high gain folded cascode operational amplifier using 1.25 um CMOS technology. Int. J. Sci. Eng. Res. **2**(11), 1–9 (2011)
3. Razavi, B.: Design of Analog Integrated Circuits. McGraw Hill, New York, NY (2001)
4. Arifuddin, M., Naaz, M.: A fully differential gain boosted folded cascode OTA with 100 dB gain. IJAREEIE **5** (2016)
5. Vallee, R.E., El-Masry, E.I.: A very high frequency CMOS complementary folded cascode amplifier. IEEE J. Solid State Circ. **29**(2), 130–133 (2004)

Electromagnetic Simulation of Optical Devices

Arvind Vishnubhatla

Abstract Remote sensing envisages the beaming of signals and the statistical processing of the inference. Traditionally, the probe signal is a radio wave, and a detector is used to anticipate the delay and direction of the reflected signal. The combination of optics (lasers, modulators, and switches) and electronics plays a key role in electro-optic devices. Analogous to electrons, the steady stream of photons in vacuity is called photonics. It is possible to construct passive devices that decompose and blend light in the optical domain. In this paper, we model the electromagnetic simulation of a set of optical devices through OptiFDTD.

Keywords Remote sensing · Optical · Signal · Electromagnetic · Simulation · OptiFDTD

1 Introduction

Remote sensing envisages the beaming and capturing of signals and the statistical processing of the inference. Traditionally, the probe signal is a radio wave, and a detector is used to anticipate the delay and direction of the reflected signal.

Once the radio waves impinge on an object, the changes which these radio waves undergo on reflection can give a clue about how far the waves traveled and the kind of obstacles encountered. The case scenario of research desires that remote sensing be achieved through light waves. The photons which constitute the optical signal are resistant to interference and parasitic coupling due to electromagnetic waves as opposed to electrical signals.

Classical optics resembles very closely to photonics, and here, lenses and mirrors are used to govern and modify electromagnetic waves. The application of photonics will refer to the evolution of signal processing devices, which use optoelectronics—a workable application of optics—where circuits made up of both electrical domain and optical domain will be made use of.

A. Vishnubhatla (✉)
Electronics and Communications Department, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_42

The combination of optics (lasers, modulators, and switches) and electronics plays a role in electro-optic devices. Analogous to electrons, the steady stream of photons in vacuity is called photonics. It is possible to construct passive devices which decompose and blend light in the optical domain. In this paper, we model the electromagnetic simulation of a set of optical devices through OptiFDTD.

2 Related Work

A large number of devices are fabricated after complex computer-aided engineering. Sensitivity and reliability based on optical properties are used to develop lab-on-a-chip [1]. Graphene oxide—a carbon-based material—is used to study optical intensity [2]. Optical interconnects using waveguides have been fabricated [3]. Things like microwave channelizers using a Fabry–Perot cavity are studied while fixing device requirements [4]. A perturbation analysis is used to determine the propagating modes in the waveguides [5]. 3D simulation of photonic components where optical signal is converted to electrical signal is studied [6]. Light propagation with nanometer resolution is analyzed through optical microscopy [7]. Conformal mapping is used to analyze the optical propagation characteristics [8]. Complex optical structures are investigated using beam propagation methods. However, this method has limitations regarding the characterization of structures that scatter radiation. The absorption coefficient is thus studied [9]. This involves the development of a CAD tool using the FDTD algorithm [10]. Substantial amount of theoretical and experimental work has been performed to describe the features of transmittable and scattering properties of optical fibers. A detailed description of quantitative application of multimode fibers is obtained by the solution of Helmholtz equation. A Fourier analysis reveals modal eigenvalues and strength of individual modes. A moving window-based Fourier transform in the axial distance shows the portion of the spectrum that is decayed [11].

3 Proposed Work

In physics, the **Poynting vector** represents the continuous change of guided energy density (the rate of energy transfer per unit area in units of watts per square meter (W m^{-2})) of an electromagnetic activity or interest.

Electromagnetic Wave Equation
The equation governing an electric field (EF) in the *x*-direction is as given below:

$$\vec{S} = \frac{1}{\mu_0} \vec{E} \ x \ \vec{B}$$

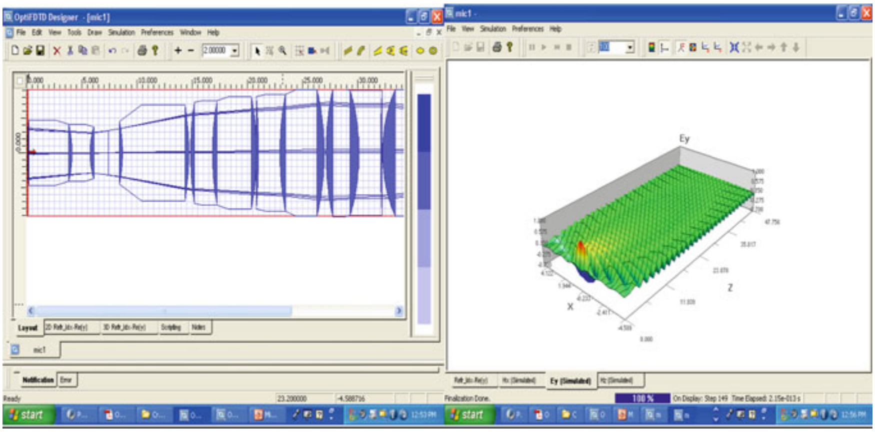


Fig. 1 Microscope

A similar form of the equation holds for a magnetic field (MF).

Both the EF and the MF are at an angle of 90° to the direction of travel x.

Maxwell's equations govern wave equations for the EF and MF.

$$S = \frac{1}{\mu_0} E B$$

To conform Maxwell's equations, the solution must reflect the below equation.

$$S = \frac{1}{c\mu_0} E_m^2 \overline{\sin^2(kx - \omega t)} = \frac{1}{c\mu_0} \frac{E_m^2}{2}$$

The simulations are carried out by a finite-difference time-domain (FDTD) method. The devices simulated are the microscope, eyepiece, wide-angle, GRIN lens, telephoto, and binocular. Data components (3D) E_y, H_x, E_x, and H_y are calculated (Figures 1, 2, 3, 4, 5, and 6).

4 Result Analysis

The Maxwell's equations for optical devices are successfully solved.

5 Discussion

Optical parts are designed and verified through OptiSystem and OptiFDTD.

An optical arrangement of lenses helps to appropriately focus the probe radiation onto the target. Proper design and testing mitigate the risks involved and ensure a reliable end product.

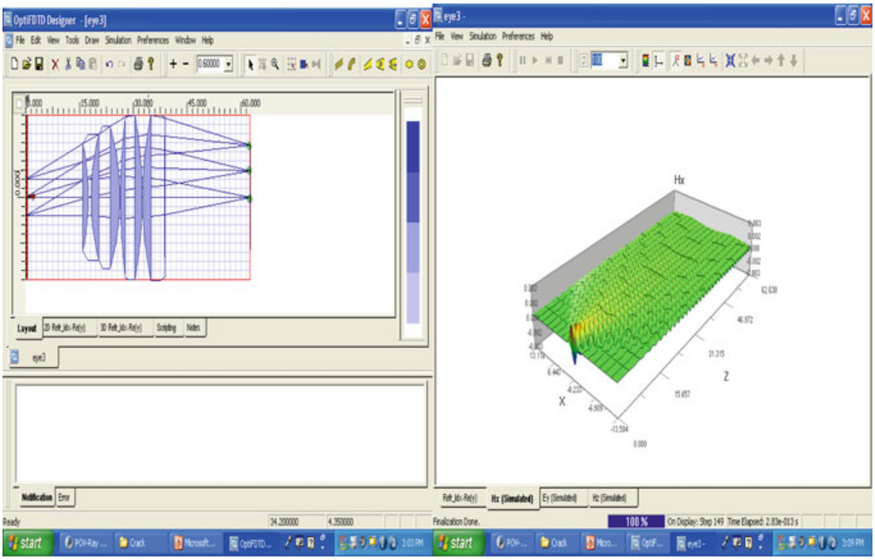


Fig. 2 Eyepiece

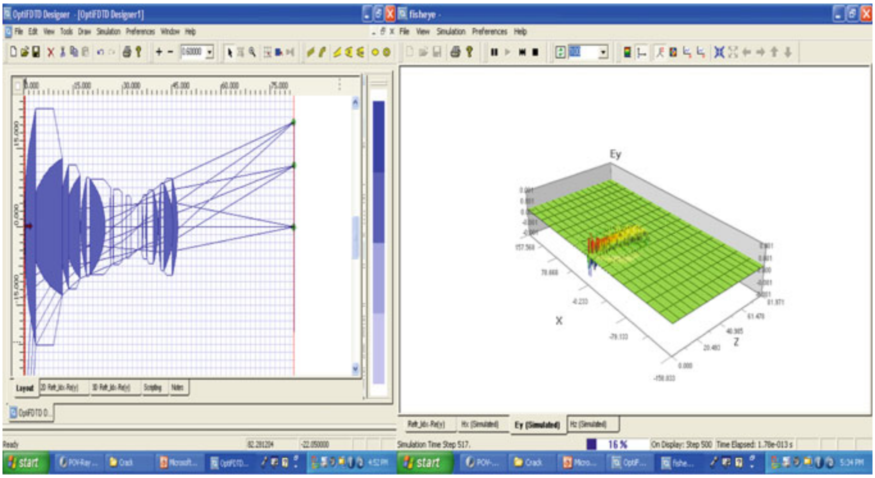


Fig. 3 Wide-angle

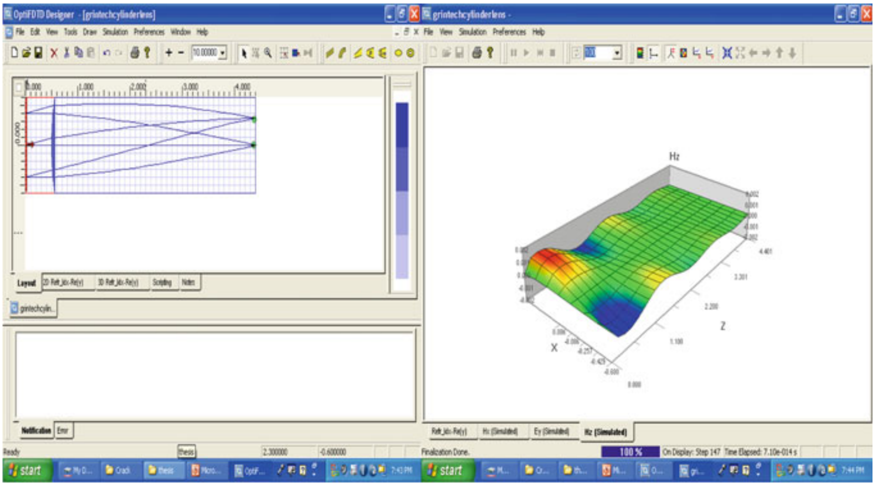


Fig. 4 GRIN lens

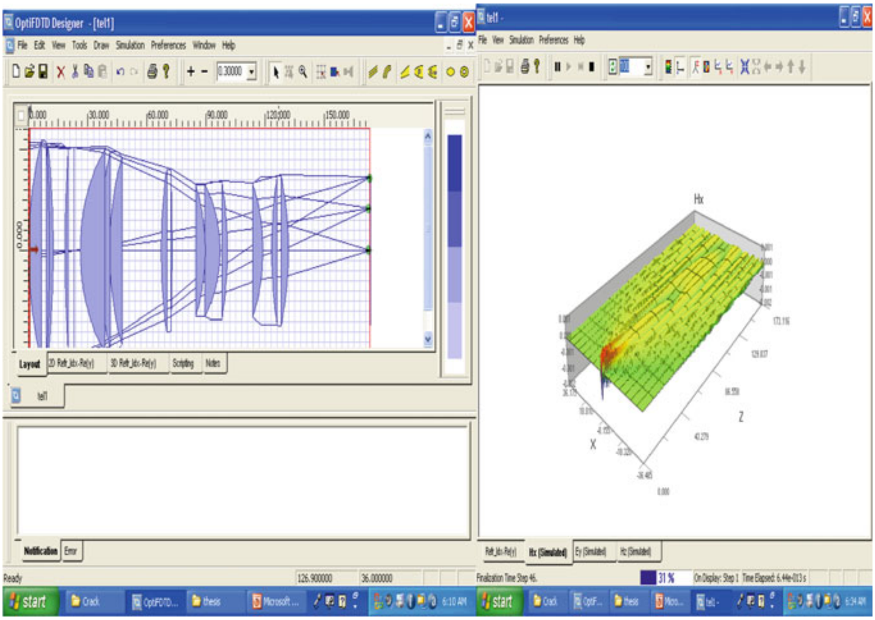


Fig. 5 Telephoto

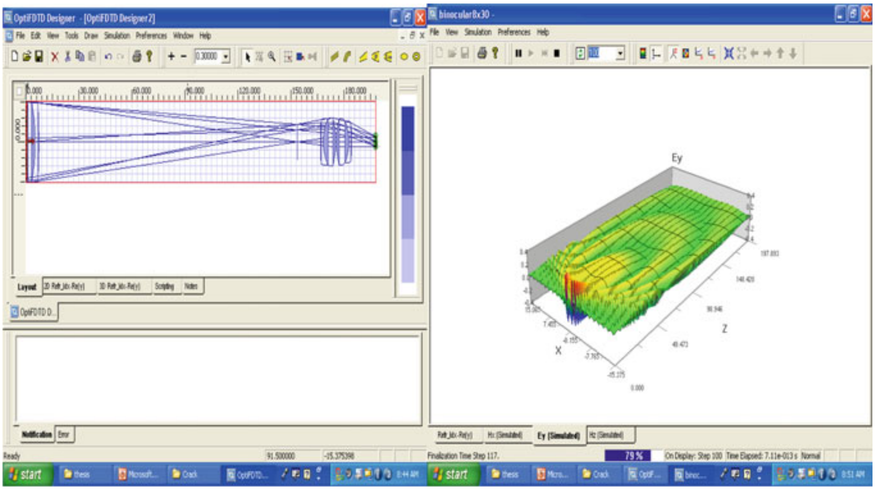


Fig. 6 Binocular

References

1. Cesare, G., Asquini, R., Buzzin, A., Caputo, D.: Electro-optical detector for lab-on-chip applications ©2017 IEEE
2. Walantina, E., Dhinesh Babu, N.R.: Band gap analysis of graphene oxide nanostructures using Opti-FDTD. In: 2016 Online International Conference on Green Engineering and Technologies (IC-GET) (2016)
3. Yoshimura, T.: Simulation of self-aligned optical coupling between micro- and nano-scale devices using self-organized waveguides. J. Light. Technol. **33**(4) 2015
4. Ren, G., Nguyen, T.G., Mitchell, A.: Gaussian beams on a silicon-on-insulator chip using integrated optical lenses. IEEE Photonics Technol. Lett. **26**(14) 2014
5. Schulkin, B., Member IEEE, Zhang, X.C.: Active balance in free-space electro-optic detection of terahertz waves. J. Light. Technol. **27**(17), 3773–3763 (2009)
6. Drummond, M.V., Student Member IEEE, Paulo, P., Monteiro Member IEEE, Nogueira, R.N., Member, IEEE: Photonic true-time delay beamforming based on polarization-domain interferometers. J. Light. Technol. **28**(17), 2492–24983
7. Agmon, A., Schrenk, B., Student Member IEEE, Prat, J., Member IEEE, Nazarathy, M., Senior Member IEEE: Polarization beamforming PON doubling bidirectional throughput. J. Light. Technol. **29**(4), 2579–2585
8. Fischer, J.K., Member IEEE, Ludwig, R., Molle, L., Schmidt Langhorst, C., Member OSA, Leonhardt, C.C., Matiss, A., Schubert, C.: High-speed digital coherent receiver based on parallel optical sampling. J. Light. Technol. **29**(4), 378–385, ISO/IEC 14496–3:2001/Amd.1:2003, Bandwidth extension
9. Borah, D.K., Member IEEE, Voelz, D.G.: Pointing error effects on free-space optical communication links in the presence of atmospheric turbulence. J. Light. Technol. **27**(18), 3965–3973 (2009)
10. Vu, K., Hua, K.A., Senior Member IEEE, Tavanapong, W., Member IEEE: Image retrieval based on regions of interest Aug 2003. IEEE Trans. Knowl. Data Eng. **15**(4), 1045–1049
11. Lu, Y.Y., Zhu, J.: Propagating modes in optical waveguides terminated by perfectly matched layers. IEEE Photonics Technol. Lett. **17**(12) (2005)

Review on Radio Frequency Micro Electro Mechanical Systems (RF-MEMS) Switch

R. Karthick and S. P. K. Babu

Abstract Miniaturization of mechanical or electromechanical systems has paved the way to develop Micro Electro Mechanical Systems (MEMS), and they have the potentials for application in communication systems. Radio Frequency MEMS (RF-MEMS) switches can be used as an alternative to mechanical and semiconductor devices-based switches such as PIN diodes or varactor diodes for their better isolation, reduced insertion loss, low-power consumption and higher-power handling capabilities. There are various constraints involved in designing RF-MEMS switch like finite or limited time to toggle, prone to failure, power handling capacity, RF performance, material selection, etc. Hence, it is necessary to properly select key parameters and optimize the switch to achieve desired outcome for specific applications. This paper discusses design constraints and various parameters involved in designing RF-MEMS switch. From the review, it is found that shunt-type configuration of RF-MEMS switch with electrostatic actuation, capacitive contact type and bridge structure are suitable for millimetre wave applications which are explored for future bandwidth hungry communication systems.

Keywords RF-MEMS · Switches

1 Introduction

A Micro Electro Mechanical System (MEMS) is a technology used for producing systems, devices and products with the combination of electric and mechanical components without losing their characteristics and efficiency in microscale. The size of the MEMS devices is of the order 20 μm to a millimetre. Since it is in microscale, it requires low power, less space and low cost of production, and on the other hand more reliable, much faster can be batch processed and able to incorporate

R. Karthick (✉)
Department of Physics, Periyar Maniammai Institute of Science and Technology, Thanjavur, India
e-mail: karthickmtech@gmail.com

S. P. K. Babu
Department of Electronics and Communication Engineering, Periyar Maniammai Institute of Science and Technology, Thanjavur, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_43

multiple and complex functions [1]. RF-MEMS devices can be classified into two: active MEMS devices (tuners, variable capacitors and switches) and passive MEMS devices (couplers, filters and transmission lines) [2, 3]. Recent drastic development in the telecommunication industry specifically in RF, micro and mm-wave applications, requires devices with high linearity and reliability, low-power consumption and compatible with existing fabrication techniques. RF-MEMS are a better solution for the above-mentioned requirements [4–7].

RF-MEMS devices are specially designed and fabricated for RF circuits, which are entirely different from the MEMS devices working in radio frequencies [4]. RF-MEMS devices are classified as extrinsic type (transmission lines) and intrinsic (switches) based on either they are placed inside or out of the RF circuit [3]. In this paper, RF-MEMS switches are discussed and particularly for millimetre wave applications. The analysis is important as it becomes more complicated at RF frequencies.

2 Classification of RF Switches

RF switches' routes signal through various transmission paths. When a switch is placed in the path of RF signal, it can offer resistance in the signal path, capacitance between signal to signal or signal to ground path and crosstalk as well [8]. Since RF switches are integral and fundamental part of RF system, proper tuning of switch specifications such as isolation, insertion loss, switching time and power handling promote considerable and desired changes in RF performance and output of RF system [2, 9]. Hence, we can infer that RF switch is more than just common switch. There are four categories of RF switches as shown in Fig. 1. The selection of a RF switch is based on application, type of platform, signal level, switching speed, etc. Classified switches perform better than one another in different switch parameters, even though all switches have their advantages and disadvantages [4].

Mechanical switch: High-power handling broadcast applications like TV, AM, FM and HF use mechanical switches, available with coaxial connectors and waveguides. The switching is carried out by interrupting the electrical path or transmission line physically by externally applied actuation voltage. Even though mechanical switches

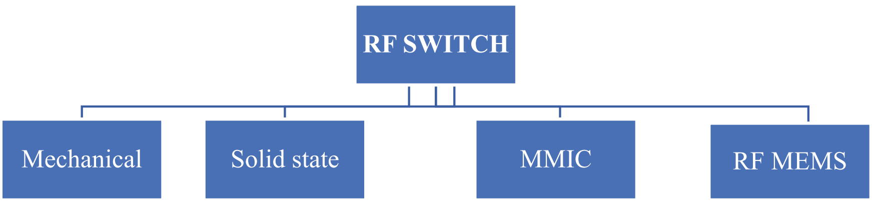


Fig. 1 Types of RF switches

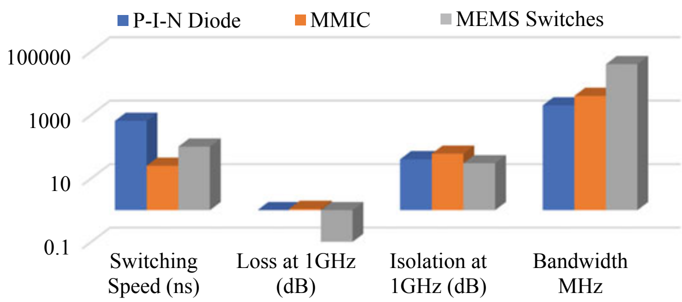


Fig. 2 Comparison of switches: solid state, MMIC and MEMS [2]

exhibit high-power handling, high isolation and low insertion loss, their switching speed is very low of the order of 2–50 ms [2, 10].

Solid-state switch: When semiconductor components like diode or FET are placed in series with RF path, it acts as a switch and exhibits low insertion loss but gives poor isolation. These switches are widely used in the field of communication where high switching speed is required like digital modulator, radar, multiplexing and demultiplexing. [11]. However, the shunt configuration provides high isolation with increased insertion loss.

MMIC switch: Compared to solid-state switches, MMIC switches work over wider bandwidth with faster switching speed. They are widely used because it eradicates the demand for attaching discrete components by simultaneous fabrication of both active and passive components on a semi-insulating semiconductor substrate [12]. Major disadvantages are poor isolation and large insertion loss (Fig. 2).

RF-MEMS switch: The usage or application of MEMS switch in radio frequency, millimetre wave and microwave systems is termed as RF-MEMS switch. In other words, designing and fabrication of MEMS switches specifically for RF circuits are called RF-MEMS switches [13, 14]. As communication industry focuses on handheld and miniaturized devices, RF performance, low-power consumption, size and cost are also to be considered while choosing or designing a switch [15]. RF-MEMS switch [16] is a fare solution that makes use of the advantages of both mechanical and electronic switches and overcomes most of their disadvantages [17] (Fig. 3).

3 Classification of MEMS-Based RF Switches

Mechanical section concerns about movement of the switch (either vertical or lateral) and the force required to make the movement (actuation mechanism). The electrical part deals with configuration of switch either series or shunt with contact type of capacitive or ohmic. Depending on the requirement, one can choose cantilever or bridge structure with different combination of actuation mechanism, mechanical contact and circuit configuration [18].

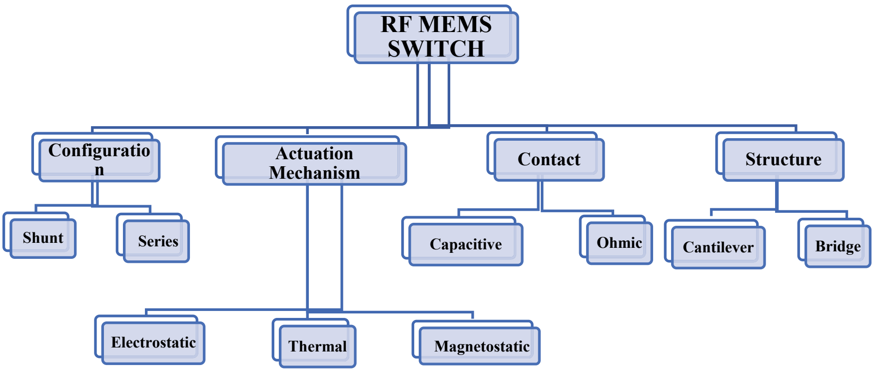


Fig. 3 Classification of RF-MEMS switches

3.1 Configuration

Series: The RF series switch consists of thin metal strip suspended over discontinuous transmission line as shown in Fig. 4. The switch is in ON state when the induced electrostatic force by electrodes pulls the strip to move downwards and short-circuits the input and output of the transmission line which make RF signal to flow through it. Removal of applied voltage results in moving back of the strip to its original position which disconnects the transmission line and blocks the RF signal flow. Then, the switch is driven to OFF state [2, 19].

C_s—Series capacitance, C_p—Parasitic capacitance, R_c—Contact resistance, Z_0—Characteristic impedance, Z_h—High impedance, β—Phase constant, βl—Electrical length of the line.

Shunt: Unlike series configuration, shunt switch consists of a thin metal strip hanging over a coplanar waveguide transmission line which contains a central signal and two ground lines parallel to it. The construction of shunt switch is given in Fig. 5. A small dielectric layer over the *t*-line forms the capacitor between the metal strip and transmission line. Without any applied bias voltage, the switch is in ON state. Since the strip is supported by its stiffness and spring constant, it remains in up-state and allows the RF signal to flow through the central signal line [2].

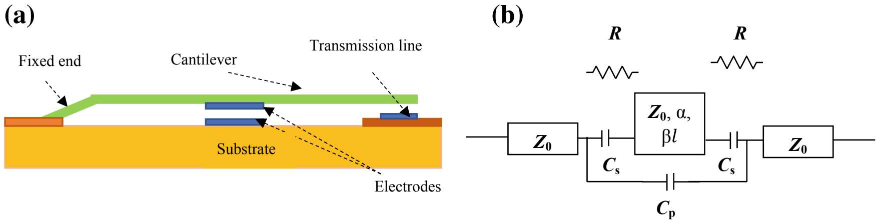


Fig. 4 **a** Series switch: cross-sectional view. **b** Electrical equivalent circuit model

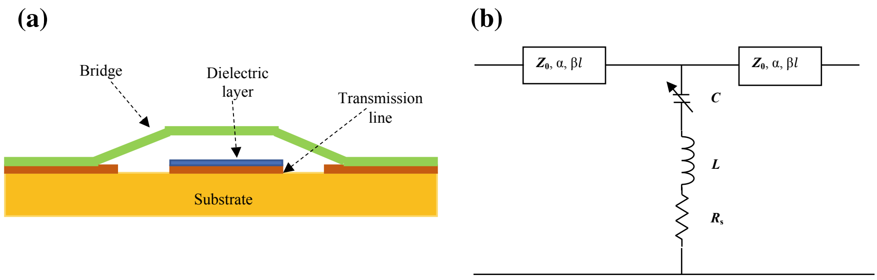


Fig. 5 **a** Shunt switch: cross-sectional view. **b** Electrical equivalent circuit model

Z_0—Impedance, α—Attenuation constant, β—Phase constant, βl—Electrical length of the line. C—Variable capacitance, R_s—Shunt resistance, L—Inductance.

When the bias voltage is applied to the membrane, the switch moves to OFF state by snapping down the metal strip to the central signal line. Now, the RF signal in the central signal line is coupled to the ground lines of CPW. This leads to decrease in RF resistivity and increases capacitance between metal strip and signal line of CPW. Hence, the RF signal flow from input to output through signal line is terminated [18]. Switches with different configurations are studied in [20–23].

3.2 Actuation Mechanism

Actuation is the process of making the switch ON or OFF with respect to an applied controlling parameter. Based on the mechanism and controlling parameter, actuation mechanism can be classified as follows: electrostatic, electromagnetic, thermal and piezoelectric [2]. Based on the application and requirement such as power consumption, actuation voltage, size, contact force, power and switching speed, specific actuation mechanism can be implemented.

Electrostatic actuation mechanism uses DC bias for actuation. The applied DC bias on the switch membrane induces an electrostatic force which collapses the beam to the down state. This mechanism leads to low-power consumption, small size and linearity. Disadvantages include high actuation voltage and low switching speed [24] (Table 1).

Thermal actuation technique uses thermal micro actuators to ON and OFF the switch by applying voltage or current pulse. Thermally actuated switches have high reliability and power handling capacity but not preferred commonly due to its high-power consumption [25].

Electromagnetic actuation uses coils and ferromagnetic materials as electromagnetic actuators to actuate the switch. Though this technique requires only very low voltage for actuation but consumes high current and size also large compared with other techniques [26].

Table 1 Actuation mechanism [2]

	Voltage (V)	Current (mA)	Power (mW)	Size	Switching time (μs)	Contact force (μN)
Electrostatic	20–80	0	0	Small	1–200	20–1000
Thermal	3–5	5–100	0–200	Large	300–10,000	500–4000
Electromagnetic	3–5	20–150	0–100	Large	300–1000	50–200
Piezoelectric	3–20	0	0	Large	50–500	50–200

Piezoelectric mechanism uses specially designed piezoelectric actuator which is attached to the switch membrane to move downwards or upwards. The property of elastic deformation due to the electric field makes this mechanism suitable for low actuation voltage and high force requirements. Major disadvantages are its size and switching time [27–29].

3.3 Contact Type

Based on the contact mode between the electrodes, RF-MEMS switches can be categorized into ohmic (metal–metal) which is used to attain resistive contact and capacitive (metal–dielectric–metal) contact type that gives capacitance ratio between ON and OFF states [2].

Figure 6 shows cross-sectional view of ohmic contact switch. When an actuating voltage is applied, the induced electrostatic force causes the beam to bend and come in contact with bottom transmission line directly. The parasitic capacitance between the top electrode and the bottom line gives isolation in OFF state [30].

There are many disadvantages in this configuration like adhesion of layers, interfacial stresses and stiction which lead to failure of the switch life time and reliability. The impact force between top and bottom metal contacts causes damage, pitting and hardening of contact area [31, 32]. Capacitive switch is shown in Fig. 7, and it incorporates dielectric between the upper electrode and the transmission line. This improves isolation and reduces the problems of stiction and adhesion between the layers. When the switch is actuated, unlike ohmic contact, the beam directly rests

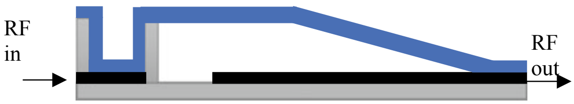


Fig. 6 Ohmic (metal–metal) contact switch

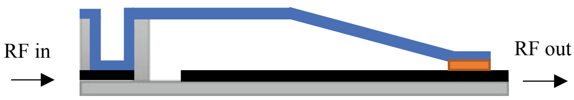


Fig. 7 Capacitive (metal–dielectric–metal) contact switch

on the dielectric layer and capacitively coupled to the bottom transmission line. This causes low impedance between the electrodes, and its ratio in OFF-to-ON is directly proportional to the ratio of capacitance in ON and OFF state [33–35].

3.4 Structure

A typical MEMS switch structure consists of thin membrane or beam named as bridge suspended over transmission line. The beam is actuated in such way that it can allow or block the signal through the transmission line by mechanical movement. Physical arrangement of the membrane makes the difference in structure either cantilever beam or bridge structure [2].

In cantilever beam structure, one end of the membrane is fixed, and other end is free to move in transverse direction, i.e. up and down. During actuation, the electrostatic force pulls the cantilever to snap down on the transmission line using electrode, which make or break the RF signal flow depends on the configuration either series or shunt [36]. The movement of the cantilever is controlled by the applied pull-in voltage. This structure has various advantages like high switching speed, requires low actuation voltage but has the issues of stiction, self-latching due to interfacial stress and complicate fabrication steps [37] (Fig. 8).

Bridge or fixed–fixed beam structure is shown in Fig. 9, in which, ends of the membrane are fixed which forms a metal bridge suspended over a transmission line. During actuation, the applied voltage exceeds the pull-in voltage of the bridge snapped down on the transmission line to deal with the signal flow. The membrane

Fig. 8 Construction of cantilever-based beam structure

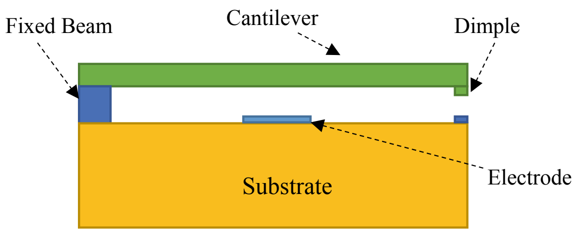
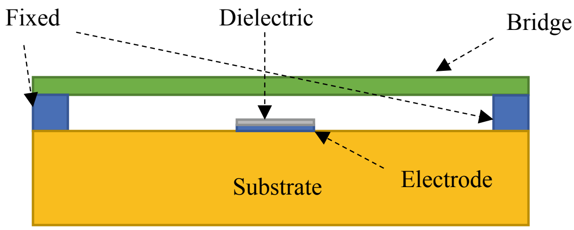


Fig. 9 Construction of bridge-type RF-MEMS switch



falls back to its original position when the applied voltage is removed [38]. Both the structures are optimized and analysed for RF applications [39–42].

4 Design Considerations of RF-MEMS Switch

4.1 Effect of Beam

The beam which may be either fixed–fixed or cantilever structure plays a key role in mechanical and electromagnetic modelling of MEMS switches. By properly modelling length, thickness and thickness-to-length ratio of the beam, major parameters can be optimized. The first step in mechanical modelling of switch is to determine spring constant. In both cases of bridge and cantilever structure, the spring constant for evenly distributed beam is given by Eq. (1) [2, 43].

$$k = 32Ew\left(\frac{t}{l}\right)^3 \tag{1}$$

From Eq. (1) it is inferred that (*t*/*l*) determines the spring constant, optimization its value will increase the spring constant and helps in reducing residual stress due to spring stiffness occurs in fixed–fixed beam structure. Cantilever beam does not contain any residual stress, at the same time an undesirable effect that is deflection upon release occurs due to built-in stress gradient [44]. For same (*t*/*l*) value, cantilever beam shows much smaller value of spring constant (48 times) than the fixed–fixed beam structure.

$$V_{\mathrm{p}} = \sqrt{\frac{8\,\mathrm{kg}^3}{27\varepsilon_0 W w}} \tag{2}$$

The above equation relates pull-in voltage with spring constant, beam dimensions and permittivity in free space [45]. Lower spring constant results in lower pull-down voltage. Different beam structures are proposed by many researchers to reduce the pull-down voltage, by implementing meanders or flexures to support the beam [46]. One more component which is present only in fixed–fixed beam is the compressive stress. To withstand larger compressive stress, the length of the beam can be reduced to increase the stiffness which intern increases the pull-down voltage during actuation [44] (Table 2).

Table 2 Various types of support beams and their spring constant

Beam support	Spring constant
Fixed–fixed flexure [43]	$k = 4Ew\left(\frac{t}{l}\right)^3$
Crab-leg flexure [47]	$k = \frac{4Ew\left(\frac{t}{l_c}\right)^3}{1+\frac{l_s}{l_c}\left[\left(\frac{l_s}{l_c}\right)^2+12\frac{1+\nu}{1+\left(\frac{w}{t}\right)^2}\right]}$; $k \approx 4Ew\left(\frac{t}{l_s}\right)^3$ for $l_s \gg l_c$
Folded flexure [48]	$k \approx 2Ew\left(\frac{t}{l}\right)^3$
Serpentine flexure [48, 49]	$k \approx \frac{48GJ}{l_a^2\left(\frac{GJ}{EI_x}l_a+l_b\right)n^3}$ for $n \gg \frac{3l_b}{\frac{GJ}{EI_x}l_a+l_b}$

4.2 Effect of Dielectric

MEMS switches use dielectric material of thickness around 1000–2000 Å between the beam and lower electrode. The properties of dielectric material such as thickness, dielectric constant and roughness play vital role in electromagnetic modelling of switch [2]. From Eq. (3), it is inferred that the effect due to dielectric thickness and relative dielectric constant becomes unavoidable while calculating capacitance of MEMS switch.

$$c = \frac{\varepsilon\, \varepsilon_0 A}{g + \frac{t_d}{\varepsilon_r}} \tag{3}$$

t_d thickness and ε_r dielectric constant of dielectric layer. ε accounts for capacitance degradation due to surface roughness of layer.

$$\frac{C_d}{C_u} = \frac{\left(\frac{\varepsilon_0 \varepsilon_r A}{t_d}\right)}{\left(\frac{\varepsilon_0 A}{g+\frac{t_d}{\varepsilon_r}}\right)} \tag{4}$$

For better RF performance, capacitance ratio (4) should be increased [2], and it can be achieved by using thin dielectric layer with high dielectric constant. But due to pinhole problems in fabrication and thinner layer cannot withstand higher actuation voltage issues, it is difficult to deposit dielectric layer thinner than 1000 Å [50, 51]. Different dielectric materials with high dielectric constant are proposed in [52–54], to achieve RF performance. The dielectric layer also affects the actuation voltage and down-state capacitance due to its surface roughness or dielectric layer surface is not perfectly flat. The electrostatic force equation contains an additional factor (ε_r) [55].

$$F_e = \varepsilon_r\left(\frac{QE}{2}\right) \tag{5}$$

Increase in surface roughness reduces the contact area, which intern reduces the down-state capacitance. The surface may be very rough and contact area may be very small, but it is acceptable since the down-state capacitance is always high [56, 57].

4.3 Gap Between Membrane and Dielectric

The electric field due to applied voltage is given by Eq. (6) [2]. Reduced gap height can decrease the required pull-down voltage [58].

$$E = \frac{V}{g} \tag{6}$$

Gap height deals with various RF parameters, which are derived from switch capacitance like insertion loss, isolation loss and return loss. Increasing or decreasing gap height makes a huge variation in MEMS capacitance, and hence, RF performance [58].

The switching time of the RF-MEMS switch can be optimized with Knudsen number and Squeeze number. Knudsen number deals with collision of gas molecules, and it is desired to be high to reduce collision. Reducing gap height will increase the Knudsen number [2, 59].

$$K_n = \frac{\lambda}{g} \tag{7}$$

Squeeze number determines how fast the gas escapes from the gap without compression. Increasing the gap may decrease the Squeeze number but it decreases the Knudsen number. The damping coefficient is strongly dependent on gap height. The gap height also determines the diameter of the holes to be made on the beam, that is the hole diameter should be always less than $3g$ [2, 60].

5 MEMS Capacitance and RF Parameter

RF performance of RF-MEMS switches is estimated by parameters such as return loss, insertion loss and isolation loss for given capacitance and the ratio of up- and down-state capacitance of the switch [61]. An ideal series switch shows very low insertion loss in down state and infinite isolation in up state. An ideal shunt capacitive switch delivers acceptable isolation in down state and low insertion loss in up-state position [62].

The up-state capacitance, i.e. at $g \neq 0$; $\varepsilon = 1$ (8a); the down-state capacitance, i.e. at $g = 0$; $\varepsilon \neq 1$; (0.4–0.8) (8b)

$$C_{\mathrm{u}} = \frac{\varepsilon_0 A}{g + \frac{t_{\mathrm{d}}}{\varepsilon_{\mathrm{r}}}}; \tag{8a}$$

$$C_{\mathrm{d}} = \frac{\varepsilon_0 \varepsilon_r A}{t_{\mathrm{d}}}; \tag{8b}$$

The up-state/down-state capacitance ratio

$$C_{\mathrm{r}} = \frac{C_{\mathrm{d}}}{C_{\mathrm{u}}}; \quad \frac{C_{\mathrm{d}}}{C_{\mathrm{u}}} = \frac{\left(\frac{\varepsilon_0 \varepsilon_r A}{t_{\mathrm{d}}}\right)}{\left(\frac{\varepsilon_0 A}{g + \frac{t_{\mathrm{d}}}{\varepsilon_{\mathrm{r}}}}\right)} \tag{9}$$

The switch is modelled as CLR circuit, and the switch impedance is obtained by resistance R, inductance L and variable capacitance $C = C_{\mathrm{d}} + C_{\mathrm{u}}$ [2]

$$Z_{\mathrm{s}} = R_{\mathrm{s}} + j\omega L + \frac{1}{j\omega C}; \quad Z_{\mathrm{s}} = \begin{cases} \frac{1}{j\omega C}, & f \ll f_0 \\ R_{\mathrm{s}}, & f = f_0 \\ j\omega L, & f \gg f_0 \end{cases} \tag{10}$$

The impedance depends on the resonant frequency of the switch. The CLR circuit behaves as capacitor for frequency less than resonant frequency, as inductance is greater than resonant frequency and series resistance when the frequency is equal to resonant frequency [2, 51].

The electrical performance is described by the S-parameter analysis of the switch in its up and down state. S parameters are effective indicator of RF performance of MEMS switch [63].

The parameter S_{11} is analysed in up-state position to estimate return loss and given by Eq. (11),

$$S_{11} = \frac{-j\omega C_{\mathrm{u}} Z_0}{2 + j\omega C_{\mathrm{u}} Z_0} \tag{11}$$

The parameter S_{21} given by Eq. (12) is analysed in both up and down states to estimate insertion loss and isolation.

$$S_{21} = \frac{1}{1 + j\omega C_{\mathrm{d}} Z_0/2} \tag{12}$$

The loss in MEMS switch is also determined by S-parameter, and it is represented in Eq. (13) [2, 64].

$$\mathrm{Loss} = 1 - |S_{11}|^2 - |S_{21}|^2 \tag{13}$$

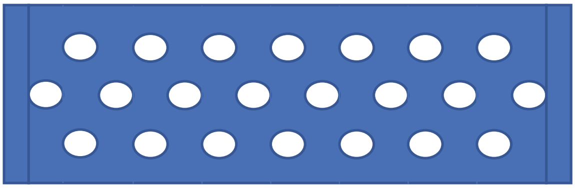


Fig. 10 Beam with holes

6 Effect of Perforation on the Beam

The switch membrane is provided with holes of small diameter (3–8 μm) spaced at a period of 5–6 μm in lattice plane and can occupy around 60% of the total membrane area [2]. Physically, these holes reduce total mass which yield better mechanical resonant frequency, and reduced Young's modulus leads to decrease in pull-in voltage of the membrane [65]. If the diameter of the hole is less than 3*g*, it does not disturb the up-state capacitance since the fringing field envelopes the hole area but it has an effect on down-state capacitance and affects the electrostatic force (Fig. 10).

The holes play a major role in reducing damping coefficient, since the gas underneath membrane escapes through the holes when actuated, and there is no possibility for Squeeze number to exist [66]. During fabrication, holes help to remove the sacrificial by etching techniques [67, 68].

7 Material Selection

Material selection for dielectric layer: The physical structure and properties of material chosen as dielectric layer directly affect the performance of switches [69]. Even though optimizing the physical structure of dielectric layer, surface roughness, thickness and width improve the switch performance but have their own limitations [70]. The choosing proper dielectric material with specific properties listed in Table 3

Table 3 Properties of materials used for dielectric layer [72, 73]

Material	Dielectric constant	Resistivity
SiO_2	3.9	10^{14}
Si_3N_4	7.5	10^{14}
Al_2O_3	9.8	10^{14}
PZT	3850	7×10^4
HfO_2	25	10^{14}
BST	800	10
TiO_2	100	10^{12}

Table 4 Properties of materials considered for membrane [72, 73]

Material	Thermal conductivity (K)	Electrical resistivity (ρ)	Young's modulus (E)	Poisson's ratio (ν)	Thermal expansion coefficient (α)
Silicon nitride	29	$>10^{12}$	304	0.3	2.7
Aluminium oxide	39	$>10^{12}$	69	0.33	7.4
Nickel	70	9.50×10^{-8}	204	0.31	13.3
Platinum	71	10.60×10^{-8}	171	0.39	9.1
Molybdenum	142	5.20×10^{-8}	320	0.32	4.9
Aluminium	222	2.90×10^{-8}	69	0.33	23.6
Copper	315	1.72×10^{-8}	115	0.33	17
Gold	388	2.35×10^{-8}	77	0.42	14.2

gives promising effect on performance of switch without affecting the other parameters [71]. Out of the various properties, relative permittivity deals with pull-down voltage and hold on voltage. Resistivity of dielectric materials gives the measure of AC losses.

To reduce pull-down voltage and hold voltage material with high dielectric constant must be selected and to minimize AC losses during actuation, low resistivity materials are preferred [73, 74].

Material selection for beam membrane: The physical structural optimization of switches has their own limitations, but by proper selection of material for switch membrane will help in optimizing various important parameters life pull-in voltage, RF loss and residual stress [69]. The important material parameters which taken into consideration for designing better switch membrane are detailed in Table 4 [75].

Choosing material with higher value of thermal expansion coefficient (α) and Poisson's ratio (ν) reduces pull-in voltage. RF losses can be minimized by preferring material with low value of Young's modulus (E) and electrical resistivity (ρ). Material with high thermal conductivity (K) and low electrical resistivity (ρ) provides minimum thermal residual stress [76, 77].

8 Conclusion

RF-MEMS switches at frequencies in the range of tens of GHz will be needed for future communication systems such as 5G mobile cellular communication. As we discussed earlier, mechanical switches have the ability to handle high-power devices but only at lower RF frequencies that too at a much slower speed. On the other hand, solid-state switches work at high speed with disadvantage of low-power handing and resistive losses. RF-MEMS give better solution for millimetre wave requirement. Electrostatic mechanism is most predominantly preferred in switching mechanism because of its unique property of low-power consumption, smaller size and suitable for complex circuits.

Series switch has problem of poor isolation, stiction and consumes more power compared to shunt, which makes the series type suitable only for low-frequency operations. The advantage of implanting shunt configuration is its low parasitic because of continuous *t*-line, higher RF power handling capability, ease of fabrication and shows better RF performance. The isolation and insertion losses are two parameters which have to take into account primarily at higher frequencies. Isolation in case of contact-type switches reduces with increase in frequency with huge reduction in reliability when used at RF power levels of 50–100 mW. Capacitive type is preferred for high-frequency application.

Cantilever structure is preferred for series switch configuration while bridge is most commonly used for shunt switch configuration. Even though fixed–fixed beam structure has disadvantages of higher spring constant, they preferred over cantilever because of their stability, lower sensitive to stress and ease of fabrication.

A typical MEMS switch, balancing all constraints, may have gap height of 2–3 μm. The (t/l) ratio determines spring constant, pull-in voltage, residual stress and compressive stress. Beam with length $(l) = 250$–350 μm, thickness $(t) = 0.5$–2 μm and width $(w) = 80$–120 μm and dielectric layer of thickness 1000–1500 Å are suggested for RF-MEMS switches. Holes in the membrane with specific diameter (less than $3g$) and gap between them also help to reduce actuation voltage and spring constant. In case of material selection, Barium Strontium Titanate (BST) seems to be a better material for dielectric layer of MEMS switch compromising for low loss and pull-down voltage, aluminium and gold are found to be best-suited materials for bridge membrane design.

References

1. Hsu, T.-R.: MEMS and Microsystems: Design and Manufacture. McGraw Hill Education Private Limited (2002)
2. Rebeiz, G.M.: RF MEMS Theory, Design, and Technology. Wiley, New York (2003)
3. Brown, E.R.: RF MEMS switches for reconfigurable integrated circuits. IEEE Trans. Microw. Theory Tech. **46**, 1868–1880 (1998)
4. Varadan, V.K., Vinoy, K.J., Jose, K.A.: RF MEMS and Their Applications. Wiley, New York (2011)
5. Nguyen, C.T.C.: Microelectromechanical system for wireless communication. In: The 11th Annual International workshop on Micro Electro Mechanical Systems. Heidelberg, Germany, pp. 1–7 (1998)
6. Zhang, S., Ying, Z.N., Xiong, J., He, S.L.: Ultrawideband MIMO/diversity antennas with a tree-like structure to enhance wideband isolation. IEEE Antennas Wirel. Propag. Lett. **8**, 1279–1282 (2009)
7. Zhu, Y.-Q., Han, L., Tang, J.-Y.: MEMS Switch Manages Millimeter-Wave Signals", Microwaves & RF, pp. 58–65, Nov 2013

8. Chang, K.: Microwave Solid State Circuits and Application. Wiley, Chichester, Sussex (1994)
9. Mafinejad, Y., et al.: Design and Simulation of a RF Mems Shunt Switch for Ka and V Bands and the Impact of Varying Its Geometrical Parameters. IEEE, pp. 823–826 (2009)
10. Rebeiz, G.M., Muldavin, J.B.: RF MEMS switches and switch circuits. IEEE Microw. Mag. **2**, 59–71 (2001)
11. Mansour, R.R., et al.: RF MEMS devices. In: Proceedings of the International Conference on MEMS, NANO and Smart Systems, pp. 103–107 (2003)
12. Hindle, P.: The state of RF/microwave switch devices. J. Microw. **53**(11), 20–36 (2010)
13. Wipf, S.T., et al.: D-band RF MEMS SPDT switch in a 13 μm SiGe BiCMOS technology. IEEE Microw. Wirel. Compon. Lett. **26**, 1002–1004 (2016)
14. Peterson, K.E.: Micromechanical membrane switches on silicon. IBM J. Res. Dev. **23**, 376–385 (1979)
15. Brown, E.R.: RF MEMS for reconfigurable integrated circuits. IEEE Trans. Theory Technol. **46**, 1868–1880 (1998)
16. Vinoy, K.J., et al.: Surface micromachined capacitive RF switches with low actuation voltage and steady contact. J. Microelectromech. Syst. **26**, 643–652 (2017)
17. Persano, A., et al.: Influence of design and fabrication on RF performance of capacitive RF MEMS switches. Microsyst. Tachnol. **22**, 1741–1746 (2016)
18. Muldavin, J.B., Rebeiz, G.B.: High isolation CPW MEMS shunt switches—part I: modeling. IEEE Trans. Microw. Theory Tech. **48**, 1045–1052 (2000)
19. Khodaddy, K., et al.: Design and modelling of a novel RF MEMS series switch with low actuation voltage. Microsyst. Technol. **22**(12), 2921–2929 (2015)
20. Muldavin, J.B., Rebeiz, G.B.: High isolation CPW MEMS shunt switches—part I: design. IEEE Trans. Microw. Theory Tech. **48**, 1053–1056 (2000)
21. Jung, C.W., De Flaviis, F.: RF-MEMS capacitive series of CPW&MSL configurations for reconfigurable antenna application. In: IEEE Antennas and Propagation Society International Symposium, vol. 2A, pp. 425–428, July 2005
22. George, R., et al.: Design of series RF MEMS switches suitable for reconfigurable applications. In: IEEE-Proceedings of ICCPCT (2017)
23. Zhang, L.X., Zhao, Y.P.: Electromechanical model of RF MEMS switches. Microsyst. Technol. **9**, 420–426 (2003)
24. Cho, I.J., et al.: A low voltage and low power RF MEMS series and shunt switches actuated by combination of electromagnetic and electrostatic forces. IEEE Trans. Microw. Theory Tech. **53**(7), 2450–2457 (2005)
25. He, X.J., et al.: Electrothermally actuated RFMEMS capacitive switch with atomic layer deposited dielectric. In: 16th International IEEE Conference, pp. 2470–2473 (2011)
26. Bachman, M., et al.: High power magnetically actuated microswitches fabricated in laminates. IEEE Electron. Dev. Lett. **33**, 1309–1311 (2012)
27. Guerre, R., et al.: Wafer level transfer technologies for PZT based RF MEMS switches. J. Microelectromech. Syst. **19**, 548–560 (2010)
28. Molaei, S., Ganji, B.A.: Design and simulation of a noval RF MEMS shumt capacitive switch with low actuation voltage and high isolation. J. Microsyst. Tech. **23**(6), 1907–1912 (2016)
29. Li, M., et al.: Design and fabrication of a low insertion loss capacitive RF MEMS switch with noval micro structures for actuation. Solid State Electron. **127**, 32–37 (2016)
30. Lee, H.C., et al.: Design, fabrication and RF performance of two different types of piezoelectrically actuated Ohmic MEMS switches. J. Micromech. Microeng. **15**, 2098–2104 (2009)
31. Van Spenger, W.M., et al.: On the physics of stiction and its impact on the reliability of microstructures. J. Adhes. Sci. Technol. **17**, 563–582 (2003)
32. Czaplewski, D.A., et al.: Lifeime limitations of Ohmic, contacting RF MEMS switches with Au, Pt and Ir contact materials due to accumulation of 'friction polymer' on the contacts. J. Micromech. Microeng. **15**, 2098–2104 (2009)
33. Goldsmith, L., et al.: Performance of low loss RF MEMS capacitive switches. IEEE Microw. Guided Wave Lett. **8**, 269–271 (1998)

34. Mafinejad, Y., et al.: Low insertion loss and high isolation capacitive RF MEMS switch with low pull-in voltage. Int. J. Adv. Manuf. Technol. **93**(1), 661–670 (2017)
35. Sawant, B., et al.: Modeling and analysis of low voltage, high isolation capacitive type RF MEMS switches. In: IEEE, ICCCNT 2018, IISC India, July 2018
36. Rahman, H.U., et al.: Cantilever beam design for RF MEMS switches. Micromech. Microeng **20**, 1–12 (2010)
37. Vakilian, M., et al.: Optimization of cantilever based MEMS switch used in reconfigurable antennas. In: IEEEICSE 2012 Proceedings, 2012, Kuala Lumpur, Malaysia
38. Hu, G.-W., Liu, Z.-W., Hou, Z.-H., Liu, L.-T., Li, Z.-J.: A dielectric bridge type series contact switch 0–10 GHz applications. In: 2006 8th International Conference on Solid State and Integrated Circuit Technology Proceedings, pp. 542–544, 23–26 Oct 2006
39. Saha, S.C., et al.: Modeling of spring constant and pull down voltage of non-uniform RF MEMS cantilever incorporating stress gradiant. J. Sens. Trans. **11**, 54–68 (2008)
40. Liu, Y., et al.: A compact single-cantilever multicontact RF MEMS switch with enhanced reliability. IEEE Microw. Wirel. Compon. Lett. **28**, 191–193 (2018)
41. Sravani, K.G., et al.: Role of dielectric layer and beam membrane in improving the performance of capacitive RF MEMS switches for Ka band applications. Microsyst. Technol. (2018)
42. Zhang, N., et al.: Design and performance of a J band MEMS switch. MDPI Micromach. **10**(7), 467 (2019)
43. Roark, R.J., Young, W.C.: Formulas for Stress and Strain, 6th edn. McGraw-Hill, New York (1989)
44. Gere, J.M., Timoshenko, S.P.: Mechanics of Materials, 4th edn. PWS Publishing Company, Boston (1997)
45. Fedder, G.: MEMS fabrication. In: Proceedings of the IEEE International Test Conference, 30 Sept–2 Oct 2003
46. Gupta, A.K., Sharma, N.: Investigation of actuation voltage for non-uniform serpentine flexure design of RF MEMS switch. Springer Microsyst. Technol. **20**, 413–418 (2014)
47. Yun, W.: A surface micromachined accelerometer with integrated CMOS detection circuitry. Ph.D. thesis. University of California, Berkeley, CA (1992)
48. Fedder, G. K.: Simulation of microelectromechanical systems. Ph.D. thesis. University of California, Berkeley, CA (1994)
49. Pacheco, S.P., et al.: Design of low actuation voltage MEMS switch microwave symposium digest. IEEE MTT-S International, vol. 1, pp. 165–168 (2000)
50. Badia, M.F.B.: RF MEMS shunt capacitive switches using AIN compared to Si_3N_4 dielectric. J. Microelectromech. Syst. **21**(5), 1229–1240 (2012)
51. Wei, H., et al.: High on/off capacitance ratio RF MEMS capacitive switches. Micromach. Microeng. **27**(5), 055002 (2017)
52. Park, J.Y., et al.: Monolithically integrated micromachined RF MEMS capacitive switches. Sensors **89**, 88–94 (2001)
53. Persano, A., et al.: Ta_2O_5 thin films for capacitive RF MEMS switches. J. Sens. **2010**, 5 (2010)
54. Wang, G., et al.: Novel reliable RF capacitive MEMS switches with photodefinable metal-oxide dielectrics. J. Microelectromech. Syst. **16**, 550–555 (2007)
55. Kogut, L.: The influence of surface topography on the electromechanical characteristics of parallel-plate MEMS capacitors. J. Micromech. Microeng. **15**, 1068–1075 (2005)
56. Yu, A.B., Liu, A.Q., Zhang, Q.X., Hosseini, H.M.: Effect of surface roughness on electromagnetic characteristics of capacitive switches. J. Micromech. Microeng. **16**(10), 2157 (2006)
57. Goldsmith, C.L., Forchand, D.I.: Temperature variation of actuation voltage in capacitive MEMS switches. IEEE Microw. Wireless Compon. Lett. **15**, 718–720 (2005)
58. Hosseinzadeh, S., Zehtabchi, A.R., Habibnejad. M.: Determination the effects of structural parameters on pull down voltage of RFMEMS switches. In: Microwave Conference. IEEE (2007)
59. Agarwal, S., Kashyap, R., Guha, K., Baishya, S.: Modeling and analysis of capacitance in consideration of the deformation in RF MEMS shunt switch. Superlattices Microstruct. (2016). https://doi.org/10.1016/j.spmi.2016.10.022

60. Philippine, M.A., et al.: Experimental validation of topology optimization for RF MEMS capacitive switch design. J. Microelectromech. Syst. **22**, 1296–1309 (2013)
61. Muldavin, J.B., Rebeiz, G.B.: High isolation CPW MEMS shunt switches—part I: modeling. IEEE Trans. Microw. Theory Techn. **48**, 1045–1052 (2000)
62. Ansari, H.R., et al.: Design and simulation of a novel RF MEMS shunt capacitive switch with a unique spring for Ka-band application. J. Microsyst. Technol. **25**(2), 531–540 (2018)
63. Mafinejad, Y., et al.: Design and simulation of a high isolation RF MEMS shunt capacitive switch for C-K band. IEICE Electron. Exp. **10**, 1–8 (2013)
64. Rebeiz, G.M., Entesari, K., Reines, I.C., Park, S.-J., El-Tanani, M., Grichener, A., Brown, A.R., et al.: Tuning into RF MEMS. IEEE Microw. Mag. **10**, 55–72 (2009)
65. Ma, L.Y., et al.: A novel design of a low-voltage low-loss T-match RF-MEMS capacitive switch. In: Microsystem Technologies (2017)
66. Ravirala, A.K., et al.: Design and performance analysis of uniform meander structured RF MEMS shunt switch along with perforations. J. Microsyst. Technol. **24**(2), 901–908 (2017)
67. Jayavardhani, K., et al.: Design and simulation of low actuation voltage shunt RF MEMS shunt capacitive switch with serpentine flexures & regular perforations. Int. J. Eng. Technol. **7**, 4–8 (2018)
68. Sharma, A., Shah, A., Bharti, R.: Design & simulation of low actuation voltage perforated shunt RF MEMS switch. Int. J. Eng. Tech. Res. (IJETR) **3**(6) (2015)
69. Guha, K., Laskar, N.M., Gogoi, H.J., Borah, A.K., Baishnab, K.L., Baishya, S.: Novel analytical model for optimizing the pull-in voltage in a flexured MEMS switch incorporating beam perforation effect. Solid-State Electron. **137**, 85–94 (2017)
70. Ashby, M.F.: Material Selection in Mechanical Design, 2nd edn. Butterworth-Heinemann, Oxford, UK (1999)
71. Guisbiers, G., et al.: Material selection procedure for RF-MEMS. Microelectron. Eng. **87**, 1792–1795 (2010)
72. Lahiri, S.K., Saha, H., Kundu, A.: RF switch: an overview at a glance. In: 4th International Conference on Computer and Devices for Communication, Kolkata, Dec 2009
73. Callister, W.D.: Material Science and Engineering: An Introduction, 7th edn. Wiley, New York (2007)
74. Jlassi, B., Merdassi, A.: Design methodology of a high power RF MEMS switch for wireless communication. In: 4th Annual Caneus Fly by Workshop, Montreal QC, June 2011
75. Wang, G., et al.: A high performance tunable RF MEMS switch using barium strontium titanate (BST) dielectrics for reconfigurable antennas and phased arrays. In: IEEE Antennas and Wireless Propagation Letters, vol. 4, Aug 2005
76. Tan, S.G., et al.: Electromechanical modelling of high power RF-MEMS switches with ohmic contact. In: 2005 European Microwave Conference, Oct 2005
77. Renies, I., Pillans, B., Rebeiz, G.M.: Thin-film aluminium RF MEMS switched capacitor with stress tolerance and temperature stability. J. Microelectromech. Syst. **20**, 193–202 (2011)

Design of Generalized Rational Sampling Rate Converter Using Multiple Constant Multiplication

K. Gayathri, B. Aravind Krishna and Navin Kumar

Abstract In this paper, a multirate sampling filter is designed by using multiple constant multiplication (MCM). This is achieved by replacing the multipliers in generalized rational sampling rate converter (GRSRC) with codes derived from MCM algorithm. Using the existing generalized rational sampling rate converter structure, we achieved reduced computational complexity since the delay requirements are significantly reduced. However, in order to lower the cost of the hardware, we used a multiplierless approach. In addition, the area utilization of this modified structure has been seen to be significantly reduced in comparison to that of the unintegrated structure with a comparable trade-off in terms of the speed and power requirement. Simulink model is developed, and FPGA implementation is completed. It is seen that the computational performance and many other parameters are improved.

Keywords Multirate sampling · Polyphase FIR filter · Generalized sampling rate converter · Multiple constant multiplication

1 Introduction

Multirate digital signal processing is very important in some of the applications [1–3]. For example, conversion between digital audio tape and compact disc format requires sampling rate conversion from 48 to 44.1 kHz. Different sampling rates are often required to increase the system efficiency. Multirate digital signal processing mostly deals with linear filtering. For applications involving nonlinear filters and sampling rate conversions, linear multirate method would be difficult. In this situation, nonlinear filters must be used. However, they are computationally complex [4].

K. Gayathri (✉) · B. A. Krishna · N. Kumar
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

N. Kumar
e-mail: navinkumar@ieee.org

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_44

On the other hand, finite-length impulse response (FIR) filters are employed for sampling rate conversions as it is simple to obtain efficient realizations using polyphase decomposition [5].

The basic structure of multirate converters comprises polyphase FIR filters. Nth-order polyphase FIR filters are decomposed into sub-banks which obtain output at different delays. The sub-banks in the polyphase decomposition of a FIR transfer function also result in FIR filters [6]. This decomposition leads to a parallel structure. Some of the applications of polyphase digital filters are multirate nonlinear filters and nonlinear echo cancellation.

Polyphase decomposition [7] is vital in the design of decimation or interpolation FIR filter. Decimation [8] and interpolation FIR filters are realized using down sampler, up sampler, adders, multipliers and delay elements. Down sampling is the method of lowering the sampling rate by discarding $(M - 1)$ samples. Up sampling is the technique of increasing the sampling rate by inserting $(L - 1)$ zeroes in between existing samples, where M and L are down sampling and up sampling factors, respectively. However, up/down sampling sometimes lead to aliasing and imaging. Filters in sampling rate converters are used to reduce this effect. Additionally, polyphase decomposition of multirate converters improves the efficiency of the system as the computational complexity is reduced.

As M and L are considered to be integer values, decimation and interpolation operations are fairly straightforward. However, if the system specification calls for sampling rate change by a rational factor [9], then design of rational sampling rate converter becomes necessary. The conventional rational sampling rate converter consists of an interpolator followed by a decimator. For an input sampling rate of f_m, the interpolator creates an intermediate sampling rate of Lf_m. This is further decimated by a factor of M, resulting in an output sampling rate of $(L/M)f_m$. Rational sampling rate converter mandates L and M to be co-prime numbers [10]. In a bid to enhance the efficiency of rational sampling rate converter, few structures have been proposed. One such structure states that polyphase rational sampling rate converters can be designed as matrix multiplications and parallel delay chains [6]. Another such structure exploits the coefficient symmetry of linear-phase FIR filter. In this case, the overall complexity is greatly reduced in comparison with cases where the coefficient symmetry is not exploited [11]. However, these structures are either lacking a general technique for the two cases of $L > M$ and $L < M$ or are made only for system-specific applications. To address these challenges, Kumar et al. in [12] discuss a generalized structure which can implement arbitrary and co-prime values of L and M. Furthermore, the number of delay elements required is reduced significantly. A structure [13] exploits the use of expensive multipliers. In addition, according to the hardware specifications, the area utilized by the multipliers is also expensive [14].

In this work, we have attempted to improve upon the existing structure by integrating the structure with multiple constant multiplication (MCM) concept, effectively making the system multiplierless. This can be achieved by replacing the multipliers in generalized rational sampling rate converter (GRSRC) with codes derived from

MCM algorithms. MCM algorithms are used for the realization of FIR filters, cryptosystems, LTI controllers, etc. The MCM-integrated GRSRC structure was implemented in Simulink environment, followed by VHDL simulation and synthesis in Xilinx ISE design suit. The area, adder, power and speed requirement of the GRSRC structure are compared with the MCM-integrated GRSRC structure. To the best of the knowledge on available literature, this has not been used.

In Sect. 2 of this paper, we elaborate on the GRSRC structure, while Sect. 3 presents the MCM method used to improve upon this structure. Results are presented and discussed in Sect. 4, while conclusion is summarized in Sect. 5.

2 Overall System Model

The generalized rational sampling rate converter structure can be implemented for all co-prime values of L and M. This structure yields an output sample rate of $(L/M)f_m$, where the input sample rate is f_m [11]. The input is required to pass through a tapped shift register or a uniform discrete shift register, after which it is divided into M branches, as shown in Fig. 1. Each of the M branches comprises a down sampler, whose decimation factor is M. There are L subfilters in every down sampler branch, followed by an up sampler, whose interpolation factor is L. The L subfilters within each branch consist of a suitable number of adders and multipliers, which perform appropriate mathematical operations on the input signal with the filter coefficients. The output of the M branches is delayed and summed appropriately, to generate the desirable output. As the representation of the filter coefficients can be obtainable

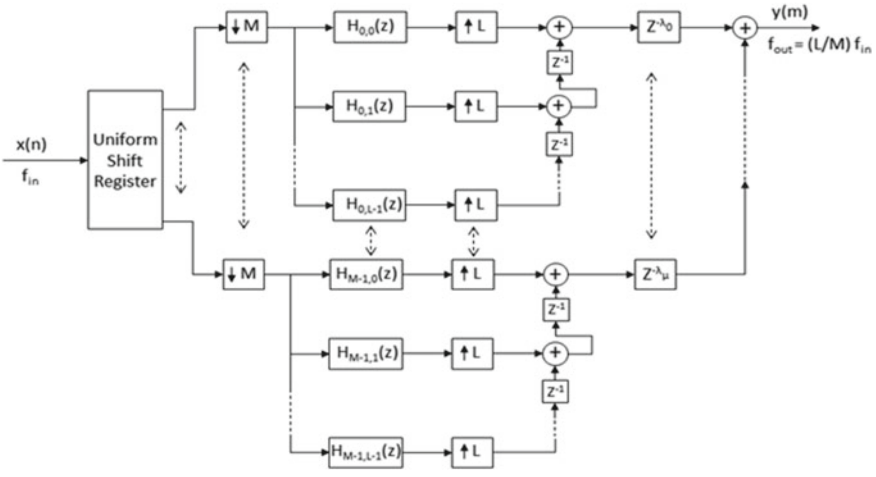


Fig. 1 Generalized rational sampling rate converter for M-valued decimator and L-valued interpolator

from a single equation, this structure is easy to implement. The advantages of this structure also include a reduction of the delay requirement from the original limit of $(L+1)$.

However, the requirement of multipliers in each of the subfilter block makes the hardware implementation of this structure fairly expensive to realize. For this reason, we attempt to modify the structure using a multiplierless approach. This is accomplished using the MCM concept. The multiplications in each of the L subfilters of GRSRC structure are now realized using an MCM algorithm. This makes the GRSRC structure multiplierless. MCM-based design is further discussed in the next section.

3 MCM-Based Design

An efficient way of implementing a multiplierless system is by using the method of MCM. This method essentially uses only shift, addition and subtraction operations for the multiplication of a variable input data with multiple fixed-point constants. Subsequently, a reduction in the hardware cost and system complexity can be expected, as it eliminates the usage of costly multipliers. Additionally, the MCM method also provides area and power efficiency in the system with less computational complexity [15].

Figure 2 shows the usage of MCM in the direct-form structure and the transposed direct-form structure of a FIR filter [15]. The dashed line shows the MCM

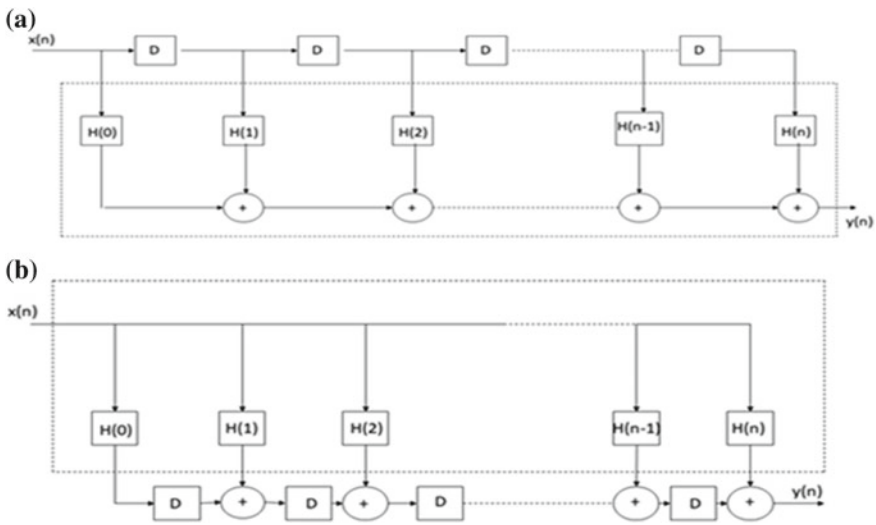


Fig. 2 **a** Direct-form FIR filter, **b** transposed direct-form FIR filter

block, which distinctly shows the multiplication of the variable input with the constant coefficients of the filter. Thus, all the addition, subtraction and shift operations required to realize a multiplication will be referred to as a multiplier block. Since the hardware required for addition and subtraction is comparable, the subtractors will also be referred to as adders. Normally, four different algorithms are available to realize MCM [12]: (i) digit-based recoding, (ii) common subexpression elimination (CSE) algorithms, (iii) graph-based algorithms and (iv) hybrid algorithm. We have used the graph-based algorithms in this work to implement the MCM block. Graph-based algorithms normally use iterative technique to construct the graph representing the multiplier block following a bottom-up approach. Some common algorithms of graph-based are BHM algorithm [16], RAG-n algorithm and HCUB algorithm. The BHM algorithm uses an inferior heuristic and hence is not very precise. The RAG-n algorithm is heavily dependent on a precomputed table to obtain the heuristic portion of the algorithm. The best available method also only works up to 19 bits. The working of this algorithm is also limited to 19 bits. Thus, the BHM and the RAG-n algorithm are not very effective when the constants are more than 19 bits [12]. This disadvantage is overcome by the HCUB algorithm. Unlike the RAG-n algorithm, the heuristic part of the HCUB algorithm does not require a precomputed table which makes the latter computationally efficient and less time consuming. Hence, we have used the HCUB algorithm in this work. The HCUB algorithm is shown in Fig. 3, where T is the set of target constants, S is the successor set, W is the intermediate set and R is the set of constants to derive. $H(R, S, T)$ stands for the heuristic of (R, S, T).

Using the HCUB algorithm, multiplication operation can now be realized using only addition, subtraction and shift operations. Applying this algorithm, the MCM problem is incorporated into the L subfilters of the GRSRC structure. The 17-tap generalized rational sampling rate converter designed for $M = 2$ and $L = 3$ is shown in Fig. 4. The filter coefficients of each branch can be easily represented as

$$H_{\mu,\gamma}(z) = \sum_{j=0}^{b} h_{\left(M(\gamma + jL) + M\lambda_{\mu} - L\mu\right)z^{-j}} \tag{1}$$

for $0 \le \mu \le (L - 1)$ and $0 \le \gamma \le (M - 1)$
where $\lambda_{\mu} = (\mu L + k)/M$ and $0 \le k \le M - 1$

The delay requirements of this structure are thus significantly reduced to $\sum_{\mu=0}^{M-1} \lambda_{\mu}$, which can be further optimized by minimizing λ_{μ} to have an upper limit of $\lambda_{\text{new}} = \max(\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_{M-1})$ [5].

To obtain the filter coefficients for our analysis, we have used the Simulink digital filter design tool with a cut-off frequency, f_c, of 10 Hz and a sampling frequency, f_s, of 8000 Hz. The standalone GRSRC structure was designed in Simulink following which the generated VHDL code was implemented in Xilinx ISE with ModelSim PE 10.4a as the simulator. The expanded view of subfilter 3, $H_{0,2}(z)$ in this structure is shown in Fig. 5a. The MCM-integrated GRSRC structure replaces all the multipli-

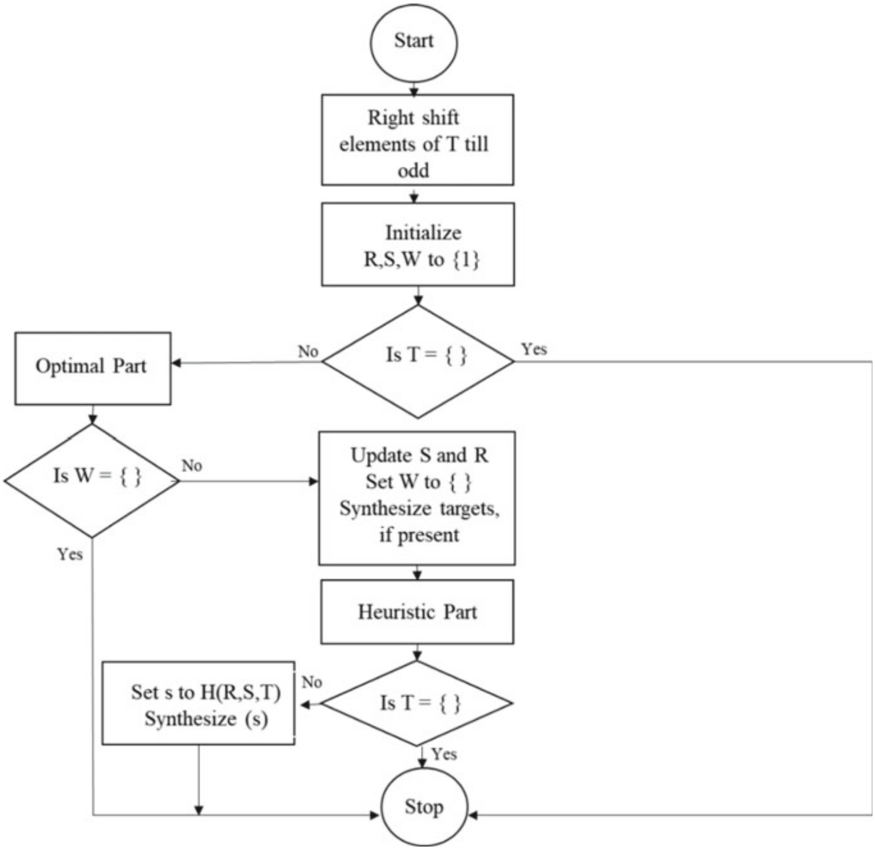


Fig. 3 Flowchart of HCUB algorithm

cations in a subfilter by an MCM block, as shown in Fig. 5b for subfilter 3, $H_{0,2}(z)$. The code for MCM block for each subfilter is derived using the HCUB algorithm and is realized in MATLAB. The multiplications in subfilter 3, $H_{0,2}(z)$, are realized using addition, subtraction and shift operations which are represented as a graphical flow in Fig. 5c.

In order to further reduce the computational complexity as well as the area and the number of adders involved, we have derived the outputs of repetitive subsystems from the same MCM block.

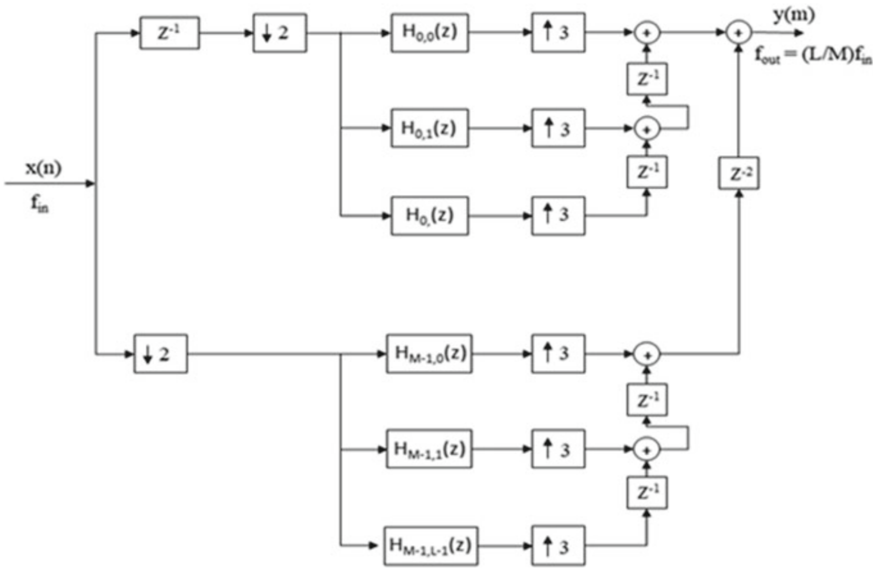


Fig. 4 Generalized rational sampling rate converter for $L = 3$ and $M = 2$

4 Results and Analysis

To understand the working of the 17-tap generalized rational sampling rate converter structure, a digital sine wave input was taken with an input signal frequency, f_{in} of 1 Hz, and a sampling period of 1/100 s. For an $L = 3$ and $M = 2$ structure, we obtained an output as shown in Fig. 6, using ModelSim and Xilinx ISE.

The simulation of the MCM-integrated GRSRC structure with $M = 2$ and $L = 3$ resulted in an output as is shown in Fig. 7a. The digital sine wave input in this case has a sampling period of 1/10 s, while f_{in} is still 1 Hz. The magnified output of this structure depicted in Fig. 7b clearly shows that for every two input samples, three output samples are obtained. The sampling rate is thus successfully changed by a factor of 3/2.

For the synthesis of the two structures, Xilinx Virtex4 XCVSX35—FF68 has been used. The multiplier in this FPGA device is implemented using DSP48s which is a multiply and accumulate circuit. Each DSP48s consists of one 18×18 signed pipelined parallel multiplier and one 48-bit signed adder. The speed and power comparisons are shown in Table 1. Table 1 shows that the speed of the system has reduced and the power consumption is comparable.

The area utilization of both the structures can be seen in Table 2. It can be deduced from Table 2 that the area required for the MCM-integrated GRSRC structure is less than that of the standalone structure (Table 3).

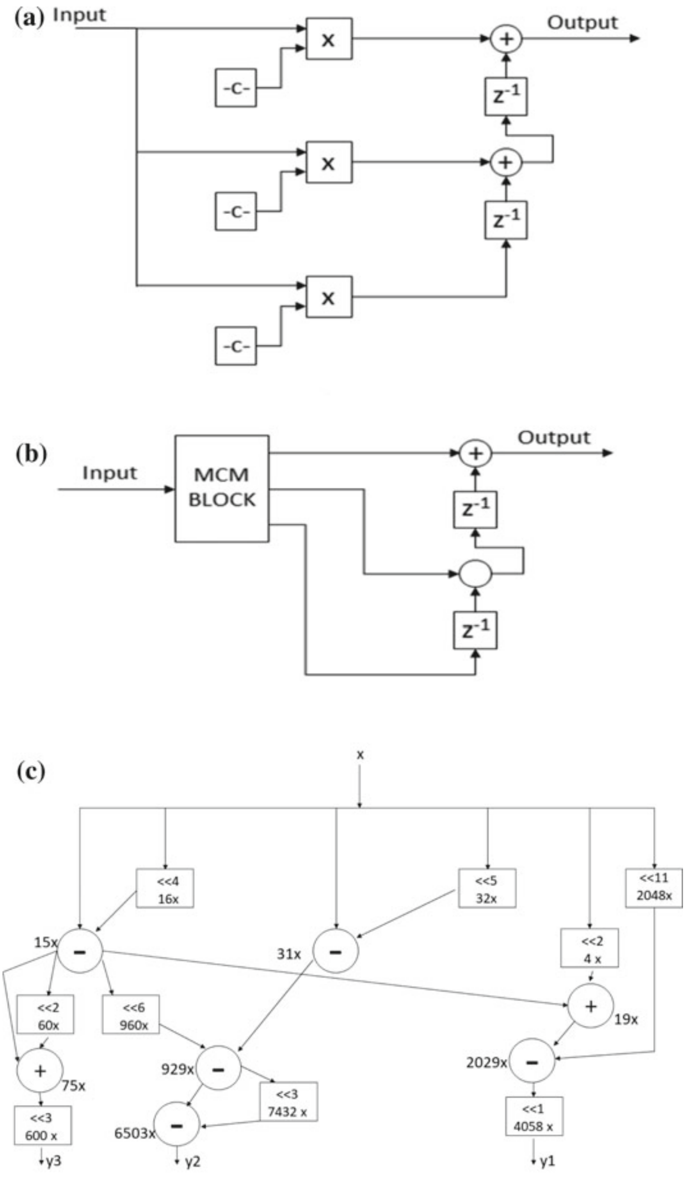


Fig. 5 **a** Contents of subfilter 3, $H_{0,2}(z)$, without using the MCM block, **b** the contents of subfilter 3, $H_{0,2}(z)$ using MCM, **c** HCUB algorithm graph for the subfilter 3

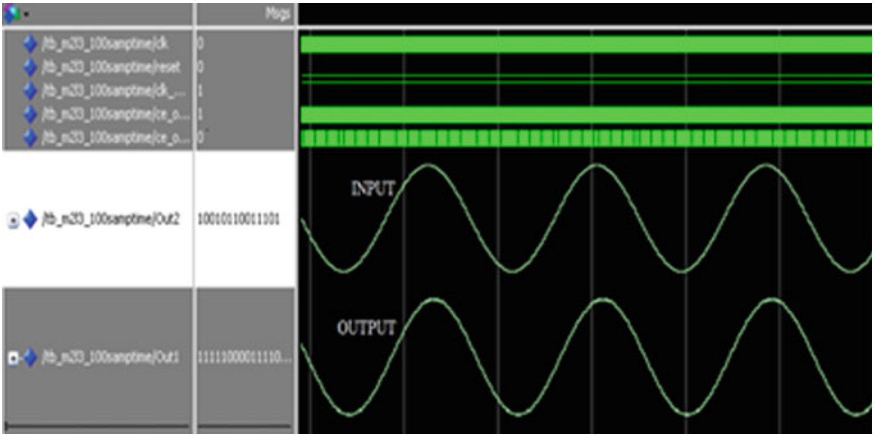


Fig. 6 Generalized sampling rate converter output waveform

5 Conclusion

From the investigative research carried out, we see that the MCM-integrated generalized rational sampling rate converter structure has significantly improved hardware structure than its standalone counterpart in terms of area utilization in the FPGA device. The cost of the structure is evidently reduced as the usage of the costly multipliers has been avoided. However, this improvement is in accordance to a comparable trade-off in terms of the speed and power requirements of the structure. The structure can be further improved by using matrix MCM method and more efficient MCM algorithms.

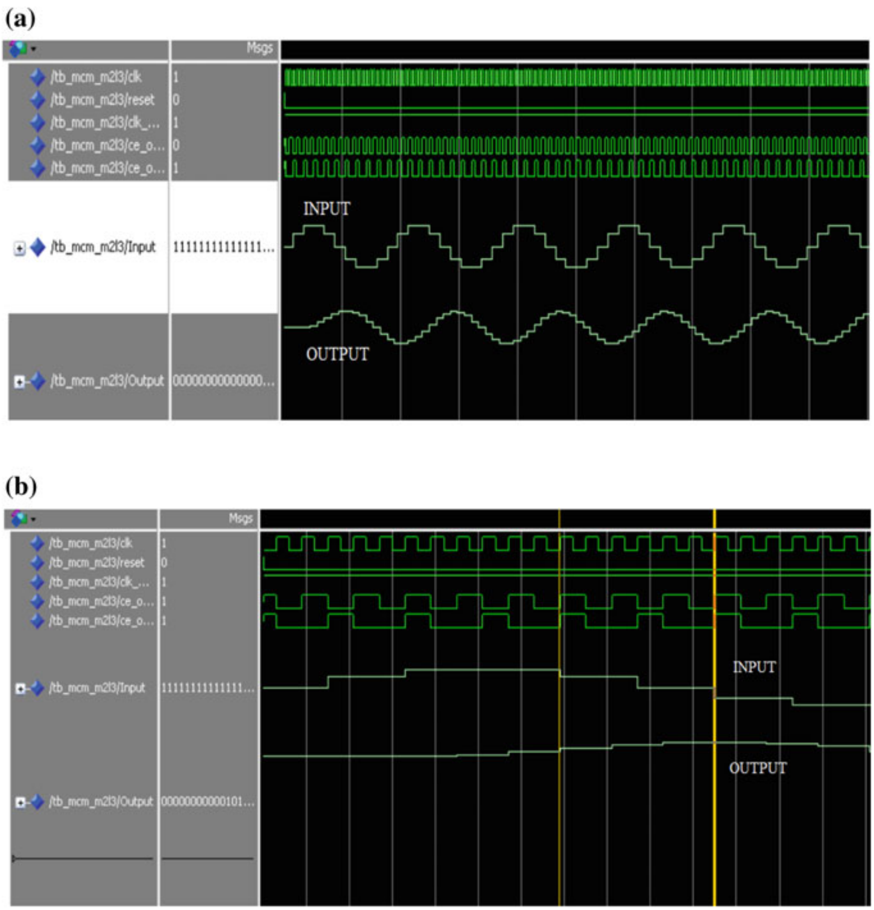


Fig. 7 **a** MCM-integrated generalized sampling rate converter output waveform, **b** magnified output waveform of MCM-integrated generalized sampling rate converter

Table 1 Speed and power comparisons of the structures

Property	GRSRC	MCM-integrated GRSRC
Total power (mW)	462.70	467.76
Speed (MHz)	83.503	76.187

Table 2 Area utilization of the two structures

	Used		% Used		Available
	GRSRC	MCM-GRSRC	GRSRC	MCM-GRSRC	
Occupied slices	381	1121	2	7	15,360
Slice flip-flop	375	825	1	2	30,720
Four-input LUTs	692	2102	2	6	30,720
DSP48s	30	0	15	0	192

Table 3 Adder utilization of the two structures

GRSRC		MCM-integrated GRSRC	
Type of adder	No. of full adders	No. of adders	Type of adder
12-bit adders	192 (16 × 12)	32 bit adders	512 (16 × 32)
16-bit adders	7200 (15 × 32 × 15)	33 bit adders	396 (12 × 33)
		33 bit adder	363 (11 × 33)
Total no. of full adders	7392	Total no. of full adders	1271

References

1. Bregovic, R., Yu, Y.J., Lim, Y.C., Saramaki, T.: Implementation of linear-phase FIR filters fro a rational sampling-rate conversion utilizing the coefficient symmetry. IEEE Trans. Circuits Syst. I Reg. Papers **58**(3), 548–561 (2011)
2. Bellanger, M.G., Bonnerot, G., Coudreuse, M.: Digital filtering by polyphse network: application to sample rate alteration and filter banks. IEEE Trans. Accoust. Speech Signal Process. ASSP-**24**(2), 109–114 (1976)
3. Heightley, J.D.: Review of digital filtering. In: Annual Symposium on Frequency Control, April 1970
4. Schwinscackl, D., Kubin, G.: Polyphase representation of multirate nonlinear filters and its applications. IEEE Trans. Signal Process. **55**, 2145–2157 (2007)
5. Fiala, P.: High performance polyphase FIR filter structures in VHDL language for software defined radio based on FPGA
6. Gustafsson, O., Johansson, H.: Efficient implementation of FIR filter based rational sampling rate converters using constant matrix multiplication. In: Proceedings of 40th Fortieth Asilomar Conference on Signals, Systems and Computers, pp. 888–891, Oct 2006
7. Yim, W.H., Coakley, F.P., Evans, B.G.: Extended polyphase structures for multirate DSP. In: IEE Proceedings F (Radar and Signal Processing), vol. 139(4), pp. 273–277 (1992)
8. Jacob, P., Anoop, B.N.: Design and implementation of polyphase decimation filter. Int. J. Comput. Netw. Wirel. (April 2014)
9. Kaushal, P., Mehra, R.: Performance analysis of fractional sample rate converter using audio applications. IOSR J. Electr. Electron. Eng. (IOSR–JEEE) (Dec 2015)
10. Johannsson, H., Gockler, H.: Two stage based polyphase structures for arbitrary integer sampling rate conversion. IEEE Trans. Circuits Syst. (May 2015)
11. Bregovic, R., Saramaki, T., Yu, Y.J., Lim, Y.C.: An efficient implementation of linear-phase FIR filter for a rational sampling rate conversion. In: Proceedings of IEEE International Symposium on Circuits and Systems, pp. 5395–5398. Island of Kos, Greece (2006)
12. Kumar, A., Yadav, S., Purohit, N.: Generalised rational sampling rate conversion polyphase filters. IEEE Signal Process. Lett. **24**(11) (2017)
13. Voronenko, Y., Puschel, M.: Multiplier less multiple constant multiplication. ACM Trans. Algorithms **3**(2) (2007)

14. Marcos Peiro, M., Wanhammer, L.: High speed, low complexity FIR filter using multiplier block reduction and polyphase decomposition. In: IEEE International Symposium on Circuits and Systems, May 2000
15. Gustafsson, O., Dempster, A.G.: On the use of multiple constant multiplication in polyphase FIR filters and filter banks. In: Proceedings of 6th Nordic Signal Processing Symposium, 9–11 June, 2004
16. Dempster, A.G., Macleod, M.D.: Use of minimum-adder multiplier blocks in FIR digital filters. IEEE Trans. Circuits Syst. II, Analog. Digit. Signal Process. **42**(9), 569–577 (1995)

Comparison of Decoupled and Coupled PWM Techniques for Open-End Induction Motor Drives

M. Rama Prasad Reddy, Karanam Deepak and M. Venkateswaralu

Abstract In this paper, multilevel inverter configuration called dual-inverter (DI) topology is presented in favor of asynchronous motor drive. The topology is easy in constriction as well as easy to operate when compared with other multilevel inverter configurations. Two special kinds of PWM methods are presented in favor of DI topology to recover the excellence of production voltage as well as decrease the common-mode voltage value. To test the concert of the PWM methods in favor of DI topology, first theoretical studies are carried and next model analysis is carried; here, MATLAB simulation model as well as outcome is presented.

Keywords Induction motor drives · Pulse-width modulation (PWM) · Multilevel inverter · DI topology

1 Introduction

Through the advancement during power electronic tools, electrical drive is gaining significance into hybrid electric vehicles (HEVs) and industrial applications. Among the electrical drives, VSI-fed induction motor drives are popular. Different PWM methods are worn for the controlling of production voltage source as well as frequency of inverters [1–6]. At high switching frequencies sharp edges of common mode voltage (potential across neutral point of induction motor and DC link) causes common mode currents. Which reduce the life of motors and bearings [7–9].

To reduce the general method voltage as well as improve the feature of production voltage sources, multilevel inverter (MLI)-fed drives are gaining importance. Various MLI methods like diode-clamped MLI (DC MLI), capacitor clamp, H-bridge MLI and dual-inverter methods are discussed in related work [10–15]. Among this dual-based inverter OEWIM drives (DI-fed open-end winding induction motor drive) are gaining importance [15]. Two two-level inverters can be controlled either independently or dependently. Based on this, PWM methods are divided into decoupled and coupled PWM methods [16–19]. The comparisons of both these PWM methods

M. Rama Prasad Reddy (✉) · K. Deepak · M. Venkateswaralu
G. Pullaiah College of Engineering and Technology, Kurnool, Andhra Pradesh, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_45

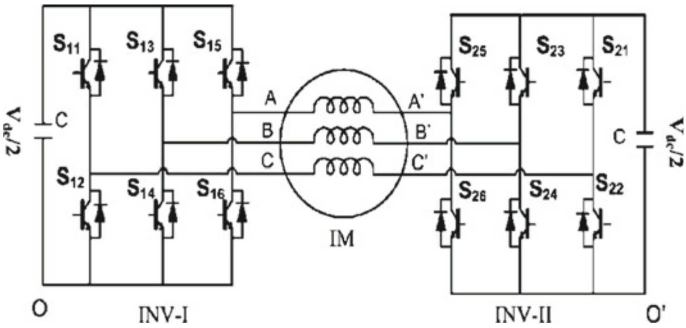


Fig. 1 Dual-inverter-fed OEWIM drive

meant for dual inverters are presented in this paper. The implementation of both PWM techniques was presented in simplified scalar carrier comparison approach.

2 Dual-Based Inverter Configuration

The circuit connection of DI-fed OEWIM drives is revealed in Fig. 1. In Fig. 1, both the MLIs are connected from a remote DC voltage source. Hence, circulating zero-sequence current (ZSC) will not be present flowing in the drive, but ZS voltage appears across the unenthusiastic rail of DC-side voltage sources, i.e., diagonally the potential across o and o'. The voltage side across the ao, bo, co, ao', bo', co' is called the pole-type voltages, and aa', bb' and cc' are called the phase voltage.

Hence, the efficient pole-type voltage source can be calculated as in Table 1. The helpful pole-type voltage source has three different levels. Hence, the dual-type inverter pattern is called as three-level inverter configuration.

3 PWM Methods Designed for Dual-Inverter (DI) Configuration

Each inverter can be controlled independently or dependently to generate three-level voltage source. In this, PWM methods are classified into decoupled and coupled PWM techniques. In decupled PWM technique, each inverter is treated separately whereas in coupled PWM techniques both the inverters are considered as single unit.

Table 1 Switches to be turned ON and corresponding pole voltages and effective pole voltage

Switching states in Inverter-1	Pole voltage of Inverter-1	Switches ON position Inverter-2	Pole voltage of Inverter-2	Calculated effective pole-type voltage source
S_{14} or S_{16} or S_{12} is ON S_{11} or S_{13} or S_{15} is OFF	*0*	S_{25} or S_{23} or S_{21} is ON S_{22} or S_{24} or S_{26} is OFF	$V_{dc}/2$	$-V_{dc}/2$
S_{14} or S_{16} or S_{12} is ON S_{11} or S_{13} or S_{15} is OFF	*0*	S_{22} or S_{24} or S_{26} is ON S_{25} or S_{23} or S_{21} is OFF	0	0
S_{11} or S_{13} or S_{15} is ON S_{14} or S_{16} or S_{12} is OFF	$V_{dc}/2$	S_{25} or S_{23} or S_{21} is ON S_{22} or S_{24} or S_{26} is OFF	$V_{dc}/2$	0
S_{11} or S_{13} or S_{15} is ON S_{14} or S_{16} or S_{12} is OFF	$V_{dc}/2$	S_{22} or S_{24} or S_{26} is ON S_{25} or S_{23} or S_{21} is OFF	0	$V_{dc}/2$

3.1 Decoupled PWM Techniques

In a dual inverter is a 3-∅ inverter, considered 3- modulating pointers to each inverter as given in (1) and (2). Because of separate control (decoupled) of both the inverters, separate set of modulating signals are considered. To get maximum output voltage, a phase shift of π is considered between modulating signals of Inverter-1 and Inverter-2.

$$V_{\text{ref}\,a} = \cos(\omega t); \quad V_{\text{ref}\,b} = \cos(\omega t - 2\pi/3); \quad V_{\text{ref}\,c} = \cos(\omega t - 4\pi/3)R \tag{1}$$

$$V_{\text{ref}\,a'} = \cos(\omega t - \pi); \quad V_{\text{ref}\,b'} = \cos(\omega t - 5\pi/3); \quad V_{\text{ref}\,c'} = \cos(\omega t - \pi/3) \tag{2}$$

These modulating signals are modified to new modulating pointers by addition of a ZSC signals toward the old modulating signals (1) and (2).

$$V_{zs} = \frac{V_{dc}}{2}(2a_0 - 1) - a_0 V_{\max} + (a_0 - 1) \tag{3}$$

Hence, new modulating signals are obtained by adding old modulating signals and zero sequence as given (4).

$$V^{*}_{\text{ref}} = V_{\text{ref}} + V_{zs} \tag{4}$$

3.2 Coupled PWM Methods

In coupled PWM method, the DI configuration is treated as signal of three-phase three-level inverter configuration. Hence in this type of coupled PWM techniques, one set of modulating signal as given in (5) is considered. The modulating pointers are compared through two-stage shifting carrier pointers toward produce control signals in favor of DI configuration.

$$\mathrm{V}_{\mathrm{ref}\,a} = \cos(\omega t); \quad \mathrm{V}_{\mathrm{ref}\,b} = \cos(\omega t - 2\pi/3); \quad \mathrm{V}_{\mathrm{ref}\,c} = \cos(\omega t - 4\pi/3) \tag{5}$$

$$V_{zs} = \frac{V_{\mathrm{dc}}}{2}(2a_0 - 1) - a_0 V_{\max} + (a_0 - 1) \tag{6}$$

$$V^{*}_{\mathrm{ref}} = V_{\mathrm{ref}} + V_{zs} \tag{7}$$

The modulating signals given in (5) are modified to new modulating signals by adding a ZSC to the old modulating signals (6) and (7). *ao* is constant among zero as well as one. Here for the modulating signal shown in Figs. 2 and 3, *ao* is taken as 0.5. *ao* can take any value between 0 and 1 to yield different types of modulating signals. In this paper, only one type is considered.

The invention of PWM method of signals to a phase with coupled PWM method is revealed in Fig. 3. From the control signals, it is experiential that through coupled

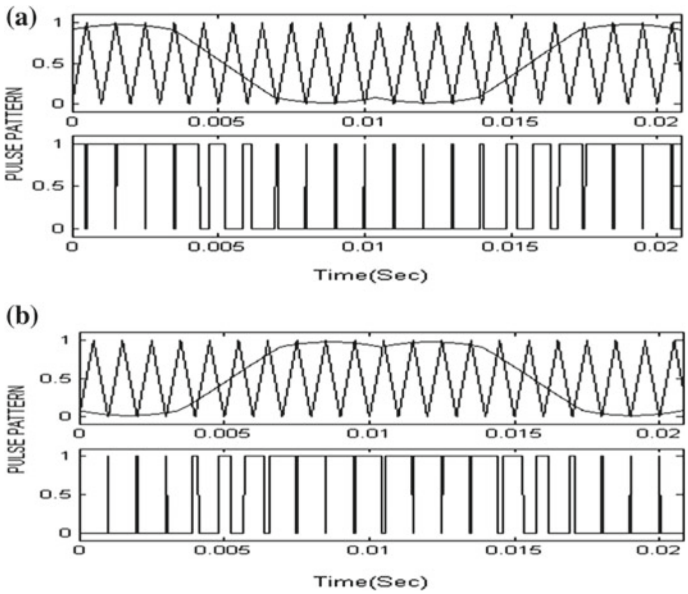


Fig. 2 Understanding of decoupled PWM method to **a** Inverter-1 and **b** Inverter-2

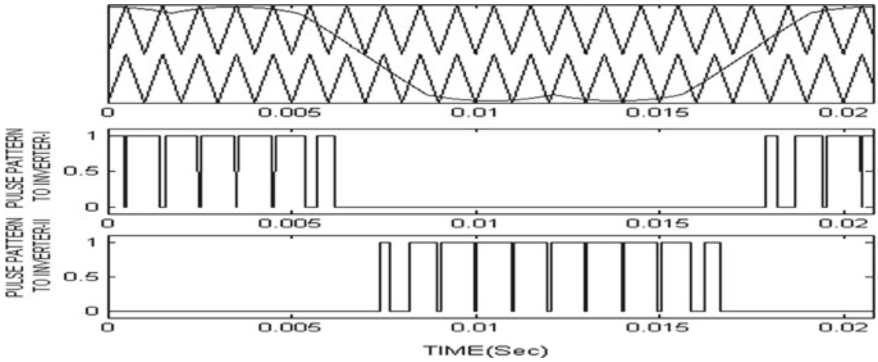


Fig. 3 Realization of coupled PWM technique to **a** Inverter-1 and **b** Inverter-2

PWM method while individual inverter is switched further inverter is clamped but with decoupled PWM techniques both inverters are continuously switched. Hence, it is observed with coupled PWM techniques switching losses are reduced.

The performance of decoupled and coupled PWM techniques for DI-fed OEWIM drive is tested in MATLAB simulation setting. The simulation limit use during the analysis is DC link voltage 540 V, switching frequency 3 kHz and induction motor parameters 4 Hp, 400 V, 50 frequency Hz, 1430 speed rpm. The results of pole voltages, effective pole-type voltage source of Inverter-1 as well as Inverter-2, efficient phase voltage sources of the drive and common-mode voltage with decoupled and coupled PWM methods are revealed in Fig. 4 as well as Fig. 5.

With the application of key DC-side voltage source of 540 V (V_{dc}) to the DI topology, the pole voltages with both the PWM techniques are having instantaneous values of 0 or 270 ($V_{dc}/2$). The efficient pole-type voltage source of the drive has

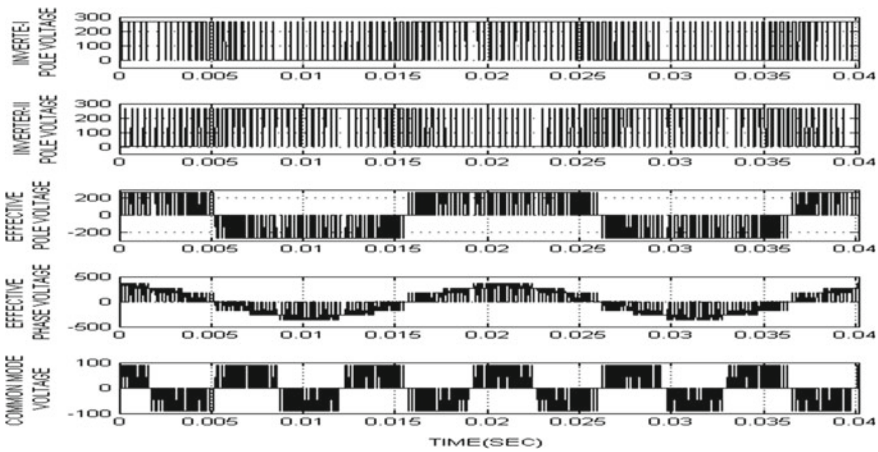


Fig. 4 Different voltage source plots with decoupled PWM

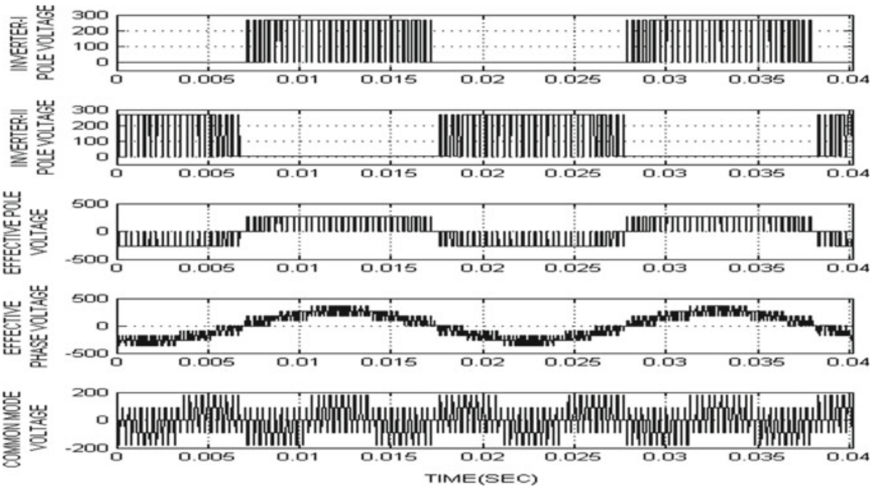


Fig. 5 Different voltage plots with coupled PWM

voltage levels of −270 ($-V_{dc}/2$), 0, 270 ($V_{dc}/2$). Hence, the DI configuration is termed three-level inverter configuration. Though the effective pole voltages have related no. of voltage levels, the efficient point voltage source has different voltage levels because of continuous switching of both the inverters in decoupled PWM technique.

The number of voltage sources of level created during effective pole voltage source is identical through the PWM methods. But the magnitude of voltage levels is different. The voltage levels in efficient phase voltage source have a level of 360 ($2V_{dc}/3$), 180 ($V_{dc}/3$) and 270 ($V_{dc}/2$) with decoupled PWM technique, whereas with coupled PWM method the levels of voltage source created in efficient phase voltage are 180 ($V_{dc}/3$) and 90 ($V_{dc}/6$). The high magnitude of voltage levels created with decupled PWM technique is the main cause of increasing the harmonics. Hence, the total harmonic distortion (THD) plots exposed in Fig. 6 with decoupled PWM technique are high when compared with coupled PWM technique. Moreover because of continuous switching of both the inverters at frequency of 3 kHz, the voltage pulses are created at frequency of 6 kHz. Hence in THD plots, high magnitude of harmonics is created next to as well as approximately harmonics of 6 kHz. Through decoupled PWM method, CMV has a scale of $-V_{dc}/6$, 0 in addition to $V_{dc}/6$. Excluding through coupled PWM technique, the voltage levels are $-V_{dc}/3$, $-V_{dc}/6$, 0, $V_{dc}/6$ and $V_{dc}/3$. The high magnitude of common mode voltage (CMV) with coupled PWM causes circulating current which have motor bearings and reduce the life of motor.

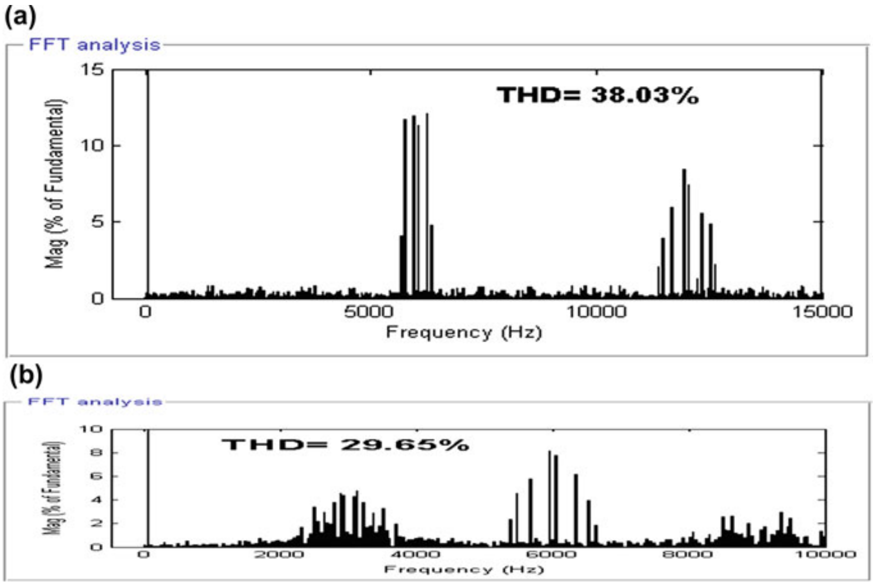


Fig. 6 Total harmonic distortion of effective phase voltage with **a** decoupled PWM and **b** coupled PWM

4 Summary

The implementation of coupled and decoupled PWM techniques for DI fed OEWIM drive is presented. The implementation of coupled and decoupled PWM techniques is carried out in carrier comparison approach. From the results it is concluded that coupled PWM techniques improve the quality of output voltage but generate high CMV, whereas decoupled PWM techniques reduce the CMV but also reduce the quality of output voltage.

References

1. Holtz, J.: Pulse width modulation—a review. IEEE Trans. Ind. Electron. **39**(5), 411–429 (1991)
2. Willi, H., Broeck, V.D., Skudelny, H.C., Stanke, G.V.: Study along with recognition of a PWM base on voltage source space vectors. IEEE Trans. Ind. Appl. **24**(1), 142–150 (1988)
3. Seung-Ki, J.S.K.: A New Voltage Intonation Method of the Space Vector PWM, pp. 742–747. IEEE (1995)
4. Narayanan, G., Ranganathan, V.: Triangle evaluation as well as hole vector approach toward PWM in inverter fed drive. JISC **80**, 409–427 (2000)
5. Reddy, M.R., Reddy, T.B., Subba Rayudu, D.: Space vector base novel as well as easy combined PWM algorithm in favor of asynchronous motor drives. IEEE **2**, 9–32 (2010)
6. Da Silva, E., Santos, E., Jacobina, C.: PWM strategy. IEEE Ind. Electr. Magnet. **5**, 31–45 (2011)
7. Mutze, A., Binder, A.: Don't mislay your bearings Mitigation method in favor of behavior currents in inverter-supplied drive system. IEEE Trans. Mag. **12**, 22–31 (2006)

8. Muetze, A., Tamminen, J., Ahola, J.: Manipulate of motor running parameter happening liberation bearing current movement. IEEE Trans. Ind. Appl. **47**(4), 1767–1777 (2011)
9. Willwerth, A., Roman, M.: Electrical manner spoil a lurk difficulty in inverter driven traction induction motors. In: Proceedings of the ITEC, pp. 1–6, June 2013
10. Nabae, I., Takahashi, I., Akagi, H.: A novel unbiased tip clamped PWM inverter. IEEE Trans. Ind. Appl. **17**(5), 518–522 (1981)
11. Rodríguez, J., Lai, J.: MLI: a review of methods, controls, as well as application. IEEE Trans. Electron. **49**, 724–737 (2002)
12. Kouro, S., Malinowski, M., Gopakumar, K., Pou, J., Franquelo, L.G.: Current advance as well as engineering application of multilevel converters. IEEE Trans. Ind. Electron. **57**(8), 250–255 (2010)
13. Zhang, Y., Zhao, Z., Zhu, J.: A fusion PWM functional toward high-power 3-level inverter-fed asynchronous motor drives. IEEE Trans. Ind. Electron. **58**(8), 3409–3420 (2011)
14. Adam, G.P., Finney, S.J., Massoud, A.M., Williams, B.W.: Capacitor balance issues of the diode-clamped multilevel inverter operated in a quasi two-state mode. IEEE Trans. Ind. Electron. **55**(8), 3088–3099 (2008)
15. Stemmler, H., Guggenbach, P.: Configurations of high-power voltage-source inverter drives. In: Proceedings of the EPE Conference, pp. 7–14. Brighton, UK, Sept 1993
16. Somasekhar, V.T., Srinivas, S., Kumar, K.K.: Effect of zero-vector placement in a dual-inverter fed open-end winding induction motor drive with alternate sub-hexagonal center PWM switching scheme. IEEE Trans. Power Electron. **23**(3), 1584–1591 (2008)
17. Reddy, B.V., Somasekhar, V.T., Kalyan, Y.: Decoupled space-vector PWM strategies for a four-level asymmetrical open-end winding induction motor drive with waveform symmetries. IEEE Trans. Ind. Electron. **58**(11), 5130–5141 (2011)
18. Reddy, M.H.V., Reddy, T.B., Reddy, B.R., Suryakalavati, M.: Reduction of common mode voltage in asymmetrical dual inverter configuration using discontinuous modulating signal based PWM technique. J. Power Electron. **15**(6), 1524–1532 (2015)
19. Reddy, M.H.V., Reddy, T.B., Reddy, B.R., Suryakalavati, M.: Random PWM technique for dual-inverter-fed vector-controlled induction motor drive. J. Control Autom. Electr. Syst. **27**(1), 60–68 (2016)

Computer Tools for Energy Systems

Atyam Nageswara Rao, P. Vijayapriya, M. Kowsalya and S. Suman Rajest

Abstract This manuscript comprises a brief review of distinct tools that are used for analysing the renewable integration. Though numerous tools are used, a few are considered for explanatory purposes using the web sources of various tool developers. The details in this manuscript give the reader the necessary information to select and identify a suitable tool for renewable energy integration and its analysis for diverse objectives. This manuscript reveals that there is no tool exclusively which addresses all the problems that are related to renewable energy integration. Every objective has its own tool fulfilling its criterion. All the tools mentioned in this manuscript are related to typical applications for analysing the energy system from the state level to the national level. The details of the tools mentioned for analysis are looked at various factors like their energy sector, accounted technology, parameters, availability of tools, etc. Lastly, this manuscript provides information related to direct the decision-maker.

Keywords Energy tools · Renewable energy · Renewable energy integration · Power energy system

1 Introduction

The electrically powered industry plays its role in meeting the enormous demands of the nation. The experts predicted that the capital expansion for the infrastructure and the operations along with our set constraints are vital to the electric companies and utilities. The complexity increases at the time of operation so there is a tremendous need in improving the operation with the available tools. Usually, training facilities

A. Nageswara Rao (✉)
Department of Electrical and Electronics Engineering, Presidency University, Bangalore, India

P. Vijayapriya · M. Kowsalya
School of Electrical Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

S. Suman Rajest
Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamil Nadu, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_46

for the power system operation requires the hardware and software setup which are expensive in terms of growth are mainly because of distinct platforms integration. Years back, simulation has extensively adopted as a means to comprehend the power system operation and control. Today, computer-based tools for handling the problems on the specific application are a basic requirement. Efforts and expectations concerned with the tool designer is a sign of development. The software tool packages for energy systems are classified into two classes such as commercial software and academic aimed software which are well tested computationally and readily available in market following an all in one philosophy. In the face of their fullness, the aforementioned software can result cumbersome for academic purposes. More importantly, the commercial software does not allow the addition of new algorithms or alteration of source code. The flexibility and easy prototyping are the utmost crucial phases in academic purpose than the efficacy in computation. Then again, we have a variety of open-source academic research tools which are aimed at a particular aspect in analysing the power system. In the earlier era, quite a few high-level languages have become prevalent in academic purposes.

Only just, various challenges have arisen in present society such as change in climate, energy supply security, as well as economic slump. Consequently, the energy sector, exclusively renewable energy, is being beleaguered to combat these challenges concerned. In the 1970s, because of oil crisis, the interest has been focused considerably upon renewable utilisation. A vital element in this transfer of interest is to articulately show the technical analyses of how the implementation of renewable energy can be done, and its effect on other portions of the power system. For such technical analyses, computer-based tools give the creative solutions by modelling the predefined power systems. Creating novel tool is a time-consuming aspect for analysing each and every problem; hence, it would be highly feasible in accessing the existing tool. Nevertheless, when commencement of a study into the possibilities of renewable energy, it is hard to recognise which energy tool is the utmost appropriate for the study, even with existing literature. Renewable energy systems are relied on one or multiple sources. A hybrid renewable system comprises of two generating options based on units of renewable or fossil fuel or its combination. A few softwares are also developed based on hybrid technology of renewable energy to simplify its design and maximise its use. Therefore, this paper tells about the selection of an appropriate energy tool which provides a brief comparison of a few energy tools used for energy system as well as in distinct platforms of renewable energy.

2 Computer Tools for Energy System

A wide range of understanding is crucial about the features, shortcomings, usage and the choice for the available software tools for the academic studies. Experts have identified seven distinct types of tools such as simulation tool (it is mainly used for operating a given power system for given set of demands in power sector. The time steps in this tool are hourly over a time period of one year), scenario tool (it is the tool pertained to long-term scenario. The time steps for its functionality are one year which combines such one year into a series of years and finally a scenario), equilibrium tool (this tool is to brief the demand–supply behaviour, prices of the economy with the market and also it is to be identified whether the equilibrium is maintained), a top-down tool (it uses a macroeconomic data to find the growth in energy demand and also in the price of energy), a bottom-up tool (it is used for identifying and analysing various energy technologies along with the investment options and its alternatives), operation optimisation tool (it is used to optimise the given power system operation), investment optimisation tool (it is used to optimise the power system investments). In this section, a few tool features are reviewed individually. The tools information is described in the aspects of technology, cost, methodology, etc. Also, history and functionality are also presented for the tools.

2.1 AEOLIUS

In Germany at an institute for industrial production, this simulation tool has come into existence for the power plant dispatch [1]. The first version of the tool is still utilised to till date. The functionality is to analyse the high penetration rates influence of energy carriers which are of fluctuating nature. This tool is performed typically on an energy system with the time step of 15 min over a maximum of one year. Thermal generation and renewable energy technologies are the areas of simulation. It does not simulate the few areas like heat and transport in the energy system. In [2], combining PERSEUS-CERT, this tool analyses the effect of wind integration.

2.2 BALMOREL

This tool mostly emphasises on CHP. Since 2000, this tool development and maintenance is an open source [3]. GAMS' modelling is the formulated language for the tool. To till date, we have ten distinct versions. To perform the analyses with this tool, at least a week's training is a prerequisite. It has a flexible time aspect. It can simulate heat but not transport of the energy system. This tool has been applied to many projects across the globe [4–11].

2.3 BCHP Screening Tool

This tool is for evaluating the power systems in the aspect of saving. It was industrialised in the USA by Oak Ridge National Laboratory [12]. To till date two versions were released. Based on the earlier experience with this tool, it needs nearly two weeks of training to upgrade from basic analysis to advanced analysis. It is particularly designed for the study of commercial buildings. The database is inbuilt for this tool. It is mainly focussed on single project investigations so large-scale models of electricity are not modelled. The tool is well thought-out to achieve the parametric studies amongst a baseline building, naturally a conventional building deprived of a CHP system, and up to 25 other CHP scenarios.

2.4 COMPOSE

Compare Options for Sustainable Energy is a tool based on evaluating a techno-economic energy project which is brought by Aalborg University in Denmark [13]. This tool's purpose is to assess the projects which support the intermittency of cost distribution and the profits under ambiguity. The current version of this tool can be downloaded at free of cost. A minimum of three to four days' training in this tool is a prerequisite. It is a user-defined system and a single project investigator. The main focus is deliberately on cogeneration. Uncertainties by the user defined may allow wide range of risks. It uses a one-hour time step over a number of years which are user defined. It also mainly focusses on framework modelling. In analysing the aids of energy storage and relocation options, this tool can be used.

2.5 E4cast

This tool is mainly developed by the Australian Bureau of Agricultural and Resource Economics (ABARE) for the Australian power system for its long-term projects [14]. Since the year of inception, the tool is regularly updated. The software is not for sale in its place, and the customers are supposed to pay for the analysis at a rate of AU$1500 per day. The time step for this tool is up to a maximum of 30 years. Naturally, this tool is used for simulating the future energy requirements within the Australian power system and recognises how these can be met. It has also used by the climate change department of Australia to assess the impacts of the various policies on emissions and renewables.

2.6 EMCAS

Electricity Market Complex Adaptive System uses a new modelling method for simulating the power system [15]. This tool was initiated in 2002 in the USA and Argonne National Laboratory is the one which updates to the latest version regularly. This tool is utilised approximately in twenty countries by the universities, transmission and power companies and also by various system operators. To obtain knowledge about this tool, a minimum of three weeks is a prerequisite to reach advanced analysis. Maximum of one hour is required to complete the entire analysis. Also, it is used to investigate the possible impacts in the aspects of operation and economic of various events in power system. This tool also has the ability to analyse the energy system investments and issues based on expansion by using a multi-agent-based profit maximisation method.

2.7 EMINENT

To allow new solutions faster in the market, this tool has come into existence. It was developed by the Netherlands organisation for applied scientific research in the Netherlands in the year 2005 [16]. As the tool is not developed up to the mark, the access is limited. Nearly, a month's training is required to learn the tool. It has an inbuilt database and assessment tool for evaluating new solutions for the new technologies of power systems. This tool constitutes two databases such as one from the national energy and the other which contains main information about the projects that are under development. It has capability in assessing a technology in financial environment. In [17, 18] we can find the overview, case studies and its comparison with other tools.

2.8 EMPS

EFI's Multi-area Power-market Simulator is a consistent refinement of SINTEF Energy Research in Norway [19]. Its development is done since 1975. Its tool mainly meant for optimised power system operation of hydrosources and thermal sources. In addition to thermal generation technology, the renewables are also one. Over the years to till date, we have up to ten versions updated and cumbersome of users have brought it. Minimum of a month's training is required for doing an advanced analysis with this tool. This tool has two parts, the first one is the strategic evaluation part and the other is simulation part. The time step for this tool is a week and the analysis can be carried out up to a period of 25 years. In [20–24], various case studies are undertaken using this tool.

2.9 Energy PLAN

Since 1999 at Aalborg University, Denmark, this tool has been continuously developed and maintained [25]. As it is easy to download, we can find a greater number of users with the latest updated versions. The period of learning this tool is up to a month and may extend based on the level of complexity. It is a user-friendly tool and the main purpose is to assist the design of energy planning strategy by simulation of the whole power system. Mostly, all the areas in the electric sector can be modelled by this tool. This tool optimises the given system operation as opposed to tools where the investments of the energy system have been optimised. Numerical analysis is done by this tool in electrical sector in large-scale integration of renewables in various countries like Estonia, Poland, Spain Germany and the UK [26, 27].

2.10 Energy PRO

The tool is a closed package of modelling, design, analysis and optimisation of fossils and renewables. EMD International A/S at Denmark industrialised this tool and also maintained by themselves [28]. From the inception of this tool, they have got over fifty versions. This package is a module chosen and used by sixteen countries with approximately a hundred users. Minimum of one-day training is used to learn this tool. This tool specialises mostly in thermal plant studies. It also involves models pertaining to heating sector but not in the technology of transport. It takes a minute time step for a duration of 40 years maximum for carrying out the analysis. Today, this tool is used for analysing the CHP plants at fixed tariffs [29] and also stimulates the compressed air energy storage in the spot market. Also, in [30, 31], analysis about heating and optimal sizing on the spot market is carried out.

2.11 ENPEP-BALANCE

The tool was established by Argonne National Laboratory in the USA in the year 1999. The nonlinear software matches the power demand with the existing resources and technologies. Approximately, this tool is used by fifty countries all over the world. In [32], this tool can be freely accessed. For learning this tool in a basic and advanced way, a person requires a period of two weeks. This tool is simulated based on market to determine the price changing and energy demand levels in the power market. The detailed analysis is done on annual basics to an extent of 75 years maximum. This tool depends on a decentralised process of decision making in the power sector. All types of generations can be simulated along with the financial aspects. Simultaneously, this tool finds the supply intersection and energy demand curves for all forms of energy in the network. The tool also services the iterative technique for a solution which is within the user-defined convergence tolerance. In [33, 34], a few case studies are mentioned.

2.12 GTMax

The Generation and Transmission Maximisation tool simulates the generating units dispatch and economic trading amongst the utilities with a network representation of grid [35]. In 1995 at Argonne National Laboratory, USA, this tool was established, and also, this is used in twenty-five countries across the globe. The price can be known only by contacting them. According to the expert's knowledge, learning the tool needs a week's training and also to adopt the advanced feature's implementation. The loads located at various locations can be served by this tool. All the technologies can be simulated by this tool. Also, in [36–39], this tool has a numerous investigation.

2.13 H2RES

This simulation tool mainly stimulates the renewable sources integration into power system. It was industrialised in the year 2000 by the amalgamation of Instituto Superior Tecnico in Lisbon, Portugal, and the University of Zagreb, Croatia [40]. To get an expertise in this tool, two months are a prerequisite, and also, this tool is used internally for research rather than external usage. This tool has been explicitly designed for renewable integration when operated at standalone condition. It can aid as a renewable planning tool along with the larger power systems. The tool considers all thermal generations but not nuclear and also it considers all renewable generations leaving tidal as exception. All the storage and conversion technologies are considered in this tool except compressed air energy storage. Using the criteria of satisfactory proportion of intermittent and renewable electricity in energy systems, this tool integrates as much as possible renewable and intermittent energy into the power system. The financial aspect is not included in this tool which may be added later. Finally, numerous investigations are done by various countries by this tool [41–43].

2.14 HOMER

The HOMER is an accessible micro-power design tool which is brought up in the year 1992 by the National Renewable Energy Laboratory in the USA. To till date, it has reached forty-two updated versions which have access [44]. Experts state the tool has used by thirty-two thousand people who learnt this tool in the span of one day. This tool simulates and optimises the system at standalone and grid-connected scenarios with various combinations of renewables. A few financial aspects are involved. It requires a time step of one minute for analysis considering a period of one year. It can also perform sensitivity analysis where the investigation can be done if there are any uncertainties in input variables. Finally, in [45, 46], various investigations are involved using HOMER as a software tool.

2.15 HYDROGEMS

As the name states, this software tool is based on hydrogen energy which is suitable for simulating the hydrogen energy system integration, specifically renewable-based standalone energy systems [47]. This tool is first developed in the year 1995 at the Institute for Energy Technology, and also, it was used in various research studies [48, 49]. Training of three months minimum is required for learning this tool. A period of one month is sufficient for the TRNSYS users. The performance analysis can be done in multiple time steps. These tool components are based on various streams of engineering like electrical, mechanical and heat transfer engineering. The pragmatic parts of this tool are designed to find the possible parameters based on the available data in the literature. This tool is accounted for the financial aspects also. This tool is used to analyse the operation of various renewable sources in standalone system as well as the simulation is also done based on hydrogen-based renewable energy [48, 49].

3 Conclusion

This research work achieves high efficiency in all aspects of information and technology. This will be applied in typical applications. The implemented proposed tool provides numerous effective and efficient information and especially for energy tool. Energy tool is used evidently and it is accessible in terms of region, analysis, technologies and objectives.

References

1. Universität Karlsruhe: Institute for Industrial Production. http://www.iip.kit.edu/65.php. Accessed 18.06.09
2. Karlsson, K., Meibom, P.: Integration of Hydrogen as Energy Carrier in the Nordic Energy System. Risø National Laboratory (2006)
3. Ravn, H.: Balmorel. http://www.balmorel.com/. Accessed 22.04.09
4. Ball, M., Wietschel, M., Rentz, O.: Integration of a hydrogen economy into the German energy system: an optimising modelling approach. Int. J. Hydrogen Energy **32**(10–11), 1355–1368 (2007)
5. Ea Energy Analyses: 50% Wind Power in Denmark in 2025. Ea Energy Analyses (2007)
6. Heggedal, A.M.: Investment in new transmission capacity between Estonia and Finland—effects on the electricity market and welfare. Masters thesis, Department of Economics and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway (2006)
7. Eesti Energia, Latvenergo, Lietuvos Energija, Elkraft System, COWI, Danish Energy Agency: Power Sector Development in a Common Baltic Electricity Market. Elkraft System, COWI (2005)
8. Elkraft System, COWI, Lietuvos Energija, Lithuanian Energy Institute: Economic Analyses in the Electricity Sector in Lithuania. Elkraft System, COWI, Lietuvos Energija, Lithuanian Energy Institute (2002)

9. Ea Energy Analyses: Large Scale Wind Power in New Brunswick—A Regional Scenario Towards 2025. Ea Energy Analyses (2008)
10. Morthorst, P.E., Jensen, S.G., Meibom, P.: Investering og prisdannelse på et liberaliseret elmarked (Investment and Pricing in a Liberalised Electricity Market). Risø National Laboratory (2005)
11. Jensen, S.G., Meibom, P.: Investments in liberalised power markets: gas turbine investment opportunities in the Nordic power system. Int. J. Electr. Power Energy Syst. **30**(2), 113–124 (2008)
12. Oak Ridge National Laboratory: Whole-Building and Community Integration Program
13. Aalborg University: EnergyInteractive.NET. http://energyinteractive.net. Accessed 11.06.09
14. ABARE: ABARE Models
15. Argonne National Laboratory: Electricity Market Complex Adaptive System (EMCAS)
16. EMINENT2: Welcome to EMINENT
17. Segurado, R., Pereira, S., Pipio, A., Alves, L.: Comparison between EMINENT and other energy technology assessment tools. J. Cleaner Prod. **17**(10), 907–910 (2009)
18. Jansen, P., Koppejan, J., Hetland, J., Klemeš, J., Phuengphaeng, T., Pipio, A.: EMINENT accelerates market introduction of promising early stage technologies for transport and energy. In: Proceedings of the CISAP1—1st Italian Convention on Safety and Environment in Process Industry, Palermo, Italy, 28–30 Nov 2004
19. SINTEF: EOPS and EMPS
20. Warland, G., Haugstad, A., Huse, E.S.: Including thermal unit start-up costs in a long-term hydro-thermal scheduling model. In: Proceedings of the 16th Power Systems Computation Conference, Glasgow, Scotland, 14–18 July 2008
21. Sedaghati, A.: Evaluating the consequences of investment in distributed power production. In: Proceedings of the 2005 IEEE International Symposium on Intelligent Control and 13th Mediterranean Conference on Control and Automation, Limassol, Cyprus, 27–29 June 2005
22. Fosso, O.B., Gjelsvik, A., Haugstad, A., Mo, B., Wangensteen, I.: Generation scheduling in a deregulated system. The Norwegian case. IEEE Trans. Power Syst. **14**(1), 75–81 (1999)
23. Haugstad, A., Rismark, O.: Price forecasting in an open electricity market based on system simulation. In: Proceedings of the EPSOM'98—International Conference on Electrical Power Systems Operation and Management, Zürich, Switzerland, 23–25 Sept 1998
24. Doorman, G., Kjølle, G., Uhlen, K., Ståle, E., Flatabø, N.: Vulnerability of the Nordic Power System: Main Report. SINTEF, Nordic Council of Ministers (2004)
25. Lund, H., Munster, E.: Modelling of energy systems with a high percentage of CHP and wind power. Renew. Energy **28**(14), 2179–2193 (2003)
26. Lund, H., Salgi, G.: The role of compressed air energy storage (CAES) in future sustainable energy systems. Energy Convers. Manage. **50**(5), 1172–1179 (2009)
27. Mathiesen, B.V., Lund, H.: Comparative analyses of seven technologies to facilitate the integration of fluctuating renewable energy sources. IET Renew. Power Gener. **3**(2), 190–204 (2009)
28. EMD International A/S: http://www.emd.dk/. Accessed 23.04.09
29. Sauer, C., Erge, T., Barnsteiner, M.: Demonstration of innovative electricity marketing options from decentralised generation—the Badenova showcase. In: Proceedings of the CISBAT 2009: Renewables in a Changing Climate, Lausanne, Switzerland, 2–3 Sept 2009
30. Lund, H., Siupsinskas, G., Martinaitis, V.: Implementation strategy for small CHP-plants in a competitive market: the case of Lithuania. Appl. Energy **82**(3), 214–227 (2005)
31. Strckiene, G., Andersen, A.N.: Analyzing the Optimal Size of a CHP-Unit and Thermal Store When a German CHP-Plant is Selling at the Spot Market. EMD International A/S, Market Access for Smaller Size Intelligent Electricity Generation (MASSIG) (2008)
32. Argonne National Laboratory: Energy and Power Evaluation Program (ENPEPBALANCE)
33. International Atomic Energy Agency: Comparative Assessment of Energy Options and Strategies in Mexico Until 2025. International Atomic Energy Agency (2005)
34. Conzelmann, G., Koritarov, V.: Turkey Energy and Environmental Review. Argonne National Laboratory (2002)

35. Argonne National Laboratory: Generation and Transmission Maximization (GTMax) Model
36. Argonne National Laboratory: Power Systems Analysis Program
37. Koritarov, V.S., Veselka, T.D.: Modeling the Regional Electricity Network in Southeast Europe. Argonne National Laboratory (2005)
38. Kostova, B., Poprea, L., Popescu, V., Veselka, T.D.: Simulation of regional power markets in the planning of trans-border interconnections. In: Proceedings of the IEEE PES PowerTech 2009, Bucharest, Romania, 28 June–2 July 2009
39. Argonne National Laboratory: The Economic Cost of the March 2008 Glen Canyon "Flush"
40. Instituto Superior Técnico, University of Zagreb: H2RES
41. Fowler, P., Krajacic, G., Loncar, D., Duic, N.: Modeling the energy potential of biomass—H2RES. Int. J. Hydrogen Energy **34**(16), 7027–7040 (2009)
42. Krajacic, G., Duic, N., da Graça Carvalho, M.: H2RES, energy planning tool for island energy systems—the case of the Island of Mljet. Int. J. Hydrogen Energy **34**(16), 7015–7026 (2009)
43. Duic, N., da Graça Carvalho, M.: Increasing renewable energy sources in island energy supply: case study Porto Santo. Renew. Sustain. Energy Rev. **8**(4), 383–399 (2004)
44. HOMER Energy LLC: HOMER
45. Bekele, G., Palm, B.: Wind energy potential assessment at four typical locations in Ethiopia. Appl. Energy **86**(3), 388–396 (2009)
46. Rehman, S., El-Amin, I.M., Ahmad, F., Shaahid, S.M., Al-Shehri, A.M., Bakhashwain, J.M., et al.: Feasibility study of hybrid retrofits to an isolated off-grid diesel power plant. Renew. Sustain. Energy Rev. **11**(4), 635–653 (2007)
47. Institute for Energy Technology: HYDROGEMS. http://www.hydrogems.no/. Accessed 25.04.09
48. Zoulias, E.I., Glockner, R., Lymberopoulos, N., Tsoutsos, T., Vosseler, I., Gavalda, O., et al.: Integration of hydrogen energy technologies in stand-alone power systems analysis of the current potential for applications. Renew. Sustain. Energy Rev. **10**(5), 432–462 (2006)
49. Ulleberg, Ø., Ito, H., Maack, M.H., Ridell, B., Miles, S., Kelly, N., et al.: Hydrogen Implementing Agreement (HIA); Task 18—Integrated Systems Evaluation; Subtask B—Demonstration Project Evaluations; Final Report. International Energy Agency (2007)

Three-Interacting Tank Controlled with Decentralized PI Controller Tuned Using Grey Wolf Optimization

K. Anbumani and R. Rani Hemamalini

Abstract In most of the process industries, modelling and control is the common problem for a multi-input–multi-output nonlinear process. In this chapter, for benchmark system the three-tank system is considered. Three-tank system is majorly used for many industrial applications in various domains. The main objective is to model the system and design a controller with good transient and steady-state performance for a nonlinear MIMO process, which is a challenging task. Decentralized PI controller is applied to three-interacting cylindrical tank process. The process under study is a MIMO process with two manipulated variable and three process outputs, height of the tanks. Decentralized controller needs controller design between the most interacting pair. Mathematical modelling is obtained from the first principle theorem, state-space and transfer function methods. The interaction among all the inputs and outputs are computed by relative gain array (RGA). Condition number is computed to check whether the process is ill-conditioned or not. The PI controller applied between the most interacting loops is tuned using grey wolf optimizer and the performance studies done for servo and regulatory operation. Simulation results show the proposed controller can be implemented for a non-square matrix.

Keywords Three-interacting cylindrical tank · Decentralized controller · RGA analysis · Condition number · Grey wolf optimizer

K. Anbumani (✉)
Sathyabama Institute of Science and Technology, Chennai, India
e-mail: anbumani.ei@sairam.edu.in

Sri Sairam Engineering College, Chennai, India

R. Rani Hemamalini
Department of Electrical and Electronics Engineering, St. Peter's Institute of Higher Education and Research, Chennai, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_47

1 Introduction

In industry, many processes are complex and multivariable in nature or MIMO systems. The ageing in process, load changes, nonlinearities and time-dependent characteristics cause plant parameter changes which changes in dynamic behaviour of the system. Cylindrical tanks used in process industries like petrochemical industry, paper mill and mixing process for storage and discharge of fluids. The cylindrical tank finds wide applications in process industries like water purification, chemical, food and pharmaceutical industries. In process industries, the liquid flow and level in the tank are a challenging problem. The tank liquid level should be maintained properly for chemical equilibrium in various industry processes. Regulation of liquid level in the tank is extremely complex and challenging in process industry for a long period. The classical three-tank interacting process is considered as experimental set-up. The three-tank experimental set-up system is available with the aid of sensor and pump for flow distribution system. The experimental set-up consists of three separate cylindrical tanks interconnected, has a common flow path with drain valves to a reservoir tank. The level of the water in the tank is visible through glass tank. The issues like couplings, interactions, unmodelled dynamics and uncertainties are present in the three-tank system. It increases the need of multivariable controller design with better transient performance.

In industries, proportional-integral-derivative controllers are commonly used for more than few decades for process control applications because of their performance in slow process plants. For process with interaction, the multiloop PI controller is a normal controller. The multiloop PID controller operation gives improved performance in multivariable process [1, 2]. In MIMO process, the main gain of implementing multiloop control system is that, if any loop flops, control engineer can easily recognize and substitute it, similarly easy to tune it.

RGA is cast-off to find interaction between input and output of a system. It is a powerful tool employed to identify manipulated variable and controlled variable that produces desired response. For multiloop controller design, the clear knowledge has to be created by control engineer which manipulated variable is matched with controlled variable. The effect among manipulated variable and controlled output at steady state is identified by using RGA (Bristol 1966). RGA decides input–output pairs for the SISO controllers in MIMO system. It measures the changes in the gain between one input and one output when closing the loops between all other inputs and outputs by means of perfect control. If no changes in the gain, RGA recommends the inputs and outputs combination as a pair of SISO controller. The gain changes tested for all other remaining input and output pairs.

Singular value decomposition (SVD) reduces error and noise level of the system. SVD is used to find linear combination in set of variables such as condition number. The plant decomposition level can be identified by using condition number. The rank of ill-conditioned plant valued in terms of condition numbers and some singular values. Gain matrix by big condition number signifies ill-conditioned [3]. If the condition number is 50 and above, the minor error in system can alter huge in control action, consequently then the system cannot be disintegrated.

The interaction between the input and output variables causes difficulty in multi-loop controller for a multivariable process. In this paper, entire MIMO process disintegrated into single input–single output process to minimize the interaction effect. Since the straight forward approach and the features of modest design, relaxed tune, decentralized controller is favoured in many MIMO dynamic system also the disappointment of one loop will not disturb the whole system [4, 5]. The loop interface of dynamic nonlinear system is very difficult to understand [6–8]. The optimal controller plays vital role in process interaction. Data collected from various processes contains a large amount of uncertainties due to human errors, and hence, it is difficult to analyse the behaviour of such system for a desired accuracy. If the behaviour of the system calculated, then they have high uncertainty, also it is not easy to analyse the behaviour of the system. In order to implement the intelligence in controller fuzzy logic, neural network and genetic algorithm, similar soft computing techniques and other evolutionary algorithms are adapted. Compared with other heuristics system, grey wolf optimization (GWO) does not employ crossover operators to produce new solution from the present one. The frequency-domain specification and time-domain specification corresponding settling time, rise time, maximum over shoot, peak time, gain margin and phase margin are obtained neighbouring values using the GWO optimization algorithm which will fine-tune the constraints of the controller so that the desired specification can be easily obtained. With ever increasing growth in the heuristic algorithms, GWO algorithm is used for optimal tuning of the controller parameters.

This paper is planned as follows: Sect. 2 describes the interaction of three-tank system. The interaction study and loop coupling are discussed in Sect. 3. Decentralized controller is debated in Sect. 4. Last, conclusions are specified in Sect. 5.

2 Cylindrical Three-Tank Interacting System

In different configuration of three-tank systems, the three interconnecting tank configuration is the attractive one. The level control involves complexity, and implementation difficulty due to the level of one tank depends upon the level of another tank. This configuration possesses high nonlinearity, complexity and coupling. The three cylindrical tanks with same cross-sectional area (A) are serially interconnected with two cross-sectional pipes. Figure 1 shows the plexiglass tank. The interaction between tank 1, tank 2 and tank 3 denoted T1, T2, T3 is equipped with manual controllable value (HV1 and HV2). The interaction of process is desired by the position of this valve (HV1 and HV2). Hand valves, HV1 and HV2, are used to change the position of the valve once fixed the position should not be changed. Changing the position will result in different dynamics, thereby resulting in controller failure.

To store the water, the three-tank system has a reservoir. Double DC motor with adjustable speed is used to pump the water. The water is pumped through the pumps pump 1 and pump 2 to the tanks through control valves. The voltage applied to DC

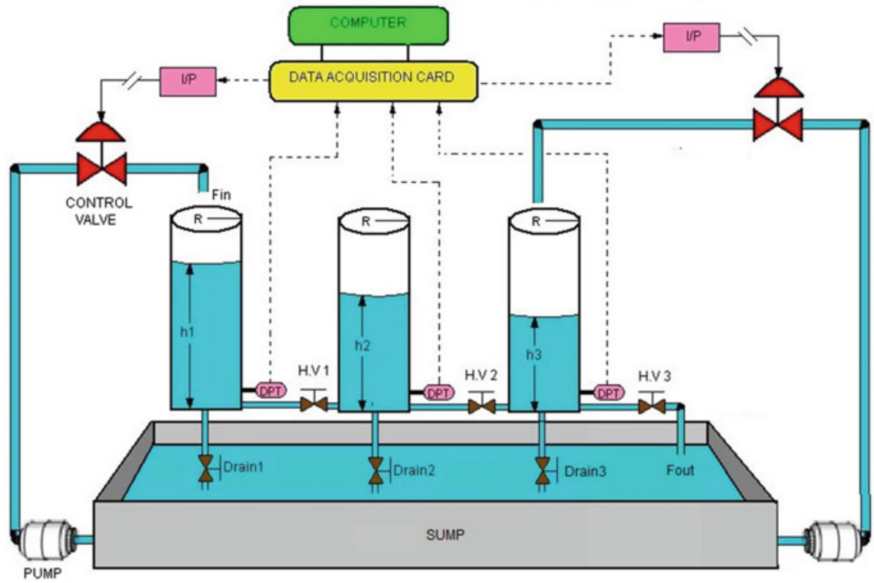


Fig. 1 Three-interacting tank system with two inputs

motor is proportional to the flow rate of the pump. There are manual valves HV3, HV4, HV5, HV6 which are used to introduce dynamics to the process. The flow can be measured by rotameter. On the highest and lowest part of the tank, the water incursion and drainage are providing, respectively. To keep the level of water in the tanks, flow regulators are placed at the drainage of the tank 1, 2 and 3. The lowermost pressure formed by water level is measured by DPT and it stretches the elevation in terms of milliamps also can be adjusted in terms of water level (cm). The drainage of the systems is controlled by HV3, HV4, HV5 and HV6. The valve coefficient of the system is β_i and the interacting control valves HV1, HV2 having valve coefficient of β_{ij}. Maintaining the tank level in tank 1 and tank 3 by varying the inflow of tank 1 (F_{in1}) and tank 3 (F_{in3}) is the main objective of this work. The three-interacting tank process schematic diagram is shown in Fig. 1,

where

- h_i Level of tank i (cm)
- u_i Control input to control valve cv_i (v)
- A_i Area of tank i (cm)
- α_i Cross section area of pipe linking tank i (cm^2)
- β_i Valve co-efficient between tank i
- β_{ij} Valve co-efficient among tank i and tank j
- K_i Gain of valve cv_i $\left(\text{cm}^3/\text{vs}\right)$
- g Gravity.

Table 1 System parameters

A_1, A_2, A_3 (cm^2)	$\alpha_1, \alpha_2, \alpha_3$ (cm^2)	β_{12}	β_{23}	$\beta_1, \beta_2, \beta_3$	K_1, K_2 (cm^3/vs)
615.75	5.0671	0.9	0.8	0.3	75

The liquid can flow from tank 3 to tank 1 or from tank 1 to tank 3 via tank 2. The ultimate aim of this paper is to maintain the level of tank 1 and tank 3 by changing the corresponding valve position simultaneously. The difficulty here is that change in valve position causes change in dynamics of the system which will affect the linearity of the system. Based on the liquid ideal properties, the mathematical model of the system is obtained.

To support the clarification of experimental data, to predict the significances of variations of system contribution or operating condition, to identify optimal process or working conditions and for control application, the mathematical model of the system was developed. But the process dynamics is the main problem in modelling that must be taken or else not any usage in modelling the system. Since the bid of essential physical and chemical moralities to the process, the dynamic model of the system has been evaluated, using conventional mathematical modelling approach. Due to the presence of nonlinear terms in the equations, the equations first must be linearized. Many methods are available for modelling MIMO process. The state-space realization of the system is much more acceptable than any other realization, because of the high interaction found between the subsystems (i.e. tanks). Hence using law of conservation of mass, the differential equations describing the system are given as (1), (2), (3). Level in tank 1 and tank 3 is considered as output variable. The inflow rate of tank 1 and tank 3 is considered as input variable. The parameter of the system and the constrained variables are given in Table 1.

The process consists of three interconnected cylindrical tank T1, T2, T3. The cross-sectional area of the system is A_1, A_2, $A_{3,}$ respectively, and equal in size. At the lowermost part of the tank, the drain valve is provided. The cross-sectional area of the pipe to the drain valve is ‘a’. The flow of liquid from the tank can be derived from Bernoulli’s theorem. The values of different process parameter of the actual three-tank system are shown in Table 1.

The open-loop information was obtained in the three-interacting cylindrical process. The variation of the inflow amount F_{in1} then F_{in3} gives the open circle information. The working facts initiate out from the input–output features. Applying the Jacobian function for the differential first principle model, state-space model remains attained. Table 2 gives the operation conditions, state-space model and transfer function matrix intended for the three-interacting cylindrical tank system. To understand the behaviour of the system, the state-space model is used. The transfer function remains used to find the interaction of the process. But the mathematical model is nonlinear in nature.

Table 2 Predictable transfer function and state-space prototypical of the three-interacting cylindrical tank system and working conditions

Operating facts
$V_{1s} = 5.5, V_{2s} = 5.5, h_{1s} = 16.8, h_{2s} = 16.34, h_{3s} = 16.92$
State-space model
$A = \begin{bmatrix} -0.2551 & 0.2418 & 0 \\ 0.2418 & -0.4467 & 0.1914 \\ 0 & 0.1914 & -0.2047 \end{bmatrix}$
$B = \begin{bmatrix} 0.1218 & 0 \\ 0 & 0 \\ 0 & 0.1218 \end{bmatrix}$
$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$
Transfer function matrix G(s)
$\begin{bmatrix} \frac{0.1218\,s^2+0.07935\,s+0.0066706}{s^3+0.9066\,s^2+0.1626\,s+0.002016} & \frac{0.005638}{s^3+0.9066\,s^2+0.1626\,s+0.002016} \\ \frac{0.02945\,s+0.006029}{s^3+0.9066\,s^2+0.1626\,s+0.002016} & \frac{0.02331\,s+0.005949}{s^3+0.9066\,s^2+0.1626\,s+0.002016} \\ \frac{0.005638}{s^3+0.9066\,s^2+0.1626\,s+0.002016} & \frac{0.1218\,s^2+0.08549\,s+0.006762}{s^3+0.9066\,s^2+0.1626\,s+0.002016} \end{bmatrix}$

$$\frac{\mathrm{d}h_2}{\mathrm{d}t} = -\mathrm{sgn}(h_1 - h_2)\frac{\beta_{12}\alpha_{12}}{A_1}\sqrt{2g(h_1(t) - h_2(t))} + \frac{K_1}{A_1}u_1\beta_1\alpha_1\sqrt{2g(h_1(t))} - \frac{\beta_1\alpha_1}{A_1}\sqrt{2g(h_1(t))} \tag{1}$$

$$\frac{\mathrm{d}h_2}{\mathrm{d}t} = \mathrm{sgn}(h_1 - h_2)\frac{\beta_{12}\alpha_{12}}{A_2}\sqrt{2g(h_1(t) - h_2(t))} - \mathrm{sgn}(h_2 - h_3)\frac{\beta_{23}\alpha_{23}}{A_2}\sqrt{2g(h_2(t) - h_3(t))} - \frac{\beta_2\alpha_2}{A_2}\sqrt{2g(h_2(t))} \tag{2}$$

$$\frac{\mathrm{d}h_2}{\mathrm{d}t} = \mathrm{sgn}(h_1 - h_2)\frac{\beta_{12}\alpha_{12}}{A_2}\sqrt{2g(h_1(t) - h_2(t))} - \mathrm{sgn}(h_2 - h_3)\frac{\beta_{23}\alpha_{23}}{A_2}\sqrt{2g(h_2(t) - h_3(t))} - \frac{\beta_2\alpha_2}{A_2}\sqrt{2g(h_2(t))} \tag{3}$$

3 Interaction Analysis

In MIMO process, one manipulated variable can affect the number of controlled variables in a loop or other loops present in the system. MIMO control scheme is important for a process which has multiple interactions between different variables and has multiple dependencies. Understanding the relation of manipulated and controlled variable in a MIMO system will be helpful for designing and implementing a control scheme for the system. In a three tank, where the level of the different tanks is affected by the in and out flow rate of the process. The effect of inflow and outflow rate on the level of the tank or the interaction between the tanks is considered as MIMO variables. A good MIMO control means the individual variables can be decoupled without affecting the other variables. It is also stable during interaction, load disturbances and dynamic situation. The interaction is analysed using relative gain array (RGA). RGA is the well-established tool for analysis of different control system. RGA provides quantitate analysis of interaction between the input and output variables. RGA provides paring of manipulated and control variable for control scheme. RGA designs single-input–single-output controller for each input and output pair. RGA has many properties. One of the properties is used to pair the input and output for controlling the loops individually. The control scheme also stables with respect to dynamic situation, load changes and random disturbances. RGA is computed under steady-state conditions. In other words, RGA is the normalized gain matrix that describes the impact of each controlled variable on manipulated variable. The I/O close to one and positive is paired. By selecting proper input, output pair the issue of loop interaction can be reduced. In this chapter, non-square matrix stays cast-off to find towards sparing. Condition number remains another metric towards check the ill-condition of the plant. It is computed from the steady-state gain matrix. If the condition number is high, the plant is ill-conditioned. System can be controlled if the condition number is less than 50.

$$b = g11^2 + g12^2 \tag{4}$$

$$c = g11 + g21 + g12g22 \tag{5}$$

$$d = g21^2 + g22^2 \tag{6}$$

$$\sigma_1 = s_1^2 = \frac{(b+d) + \sqrt{(b-d)^2 + 4c^2}}{2} \tag{7}$$

$$\sigma_2 = s_2^2 = \frac{bd - c^2}{s_1^2} \tag{8}$$

$$\sum = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix} \tag{9}$$

Table 3 Gain matrix, RGA and condition no.

Gain matrix	Relative gain array	Condition number
$\begin{bmatrix} 0.1218 & 0.0056 \\ 0.0295 & 0.0233 \\ 0.0056 & 0.1218 \end{bmatrix}$	$\begin{bmatrix} 0.9597 & -0.0038 \\ 0.0502 & 0.0300 \\ -0.0039 & 0.9739 \end{bmatrix}$	1.1428

$$\text{Condition number CN} = \frac{s_1}{s_2} = \frac{\sigma_{\text{max value}}}{\sigma_{\text{min value}}} \tag{10}$$

where the singular values of gain matrix be situated σ_{max} and σ_{min}.

Table 3 shows the gain matrix, relative gain array and condition number for the three-interacting tank process around the operating condition specified in Table 2. On analysis of the relative gain array, h_1 is paired with $F_{\text{in}1}$ and h_3 is paired with $F_{\text{in}2}$. Condition number is found to be 1.1428 which is much lesser than 50 indicating that the process is conditioned and can be controlled using a multiloop decentralized controller.

4 Decentralized Controller Design Tuning Using Grey Wolf Optimizer

Multiloop decentralized controller converts the MIMO system into SISO systems with PI controller for each loop. Entire MIMO process is split into many SISO systems, and the controllers are tuned for the loops using GWO.

4.1 Grey Wolf Optimizer

Heuristics algorithms are used for optimization in many applications. Grey wolf optimization remains experiential algorithm presented by Mirjalili et al. in 2014. Grey wolf algorithm was designed to replicate the behaviour of grey wolves hunting. The behaviour of the wolf social hierarchy and hunting behaviour of wolf was inspired and proposed by Seyedali. An algorithm is created between set of solutions and the problems for a population of wolf chasing a prey. GWO categorized and obsessed by four groups, namely alpha (α), beta (β), delta (δ) and omega (ω). Grey wolf algorithm was designed to replicate the behaviour of grey wolves hunting.

1. Alpha is the leader of the crowd which makes decision regarding hunting, sleeping, wake up all the decision followed by the crowd. Alpha needs not be the strongest member but a good organizer.

2. Beta helps the alpha in decision-making and other activity. Beta may male or female. Beta passes the information to the crowd and gets the feedback to alpha. Beta is the right candidate to get the position of alpha when the situation is unable to handle by alpha.
3. Delta performs the operation of sentinels, elders, scouts, hunters and caretaker. Scouts take the responsible for watching the boundary and warn the crowd if danger occurs. Sentinels provide a guarantee of safety. Elders are used by alpha and beta for their experience. Hunters help the alpha and beta to hunt the prey and provide food for the crowd. Caretakers take care of weak, ill and health of the crowd.
4. Omega assists the entire crowd to maintain the dominance structure. It gives the response to all the three members. Omega follows the alpha, beta and delta.

The hunting is done by three steps: searching, encircling and attacking preys. For optimizing the controller, the parameters alpha (first fit solution), beta (next fit solution), delta (third fit solution) and omega (remaining solution) number of search agents, iterations, site number and stopping criterion needs to be initialized. Omega wolves follow alpha, beta and delta wolves.

Equations for encircling are given by:

$$\vec{D} = \left|\vec{C}\vec{X}_{\mathrm{p}}(t) - \vec{X}(t)\right| \quad \text{and} \quad C = c_1 \cdot r_2 \tag{11}$$

$$\vec{X}(t+1) = \vec{X}_{\mathrm{p}}(t) - \vec{A}\vec{D} \tag{12}$$

$\vec{A} = 2\vec{a}\vec{r}_1 - \vec{a}$ and $\vec{C} = 2\vec{r}_2$ are coefficient vectors

wherever t is the present iteration, X_{p} remains the prey position, X stands the wolf position and C is a coefficient element evaluated using random vector, r_2, C_1 is the constant. The difference in result vector D is used to move the crowd towards and away from the region where the prey is located. $\vec{a}$ decreased from 2 to 0 all over the iteration. $\vec{r}_1, \vec{r}_2$ stay random vectors it can vary in the interval of 0–1. Then, the value of A is lesser than 1 it corresponds to hunting behaviour and if A is greater than 1 exploration behaviour. Search agents omega solutions (least solution) are forced to move towards alpha, beta and delta solutions.

$$\vec{D}_\alpha = \left|\vec{C}_1\vec{X}_\alpha - \vec{X}\right|, \vec{D}_\beta = \left|\vec{C}_2\vec{X}_\beta - \vec{X}\right|, \vec{D}_\delta = \left|\vec{C}_3\vec{X}_\delta - \vec{X}\right| \tag{13}$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \left(\vec{D}_\alpha\right), \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \left(\vec{D}_\beta\right), \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \left(\vec{D}_\delta\right) \tag{14}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1(t) + \vec{X}_2(t) + \vec{X}_3(t)}{3} \tag{15}$$

From Eqs. (13) and (14), the prey positions correspond to alpha, beta and delta which are obtained. Wolf position is updated using Eq. (15). Final solution is obtained by updation of the alpha, beta and delta positions depending on the victim position and omega wolves encircling the victim. The flow chart representation of GWO is shown in Fig. 2.

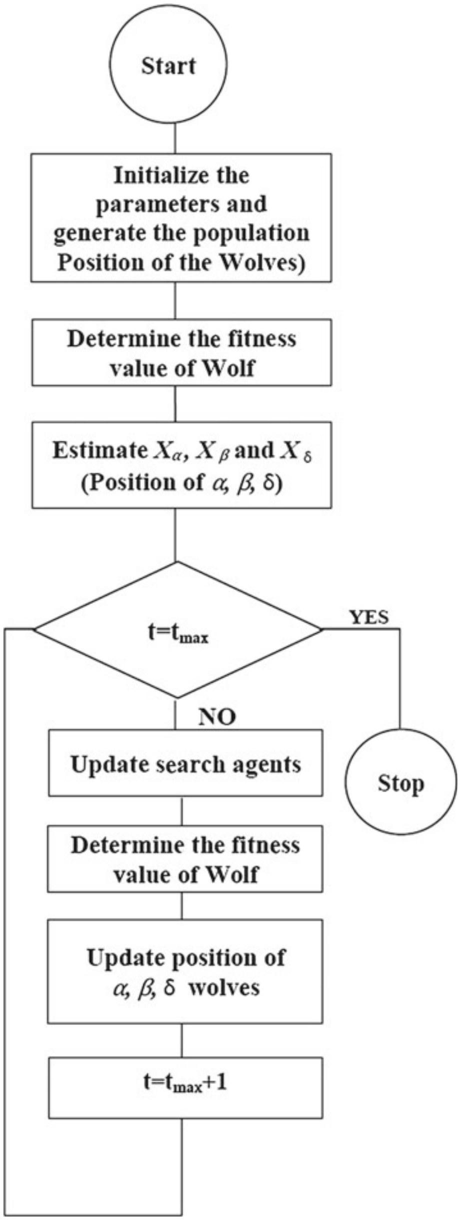


Fig. 2 GWO algorithm flow chart

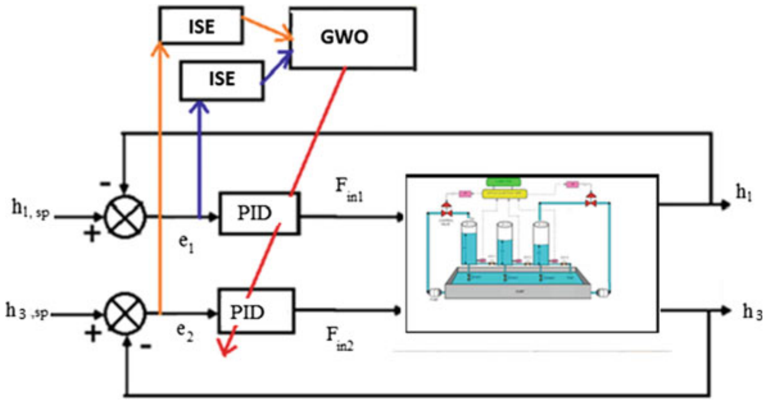


Fig. 3 Online tuning of decentralized PI controller for interacting cylindrical three-tank system

4.2 Online Optimization of the Decentralized Controller Using Grey Wolf Optimizer

In Fig. 3, the block diagram shows of the optimal decentralized MIMO PI controller using grey wolf optimizer. Conventionally, decentralized controller comes along with a decoupler to reduce the effects of the other loops. In this decentralized controller, the grey wolf optimizer takes into the interaction effects and tunes the PI controller eliminating the effects of the other process variable. The multiloop PI controller is implemented to the three-interacting cylindrical tank process. The tuneable parameters are K_p, K_i and the fitness function is the integral square error (ISE). 50% weightage is given to ISE value in each loop, and the cumulative error is used as the objective function.

Decentralized PI controller values are $K_{p1} = 37.0159$, $K_{i1} = 12.7233$ for loop 1, $K_{p2} = 8.455$, $K_{i2} = 8.455$ for loop 2. For checking the performance, set point change is given to the planned controller. The servo response of the planned controller is shown in Fig. 4. Set point is varied intended for both the loops randomly with a time interval of 10,000 s, and the response of heights h_1 and h_3 in both the loops is shown to track the set point change accurately. Disturbance is given to the process, and the response with disturbance rejection is shown in Fig. 5. Disturbance is given every 5000 s in both positive and negative directions and the response plotted. The proposed scheme is quite faster in rejecting the disturbance with a smooth response, the response is shown in Fig. 6 under servo-regulatory conditions and the response shows that the controller is capable of the tracking the set point and rejecting the disturbance.

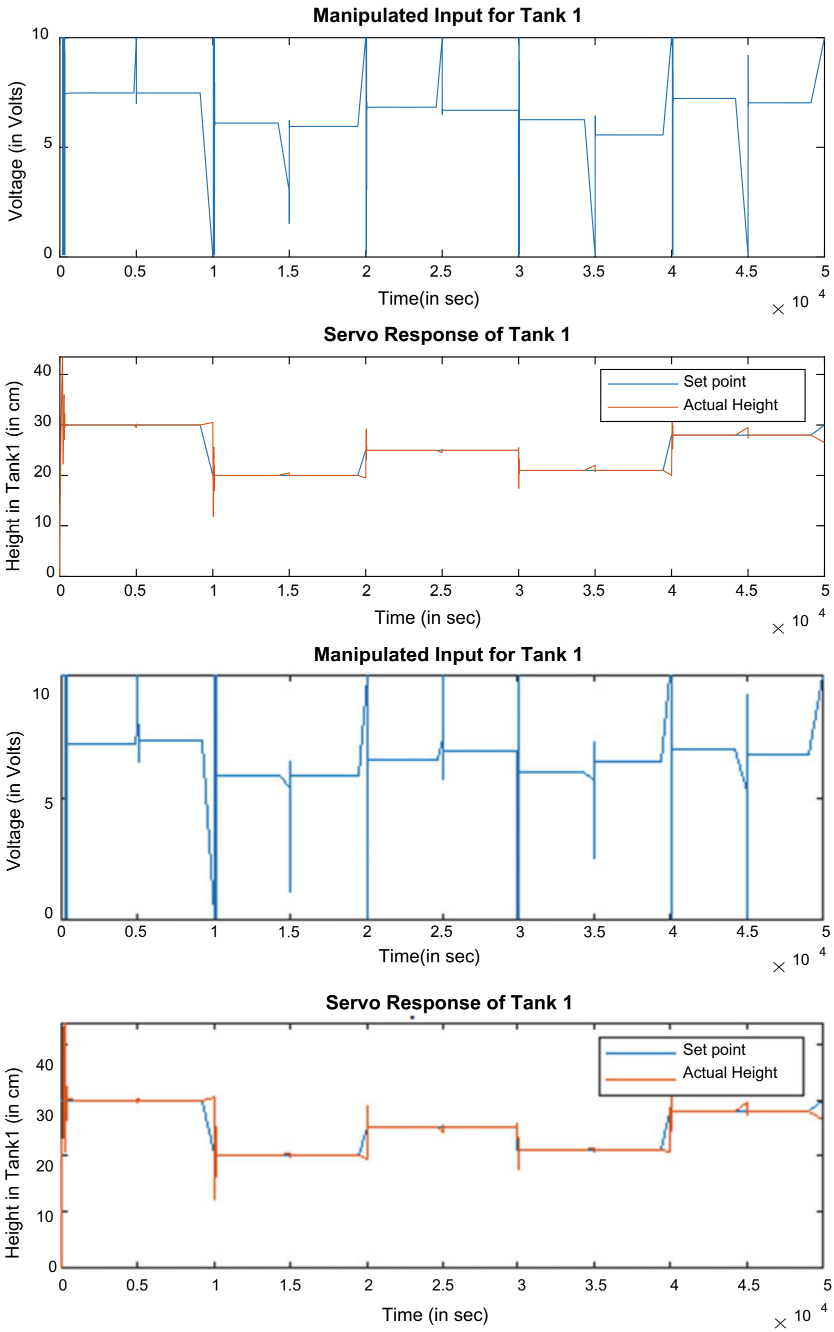


Fig. 4 Servo response of the decentralized controller tuned using grey wolf optimizer for interacting cylindrical three-tank process

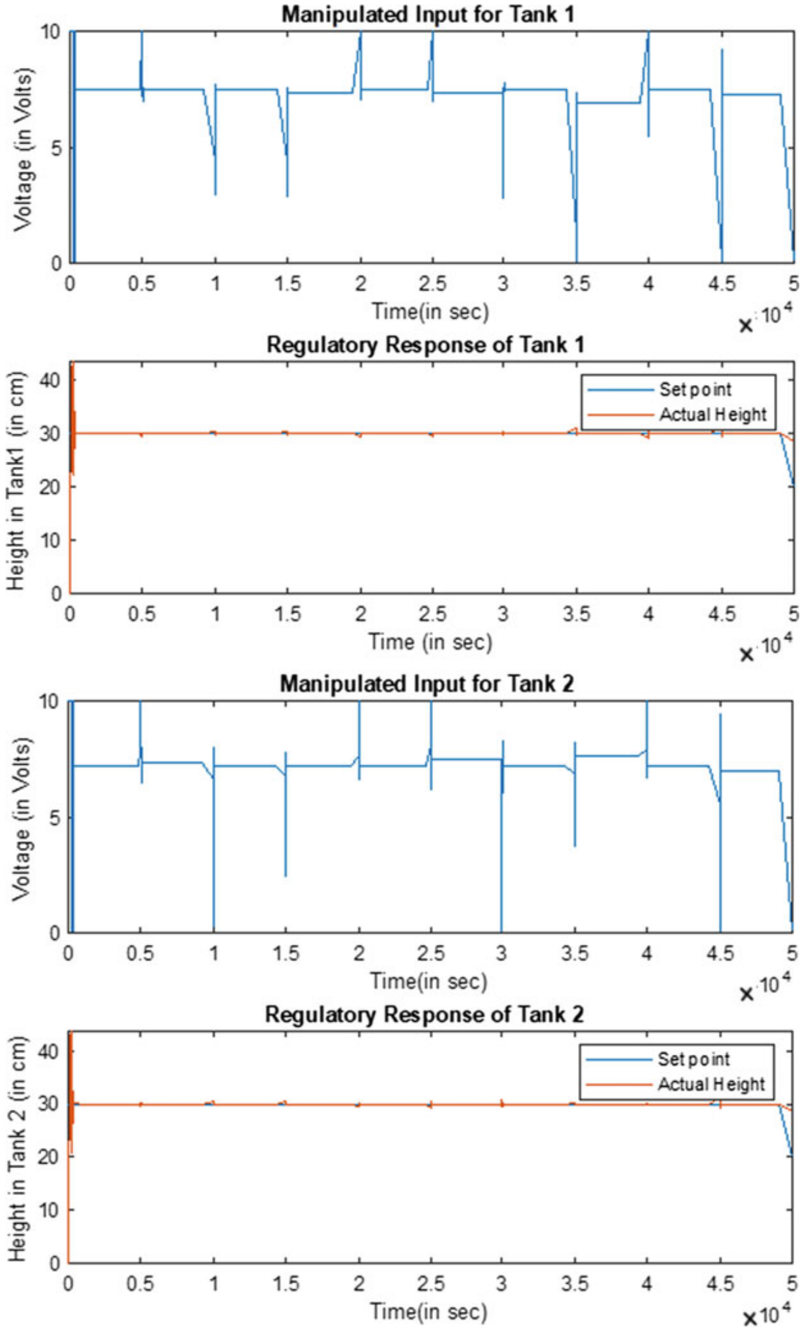


Fig. 5 Regulatory response of the decentralized controller tuned using grey wolf optimizer for interacting three-tank cylindrical process

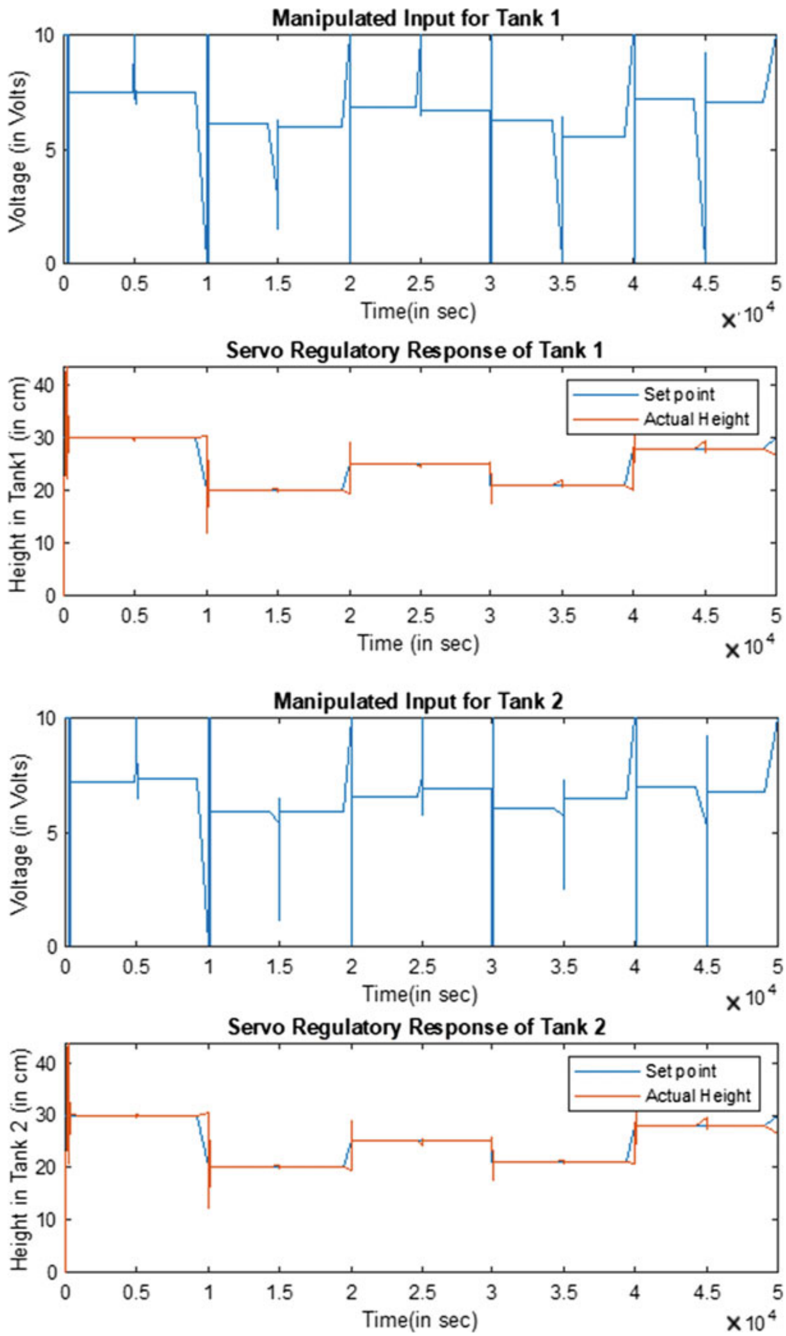


Fig. 6 Servo-regulatory response of the decentralized controller tuned using grey wolf optimizer for three-interacting cylindrical tank process

5 Conclusion

In this research work, multiloop decentralized PI controller is tuned using grey wolf optimizer. Mathematical modelling of the process is done and interaction analysis is done. On investigation of the relative gain array, it is seen that F_{in1} is highly interacting with h_1 and F_{in2} interacts with h_2. Condition number is found to be low specifying that the system is well-conditioned. Grey wolf optimizer tunes the PI parameter. The servo, regulatory and servo-regulatory response is conducted on the proposed controller after the results, and it is chosen that the planned controller remains well suited for the tuning of decentralized controller. Decoupler usage is also avoided, thereby making the controller simpler.

References

1. Astrom, K.J., Johansson, K.H., Wang, Q.-G.: Design of decoupled PI controllers for two-by-two systems. IEEE Proc. Control Theory Appl. **149**, 74–81 (2002). Special section on PID control
2. Nordfeldt, P., Hagglund, T.: Decoupler and PID controller design of TITO systems. J. Process Control **16**, 923–930 (2006)
3. Chen, J., Freudenberg, J.S., Nett, C.N.: The role of the condition number and the relative gain array in robustness analysis. Automatica **30**(6), 1029–1035 (1994)
4. Ali, A., Majhi, S.: PID controller tuning for integrating process. ISA Trans. **49**, 70–81 (2010)
5. Yeroglu, C., Tan, N.: Classical controller design techniques for fractional order case. ISA Trans. **50**, 41–72 (2011)
6. Mudi, R.K., Dey, C.: Performance improvement of PI controllers through dynamic set-point weighting. ISA Trans. **50**, 220–230 (2011)
7. Luyben, W.L., Jutan, A.: Simple method for tuning SISO controllers in multivariable system. Ind. Eng. Chem. Process Des. Dev. **25**, 654–660 (1986)
8. Stephanopoulos, G.: Chemical Process Control. Prentice Hall of India Pvt. Ltd., New Delhi (2001)

K. Anbumani was born in Tamil Nadu, India in 1980. He received his B.E. degree from Saranathan College of Engineering, Trichy, in 2002. And M.E. from Annamalai University, Chidambaram, Tamil Nadu, 2005. Currently he is working as Assistant Professor, in the Department of Electronics and Instrumentation Engineering, Sri Sairam Engineering College, Chennai, 600045, India. Also, he is part time research scholar at Sathyabama Institute of Science and Technology. His current research interests include process control, soft computing and Industrial Automation. He has published many numbers of papers in various peer review journals and conferences.

R. Rani Hemamalini received her M.Tech. and Ph.D. degree from National Institute of Technology, Trichy, Tamil Nadu, India. She has 22 years of experience in teaching and research area. Currently she is working as professor in Department of Electrical and Electronics Engineering, St. Peter's Institute of Higher Education and Research, Chennai, 600054, Tamil Nadu, India. She received various projects from eminent institute and organized a greater number of events at different level. She has published more papers in peer review international journals and conferences.

An IEC 61131-3-Based PLC Timers Module Implemented on FPGA Platform

Dhruv M. Patel, Ankit K. Shah and Yagnesh B. Shukla

Abstract In assembling units, the PLC plays a significant role in measuring and controlling various applications like motor control, blending of fluid inside tanks, fluid level control of tanks and others. The existing PLCs are offering sluggish response and poor scanning time based on sequential processing execution. This paper has proposed various delay timer modules of PLC on reconfigurable FPGA-LabVIEW platform with IEC 61131-3 standard compliant. The GUI and LD-based input user program are created with the assistance of LabVIEW, and it is assembled with a massive parallel processing-based hardware structure inside FPGA. The TON, TOFF, RTON and RTOFF timer modules are developed with various time-based values like 1–1000 ms. Others PLC functions like AND, OR, NOT, NAND, NOR, X-OR and X-NOR, PT and OT are realized and simulated with timer module. The PLCs-based delay timer modules are fully implemented inside FPGA hardware (NI myRIO-1900). Proposed algorithm is validated on DC motor, electrical oven and single tank systems along with the comparison of SIMATIC S7-200 PLC. The proposed design offers remarkable benefits like user friendliness, cost-effectiveness, miniaturization, simplicity, faster speed and scanning time.

Keywords Programmable logic controller (PLC) · Ladder diagram (LD) · Reconfigurable hardware · Field programmable gate array (FPGA)

D. M. Patel (✉)
PhD Scholar, Instrumentation and Control, Gujarat Technological University, Chandkheda, Gujarat, India

Assistant Professor, Instrumentation and Control, Sardar Vallabhbhai Patel Institute of Technology, Vasad, Gujarat, India

A. K. Shah
Assistant Professor, Instrumentation and Control, L. D. College of Engineering, Ahmedabad, Gujarat, India
e-mail: ankitshah.ic@ldce.ac.in

Y. B. Shukla
Professor, Electronics and Communication, Sardar Vallabhbhai Patel Institute of Technology, Vasad, Gujarat, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_48

1 Introduction

PLCs are a client-configurable, microcontroller (μc)/microprocessor (μp)-based computerized computational device [1, 2] utilized in industry for programmed control of various units and performs various tasks like successive rationale, timing and counting logic, arithmetic and logical functions, data manipulation, communication, analog as well as simple and/or hybrid control algorithms [3, 4]. The processing speed of the μc/μp and input client program length characterize the scanning period of PLCs. The good reliability, flexibility and robustness of PLCs made attraction toward the designer. However, in certain real-time applications, it has been noticed that μc/μp-based PLC designs are not suitable enough to cover the performance and indicate that μc/μp-based PLCs need technological improvement. This improvement can be given by financially savvy, reconfigurable hardware FPGA-based technology [5]. The point of the exhibited work is to synthesis the LD program into the consecutive and parallel structure and to execute different timer modules like timer ON (TON), timer OFF (TOFF), retentive timer ON (RTON), retentive timer OFF (RTOFF) delay with various based qualities like 1, 10, 50 ms and others on the FPGA hardware. It is additionally used to execute different modules like AND, OR, NOT, NAND, NOR, X-OR and X-NOR, pulse timer (PT) and oscillating timer (OT).

2 Materials and Methods

2.1 Materials

2.1.1 IEC 61131-3

An International Electrotechnical Commission (IEC) 61131 standard describes programming languages, concepts and guidelines for PLCs' projects [6]. An execution of industrial PLCs and industrial applications is developed with the help of textual programming languages like structured text (ST), instruction list (IL) and graphical languages like LD and function block diagram (FBD) [7, 8].

2.1.2 FPGA

FPGAs are a reconfigurable, universally useful, staggered programmable logic device that modified on the bundle with the assistance of clients [9]. Present-day FPGAs would now be able to contain over a million equal rationale logic gates and a huge number of flip-flops in the single bundle and that can be designed to actualize an assortment of programming calculations. The computerized frameworks are structured by composing programming as HDL like VHDL [10], Verilog and others.

The author [11] presents the LabVIEW structure for FPGA programming, explicitly focused to researchers and specialists. The authors [12–16] presented improved PLC configuration, utilizing FPGA dependent on parallel execution system for the development of execution and adaptability of PLCs. Consequently, we propose the methodology of usage of FPGAs rather than traditional μp and μc for PLC configuration because of its potential for critical execution upgrades due to FPGA's parallel execution.

2.2 Methods

2.2.1 Proposed Design Methodology and Architecture

This paper presented a conventional micro-PLC design that could be acknowledged using configurable FPGA-based hardware. Once FPGA is appropriately configured by LabVIEW programming, alongside a suitable graphical user interface (GUI), and the design shows the ability to work as micro-PLC with great execution and reasonable flexibility. An IEC 61131-3-based user ladder program could be an input to micro-PLC design. The LabVIEW programming is utilized for ladder logic troubleshooting. In this way, our proposed methodology requires no extraordinary preparing necessities for the plant specialists or clients, as it utilizes pre-configured FPGA as a conventional micro-PLC. The design methodology of the proposed configuration has been presented in Fig. 1. By utilizing sequential correspondence, reasonable FPGA code has been transmitted to a FPGA using GUI. This code gives data with respect to choice of specific rung, segments utilized like TON, TOFF, RTON, RTOFF, PT, OT, AND, OR, NAND, NOR, NOT, EX-OR, EX-NOR and its information sources.

The focus of proposed work is to design, simulate and validate the PLCs timer module inside the FPGA hardware. The proposed architecture of FPGA-based PLCs timer module is presented in Fig. 2. The timer modules are used to delay action and keep an output state ON or OFF for a specified time interval after or before an input state turns ON or OFF. The TON, TOFF, RON and RTOFF timer modules have information about the address of timer, base time (BT), preset time (PRT) and accumulated timing value (ATV) or elapsed time (ET) and timer done bit (TDN) of timer. The PRT of timer is defined as the time duration of the timing circuit. The FPGA-PLC timer module has timer controller, timer module like TON, TOFF, RTON, RTOFF, PT and OT. Based on the BT (1, 10, 50 ms and others) of timer, the timer controller starts accumulating ATV, compares ATV to PRT and updates TDN equal to 1 if both values are the same else 0. The results of the timing functioning would be stored in the timer parameter list of a RAM. In this design for TON, TOFF, RTON and RTOFF timer, the PRT is equal to timer delay (TD) value divided by the TB of timer . For example, if a TD of 11 s is required, the timer will have a PRT of

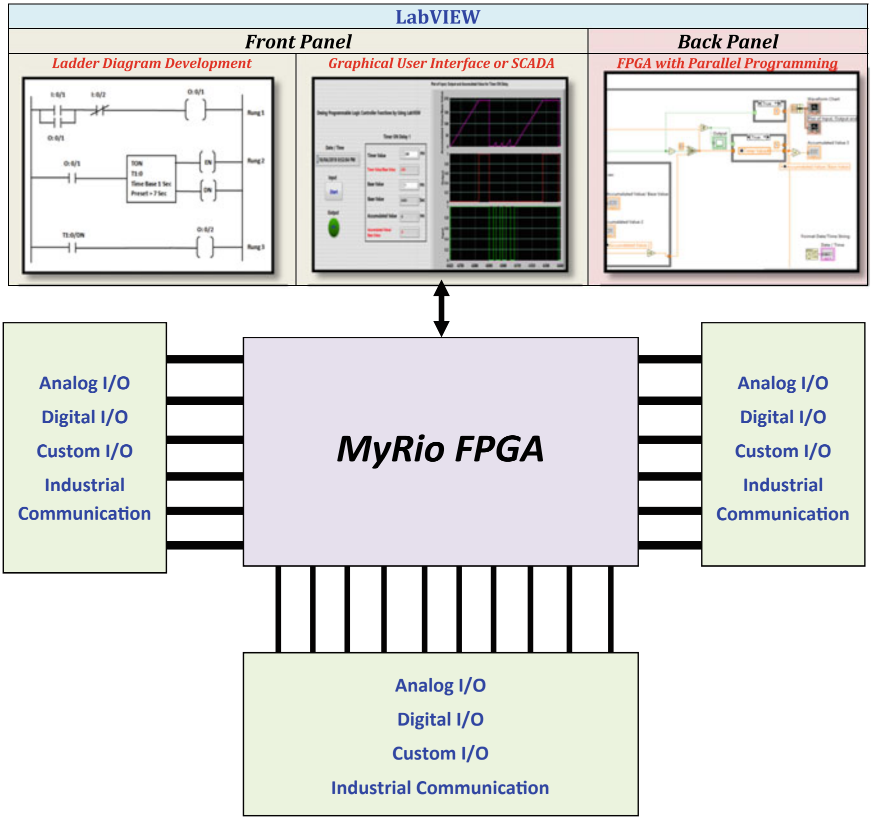


Fig. 1 Proposed design methodology

11 s and TB 1 s. The time delay is defined as the product of PRT and TB. The ATV is defined as the amount of time that has elapsed since it is energized. The timers can be programmed with different TB like 1, 0.1, 0.01 and 0.001 s. For example, if you select TB = 0.01 s and PRT = 400, the timer has a TD = 4 s (400 × 0.01 s = 4).

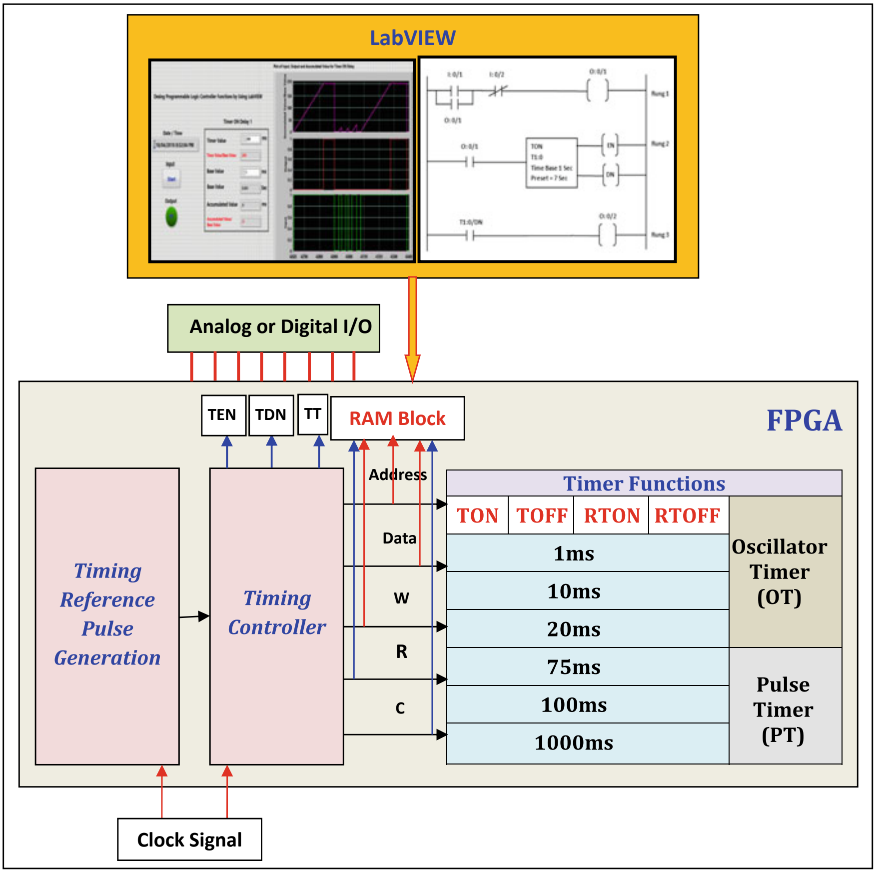


Fig. 2 Proposed architecture

3 Theory

3.1 Results of Literature Survey

To solve existing PLCs constraints, the FPGA-PLCs have been developed by many researchers. Milik [14] present a set of algorithms for the synthesis of reconfigurable PLCs on FPGA platform and programmed according to IEC1131 and EN61131. Monmasson et al. in [17] contribute design methodologies and the newest ideas and features of FPGA for industrial control applications. The authors proved that FPGA-based system offers better performance and high-speed operation. The authors in [18] presented smaller-scale PLCs configuration dependent on FPGA, containing just seven components in series and four components in parallel for each rung or system. The design [18] offers less scan time, and its each rung execution requires '2 * m'

clock cycles. For example, if the ladder program involves ‘*n*’ rungs, it will require ‘2 * *m* * *n*’ clock cycles for PLC scan, where ‘*m*’ shows numbers of segments utilized in all rungs of PLC LD. For the most extreme conceivable ladder logic, the plan exhibited in [18] needs to accomplish 2.24 μs scan time at 100 MHz clock frequency. The author in [19] supports a comparative structure approach and development of CPU dependent on bit-word design for the PLCs executed in a Virtex-4 FPGA-based development platform. Yu et al. in [20] present PLCs timer system dependent on ARM controller and FPGA hardware, in which ARM controller is utilized to execute client programs while FPGA equipment is utilized for timer operation in parallel.

4 Results and Discussions

4.1 Timer ON Delay

The input signal of the TON delay instruction is changing from 0 to 1 state, and the output TDN bit of TON timer is set to 1 after the predefined duration of the time delay. When the TON is executed, the PRT begins and TEN pin is set to 1. In TON delay timer, TDN pin stays ON as long as the input logic to TON stays ON. The timer takes action to quit timing, when the ATV equals to PRT. When the input signal of TON delay is changed from 1 to 0 states, the TEN and TDN pin of TON delay are reset to 0 and ATV is set to 0. The timer function is started again when a new positive edge of the signal is detected at the start input. Therefore, this is called a TON delay energize instruction. The LD and a timing diagram of TON are shown in Fig. 3a, b, respectively. The timing diagram is used to monitor turn ON and OFF status of input and output. The LD indicates TON instruction as timer’s address is T1:0, PRT is 200 ms, TD is 200 ms, and its TB is 1 ms.

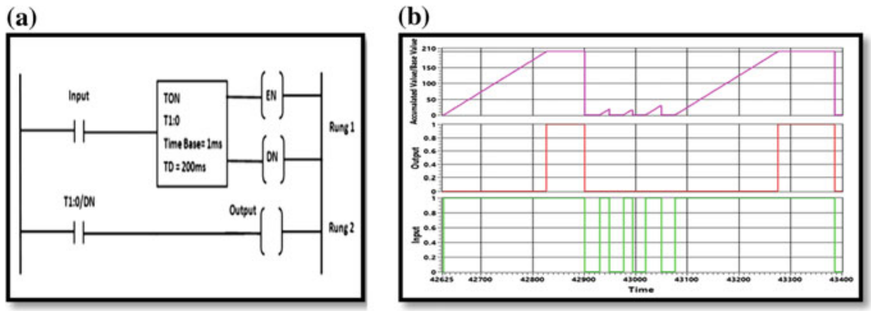


Fig. 3 TON timer module. **a** LD, **b** timing diagram results with TB = 1 ms, TD = 200 ms and PRT = 200

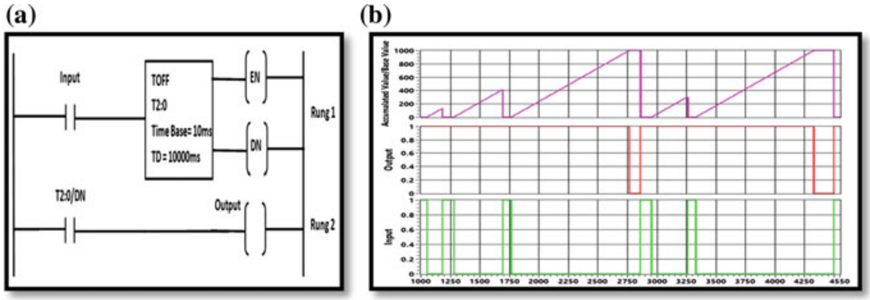


Fig. 4 TOFF timer module. **a** LD, **b** timing diagram with TB = 10 ms, TD = 10,000 ms and PRT = 1000

4.2 Timer OFF Delay

The working of TOFF module is little bit different compared to TON module with TDN and TEN outputs. TOFF has information of timer's PRT, TD value, TB value and ATV value. When the input of TOFF is putting on turns OFF condition and a specific delay has happened, the TOFF delay module de-energized its TDN output. The functioning of TOFF timer module with ladder and timing diagram is presented in Fig. 4a, b respectively. Figure 4a indicates the TOFF instruction as timer's address is T2:0, PRT is 1000, TD is 10,000 ms, and its TB is 10 ms.

The TEN and TDN outputs of TOFF module remain turn ON with respect to ON status of the input. Due to that, the accumulation of timer timing is not started. When the status of TOFF module input signal is changed from ON to OFF, the status of TEN is updated to OFF and the ATV of timer timing is started. Once the PRT equals ATV, the accumulation of timing is stopped and TDN status is updated to OFF.

4.3 Retentive Timer ON Delay

Once the input signal of the RTON delay is changing from 0 to 1 state, the TDN bit of RTON timer is set to 1 after the predefined duration of PRT. When the RTON is executed, the PRT begins and TEN pin is set to 1. The TDN pin remains set to 1 as long as the start input is still 1. Similarly, the input signal of RTON is changed from 1 to 0 states, the TEN and TDN pin of RTON delay are updated to 0, and ATV is set to last time value. The ATV is started from previous time count instead of 0 ms when a new positive edge of the input signal is detected. Once the ATV is equal to PRT value, the TDN will turn ON. When a RTON has timed out, its TDN status stays ON condition even if its TEN turns OFF. With the help of a reset instruction, the ATV and TDN are set to 0 and OFF, respectively. The LD and result of RTON timer timing diagram are presented in Fig. 5a, b, respectively.

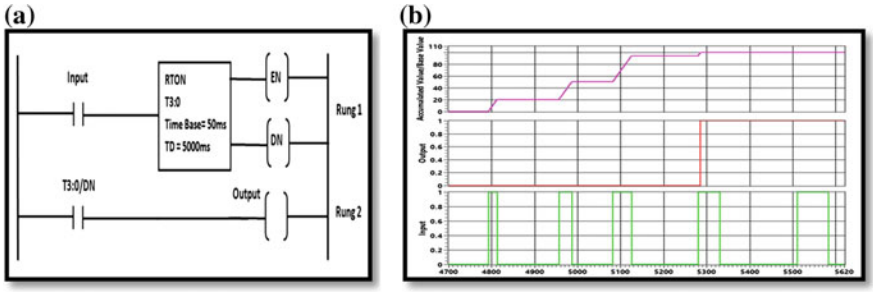


Fig. 5 RTON timer module. **a** LD, **b** timing diagram with TB = 50 ms, TD = 5000 ms and PRT = 100

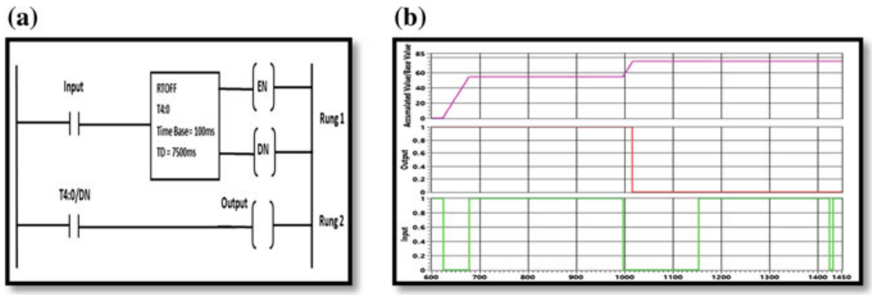


Fig. 6 RTOFF timer module. **a** LD, **b** timing diagram with TB = 100 ms, TD = 7500 ms and PRT = 75

4.4 Retentive Timer OFF Delay

The working of RTOFF is different compared to TOFF and presented in Fig. 6. A RTOFF can stop and start timing again without its ATV resetting to 0. Once the input status updated ON, then the TEN and TDN are set to ON. Due to that, the accumulation of timing is not started because RTOFF is waiting for an OFF input status in place of an ON status. When input of RTOFF is updated to OFF, the ATV of RTOFF is started from previous value and TEN will set to OFF. Once the ATV is the same as a PRT value, RTOFF stops accumulation of timing and sets TDN to OFF.

4.5 Pulse Timer

The PT module is used to create output pulses of specified time duration. Figure 7 has shown the timing diagram of PT. Once the logic of input is set to ON, the output of PT follows and stays ON for specific time delay duration. Similarly, the value of

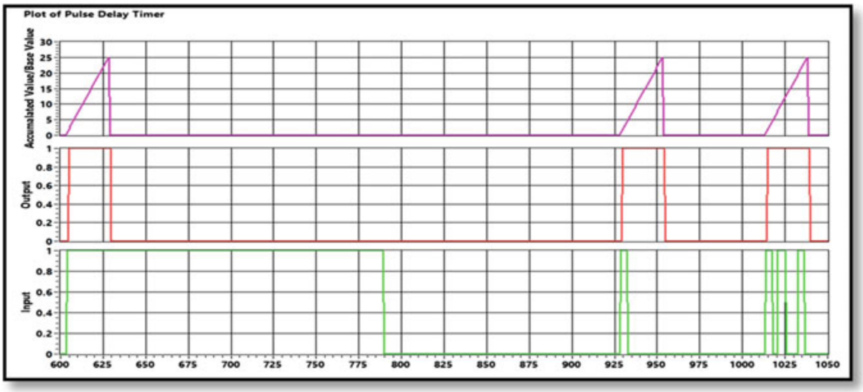


Fig. 7 Timing diagram of pulse timer

ATV or ET is increased. After termination of PT, the ET is set to 0. The status of output will stay ON until the pulse time has elapsed, irrespective of input status.

4.6 Oscillator Timer

The oscillator timer has three outputs like TDN, ET0 and ET1 and includes information about the timer's preset time 0 and 1 value, TB and ATV. When the TDN output of oscillator timer is ON, the output is energized. The OT is used for generation of train pulses of ON, OFF state with different frequencies. Similarly, it is also used for operation of motors, heaters and other appliances. For different cases, the GUI and timing diagram of OT are presented in Fig. 8a, b. The preset time 0 and preset time 1 are considered as an OFF time and ON time of PT module, respectively. Once input signal is IN set to ON, the OFF timing is started and the ET0 is increased with respect to base time. When the ET0 is equal to PRT0, the output goes ON and ET0 is reset. At the same time, IN remains ON, the ON timing function is started, and the ET1 is increased with respect to TB. When ET1 equals to PRT1, the output goes OFF and ET1 is cleared. This function is repeating continuously as long as the IN stays ON and generating trains pulses depending on PRT0 and PRT1. When IN is updated and stays OFF, the output is forced to be OFF. The GUI and timing diagram of OT with PRT0 = 500 ms and PRT1 = 800 ms are shown in Fig. 8a. The timing diagram of OT with PRT0 = 500 ms and PRT1 = 500 ms is shown in Fig. 8b.

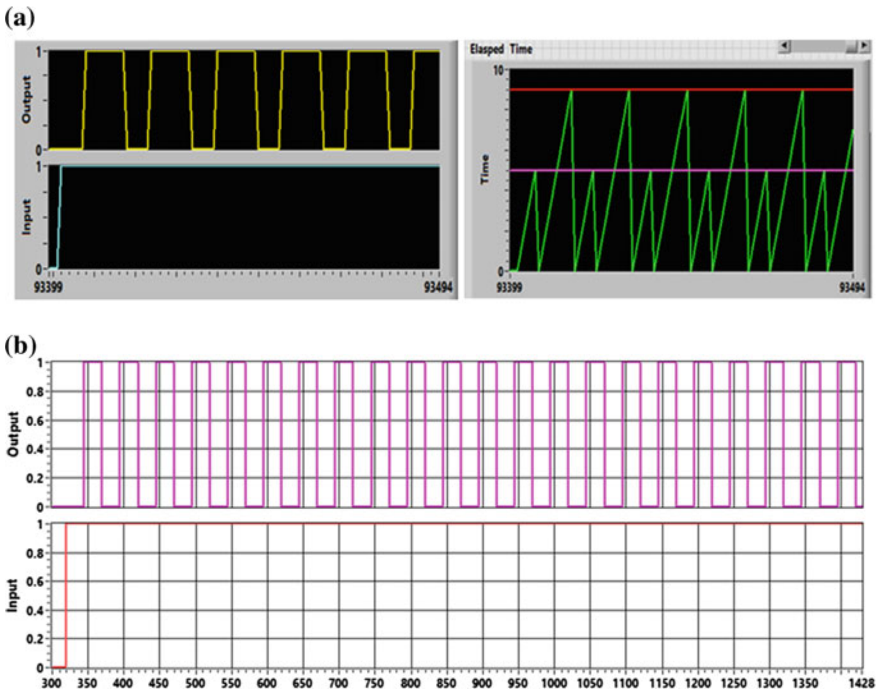


Fig. 8 **a** GUI and timing diagram of OT with PRT0 = 500 ms and PRT1 = 800 ms, **b** timing diagram of OT with PRT0 = 500 ms, PRT1 = 500 ms (square wave generation of $f = 1$ Hz)

4.7 Real-Time Applications of Timers

4.7.1 Application 1: Operation of D. C. Motor

The proposed design has been validated by implementation of D. C. motor application, presented below. A specification of inputs and outputs for application 1 has been summarized in Table 1. Pressing input button 1, motor 1 is started and after 7 seconds, motor 2 is started if input button 3 is ON. If motor 1 and motor 2 are ON and input button 5 is ON, then motor 3 is ON. If input button 6 is pressed from a Normally Closed (NC) to Normally Open (NO), motor 3 is stopped. If input button 4 is pressed from NC to NO, motor 2 and motor 3 are stopped and If input 2 is changing from NC to NO, all motors stop. The ladder logic, GUI and results for application 1 are presented in Fig. 9a, b, c respectively.

4.7.2 Application 2: Electrical Oven System

The proposed structure has been further experimentally tested on electrical oven home application. Specifications of inputs and outputs for application 2 have been

Table 1 Module inputs, outputs and timer address of Application 1

Sr. no.	Inputs	Outputs	Timer
1	Input Button 1 (Start Button) = I:0/1	Output 1 (Motor 1) = O:0/1	Timer ON delay
2	Input Button 2 (Stop Button) = I:0/2	Output 2 (Motor 2) = O:0/2	T1:0
3	Input Button 3 (Start Button) = I:0/3	Output 3 (Motor 3) = O:0/3	Base value: 1 s
4	Input Button 4 (Stop Button) = I:0/4		Preset: 7
5	Input Button 5 (Start Button) = I:0/5		Timer delay = 7 s
6	Input Button 6 (Stop Button) = I:0/6		

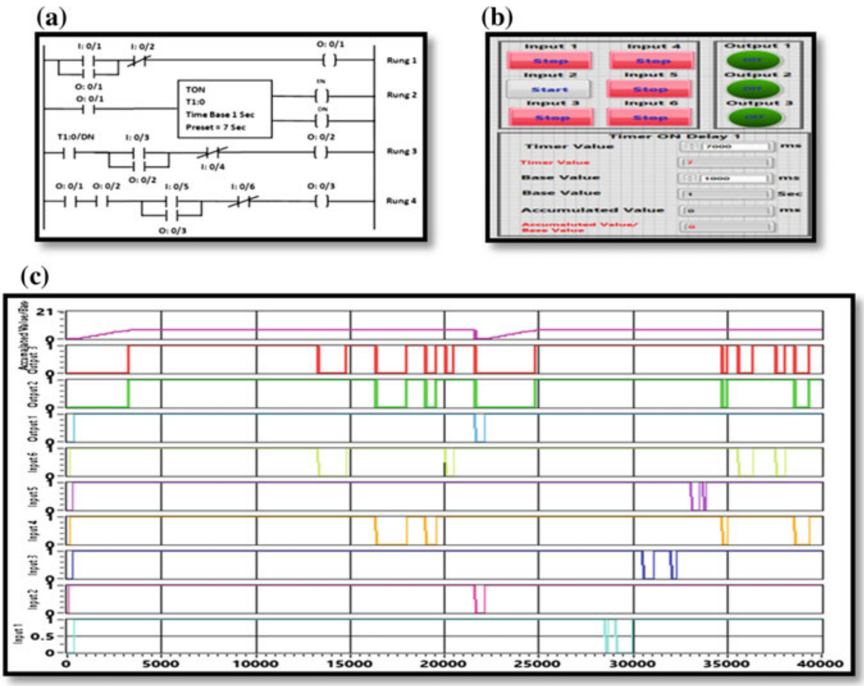


Fig. 9 Application 1. **a** LD, **b** GUI and **c** results

Table 2 Module inputs, outputs and timer address of Application 2

Sr. no.	Inputs	Outputs	Timer
1	Input Button 1 (Start Button) = I:0/1	Output 1 (electrical heater) = O:0/1	Timer OFF delay T1:0 Base value: 1 s
2	Input Button 2 (Stop Button) = I:0/2	Output 2 (cooling fan 1) = O:0/2	Preset: 8 Timer delay = 8 s
3	–	Output 3 (cooling fan 2) = O:0/3	

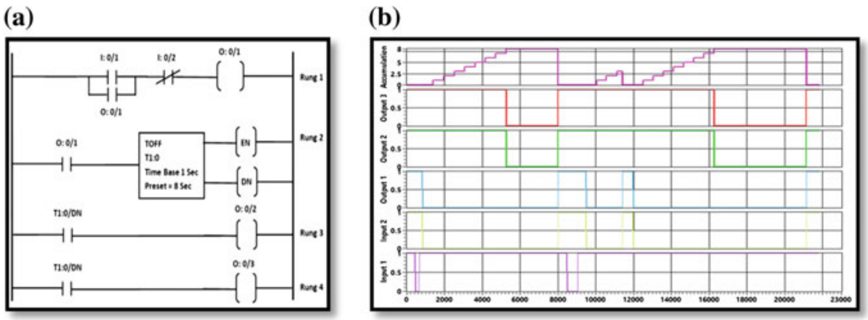


Fig. 10 Application 2. **a** LD and **b** results

listed in Table 2. The oven is heated with the help of an electrical heater and inside oven, and two ventilation fans are fitted for cooling the oven after the use. The functioning of application 2 shall work as per under-mentioned logic conditions; Heating ON Condition: Heating element and both cooling fans turn ON simultaneously. By using cooling fans, the hot air is circulated inside the oven and spreads heat; Heating OFF Condition: Heating element OFF and OFF delay timer start counting down; Timer Done Condition: Both cooling fans turn OFF and If Stop Button is pressed from a NC to NO, heater is stopped and both cooling fans are stopped after 8 s. The LD and results for application 2 are shown in Fig. 10a, b respectively.

4.8 Comparative Result Analysis of Proposed FPGA-Based PLC Design with Existing PLC

The proposed structure is validated by the execution of water level control application. Specifications of inputs and outputs for water level control application have been listed in Table 3 (SIMATIC S7-200 Smart PLC) and Table 4 (proposed design). In the proposed application, the detection of low level of water is done with the help of level switch 1 and the water pump 1 is used for pumping the water inside tank for maintaining level. Similarly, when medium water tank level is detected, after 1 s,

Table 3 SIMATIC S7-200 smart PLC inputs, outputs and timers address

Sr. no.	Inputs	Outputs	Timer
1	CPU_Input0 (I0.0, level switch 1)	CPU_Output0 (Q0.0, pump 1)	Timer ON delay T33
2	CPU_Input1 (I0.1, level switch 2)	CPU_Output1 (Q0.1, pump 2)	Base value: 10 ms Preset: 100, TD = 1000 ms

Table 4 Proposed design inputs, outputs and timers address

Sr. no.	Inputs	Outputs	Timer
1	Input 0 (I:0/0, level switch 1)	Output 0 (O:0/0, pump 1)	Timer ON delay T1:0
2	Input 1 (I:0/1, level switch 2)	Output 1 (O:0/1, pump 2)	Base value: 10 ms Preset: 100, TD = 1000 ms

pump 2 is started for pumping water level inside tank. Figure 11a–c is shown for LD, GUI and results of water level application, respectively. It indicated that when input 0 is ON, the output 0 is energized else de-energized. When input 1 is set to ON, after 1000 ms the TDN and output 1 is energized. The same program is implemented on SIMATIC S7-200 Smart PLC [21], and its result is presented in Fig. 11d. The result of Fig. 11d indicates that the inputs and outputs waveforms provide 250 ms resolution in the graph. The same ladder program is implemented in the proposed FPGA-based PLC design shown that, the inputs and outputs waveforms provide 10 ms resolution in the graph. These results also indicate that at the same time instant the inputs and outputs follow to each other.

The proposed structure contains 64 rungs and 16 components for the each rung or network. The scanning time of PLCs depends on the ladder program length, processor speed and numbers of inputs and outputs. In the proposed design, for execution of each rung requires 'm' clock cycles. For example, if the ladder diagram have 'n' rungs, it requires '$m + n + 3$' clock cycles for scan, where 'm' indicates numbers of components used in rung of PLCs ladder programs. In execution, out of three clock cycles, two are used for updating inputs and outputs and remaining is required for signal arrival before clock cycle. In this design, for highest (64 rungs and 16 elements in each rung) and lowest (1 rung, 1 element) ladder scan operation, total 83 and four clock cycles are utilized, respectively. The proposed design offers highest and lowest scan time is 2.075 μs and 100 ns, respectively, with 40 MHz clock frequency. For the greatest possible LD, the design discussed in [18] has to achieve 2.24 μs scan time at 100 MHz clock frequency. Eventually, it is apparent that the proposed approach results into faster scanning time compared to the conventional PLC.

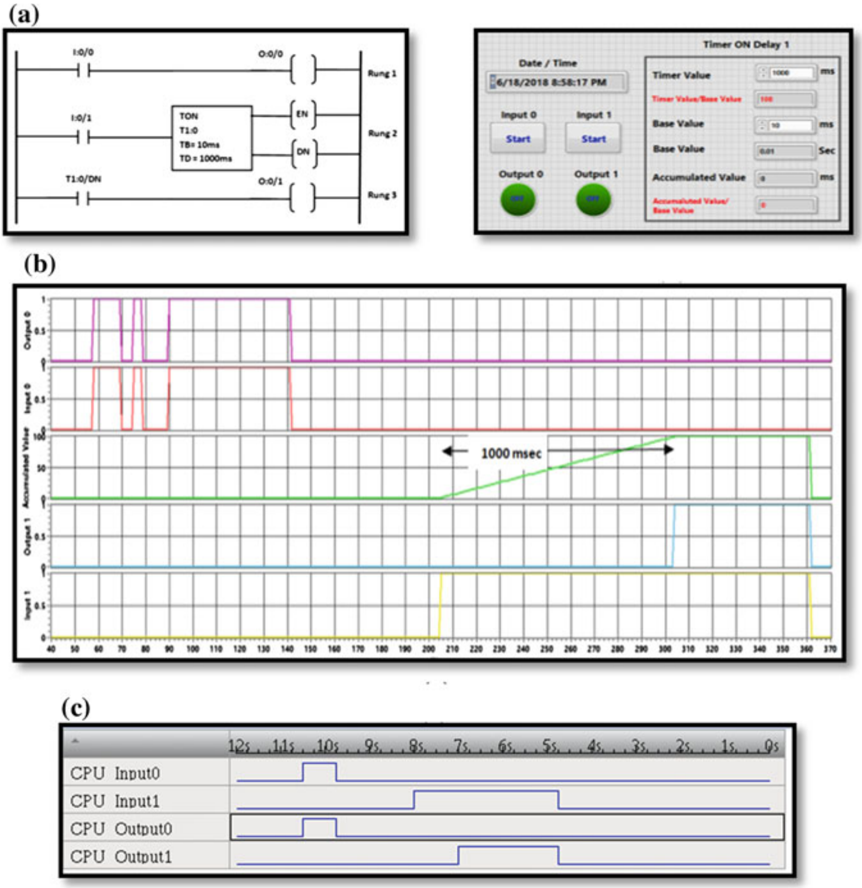


Fig. 11 **a** LD for PRT = 100, TB = 10 ms and TD = 1000 ms, **b** GUI, **c** results of proposed design and **d** results of SIMATIC S7-200 smart PLC

5 Conclusion

The proposed approach used FPGA-based computing in place of conventional microprocessor or microcontroller for PLCs design along with LabVIEW-based GUI. The proposed design also offers the parallel execution mechanism of FPGAs for parallel execution of the networks or rungs of ladder logic instead of the sequential execution of rungs by conventional microprocessor or microcontroller-based PLCs which enhance performance and flexibility. LabVIEW programming has been used for the development of ladder logic and GUI design along with FPGA hardware kit. Proposed design is validated on the real-time applications of timer module with comparative result analysis of existing PLC. Results show that the proposed approach offers less scanning time and higher flexibility compared to the conventional PLC.

Acknowledgements The authors present their heartily thanks to the management, faculty members and staff members of Sardar Vallabhbhai Patel Institute of Technology, Vasad, L. D. Engineering College, Ahmadabad and Gujarat Technological University, Chandkheda, Gujarat, India, for their cooperation, non-stoppable support and help in the development of presented work.

References

1. Webb, J.W., Reis, R.A.: Programmable Logic Controllers: Principles and Applications. Prentice Hall PTR (2002)
2. Kamel, K., Kamel, E.: Programmable Logic Controllers: Industrial Control. McGraw-Hill Education (2014)
3. Bryan, L.A., Bryan, E.A.: Programmable Controllers: Theory and Implementation. Industrial Text Company (1997)
4. Auslander, D.M., Pawlowski, C., Ridgely, J.: Reconciling programmable logic controllers (PLCs) with mechatronics control software. In: Proceeding of the 1996 IEEE International Conference on Control Applications. IEEE (1996)
5. Rodriguez-Andina, J.J., Moure, M.J., Valdes, M.D.: Features, design tools, and application domains of FPGAs. IEEE Trans. Ind. Electron. **54**(4), 1810–1823 (2007)
6. John, K.H., Tiegelkamp, M.: IEC 61131-3: Programming Industrial Automation Systems: Concepts and Programming Languages, Requirements for Programming Systems, Decision-Making Aids. Springer Science & Business Media (2010)
7. Cenelec, E.N.: 61131-3, Programmable Controller—Part 3: Programming Languages. International Standard Management Centre, Brussels (2013)
8. Chmiel, M., Czerwinski, R., Smolarek, P.: IEC 61131-3-based PLC implemented by means of FPGA. IFAC-PapersOnLine **48**(4), 374–379 (2015)
9. Trimberger, S.M. (ed.): Field-Programmable Gate Array Technology. Springer Science & Business Media (2012)
10. Hauck, S., DeHon, A.: Reconfigurable Computing: The Theory and Practice of FPGA-Based Computation, vol. 1. Elsevier (2010)
11. Andrade, H.A., Ahrends, S., Hogg, S.: Making FPGAs accessible with LabVIEW. In: FPGAs for Software Programmers, pp. 63–79. Springer, Cham (2016)
12. Chodorowski, P., Chmiel, M.: IEC 61131-3 compliant PLC structure based on FPGA multi-core solution. In: 2016 International Conference on Signals and Electronic Systems (ICSES). IEEE (2016)
13. Milik, A.: Multiple-core PLC CPU implementation and programming. J. Circ. Syst. Comput. **27**(10), 1850162 (2018)
14. Milik, A.: On hardware synthesis and implementation of PLC programs in FPGAs. Microprocess. Microsyst. **44**, 2–16 (2016)
15. Patel, D., Bhatt, J., Trivedi, S.: Programmable logic controller performance enhancement by field programmable gate array based design. ISA Trans. **54**, 156–168 (2015)
16. Patel, D., Trivedi, S., Bhatt, J.: Design and implementation of field programmable gate array based programmable logic controller. In: 2nd International Conference on Signals, Systems and Automation, G H Patel College of Engineering & Technology, Vallabh Vidyanagar, India. Universal Publishers, USA (2011)
17. Monmasson, E., et al.: FPGAs in industrial control applications. IEEE Trans. Ind. Inform. **7**(2), 224–243 (2011)
18. Gawali, D., Sharma, V.K.: FPGA based micro-PLC design approach. In: International Conference on Advances in Computing, Control, & Telecommunication Technologies, 2009. ACT'09. IEEE (2009)

19. Chmiel, M., et al.: Central processing units for PLC implementation in Virtex-4 FPGA. IFAC Proc. Vol. **44**(1), 7860–7865 (2011)
20. Yu, L., et al.: Design of PLC timer system based on ARM + FPGA. Appl. Mech. Mater. **249** (2013). Trans Tech Publications
21. Siemens AG: SIMATIC S7-200 Programmable Controller System Manual. Siemens AG, Monachium (2002)

Impact of Temperature on Circuit Metrics of Various Full Adders

M. Aalelai Vendhan

Abstract This research work emphasizes the effect of temperature variation on circuit metrics of distinct 1-bit full adder circuits operating at low voltage. This work is studied using 90 nm MOSFET technology. The design styles employed for constructing the adders are conventional complementary metal-oxide semiconductor (CMOS), complementary pass-transistor logic (CPTL) and transmission gate (TG). Cadence Virtuoso is used for designing and simulation of the circuits. The circuit metrics such as average power, delay and power delay product (PDP) are calculated from the simulation results. On comparison of the results, it is evident that average power and PDP of CPTL adder remain least affected by temperature, while the delay of CMOS adder remains least affected by temperature.

Keywords Complementary pass transistor · Propagation delay · Average power · Temperature variation

1 Introduction

In the modern era, the performance of any digital circuit is assessed by its power consumption and propagation delay. The delay of any digital circuit depends on internal factors such as the threshold voltage V_{TH} and electron mobility [1]. As the technology scales down into the sub-nanometre level, external factors such as temperature also affect the propagation delay [2]. At high supply voltages, raising the temperature increases the rate of electron scattering. As a result of this, mobility decreases, which in turn increases the delay. In contrast, when the circuit is operated in very low supply voltages, increasing the temperature reduces the value of threshold voltage V_{TH} [3]. Due to this action, the drain current increases which in turn decreases the propagation delay. This phenomenon is known as inverted temperature dependence [1].

M. Aalelai Vendhan (✉)
Department of Electronics and Communication Engineering, Panimalar Engineering College, Chennai, Tamil Nadu 600123, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_49

Temperature also plays a vital role in power dissipation of a circuit. The two main components of power dissipation are static power and dynamic power. The primary reason for static power is subthreshold leakage. The reduction of V_{TH} increases the subthreshold leakage exponentially [4]. Due to this reason, static power increases exponentially with the increase in temperature. The dynamic power of the circuit is not affected by the temperature as it depends only on supply voltage, gate capacitance and frequency [5].

The full adder forms the crux of designing larger digital circuits. It is used for constructing arithmetic circuits like multiplier, divider, subtractor and comparator [6]. A full adder can be designed using various logic styles like CMOS, CPTL, TG, gate diffusion input (GDI) [7]. Each logic style has its own advantages and disadvantages. The CPL takes very less time for computational processes, whereas it dissipates a large amount of power. The TG logic occupies less surface area than other logic styles, but it exhibits poor driving capability [8]. In this research work, the effect of temperature on circuit metrics of different full adder structures is analysed and compared. This work is structured as follows: Sect. 2 explains the different adder structures, while Sect. 3 explains the experimental set-up, and Sect. 4 analyses the simulation results. Finally, Sect. 5 is about the conclusion.

2 Distinct Full Adder Structures

2.1 Conventional CMOS Logic Adder

This type of adder is designed by the combination of both pull-up network (PMOS) and pull-down network (NMOS). It is depicted in Fig. 1. This design is implemented using 28 transistors. Strong '1' is obtained from pull-up transistors, and strong '0' is obtained from pull-down transistors. The main advantage of this design is that it

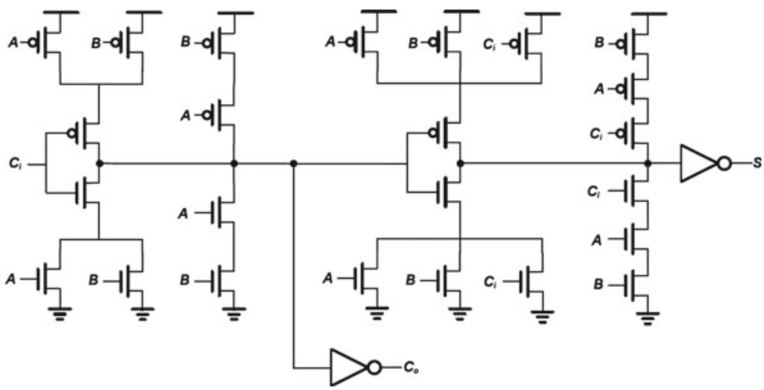


Fig. 1 Conventional CMOS logic adder

provides full-swing output [9]. But, the usage of 28 transistors leads to large area occupancy.

2.2 Complementary Pass-Transistor Logic Adder

This adder is designed by using only NMOS transistors. It is depicted in Fig. 2. The outputs in this design need to be restored, so PMOS transistors are used for restoration purposes. This circuit makes the use of both complementary inputs/outputs. This design is implemented using 38 transistors. The main advantage of this design is that its operation speed is very high [8]. But, the power dissipation in this logic is also very high.

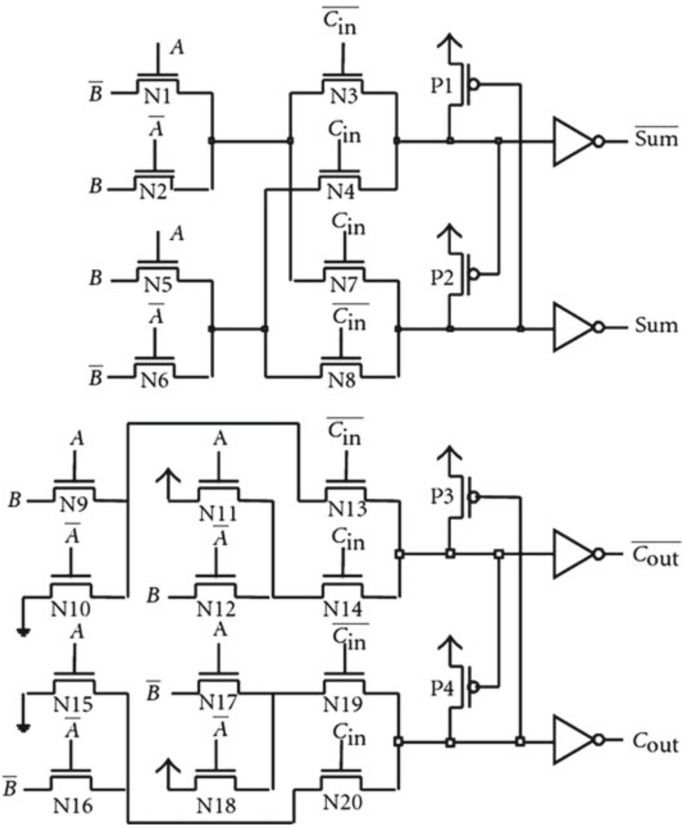


Fig. 2 Complementary pass-transistor logic adder

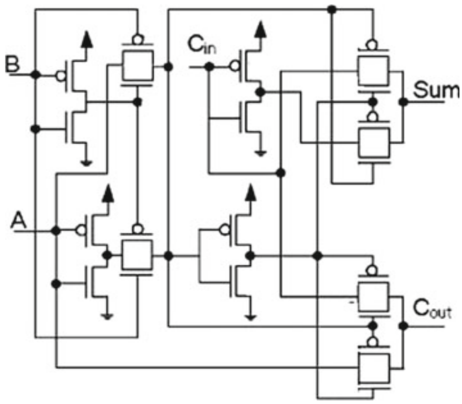


Fig. 3 Transmission gate logic adder

2.3 Transmission Gate Logic Adder

This design involves the use of parallel connection of NMOS and PMOS transistors. It is depicted in Fig. 3. This design is implemented by using 20 transistors. The main advantage of this design is that it avoids the voltage degradation problem [8]. But, this design style performs the operation very slowly.

3 Experimental Set-up

Three distinct types of logic design styles were used for designing the full adders, i.e. conventional CMOS logic-based adder, CPTL-based adder, and TG-based adder. All these structures were designed in 90 nm MOSFET technology. The simulation of these circuits was carried out using spectre simulator. These circuits were tested over the temperature range of −25 to 125 °C. The length of both NMOS and PMOS transistors are 100 nm. The width of NMOS transistor is 120 nm, whereas the width of PMOS transistor is 480 nm. The supply voltage is provided with 1 V.

4 Results and Discussion

Figure 4 illustrates the variation of average power in accordance with the temperature of various full adder circuits. From Fig. 4, it is clear that the average power of all the adder circuits increases with the increase in temperature and CPTL adder consumes high power (172 –627 nW), whereas CMOS adder consumes the least amount of power (41–183 nW). Figure 5 illustrates the variation of delay in accordance with the temperature of various full adder circuits. From Fig. 5, it is clear that the delay of all the adder circuits decreases with the increase in temperature and TG adder

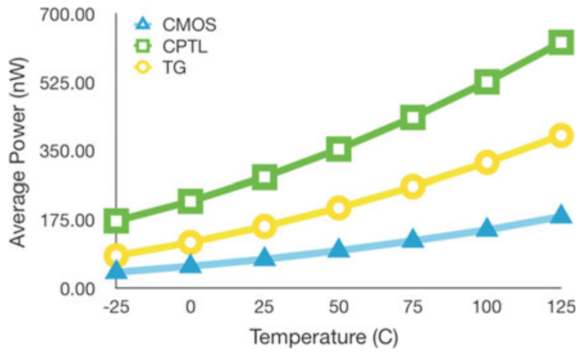


Fig. 4 Effect of temperature on average power

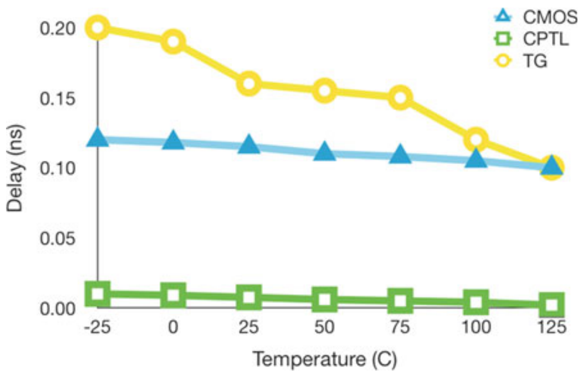


Fig. 5 Effect of temperature on delay

had the highest delay (0.2–0.1 ns), whereas CPTL adder had the least delay (0.01–0.002 ns). Figure 6 illustrates the variation of PDP in accordance with temperature of various full adder circuits. From Fig. 6, it is clear that the PDP of all the adder circuits increases with the increase in temperature and TG adder exhibits the highest PDP (16.6–38.9 aJ), whereas CPTL adder exhibits the least PDP (1.72–1.5 aJ).

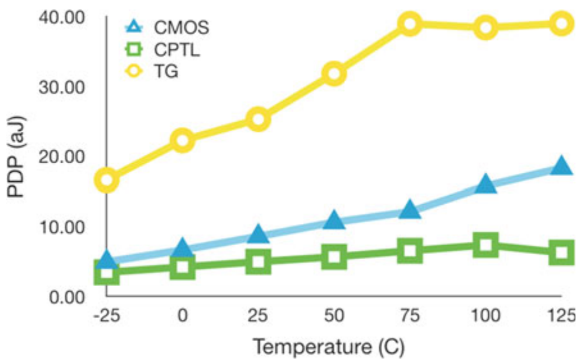


Fig. 6 Effect of temperature on power delay product

For effective analysis of the circuits, the average percentage change in circuit metrics of all the adders is calculated and depicted. Figure 7 depicts the comparative analysis of the percentage change in average power of all adder structures. From Fig. 7, it is evident that CPTL adder exhibited the smallest percentage change (22%) in average power and TG adder exhibited the highest percentage change (26%). Figure 8 depicts the comparative analysis of the percentage change in a delay of all adder structures. From Fig. 8, it is evident that CMOS adder exhibited the smallest percentage change (3%) in delay and CPTL adder exhibited the highest percentage change (23%). Figure 9 depicts the comparative analysis of the percentage change in PDP of all adder structures. From Fig. 9, it is evident that CPTL adder exhibited the smallest percentage change (14%) in PDP and CMOS adder exhibited the highest percentage change (21%). Based on the simulation results, the advantages and disadvantages of all the adder structures are presented in Table 1.

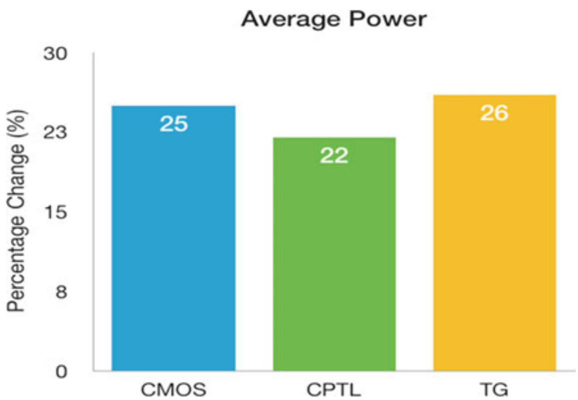


Fig. 7 Average percentage variation of average power

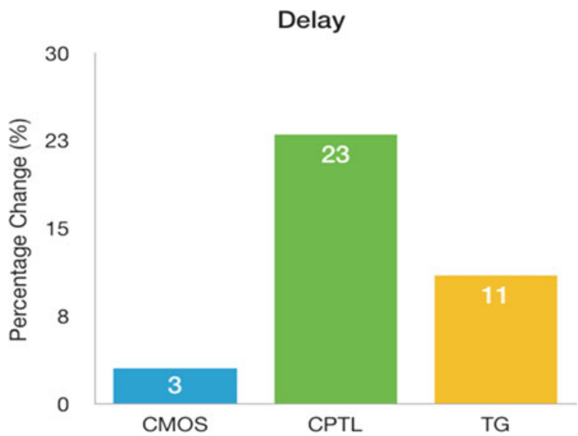


Fig. 8 Average percentage variation of delay

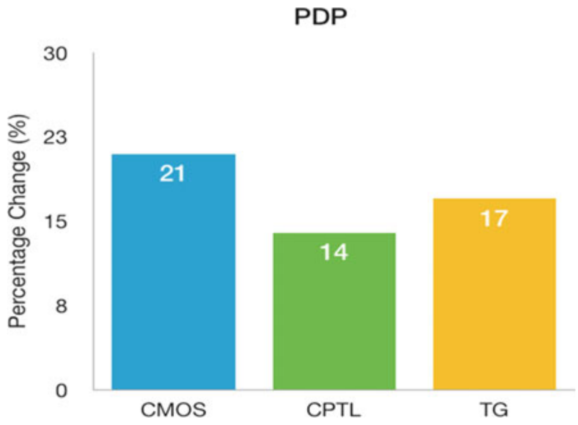


Fig. 9 Average percentage variation of power delay product

Table 1 Advantages and disadvantages of adder structures

Adder structure	Advantages	Disadvantages
CMOS	• Dissipates less power • Delay less variant against increase in temperature	• PDP more variant against increase in temperature
CPTL	• Less delay and power delay product • PDP and average power are less variant against increase in temperature	• Very high power dissipation • Delay more variant against increase in temperature
TG	• Delay is less variant than CPTL	• High delay and PDP • Average power more variant against raise in temperature

5 Conclusion

In this research work, the effects of temperature on circuit metrics of various full adder circuits which are operating on low voltages are studied. 90 nm MOSFET technology has been used for simulation purposes. The advantages and disadvantages of each adder structure have been clearly discussed in detail. From the simulation results, it can be inferred that even though CPTL adder consumes more power, its PDP and power dissipation remain less variant against temperature. Also, the delay of CMOS adder circuit is less variant against temperature. This research can be extended by using FinFET and gate-all-around FET (GAAFET) technology instead of MOSFET and also incorporating more adder design styles.

References

1. Wu, S.H., Tetelbaum, A., Wang, L.C.: How does inverse temperature dependence affect timing sign-off. In: Amara, A., Ea, T., Belleville, M. (eds.) Emerging Technologies and Circuits. Lecture Notes in Electrical Engineering, vol. 66. Springer, Dordrecht (2010). https://doi.org/10.1007/978-90-481-9379-0_13
2. Kumar, R., Kursun, V.: Reversed temperature-dependent propagation delay characteristics in nanometer CMOS circuits. IEEE Trans. Circ. Syst. II Express Briefs **53**(10), 1078–1082 (2006). https://doi.org/10.1109/TCSII.2006.882218
3. Park, C., John, J.P., Klein, K., Teplik, J., Caravella, J., Whitfield, J., Papworth, K., Cheng, S.: Reversal of temperature dependence of integrated circuits operating at very low voltages. In: Proceedings of International Electron Devices Meeting (1995). https://doi.org/10.1109/IEDM.1995.497185
4. Chinnery, D., Keutzer, K.: Overview of the factors affecting the power consumption. In: Closing the Power Gap Between ASIC & Custom. Springer, Boston, MA (2007). https://doi.org/10.1007/978-0-387-68953-1_2
5. Liao, W., He, L., Lepak, K.M.: Temperature and supply voltage aware performance and power modeling at microarchitecture level. IEEE Trans. Comput. Aided Des. Integr. Circ. Syst. **24**(7), 1042–1053 (2005). https://doi.org/10.1109/TCAD.2005.850860
6. Morris Mano, M., Ciletti, M.D.: Digital Design: With an Introduction to the Verilog HDL, 5th edn, pp. 133–150. Pearson Education, New Jersey (2013)
7. Aalelai Vendhan, M.: Analysis on circuit metrics of 1-bit FinFET adders realized using distinct logic structures. Indian J. Sci. Technol. **12**(26) (2019). https://doi.org/10.17485/ijst/2019/v12i26/145499
8. Hasan, M., Hossein, M.J., Saha, U.K., Tarif, M.S.: Overview and comparative performance analysis of various full adder cells in 90 nm technology. In: 2018 4th International Conference on Computing Communication and Automation (ICCCA) (2018). https://doi.org/10.1109/ccaa.2018.8777684
9. Venkatesan, C., Thabsera, S.M., Sumithra, M.G., Suriya, M.: Analysis of 1-bit full adder using different techniques in Cadence 45 nm technology. In: 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS) (2019). https://doi.org/10.1109/icaccs.2019.8728449

Novel Approach for Power Analysis in Microcontrollers

P. Muthu Subramanian and A. Rajeswari

Abstract Security showcases a major breakthrough in the history of the embedded systems, as the connections move beyond computing devices, from intelligent traffic management systems to missile control system. The drift of Internet protocol from version 4 to version 6 has greatly expanded the number of devices that could be connected over the Internet and it is estimated that it could accommodate three times the number of devices currently existing in the world. With this growing pace in embedded systems, security issues are an area of great concern. Objective of this approach examines the vulnerability of the commonly used ARM processor to a simple distributive embedded security—power analysis attack for difference and also making the role of service provider active by modifying the conventional security model.

Keywords Microcontrollers · Security · Embedded systems · Power analysis · Cryptography

1 Introduction

Nowadays security plays a major roll in embedded systems. This security system is expected to continue for decades. The security range required for embedded systems is from smallest RFID to satellites which are orbiting the earth. The below chapter will explain few types of security requirements and their attacks on embedded systems. In embedded systems, embedding security into devices is not a direct process. Requirement of security functionality must be determined before embedding into a device. Security requirements mainly depend upon threat models or attacks which may be fully known at that time. It has been already said that prevention is the most viable course of action when it comes to fraud. Security is an important layer to avoid fraudulent. In the embedded world, detection is an equally important layer to mitigating stolen codes. Consideration is shown on the various attacks developed

P. Muthu Subramanian · A. Rajeswari (✉)
Department of Electronics and Communication Engineering, Coimbatore Institute of Technology, Coimbatore, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_50

to take out the secret information of the cryptographic methods (cores) by analyzing the leaked information of the hardware. In the proposed technique, analysis of the attacks by means of power is said to be one of the efficient and most powerful methods for pilfering the data from the ARM processor and it is said to be a most important threat to the processor security. A simple analysis is made to compare the leaked information with the actual data to prove the ownership. Also in this proposed system, we investigate the PAA on ARM processor by means of processed data and to analyze the graphical trace for very few instructions.

2 Related Work

In the case of erroneous (faulty) attacks, the attacker attempts to takeout the secret information by applying some erroneous running in the device. Ecological parameters are influenced to generate the erroneous behavior. Non-invasive fault attacks will be based on the various parameters which influence the temperature, the voltage source, generation of clock signals and by injecting electromagnetic pulses [1–3]. Programming modifications are not required for these attacks since all these attacks are induced by means of various hardware signals. Semi-invasive fault attacks will be based on inducing faults by means of laser light. Here in this technique IC packaging to be altered for implementing the attacks. During the past years, cryptographic primitives have been reported for a large number of invasive fault attacks. Normally, attacks are common to both software (RSA and block cipher algorithms) and hardware. Fault can be identified based on either the control logic or the data path of the controller by making changes in the algorithm.

3 Methods of Fault Attacks

Following section is regarding the mechanisms of the fault injection and the types of attacks.

3.1 Fault Attack—Non-invasive

To induce the fault, modification is not required in the equipment or in the attacked device. Victim cannot recognize the chance of identifying the attack. Applying methods like sending spikes, EM and clock pulses are the distinctive methods [1]. Here, the attacked device is influenced by making some changes in the timing. Temperature and the pulses which are working externally may direct to the faults in the memory section.

3.2 Fault Attack—Semi-invasive

To induce the fault, modification is required in the chip. Here, IC is de-packaged to modify the changes in the passive layer which cannot be damaged. Laser and flashlight will be the source for inducing the attacks. Memory bits and the registers will be targeted by laser light where the flashlight covers the overall structure of the chip.

3.3 Invasive Fault Attacks

To induce the fault, bus line of the IC is modified by getting the access to the metal layer; here, the access to the metal layer is achieved by removing the passive coating of the de-packaged IC and metal layer is observed using miniprobes [4]. Quite a few analyses have been undergone earlier to identify the incorrect outputs produced in the cryptographic operations.

4 Power Analysis

Analysis on differential and correlation power is considered to be the major concern and threat to the devices operating under cryptographic conditions. Generally, devices will seep out its information on the processed records by means of power dissipated through the ground or supply channel. Gate circuit will absorb the current when there is a change in the state of the logical conditions. When the logic is 0 now the state is defined as 0 V and when the logic is 1 it will be the maximum voltage point and normally current dissipation will happen when there is a transition from maximum to minimum and this will help in recording the analysis of the power with more accuracy. [5] Thus, differential method is one of the advanced techniques compared to correlation analysis.

4.1 Fault Injection Attack

Procedure to induce an error maliciously in a computing device in order to alter the software execution behavior is defined as the fault injection attack. After fault injection, two effects are possible, i.e., execution of an instruction can be avoided and the second, corrupt the data at which the processor is working with. These effects are used to compromise the embedded device security by bypassing the security checks or leaking the private keys.

Among the several types of glitching such as optical, heat or radiation glitching, the most common glitching is the clock and voltage glitching.

4.2 Clock Glitchıng Attack

For a very short moment, a sudden increase of the system clock frequency is referred as the clock glitch. Due to the difference in the distribution paths like length, capacitance in traces and transistor gates, etc., the clock signal is not evenly distributed and does not reach every point at same time in a digital integrated circuit such as microcontroller or processor, FPGA. Based on the clock frequency specified by the manufacturer, the clock signal will be able to reach every register in average processing time. If it is beyond the specified limit, it would make the IC not to operate effectively. Therefore, if we force the IC to work beyond the set limit for a particular time period then the instruction will not be processed efficiently and it will retain back to the normal frequency. While clock glitching the entire instruction execution is avoided in the processor.

5 Proposed System

Supply voltage is connected with a resistor in the ARM processor and it is used for observing the power variation (current) of the circuit. Supply voltage should be greater than the normal voltage supplied to the ARM processor since resistance is connected across the VDD of the processor. Thus, the supply voltage should be greater than 3.3 V. Frequency with higher clock rate will have some changes in the analysis of current, and electrical signals are measured in terms of microvolt using digital oscilloscope. Bits of the controllers are varied with different logics (i.e., 0 and 1) in the ARM processor. Changes will make variations in the power pattern. Spartan 3 is enough to attack most of the microcontroller, as the Spartan 3 would produce the clock frequency of 150 MHz by modifying the signal path and it is possible to add some delay to alter the clock phase as needed. In Spartan 3, digital clock manager allows to produce different clock frequencies. To generate the clock cycle by the host microcontroller, it needs two instructions: One is to set 1, and the other is to set 0. Host microcontroller operating frequency/2 is applied to the target system to get the maximum clock frequency. To inject a clock fault, we need to generate the clock frequency much faster than the maximum operating frequency of the target system. Sometimes it is required that there is a need to generate at least 40 MHz clock signal, and for that high-speed processors like ARM series are used to achieve the required traces (Fig. 1).

When an external clock is applied to the processor, the clock rate frequency (Fmax) is defined.

No abnormal behavior can be observed if,

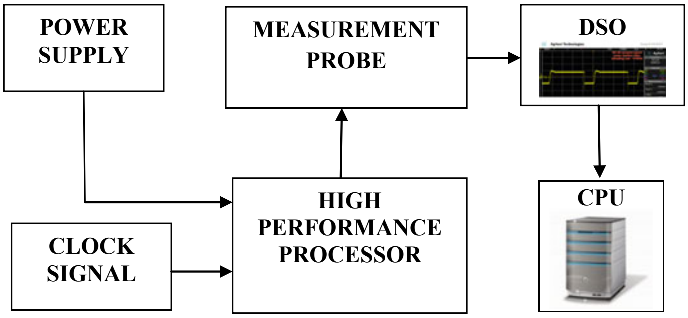


Fig. 1 Setup for power analysis

$$T\text{glitch} \geq T\text{min} \tag{1}$$

Possibility of abnormal behavior increases when,

$$T\text{glitch} < T\text{min} \tag{2}$$

While running the program, timing violation will be the major reason for the abnormal behavior.

6 Experimental Results

6.1 Power Analysis for Addition Operation

Impact of power consumption by the controller is observed from the arithmetic operation. Power variations are observed from the ARM processor while doing the arithmetic operations. One of the ports from the microcontroller is extracted with the addition output. Repeated pattern was observed from the power traces for the various instructions as shown in Figs. 2, 3 and 4.

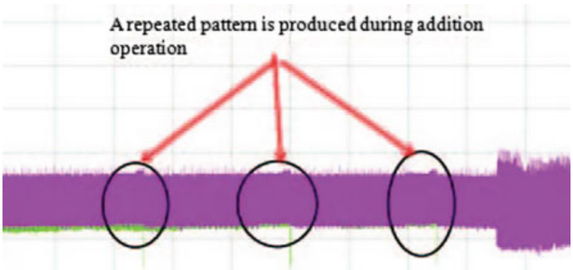


Fig. 2 Addition operation power analysis

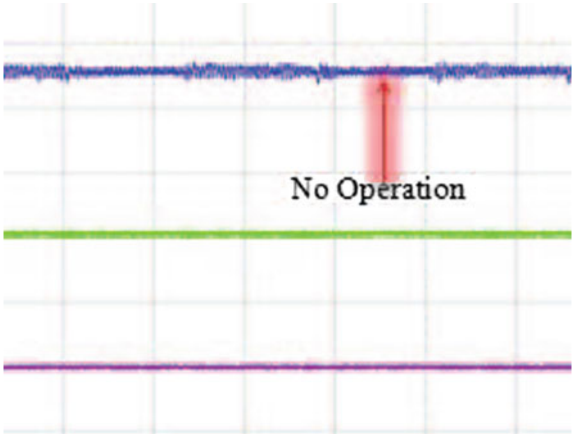


Fig. 3 NOOP power analysis

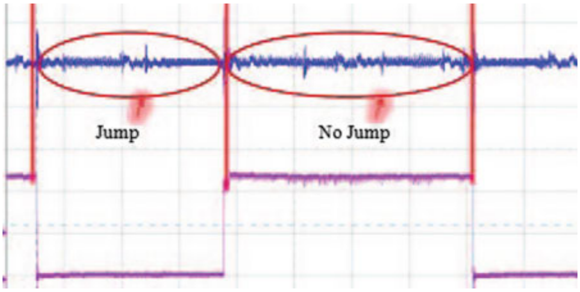


Fig. 4 JUMP and NOJUMP power analysis

It is observed from the power investigation that there is a unique correlation between the processor power utilization and the power pattern. This study is further extended into the next step as the power analysis.

7 Conclusion

Embedded security has been achieved through the power analysis. Based on the clock signal and using simple alteration to the target system by applying glitches, it is proved and tested that processors are vulnerable to power analysis attack and various instructions were tested for different data sets for the power trace correlation.

References

1. Balasch, J., Gierlichs, B., Verbauwhede, I.: An in-depth and black-box characterization of the effects of clock glitches on 8-bit MCUs. In: FDTC, 2011, pp. 105–114. IEEE (2011)
2. Hutter, M., Schmidt, J.M.: The temperature side-channel and heating fault attacks. In: CARDIS. Springer (2013)
3. Moro, N., Dehbaoui, A., Heydemann, K., Robisson, B., Encrenaz, E.: Electromagnetic fault injection: towards a fault model on a 32-bit microcontroller. In: FDTC, pp. 77–88. IEEE (2013)
4. Skorobogatov, S.P.: Semi-invasive attacks—a new approach to hardware security analysis. PhD thesis, University of Cambridge, Computer Laboratory (2005)
5. Zhai, X., Appiah, K., Ehsan, S., Howells, G., Hu, H., Gu, D., McDonald-Maier, K.D.: A method for detecting abnormal program behavior on embedded devices. IEEE Trans. Inf. Forensics Secur. **10**(8), 327–345 (2015)

Evolving Reversible Fault-Tolerant Adder Architectures and Their Power Estimation

S. Bharani Surya, C. Gokul Prasad, S. Raghul and N. Mohankumar

Abstract Fault tolerance property is incorporated in a circuit to increase its reliability. Error at the output side can be prevented by Fault-tolerant circuits. Its design enables the circuit to continue operation, at an error-free state, rather than failing completely, when some part of the circuit fails. Here, reversible adder fault-tolerant architectures of ripple carry, carry look-ahead adder (CLA) and carry skip adder (CSA) are implemented and their corresponding powers are estimated. Power efficiency and reliability of the implemented adders make them effective.

Keywords Reversible computation · Fault-tolerant circuit · Low power

1 Introduction

Reversible logic systems have the efficient computation, i.e., loss of information bits does not occur in between successive computations. Landauer [1] has proved that for a bit of information lost, $kT \ln 2$ J of heat energy is generated; here, 'k' is Boltzmann's constant and 'T' is the operating temperature. In reversible logic circuits, there is a possibility of returning to its primary state from its last state when there is some mishappenings like bit loss, i.e., input vector can be recovered from output vector.

Fault tolerance will enable a system to continue its operation correctly though there is an event where some of its components fail. These circuits have an ability to prevent error. A system is fault-tolerant if its individual components are fault-tolerant. Fault tolerance is usually system or block level. Mitra and Chowdhury [2] have shown that for a gate to be fault-tolerant if the input vector and the output vector are equal.

S. Bharani Surya · S. Raghul · N. Mohankumar (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: n_mohankumar@cb.amrita.edu

C. Gokul Prasad
PPG Institute of Technology, Coimbatore, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_51

2 Literature Survey

Any reversible fault-tolerant logic should minimize these garbage outputs and constant inputs. Bennett [3] said that when we use reversible logic, the energy loss can be decreased to a greater extent or equals zero and also proved that z power dissipation of reversible gates is null value. Clock gating technique is used to reduce the power consumption of a DTMF chip designed using conventional gates [4]. Reversible adders can also be constructed using D latch [5]. Reversible fault-tolerant subtractors are designed [6]. Power-efficient carry skip adder using Fredkin and TSK was designed [7]. Quantum-dot cellular technologies are implemented in designing adders [8]. Peres gates [9] and fault-tolerant gates [10] were also used to design adders. Binary-coded decimal adders [11], ripple carry look-ahead adder [12], and prefix hybrid adders [13] are built using reversible logic.

2.1 Reversible Gates

Reversible computation is preferred over irreversible computation to avoid power dissipation because of information bit loss. The unused outputs are called garbage outputs. Some constant value should be assigned to the extra inputs to get the desired function, so they are called constant inputs. The garbage outputs and constant outputs are produced to maintain the reversibility property. The most commonly used reversible gates are shown in Fig. 1.

The parity preserving gates such as Fredkin gate and F2G gates are shown in Fig. 2. IG gate (IG) [14] and NFT gate (NFT) [15], depicted in Fig. 3, are also being used for designing reversible circuits.

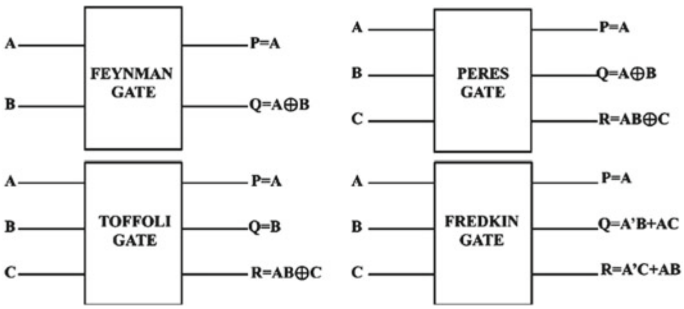


Fig. 1 Reversible gates

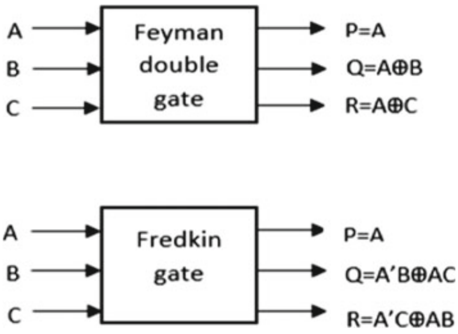


Fig. 2 Fault-tolerant gates

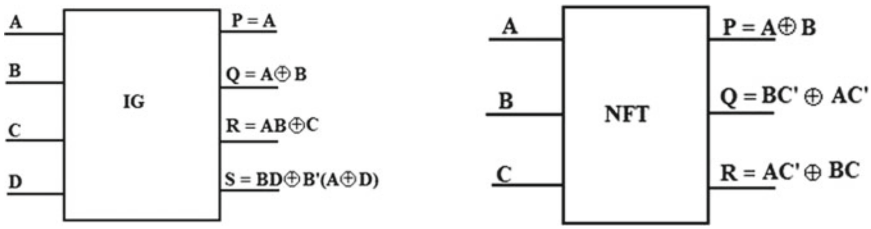


Fig. 3 IG and NFT gates

2.2 Proposed Adder Architectures

The full adder implemented using IG gates is most preferred because it is both fault-tolerant and reversible. It is found that it consumes less power when compared with the other adders. It is less complex because of which the delay will also be less. So full adder that using IG gates will be used for addition in the ripple carry adder, CLA, and CSA architectures (Fig. 4).

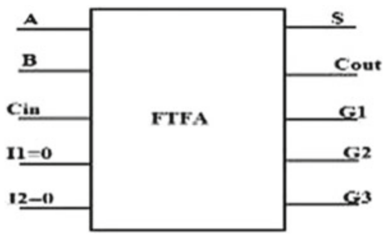


Fig. 4 Fault-tolerant reversible full adder using IG gates (FTFA)

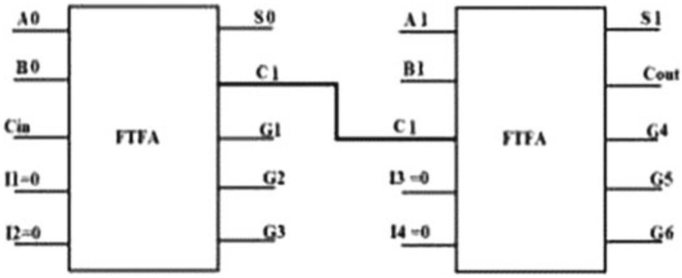


Fig. 5 Reversible ripple carry adder using FTFA

2.3 Ripple Carry Adder

The ripple carry adder is implemented and designed by using the full adder using IG gates (FTFA) [15]. The most straightforward realization of a final stage adder circuit is using this circuit. Here, we have implemented a 2-bit ripple carry adder (Fig. 5).

2.4 Carry Skip Adder

By jumping over blocks of successive adder stages, delay can be reduced by carry skip adder in carrying propagation [15]. The carry skip adder and carry look-ahead adder are comparable in speed but when it comes to power consumption and area carry requires less power and area. In Fig. 6, 4-bit carry skip adder is implemented using reversible gates such as F2G, IG, and NF.

2.5 Carry Look-Ahead Adder

The carry look-ahead adder circuit is implemented using reversible gates such as F2G, IG, and NFT gates [15]. In Fig. 6, only two-bit adder circuit is shown. In carry look-ahead adder, the carries are generated in parallel form so as to avoid the waiting time for carry propagation. Delay is avoided as carry is generated in parallel.

2.6 Power Estimation and Hardware Complexity

In Table 1, power values of three different reversible full adders are shown. The reversible full adder in Fig. 4 and IG full adder are fault-tolerant since they satisfy the parity preserving property. The full adder using Peres gate is not fault-tolerant

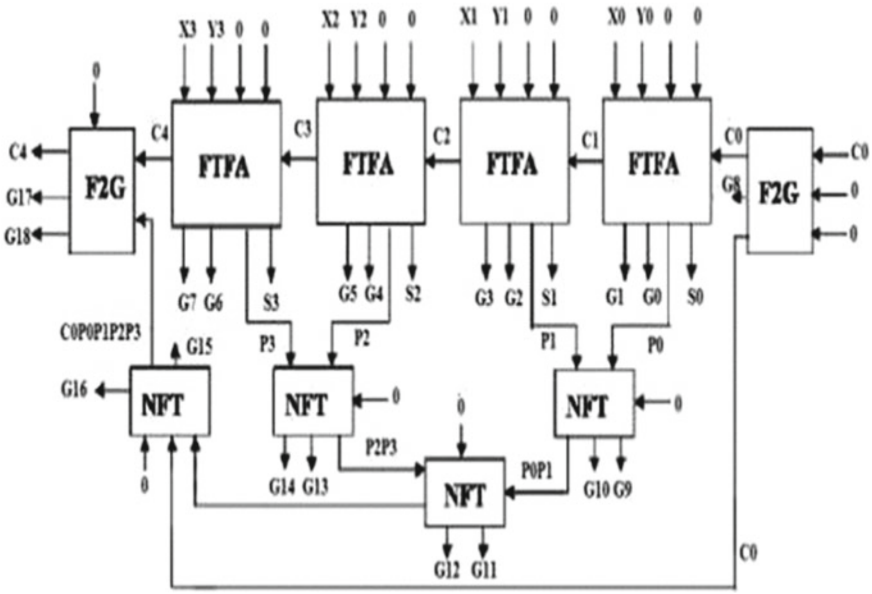


Fig. 6 Fault-tolerant reversible carry skip adder using FTFA, F2G, and NFT

Table 1 Power results of full adder

Adder circuit	Cell internal power (μW)	Shortcircuit power (μW)	Total dynamic power (μW)	Leakage power (μW)	Total power (μW)
Reversible full adder (F2G and FRG)	3.90E−06	6.04E−07	4.50E−06	1.36E−06	5.86E−06
Full adder using Peres gate	1.64E−06	2.24E−07	1.86E−06	3.92E−07	2.26E−06
IG full adder	2.38E−06	4.24E−07	2.80E−06	6.21E−07	3.43E−06

design. Maximum power can be achieved in reversible full adder using F2G and Fredkin gates. The full adder using IG gates will be most efficient because it is both reversible and fault-tolerant. Though full adder designed using Peres gate has the minimum power, it is not fault-tolerant, and hence, it is ineffective in cases where fault-tolerant feature is necessary.

In Table 2, the gate count, number of constant inputs, and number of garbage outputs of the reversible full adders were compared. Reversible full adder using F2G and FRG is the maximum in all the three. Gate count of the other two full adders is the same and number of garbage outputs and constant inputs are also low. Reversible

Table 2 Reversible full adder utilization summary

Reversible circuit	Gate count	Constant input	Garbage outputs
Ripple carry adder	2 FTFA (4 IG)	4	6
Carry look-ahead adder	2 FTFA + 10 F2G + 5 NFT	26	27
Carry skip adder	4 FTFA + 4 NFT + 2 F2G	15	18

Table 3 Power profile of reversible ripple carry, carry look-ahead, and carry skip adder

Circuit	Gate count	Constant inputs	Garbage outputs
Reversible full adder using F2G and FRG	4 FRG + 5 F2G	9	
Full adder using Peres gate	2 PG	1	11
IG full adder	2 IG	2	3

Table 4 Utilization of reversible ripple carry, carry look-ahead, and carry skip adder

Reversible circuit	Gate count	Constant input	Garbage outputs
Ripple carry adder	2 FTFA (4 IG)	4	6
Carry look-ahead adder	2 FTFA (4 IG) + 10 F2G + 5 NFT	26	27
Carry skip adder	4 FTFA (8 IG) + 4 NFT + 2 F2G	15	18

full adder using PG has the least number of constant input and garbage output but the circuit is not fault-tolerant. Among these three designs, reversible full adder using IG is the preferred full adder since it is fault-tolerant and less complex.

In Table 3, power values of ripple carry adder, CLA, and CSA are shown. The ripple carry adder has minimum power, and carry skip adder has maximum power. The power of carry look-ahead adder is 20 μW higher than that of ripple carry adder. Carry look-ahead adder is preferred because it consumes less power but also has lesser carry propagation delay than ripple carry adder.

In Table 4, the gate count, number of constant inputs, and number of garbage outputs of ripple carry adder, CLA, and CSA were compared. It can be seen that CLA has the maximum garbage output and maximum constant input. But delay will be very less when compared to ripple carry adder.

3 Conclusion

To reduce the power consumption due to bit loss, logic gates should be designed in such a way that it performs reversible computation so that information bits are not destroyed or lost. There is a small increase in power, due to more complexity, in reversible circuits. When the power constraints are taken, the carry look-ahead adder is 0.15 mW less than the carry skip adder. The area values were found to be directly proportional to the power values of the gates. The incorporation of reversibility and fault tolerance makes the reversible circuits more advantageous, and this compensates and compromises the increase in power.

References

1. Landauer, R.: Irreversibility and heat generation in the computing process. IBM J. Res. Dev. **5**(3), 183–191 (1961)
2. Mitra, S.K., Chowdhury, A.R.: Minimum cost fault tolerant adder circuits in reversible logic synthesis. In: 2012 25th International Conference on VLSI Design, Hyderabad, pp. 334–339 (2012)
3. Bennett, C.H.: Logical reversibility of computation. IBM J. Res. Dev. **17**(6), 525–532 (1973)
4. Shankar, A., Manchala, S., Mohan Kumar, N.: Low power implementation of DTMF chip by using power gating technique with merge flops. J. Eng. Appl. Sci. **13**, 6335–6340 (2018)
5. Nayak, V.S.P., Ramchander, N., Reddy, R.S., Redy, T.H.S.P., Reddy, M.S.: Analysis and design of low-power reversible carry select adder using D-latch. In: 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, pp. 1917–1920 (2016)
6. Kaur, P., Dhaliwal, B.S.: Design of fault tolerant full adder/subtractor using reversible gates. In: 2012 International Conference on Computer Communication and Informatics, Coimbatore, pp. 1–5 (2012)
7. Chiwande, S.S., Dakhole, P.K.: VLSI design of power efficient carry skip adder using TSG & Fredkin reversible gate. In: 2012 International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, pp. 370–373 (2012)
8. Kunalan, D., Cheong, C.L., Chau, C.F., Ghazali, A.B.: Design of a 4-bit adder using reversible logic in quantum-dot cellular automata (QCA). In: 2014 IEEE International Conference on Semiconductor Electronics (ICSE2014), Kuala Lumpur, pp. 60–63 (2014)
9. Peres, A.: Reversible logic and quantum computers. Phys. Rev. A **32**(6), 3266–3276 (1985)
10. Paramasivan, D., Ravikumar, V., Ramesh, D., Ramya, H., Raghavan, R., Mohankumar, N.: Low power fault-tolerant reversible full adders. In: Proceedings of International Conference on Computing and Communication, Bangalore, June 2014
11. Thakral, S., Bansal, D.: Comparative study and implementation of BCD adders for reversible logic based ALU. In: 2017 2nd International Conference on Telecommunication and Networks (TEL-NET), Noida, pp. 1–5 (2017)
12. Somani, N., Chaudhary, C., Yadav, S.: Reversible adder design for ripple carry and carry look ahead (4, 8, 16, 32-bit). In: 2016 International Conference on Computing, Communication and Automation (ICCCA), Noida, pp. 1387–1392 (2016)
13. Vudadha, C., Phaneendra, P.S., Ahmed, S.E., Sreehari, V., Muthukrishnan, N.M., Srinivas, M.B.: Design and analysis of reversible ripple, prefix and prefix-ripple hybrid adders. In: 2012 IEEE Computer Society Annual Symposium on VLSI, Amherst, MA, pp. 225–230 (2012)

14. Manoj Kumar, S.B., et al.: Design and synthesis of reversible fault tolerant carry skip adder/subtractor. Int. J. Emerg. Sci. Eng. **1**, 55–58 (2013)
15. Poornima, M., et al.: Fault-tolerant reversible logic for combinational circuits: a survey. In: Proceedings of International Conference on VLSI, Communication, Advanced Devices, Signals and Systems and Networking (2013)

A Wide-Band, Low-Power Grounded Active Inductor with High Q Factor for RF Applications

L. Bharath, D. Anila, C. N. Ajay, B. Shravani and Amit Jain

Abstract In this research work, a low-power grounded active inductor based on gyrator-C topology is proposed. The simulation results of the proposed active inductor show maximum quality factor of 341 at 2.51 GHz. The inductive bandwidth of the circuit is obtained as 0.79–2.69 GHz. The designed active inductor provides the inductance value of as high as 180 nH which makes it suitable for a wide range of applications. The circuit shows good performance in all aspects while consuming only 0.99 mW of DC power. The circuit also shows very good noise performance compared to reported works in the literature.

Keywords Gyrator · Active inductor (AI) · Cascode · Inductance · Quality factor · Transconductance

1 Introduction

The demand for RF integrated circuits has been increased significantly due to its application in wireless communication. A lot of research is going on designing efficient RF blocks such as voltage-controlled oscillators (VCOs), low-noise amplifier (LNA), filters, variable phase shifter [1–3]. The quality of the inductors used in these designs has a great impact on the performance of these blocks [4]. Usually, a spiral inductor is used for the RF designs because of its very good phase noise performance [5–8]. But the spiral inductor has several disadvantages like large chip area requirement, difficulty in achieving higher inductance value and inability to obtain higher quality factor with minimum area [9–11]. To deal with these issues, active inductor can be used in designing RF circuits. Active inductors can provide large tunable inductance, high Q factor with minimum area requirement, and higher self-resonance frequency range [12–16]. Therefore, extensive research work is going on

L. Bharath · D. Anila · C. N. Ajay · B. Shravani · A. Jain (✉)
Department of Electronics and Communication Engineering, CMR Institute of Technology, Bengaluru, India
e-mail: amit.j@cmrit.ac.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_52

designing CMOS-based active inductor (AI) for RFICs. Some of the drawbacks of the AI are power consumption, high noise, and poor linearity.

Gyrator-C configuration is the most popular approach for implementing active inductors [10, 17]. The gyrator-C architecture provides much higher value of inductance and Q factor compared to the spiral inductors. This paper represents a grounded AI based on gyrator-C approach, where one of the transconductors is realized by MOS transistor with cascode grounded configuration and the other is realized with common drain configuration.

2 Background

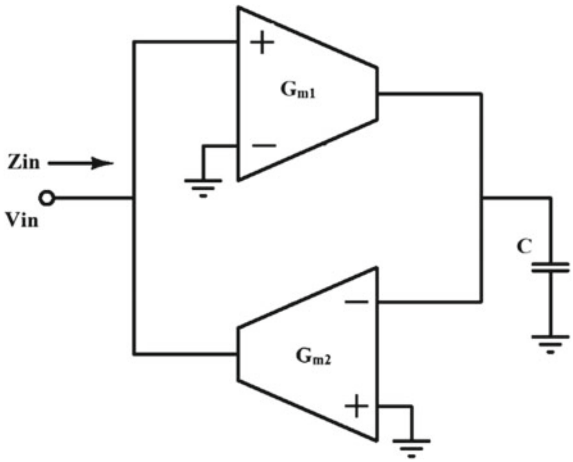
Most of the active inductors proposed in the literature are based on gyrator-C topology [18]. It consists of two transconductors (G_{m1} and G_{m2}), that are connected in feedback loop, where the input of one is derived from the output of the other as shown in Fig. 1.

The input impedance of the gyrator circuit is inversely proportional to its load impedance [19]. One port of the gyrator circuit is connected to a capacitor, giving name as the gyrator-C topology. Considering the circuit is lossless the inductive input impedance of the grounded, lossless AI is calculated as [13]

$$Z_{in} = \frac{Sc}{G_{m1} G_{m2}} \tag{1}$$

So the realized inductance of the AI circuit depends on the load capacitance as well as the product of the transconductances. Among two transconductors involved in the design, one should show a positive value, whereas the other should provide negative transconductance. The transconductors are usually implemented using MOSFETs.

Fig. 1 Gyrator-C topology



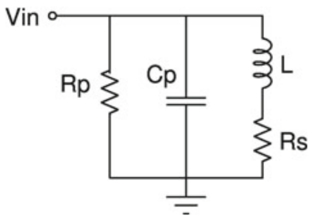


Fig. 2 Equivalent circuit of the gyrator

As these MOSFETs are non-ideal and have parasitic components, an equivalent circuit of the gyrator is shown in Fig. 2, which considers all non-ideal effects [20].

Values of the parasitic components depend on the implementation of the AI. The inductive bandwidth of the active inductors is expressed as [20].

$$\frac{R_s}{2\pi L} \leq f_{\text{inductive}} \leq \frac{1}{2\pi\sqrt{LC_p}} \tag{2}$$

3 Proposed AI Circuit

The proposed AI circuit is shown in Fig. 3. The circuit is based on the gyrator-C topology [18]. The common source and common drain configurations are used to

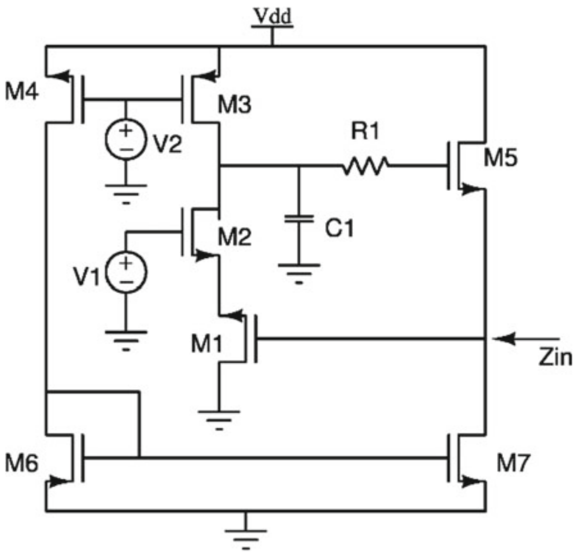


Fig. 3 Schematic of the proposed AI circuit

implement the positive and negative transconductances as required to realize inductive component. As noticed from the figure, transistors M1 and M5 are used to obtain common source and common drain configurations, respectively.

To enhance the inductance value and the Q factor, cascode structure [21–24] is used in the design which is being implemented by adding transistor M2 in the feedback loop. The feedback resistor R1 [25, 26] creates additional inductive reactance which increases the inductance of the designed AI circuit. Transistors M3 and M7 work as current sources and transistors M4 and M6 form the current mirror circuit. Two voltage sources are used to bias the respective transistors in saturation region. The biasing of transistor M2 is very important as it helps to reduce the output conductance which in turn improves the inductive behavior of the circuit. The circuit is designed for low bias current to reduce the power consumption. Also, wider transistors have not been used in the design as it reduces the self-resonant frequency [13].

4 Simulation Results and Discussions

To estimate the overall performance of the designed AI, it is simulated using 180 nm CMOS technology under cadence environment. Periodic steady-state analysis is performed to obtain the values of the two-port parameters of the designed circuit. The inductance and the quality factor of the designed circuit have been measured using Z parameters. The real and imaginary values of the impedance are shown in Fig. 4.

The imaginary value of the impedance corresponds to the inductance, and the real value is related to the series resistance of the proposed AI. The phase response of the proposed circuit is shown in Fig. 5. The designed circuit acts as an inductor between 0.79 and 2.69 GHz as observed from the phase response. The inductance plot of the

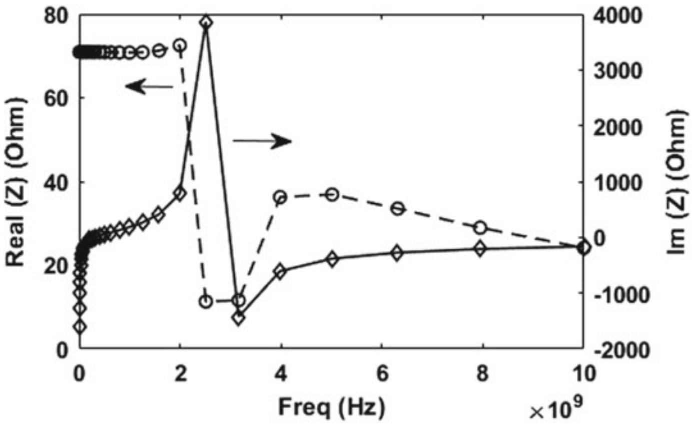


Fig. 4 Real and imaginary values of the impedance. The solid line with diamond represents the real values, and the dotted line with circle shows the imaginary values

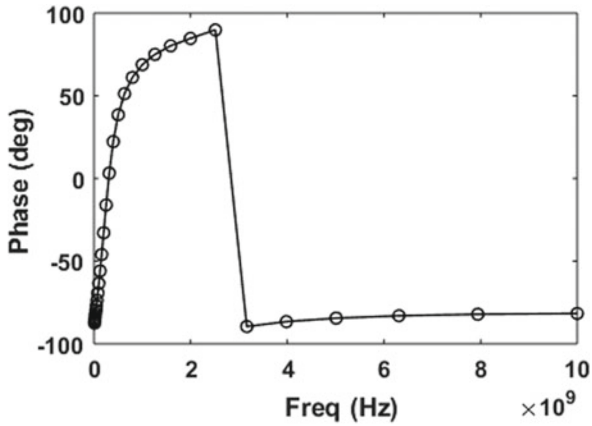


Fig. 5 Phase response of the proposed AI circuit

circuit is shown in Fig. 6. As noticed from the figure, the maximum and minimum values of the inductance considering inductive bandwidth are obtained as 25.7 and 180 nH, respectively. The maximum value of the quality factor is achieved as 341 at 2.51 GHz as observed from Fig. 7. The proposed circuit is very much immune to noise as evident from Fig. 8.

The circuit shows only $0.89\,\text{nV}/\sqrt{\text{Hz}}$ noise voltage at 2.4 GHz and remains less than $2\,\text{nV}/\sqrt{\text{Hz}}$ for the complete frequency range. The circuit consumes only 0.99 mW of DC power from 1.8 V power supply, as shown in Fig. 9. To validate the performance of the proposed AI circuit, a performance comparison has been shown with the existing work in Table 1. It can be observed from the comparison that the proposed circuit exhibits good result for most of the parameters when compared to other reported works.

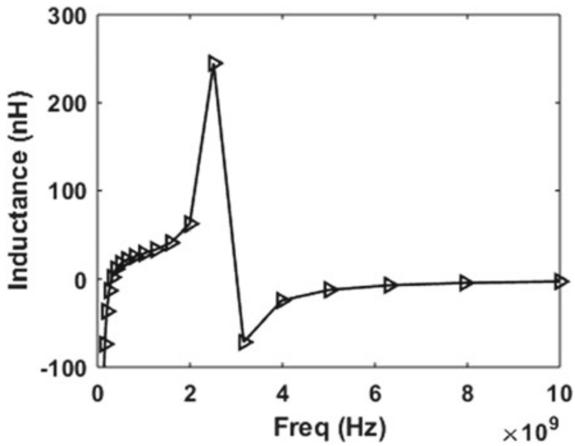


Fig. 6 Inductance versus frequency of the proposed circuit

Fig. 7 Q factor plot of the designed AI

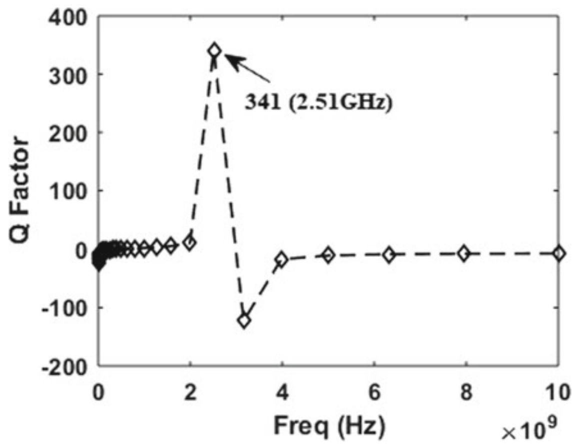


Fig. 8 Input referred noise for the designed circuit

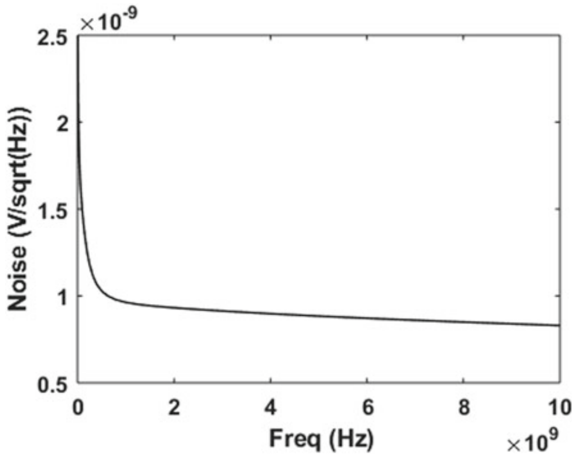


Fig. 9 Power dissipation of the proposed circuit for supply voltage of 1.8 V

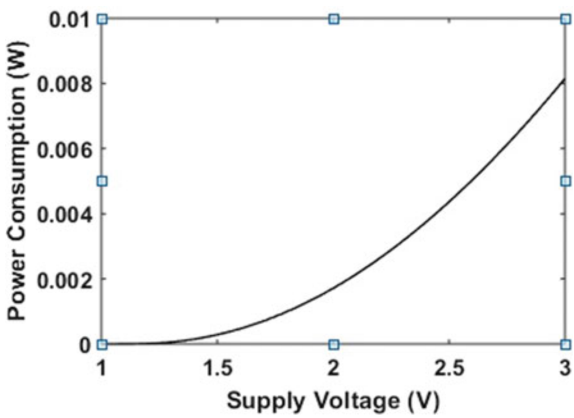


Table 1 Performance comparison with the existing work in the literature

References	V_{dd} (V)	Power (mW)	Inductance (nH)	Q_{max}	Frequency (GHz)
[3]	1.2	22.5	6.83–11.7	240	2.4–5.35
[11]	–	8.6	33	68	4
[14]	1.2	6.4	1.9	38.8	3.5
[21]	–	–	0–90	60	0.5–2.5
[26]	–	40	230	700	0.43
This work	1.8	0.99	25.7–180	341	0.79–2.69

5 Conclusion

In this work, a grounded AI circuit is designed for very low-power RF applications. The circuit shows very good inductive bandwidth while consuming only 0.99 mW of DC power. The active inductor provides very good range for inductance with the lowest value obtained as 25.7 nH, and the highest value is achieved as 180 nH. The maximum value of the Q factor is noticed to be 341 at 2.51 GHz. The proposed circuit also shows very low noise voltage of 0.89 $\mathrm{nV}/\sqrt{\mathrm{Hz}}$ at 2.4 GHz. So the proposed active inductor circuit shows good performance in all the aspects while consuming very less power from power supply.

Acknowledgements The authors thankfully acknowledge the Center of Excellence Integrated Circuit, Department of Electronics and Communication Engineering, CMR Institute of Technology, Bangalore, for the support.

References

1. Patel, D.P., Rahurkar, S.O.: Tunable CMOS active inductor using Widlar current source. J. Circ. Syst. Comput. **28** (2019). https://doi.org/10.1142/s0218126619500270
2. Jeong, Y.-J., Kim, Y.-M., Chang, H.-J., Yun, T.-Y.: Low-power CMOS VCO with a low-current, high-Q active inductor. IET Microw. Antennas Propag. **6**, 788–792 (2012). https://doi.org/10.1049/iet-map.2011.0332
3. Abdalla, M.A.Y., Phang, K., Eleftheriades, G.V.: Printed and integrated CMOS positive/negative refractive-index phase shifters using tunable active inductors. IEEE Trans. Microw. Theory Tech. **55**, 1611–1623 (2007)
4. Sachan, D., Goswami, M., Misra, P.K.: A high-Q floating active inductor using 130 nm BiCMOS technology and its application in IF band pass filter. Analog Integr. Circ. Sig. Process. **96**, 385–393 (2018). https://doi.org/10.1007/s10470-018-1196-3
5. Iniewski, K.: Advanced Circuits for Emerging Technologies. Wiley, New Jersey (2012)
6. Belini, V.L., Romero, M.A.: Design of active inductors using CMOS technology. In: Proceedings of the 15th Symposium on Integrated Circuits and Systems Design, pp. 5–10 (2002). https://doi.org/10.1109/sbcci.2002.1137674
7. Amin, M.T.: On the selection of passive elements for low phase noise LC tank VCO in 65 nm process. In: Proceedings of 3rd International Conference on Electrical Engineering and

Information Communication Technology (ICEEICT-2016), Dhaka, Bangladesh (2016). https://doi.org/10.1109/ceeict.2016.7873163
8. Nguyen, N., Meyer, R.: Si IC-compatible inductors and LC passive filters. IEEE J. Solid-State Circ. **25**, 1028–1031 (1990). https://doi.org/10.1109/4.58301
9. Faruqe, O., Saikat, M.M.M., Bulbul, M.A.K., Amin, M.T.: Comparative analysis and simulation of active inductors for RF applications in 90 nm CMOS. In: EICT (2017). https://doi.org/10.1109/EICT.2017.8275233
10. Uyanik, H.U., Tarim, N.: Compact low voltage high-Q CMOS active inductor suitable for RF applications. Analog Integr. Circ. Sig. Process. **51**, 191–194 (2007). https://doi.org/10.1007/s10470-007-9065-5
11. Lai, Q.T., Mao, J.F.: A new floating active inductor using resistive feedback technique. In: IEEE MTT-S International Microwave Symposium Digest (MTT), pp. 1748–1751 (2010)
12. Momen, H.G., Yazgi, M., Kopru, R., Saatlo, A.N.: Design of a new low loss fully CMOS tunable floating active inductor. Analog Integr. Circ. Sig. Process. **89**, 727–737 (2016). https://doi.org/10.1007/s10470-016-0784-3
13. Moezzi, M., Bakhtiar, M.S.: Wideband LNA using active inductor with multiple feed-forward noise reduction paths. IEEE Trans. Microw. Theory Tech. **60**, 1069–1078 (2012). https://doi.org/10.1109/TMTT.2012.2185947
14. Ahmed, A., Wight, J.: 6.7 GHz high-Q active inductor design using parasitic cancellation with process variation control. Electron. Lett. **46**, 486–487 (2010)
15. Reja, M., Filanovsky, I., Moez, K.: A CMOS 2.0–11.2 GHz UWB LNA using active inductor circuit. In: IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2266–2269 (2008). https://doi.org/10.1109/iscas.2008.4541905
16. Nair, M., Zheng, Y., Lian, Y.: 1 V, 0.18 μm-area and power efficient UWB LNA utilising active inductors. Electron. Lett. **44**, 1127–1129 (2008). https://doi.org/10.1049/el:20081980
17. Geiger, R.L., Sânchez-Sinencio, E.: Active filter design using operational transconductance amplifiers: a tutorial. IEEE Circuits Devices Mag. **1**, 20–32 (1985). https://doi.org/10.1109/MCD.1985.6311946
18. Tellegen, B.D.H.: The gyrator: a new electric network element. Philips Res. Rep. **3**, 81–101 (1949)
19. Thanachayanont, A., Payne, A.: VHF CMOS integrated active inductor. Electron. Lett. **32**, 999–1000 (1996). https://doi.org/10.1049/el:19960669
20. Sato, T., Ito, T.: Design of low distortion active inductor and its applications. Analog Integr. Circ. Sig. Process. **75**, 245–255 (2013). https://doi.org/10.1007/s10470-012-9955-z
21. Hammadi, A.B., Mhiri, M., Haddad, F., Saad, S., Besbes, K.: Study of wide adjustable active inductor circuits: design approaches and reconfiguration methods. In: ICEMIS (2017). https://doi.org/10.1109/ICEMIS.2017.8273063
22. Pascht, A., Fischer, J., Berroth, M.: A CMOS low noise amplifier at 2.4 GHz with active inductor load. In: 2001 Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems. Digest of Papers, pp. 1–5 (2001). https://doi.org/10.1109/smic.2001.942331
23. Reja, M.M., Moez, K., Filanovsky, I.: An area-efficient multistage 3.0- to 8.5-GHz CMOS UWB LNA using tunable active inductors. IEEE Trans. Circuits Syst. II Express Briefs **57**, 587–591 (2010)
24. Kia, H.B., A'ain, A.K., Grout, I., Kamisian, I.: A reconfigurable low-noise amplifier using a tunable active inductor for multistandard receivers. Circuits Syst. Signal Process. **32**, 979–992 (2013)
25. Hsiao, C.-C., Kuo, C.-W., Ho, C.-C., Chan, Y.-J.: Improved quality factor of 0.18 μm CMOS active inductor by a feedback resistance design. IEEE Microw. Wirel. Compon. Lett. **12**, 467–469 (2002)
26. Branchi, P., Pantoli, L., Stornelli, V., Leuzzi, G.: RF and microwave high-Q floating active inductor design and implementation. Int. J. Circuit Theory Appl. **43**, 1095–1104 (2015)

Design and FPGA Realization of Digital Lightweight Numerically Controlled Quadrature Wave Oscillator

Y. Swathi and N. Mohankumar

Abstract Modern-day DSP systems require the generation of sinusoidal or other periodic waveforms. One method of generating these signals involves "numerically controlled oscillators" (NCOs), in which a digital accumulator is used to generate sinusoidal signals using a sine/cosine lookup table for generation. This paper proposes a design based on NCO to generate a digital square wave with the provision of change of frequency and phase to any desired value with the help of a 14-bit word. The design was synthesized using the Xilinx ZedBoard FPGA development platform. The results of hardware implementation matched with the calculated (theoretical) and simulation results. This paper also presents the FPGA implementation of the proposed NCO-based quadrature wave generator that has greatly improved accuracy and precision which is also an inexpensive method.

Keywords NCO · FPGA · Digital accumulator · LUT

1 Introduction

NCO is a digital signal oscillator used as signal generator. NCOs are usually combined with a digital-to-analog converter (DAC) at its output to form a direct digital synthesizer (DDS). NCO is also known as a frequency synthesizer that generates a range of frequencies hence can be used in building digital up and down converters. NCOs cater a flexible architecture that enables easy programmability such as changing of phase and frequency on the fly. NCO has a wide range of applications in today's communication systems like spread-spectrum and phase shift keying modulation techniques, digital phase-locked loops and measurement instrumentation [1]. Frequency synthesizers are also used for constructing demodulators and execute a number of modulation techniques, like PSK, FSK, and MSK [2]. An NCO typically comprises of two components: a *phase accumulator* (PA) and a *phase-to-amplitude*

Y. Swathi · N. Mohankumar (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: n_mohankumar@cb.amrita.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_53

converter (PAC). They are greatly known for their accuracy, reliability, and stability which are exploited by the proposal of this paper.

This paper presents the generation and analysis of a quadrature wave generator that generates a quadrature wave with 14-bit words as its inputs to control its frequency and phase using the basic numerically controlled oscillator design.

2 Literature Survey

NCO plays an important role in modern-day digital communication systems. It can be used to generate different waveforms, and the output can be given as feedback to a phase detector of all digital phase-locked loop (ADPLL) to maintain the phase of the wave generated. An ADPLL usually consists of phase detector, digital low-pass filter and an NCO/DCO in the feedback. To generate a sinusoidal wave, the NCO is followed by a sine/cosine LUT [3]. One of the basic requirements in most applications is to generate and control waveforms at a wide range of frequencies. The direct digital synthesis (DDS) technique is gaining popularity and is becoming very prominent to achieve programmable analog outputs with precision and high accuracy adding upon to the advantages of its easy portability [1].

Latest NCOs uses less hardware using any of the following three prominent methods:

(1) A sine LUT
(2) Digital oscillators with round-off compensation
(3) Coordinate rotation digital computer (CORDIC)-based designs [4].

CORDIC-based digital signal processing has become an influential mechanism in communications, biomedical, and industrial products, giving motivation to the designers for creating algorithm into architecture. It can be used to calculate sinusoidal magnitude and phase values through successive rotations. The pros of CORDIC are that no cos/sin ROM needed and only a small phase LUT, shifters, and adders are needed [2]. LUTs are lightweight and power efficient hence proved to be advantageous.

Hardware description language (HDL) designs are helpful for constructing LUT-to-CORDIC ratios that are power efficient. Internal NCO parameters are usually binary word (as input frequency) and the number of CORDIC stages for a particular amplitude resolution according to the output signal [5]. ROM-based NCO architecture with an improvement for QPSK modem can be designed using the current FPGA technology [6].

Many of the commercial signal generators that are used on a daily basis in laboratories are costly and huge. NCO technology, which can provide various basic waveforms that are commonly used in experiments, has enhanced the situation since it is inexpensive, lightweight, and highly accurate [7, 8]. In Computing area or the computation time of a nonlinear activation function of an artificial neural network a twofold approach is adapted for reducing the resource requirement [9].

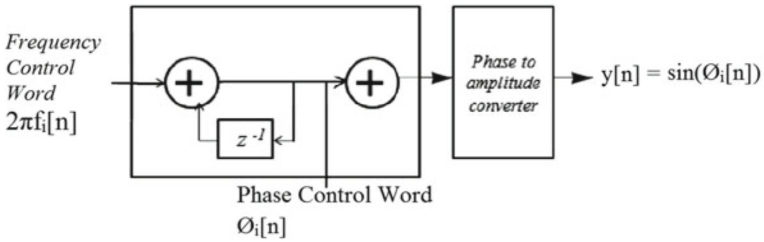


Fig. 1 Numerically controlled oscillator conceptual structure

3 NCO Architecture

Figure 1 shows the basic block diagram of a numerically controlled oscillator, which comprises of the main block of the design proposed by this paper. It consists of a phase accumulator and a phase-to-amplitude converter. The overflowing phase accumulator's (PA) output is added to the input frequency control word (FCW) at each clock sample when clocked with f_{CLK}. It generates the phase advance with time, depending on the digital frequency input word, f, thereby giving rise to the sequence $\Omega = \Sigma\mathrm{overflow}f$. This output frequency is:

$$f_{\mathrm{signal}} = f_{\mathrm{CLK}} \times f/(2^{\mathrm{FCW}}) \tag{1}$$

with resolution:

$$\Delta f_{\mathrm{signal}} = f_{\mathrm{CLK}}/(2^{\mathrm{FCW}}) \tag{2}$$

where FCW is the input frequency control word.

Typical resolutions are with FCW = 32 and for $f_{\mathrm{CLK}} = 200$ MHz, $\Delta f_{\mathrm{signal}} = 0.047$ Hz, or with FCW = 20 and for $f_{\mathrm{CLK}} = 20$ MHz, $\Delta f_{\mathrm{signal}} = 38.1$ Hz. In conventional sinusoidal wave generation techniques, this phase accumulator clock is kept in combination with phase-to-amplitude converter. It generates an output known as phase word which is generally used as an index to an LUT to provide the corresponding amplitude of the output waveform.

4 Building a Numerically Controlled Quadrature Wave Oscillator (NCQO)

See Fig. 2 and Table 1.

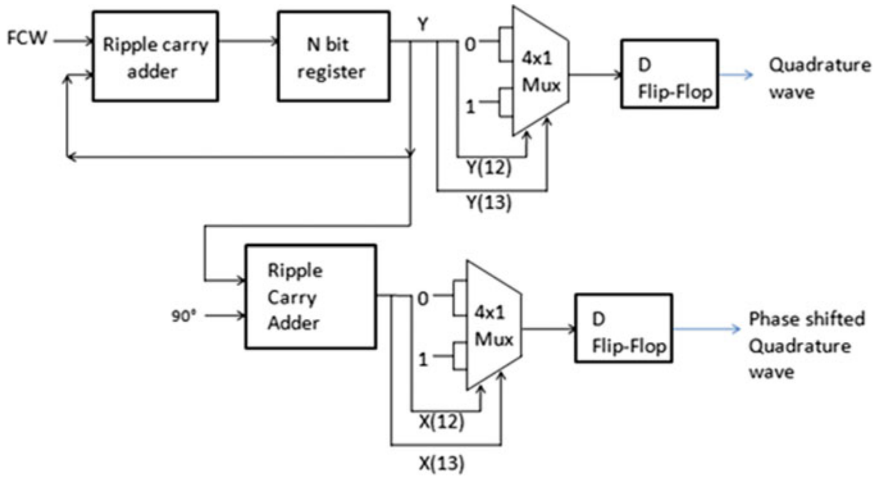


Fig. 2 Block diagram of the proposed quadrature waveform generator

Table 1 Design specification of quadrature wave NCO

Specification	Parameter values
Phase resolution	14 bits
Frequency tuning resolution	14 bits
Output signal	Square wave (phase-shifted by input phase word)
Output data width	1 bit

5 Components of NCQO

Phase Accumulator A phase accumulator comprises of a 14-bit binary full adder and a 14-bit register. Each clock cycle generates a new 14-bit output consisting of the preceding output obtained from the register summed with the input FCW. The adder is designed such that when it overflows due to the sum of the feedback value and the input 14-bit FCW, i.e., when it surpasses its maximum value of $2^N - 1$, the overflow bit is rejected to maintain the output word length equal to its input word length. The residual $Ø_n$ is kept in the 14-bit register and when the cycle reoccurs, it starts from $Ø_n$ this time as shown in Fig. 3. Since a phase accumulator is an FSM, ultimately the residual at a sample value K_0 comes back to the initial value $Ø_0$ as in Fig. 3.

Typically, the *output frequency* also known as *operating frequency* is given by the average overflow rate, which is determined using:

$$F_{\text{out}} = M * (F_{\text{clock}}/2^N) \tag{3}$$

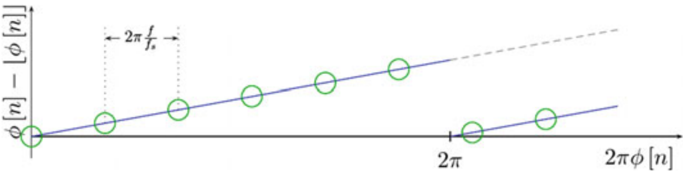


Fig. 3 Phase function. *Source* zipcpu.com [10]

The *frequency resolution*, which is the least feasible change in frequency, is given by:

$$F_{\text{res}} = F_{\text{clock}}/2^N \tag{4}$$

We can see from Eq. (1) that the phase accumulator could be seen as a frequency divider with the ratio $M/2^N$.

As mentioned earlier, the residual at any sample value K_0 will ultimately return to its initial value. The value K_0 is known as the grand repetition rate (GRR) obtained from Eq. (5).

$$\text{GRR} = \frac{2^N}{\text{GCD}(\Delta F, 2^N)} \tag{5}$$

4 × 1 Multiplexer A 4 × 1 multiplexer whose select inputs are the MSB of the output of the phase accumulator (13th and 12th bit in case of 14-bit NCO) is considered with the first two inputs being 0 and the last two inputs being 1.

D Flip-Flop When the clock is on its rising edge and the value of phase accumulator changes at some time points, race around condition occurs producing unwanted values at the output. To avoid such discrepancies, we place a D flip-flop with its input; as the output, we obtain for the multiplexer to assure a perfect square waveform as output waveform.

The above components make up the basic design for generating an in-phase square wave signal.

Phase Shifted Quadrature Waveform For example, a 90° phase shifted quadrature waveform is to be obtained, we place the above circuit in combination with another ripple carry adder whose other input is a constant 90° (01000000000000 in case of 14-bit ripple carry adder) followed by a 4 × 1 multiplexer and D flip-flop (to remove any race around conditions).

6 Simulation Results

The simulation results of the individual modules of the proposed NCO design are shown in Figs. 4, 5, and 6. Figure 5 shows that phase periodically rises and falls depicting the accumulation of phase till it sums up to 2π radians. Figure 6 shows the generated quadrature wave and its 90° phase-shifted version.

The frequency of output from the simulation results is ~10.5 ms.

The proposed design was programmed and implemented on the Xilinx Vivado tool using ZedBoard FPGA board. The obtained RTL schematic diagram of the design is shown in Fig. 7. It depicts the designed NCO quadrature waveform generator. Note that the ripple carry adder and the register form phase accumulator component

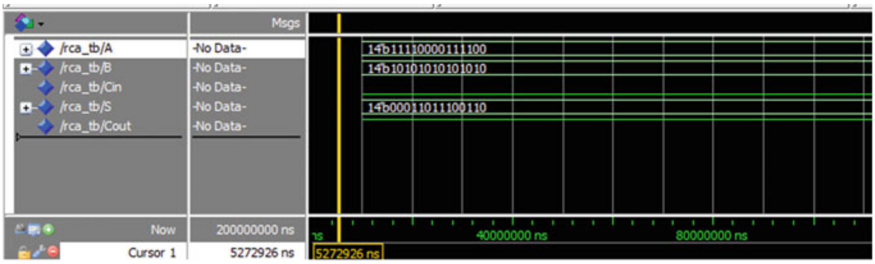


Fig. 4 Sample output of 14-bit ripple carry adder

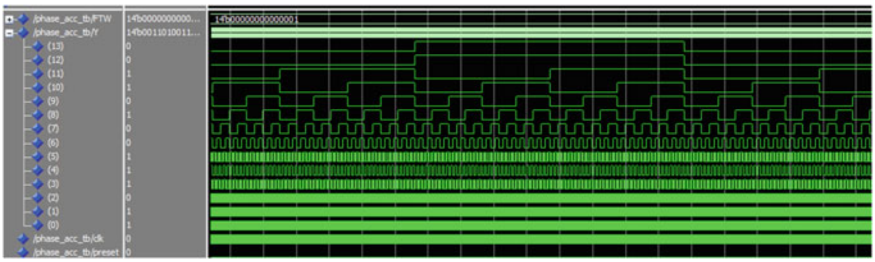


Fig. 5 Output of phase accumulator

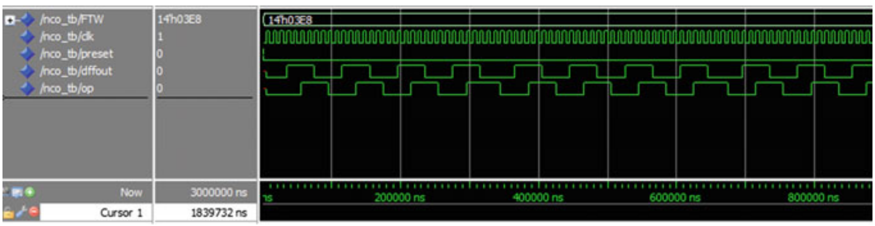


Fig. 6 Output of quadrature waveform generator

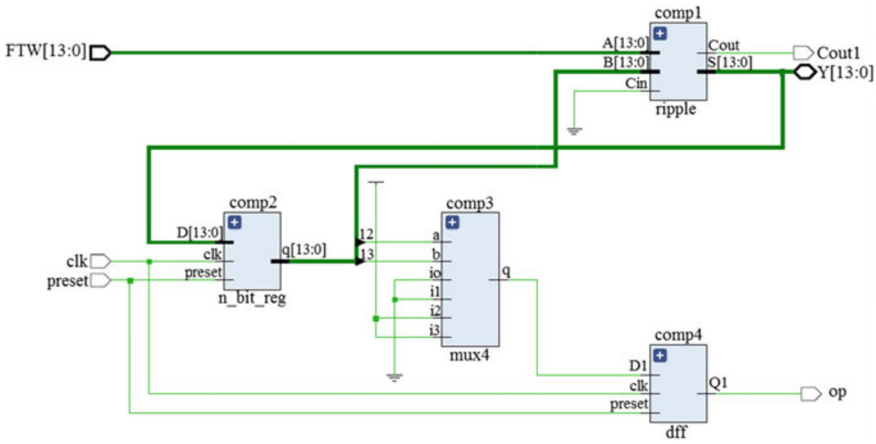


Fig. 7 RTL schematic diagram of basic NCO quadrature wave generator

Table 2 Utilization report of the entire design implemented on ZedBoard FPGA board

Logic resource	Slice LUTs	LUT as logic	Slice registers	Register as flip-flop	Bonded IOB
Resources	18	18	14	14	32
% Utilization	0.03	0.03	0.01	0.01	16.00

whose output is given as feedback to one of the inputs of ripple carry adder. Table 2 shows the utilization report of the entire design obtained after the design synthesis and implementation on the FPGA board. The device utilization summary highlights the minimal resource requirement for the proposed design.

7 Results and Analysis

A square waveform and its quadrature waveform were generated as shown in Fig. 6. The time period of the output waveform was about 10.5 ms. Theoretically obtained results are as follows:

$$M\,(\text{integer value of FCW}) = 1\,\text{kHz}$$

$$F_{\text{Clk}} = 10\,\mu\text{s} = 0.1\,\text{MHz}$$

from Eq. (5):

$$\text{GRR} = 2^{20}/\text{GCD}(2^{20}, 1\,\text{k}) = 8$$

Table 3 Power profile of the proposed NCO

Module	On-chip power utilized (W)
Entire design	14.106
Ripple carry adder	10.167
14-bit register	3.723
4×1 multiplexer	0.663

from Eq. (1):

$$\Delta F_{\text{signal}} = 0.1\,\text{MHz}/2^{20} \sim 0.095\,\text{Hz}$$

from Eq. (2):

$$F_{\text{signal}} = (0.1\,\text{MHz}/2^{20}) * 1\,\text{k} \sim 95\,\text{Hz} => T_{\text{out}} \sim 10.5\,\text{ms}$$

Hence, the calculated frequency matches with the simulation result.

From the simulation results, clearly we observe almost a zero-phase noise or phase errors. Hence, the proposed design can produce square waves with any desirable phase shift (according to the input phase word) without any phase errors or noise. This becomes the key advantage of our proposed design over analog designs like that of quadrature voltage-controlled oscillator (QVCO) which exhibits phase noise in the generated wave [11]. Table 3 shows the total as well as module wise on-chip power utilized.

8 Conclusion

The proposed design of quadrature waveform generator using NCO has been presented. The proposed design being a digital circuit ensures precision and accuracy compared to other periodic waveform generators like astable multi-vibrators that are analog circuits. The main advantage of the proposed design is that it is entirely a digital circuit which is easy and simple to implement rather than analog circuit designs. The theoretical oscillation frequency agrees with the simulated oscillation frequency. This digital module can be incorporated into any digital or mixed signal subsystems as it consumes only very less power and is area efficient. The proposed design which is lightweight and inexpensive can replace bulky equipments in the areas where a simple quadrature wave with the desired phase shift is required. It can also be used to generate clock pulses of desired frequency controlled with the input frequency control word. Future work can be focused on improving the accuracy and precision of the generated wave in practical applications. Digital phase-locked loops can be employed for accurate phase shifts.

References

1. Gaikwad, A.D., Kharat, G.U., Bodake, S.H.: A review on design and implementation of numerically controlled oscillator. Int. J. Innov. Res. Comput. Commun. Eng. (IJIRCCE) **4**(4) (2016)
2. Zhihua, L., Weilian, W.: The design of NCO based on CORDIC algorithm and implementation in FPGA (*). In: 2011 International Conference on Electronics, Communications and Control (ICECC)
3. Patil, A., Tandel, P.H.: A numerically controlled oscillator for all digital phase locked loop. Int. J. Eng. Trends Technol. (IJET) **38**(4) (2016)
4. Menon, S., Cho, G., Soderstrand, M.: An improved numerically controlled digital oscillator. In: 2003 PACRIM, 2003 IEEE Pacific RIM Conference on Communications Computers and Signal Processing, pp. 1040–1044. Print ISBN: 0-7803-7978-0
5. Janiszewski, I., Hoppe, B., Meuth, H.: Precision and performance of numerically controlled oscillators with hybrid function generators. In: 2001 IEEE International Frequency Control Symposium and FDA Exhibition, pp. 744–752
6. Ghiwala, G.D., Thaker, P.P., Amin, G.D.: Realization of FPGA based numerically controlled oscillator. IOSR J. VLSI Signal Process. (IOSR-JVSP) **1**(5), 7–11 (2013). ISSN: 2319 – 4200, ISBN No.: 2319 – 4197
7. Qi, J., Sun, Q., Wu, X., Wang, C., Chen, L.: Design and analysis of a low cost wave generator based on direct digital synthesis. J. Electr. Comput. Eng., 17 pp., Article ID 367302 (2015)
8. Janiszewski, I., Hoppe, B., Meuth, H.: Numerically controlled oscillators with hybrid function generators. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**(7) (2002)
9. Nirmaladevi, M., Mohankumar, N., Arumugam, S.: Modeling and analysis of neuro-genetic hybrid system on FPGA. Elektron. Elektrotech. (2009)
10. https://zipcpu.com/dsp/2017/12/09/nco.html
11. Raman, S., Vignesh, V., Kumar, N.: Design and implementation of quadrature voltage controlled oscillator in 65 nm CMOS. In: 7th IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), PES Institute of Technology, Bengaluru, South Campus, India (2018)

Efficient Multimedia Data Transmission Model for Future Generation Wireless Network

T. Kavitha and K. Jayasankar

Abstract To meet the resource demands of future wireless communications due to the increased usage of smart phones, smart devices and video-streaming platforms have led the future wireless communications to deploy dense heterogeneous Cloud Radio Access Network Systems (C-RANs). The heterogeneous communication environment offers fine-grained uniform experience to its subscribers and low-cost deployment irrespective of user location in the communication environment. The C-RANs have emerged as one of the promising solution to meet the operational cost, Quality of Service (QoS), and compression of baseband data requisite. This work, considers implementation of C-RAN model where baseband unit (BBU) and Remote Radio Heads (RRH) are connected through Common Public Radio Interface (CPRI) Fronthaul links. For such networks, reducing the data rate compression is very essential as the Fronthaul links capacity is limited and costly as they transport complex baseband samples. Fronthaul compression exploits the spatial and temporal behavior of time domain LTE signals for reducing the data rates has been considered by the existing models nonetheless it remains a challenge. To overcome the research challenge in building better transmission model, this work considers jointly exploiting both spatial and temporal correlations of the transmitted baseband signals to obtain efficient Fronthaul compression performance for LTE cellular networks using Refined Huffman. This work, assumes a case similar to massive Multiple-Input Multiple-Output (MIMO) communication mobile environment, where number of receiving antennas will outnumber the active user terminals. Our model applies Low-Rank (LR) approximation of complex baseband samples to obtain spatial and temporal correlations construction matrices of signals. The correlated baseband signals are then encoded using proposed refined Huffman encoder technique (RHCT) to achieve better compression. Experiments are carried out for evaluating the performance attained by the proposed method with Standard Huffman. The results obtained displays, that the proposed model attains superior performance enhancement than

T. Kavitha (✉)
Department of ECE, MVSREC, Hyderabad, India
e-mail: tkavitha_ece@mvsrec.edu.in

K. Jayasankar
Department of ECE, MGIT, Hyderabad, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_54

the existing state-of-the-art Huffman encoder model in terms of Symbol Error Rate (SER), Bit Error Rate (BER), Compression, and Throughput (Sum rate).

Keywords Cloud Radio Access Networks · Codewords · Compression · Huffman · Mobile Network · Variable Length Code · Wireless Network

1 Introduction

Future mobile wireless network systems must offer robust, flexible, and efficient communication and must guarantee Quality of Service (QoS) for provisioning different applications such as data, voice, and video. To provide these services, networks need to offer high-speed reliable communication (i.e., increased data rates). As a result, it incurs high operational cost due to rapid growth of dynamic smart equipments and Bandwidth (BW) aware apps for Service Providers (SP) to cater to the massive and dynamic requirements of data rates by the future mobile operators. The C-RAN has arisen as an efficient design and attained wide interest across various research communities [1, 2]. In cloud radio access networks, the Baseband Processing Unit (BPU) or Central Processor (CP) and upper layers operation of Base Station (BS) are migrated to a Centralized Unit (CU) (i.e., in the cloud computing environment). The BS (i.e., distributed Remote Radio Head (RRH)) or Radio Unit (RU) is associated with high capacity, low-latency Fronthaul link (FHL) for communicating with the users, which can be realized via microwave, mm wave or fiber optics technologies. As most of the BPU operation is centralized at the BPU and cloud radio access networks offer Load Balancing (LB), Energy Efficiency (EE), and Interference Management (IM) capabilities through virtual BS communication set [2, 3] along with, C-RAN offers ideal network-wide inter-cell IM. This is realizable due to multi-cell processing [4]. Nevertheless, as major BPU operation is carried out at the BPU, it necessitates huge bandwidth or communication speed on the Fronthaul (FH) links in order to transmit complex-digitized baseband signals. Thus, minimizing the bandwidth is very essential in the successful operation of cloud radio access networks. This is due to the restricted Fronthaul links size. Thus, C-RAN needs to be modeled with effective techniques of transmitting digitized BPU signals on the FHL with the restricted size.

Current link standard of open radio interface [5] and Common Public Radio Interface (CPRI) [6, 7] describes the Communication Interface (CI) between BPU and RRH on the FH communication environment. However, the elementary method modeled by Common Public Radio Interface is likely to offer Bit Rate (BR) which are hard to provision accessible FH capability size. This aided number of researchers for modeling techniques which minimize the BR of the Fronthaul communication links. It [8] conducted extensive survey of work for minimizing the Fronthaul baseband transmission bandwidth by employing encoding and decoding methods, such as optimized non-uniform quantization [9], filtering, down sampling [10], and lossless compression [11]. The lossless compression techniques like Unused Significant

Bit Removal (USBR), Elias Gamma Coding (EGC), and Adaptive Arithmetic Coding (AAC) are discussed but adaptive arithmetic coding results in better compression [11]. Further, apart from point-to-point compression, a number of transmission approaches are presented from a network-aware perspective [12–14].

A low-latency time domain compression method for Long-Term Evolution (LTE) and fourth-generation (4G) uplink signals is presented in [9] to reduce the data rate. For eliminating spectral redundancies, resampling of the signals, transform to block floating point and then quantizing the samples in respective block using non-linear quantize (NLQ) is proposed. Similar to [9, 15] presented a model with modified block scaling [16]. A Quantizer is designed by combining noise-shaped feedback coding with zero-mean Gaussian distribution for both uplink and downlink. Further, [17] exploited time correlation in the signals for building multi-stage vector quantization compression algorithm presented as existing scalar quantization model but could not exploit time correlations in LTE signals. In [18, 19], presented compress-and-forward spatial domain-based methods with Joint Decoding (JD) considering Linear Wyner (LW) mobile network using solitary antenna terminal. Similarly, [20, 21] proposed Estimate-Compress-Forward (ECF) method and Compress-and-Forward Relaying (CFR) strategies. Distributed compression design [19] utilizing Distributed Wyner–Ziv (DWZ) encoding was modeled [22]. In [23] exploited the correlation among obtained or collected signals by resolving distributed Karush–Kuhn–Tucker (DKKT) rule. The model [23] adopted single-layer compression and transmission and [24] investigated the importance of layered transmission for handling indecisions in the communication feature level of Backhaul Links (BHL).

To attain better transmission model, this work considers jointly exploiting correlation between temporal and spatial information of obtained LTE signals to attain efficient compression performance. Further, this work assumes a case similar to massive Multiple-Input Multiple-Output (MIMO) communications [25] and millimeter wave mobile environment [26] where Active User Terminals (AUT) or User Equipment (UE) will generally be less when compared to number of Receiving Antennas (RA). For such network, this work presents an efficient encoding and decoding of BPU uplink time domain signals with respect to time and space. Our model first applies low-rank approximation for obtaining correlation construction matrices with respect to space and time. Then, the obtained correlated signals are encoded using Huffman encoder [27] and transmitted. However, existing standard Huffman compression technique requires large number codewords/look-up table which in turn occupies large memory and obtains low compression ratios [28]. Further, it incurs Computation Overhead (CO) because of longer length of codewords. Thus, it requires larger bandwidth for transmission of compressed signals. Therefore, minimizing the number of codewords and their corresponding lengths, without any loss of information is required. To overcome the above said problem, this work presents a novel Refined Huffman-based encoder to perform compression of these correlated complex signals using codewords presented in [29].

The highlights and the contribution of work are mentioned below:

- This manuscript presents a novel Fronthaul compression technique that considers jointly exploiting correlation among temporal and spatial information of LTE signals.
- This paper considers a robust condition where higher number of receiving antennas at receiving radio head is considered when compared with active clients in the communication network.
- Our model applies Low-Rank approximation of the obtained baseband received signals by exploiting correlation construction matrices with respect to space and time.
- Then, the correlated signals are encoded using proposed Refined Huffman encoder to attain improved compression.

The proposed model attains superior performance considering BER, SER, throughput (sum rate), and compression ratio than the state-of-the art model using Huffman.

The manuscript is articulated as described. The proposed efficient multimedia data transmission for future generation high-speed wireless network is given in Sect. 2. Experiment analysis is presented in Sect. 3. In last section, the conclusion and future research direction are given.

2 Efficient Multimedia Data Transmission Model for Future Generation High-Speed Wireless Network with Proposed Compression Technique

This section presents an efficient multimedia data transmission model for future generation high-speed wireless networks. The architecture of proposed multimedia data transmission model is shown in Fig. 1. Consider a high-speed wireless link where each subscriber device produces symbols and are first pre-coded utilizing discrete Fourier transformation (DFT). Post that, the resultant symbols are mapped to obtain

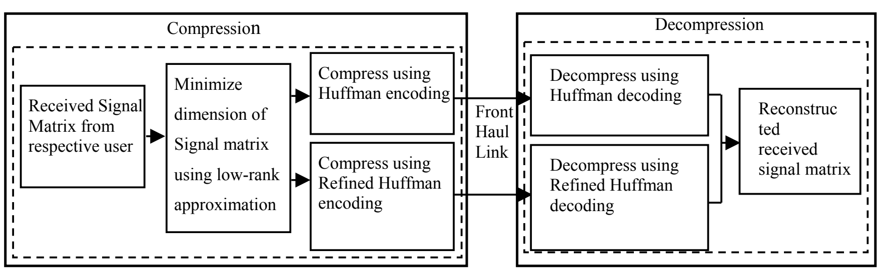


Fig. 1 Architecture of Proposed Efficient Multimedia Data Transmission Model

orthogonal frequency-division multiplexing symbols in time domain. Further, we add a cyclic prefix to the symbols before transmission of these SC-FDMA symbols. The uplink transmission uses SC-FDMA instead of OFDMA to reduce peak-to-average power ratio (PAPR). The RRH obtains orthogonal frequency-division multiplexed time domain signals from different subscribers with interference and noise because of multi-path fading and environmental conditions.

This work considers Fronthaul compression of these obtained time domain signal on high-speed wireless links, which is composed of multiple-single antenna subscribers and massive multi-antennas at RRHs. Each remote radio head can communicate with one cell and has N antennas. Further, from each active subscriber, the remote radio heads obtain signals from that communicating cell. The expected signal on a given nth antenna z_n is computed as follows

$$z_n[o] = \sum_{v} y_v[o] * i_{n,v}[o] + x_n[o] \tag{1}$$

$$n \in \{1, 2, \ldots, N\}, o \in \{0, 1, 2, \ldots\}$$

where $(*)$ depicts convolution, y_v depicts orthogonal frequency-division multiplexing symbol of vth subscriber, x_n is the additive white Gaussian noise of nth antenna, $i_{n,v}$ depicts channel reply of vth subscriber to nth antenna. Using Eq. (1), the matrix of obtained signals can be represented with dimension, $\boldsymbol{Z} \in \mathbb{D}^{N*O}$, where N depicts quantity of antennas at RRH and O depicts quantity of samples used for compression. Therefore, the matrix $\boldsymbol{Z}$ can be expressed as follows

$$\boldsymbol{Z} = \begin{bmatrix} z_1[0] & z_2[0] & \ldots & z_N[0] \\ z_1[1] & z_2[1] & \ldots & z_N[1] \\ \vdots & \vdots & \ddots & \vdots \\ z_1[O-1] & z_2[O-1] & \ldots & z_N[O-1] \end{bmatrix} \tag{2}$$

And any column in $\boldsymbol{Z}$, and is represented as follows

$$z_j = \left[z_1[0]\ z_1[1] \ldots z_1[O-1] \right]^{\boldsymbol{H}} \tag{3}$$

Since, the Z matrix is highly correlated, the column z_j, i.e., z_j, $j \in \{1, 2, \ldots, N\}$. The proposed work applies low-rank approximation to $\boldsymbol{Z}$

$$\boldsymbol{Z} = \boldsymbol{Z}_0 + \boldsymbol{F} \tag{4}$$

where $\boldsymbol{Z}_0 \in \mathbb{D}^{N*O}$ represents a no noise low-rank matrix, that gives information about the input signals (y) and channel behavior (i), and $F \in \mathbb{D}^{N*O}$ depicts a composite additive white Gaussian noise matrix. Applying low-rank approximation, results compression of input data as represented in Eq. (4) with a smaller number of coefficients. The encoded signal is then transmitted to baseband unit through its Fronthaul link. Post that, the signal obtained is then decoded at BPU/BBU for obtaining $\boldsymbol{Z}$ in the reverse manner of encoding process.

The objective of our compression method is to minimize the dimension of the matrix $\boldsymbol{Z}$, so that lesser signal sets can be transmitted over wireless (Fronthaul) link. To achieve this, our proposed work uses refined Huffman codewords proposed by author [29] to obtain improved compression.

Consider that $\boldsymbol{Z}_0$ in Eq. (4) depicts rank M matrix of size $O * N$, and $O \gg N > M$ for easiness, the low-rank estimate of $\boldsymbol{Z}$ is computed as follows as per the distance function given by Frobenius norm.

$$\boldsymbol{Z}'' = \underset{\mathcal{R}(\widehat{\boldsymbol{Z}})=M}{\operatorname{argmin}} \left\| \boldsymbol{Z} - \widehat{\boldsymbol{Z}} \right\| G \tag{5}$$

where the normalization $\|.\|G$ depicts Frobenius normalization denoted by $\widehat{\boldsymbol{Z}}$, and the ideal strategy to arrive at Eq. (5) is obtained as follows.

$$\boldsymbol{Z}'' = V_M \beta_M W_M^I \tag{6}$$

where

$$\begin{aligned} V_M &= \left[v_1 \ v_2 \ \ldots \ v_M \right] \\ W_M &= \left[w_1 \ w_2 \ \ldots \ w_M \right] \\ \beta_M &= \mathcal{D}\left[\alpha_1 \ \alpha_2 \ \ldots \ \alpha_M \right] \end{aligned}$$

where $\boldsymbol{Z}''$ denotes Singular Value Decomposition (SVD). SVD decomposes $\boldsymbol{Z}''$ into $(\cdot)^I$ conjugate transpose, $v_j \in \mathbb{D}^O$ are left eigenvector, $w_j \in \mathbb{D}^O$ are right eigenvector, $\alpha_1, \alpha_2 \ldots \alpha_M$ represent diagonal singular values and $\mathcal{D}$ represents the diagonal matrix. This paper considers the rank M is identified. For reducing the size of $\boldsymbol{Z}$, this work first obtains the matrix of the preliminary M linearly uncorrelated variables W_M identical to M eigenvectors using singular value decomposition. Further, we multiply with W_M to transform matrix $\boldsymbol{Z}$, and then map the obtained signal vectors z_j from original space of N parameter to an upcoming space of M parameter that are uncorrelated over the feature set. The matrix transformation $Q_M \in \mathbb{D}^{O*M}$ is expressed as follows

$$Q_M = Z W_M = V_M \beta_M \tag{7}$$

where Q_j, $j \in \{1, 2, \ldots, M\}$ depicts jth column of Q_M, and it is a highly decorrelated data vector.

Once we obtain the linear transformation matrix Q_M, the samples of Q_M are compressed by entropy coding. Huffman encoding method is an optimal prefix code [28] that can attain shortest average code length for a given input symbols, as the input symbols are combination of binary data 1's and 0's. The Huffman encoding/decoding can be efficiently implemented [30] using a look-up table of codewords. The author has designed codewords which are presented in [29] and are used here to compress further the matrix of values Q_M and are transmitted to the BBU through error-free capacity limited wireless (Fronthaul) channel and the estimated LR matrix $\boldsymbol{Z}''$ received by baseband unit is described using below equation

$$\boldsymbol{Z}'' = Q_M W_M^I = V_M \beta_M W_M^I. \tag{8}$$

However, the standard Huffman Compression Technique (SHCT) necessitates higher memory size due to higher quantity of codewords (183). The maximum codewords have length 13 and the compression ratio obtained using this SHCT is less [30]. Also it incurs high transmission time because of longer codewords. Thus, it requires larger bandwidth for transmission if compressed using SHCT. Therefore, reducing the average code length and codewords size without any loss of information or quality of image (lossless) is required for efficient transmission in the future crowded wireless networks.

The author has proposed ideal Huffman (Refined Huffman) compression method which is obtained by optimizing standard prefix-free code words of SHCT. The proof of ideal Huffman codewords (Refined Huffman) is given in [31]. The codewords used for experiment can be obtained from [29]. Using codewords from [29], aids in reducing the amount of bits needed for compressing the image. Also it reduces the time required for encoding process. The key for attaining better compression is to utilize longer codewords for rarely appearing symbols/feature and smaller codewords for repeatedly appearing symbols/features. The RHCT is used as Fronthaul compression, to encode the obtained BBU signals at RRH and transmit to BBU. The results are compiled between the existing model SHCT which uses Modified Huffman with proposed model which uses Refined Huffman, RHCT which author has proposed in [29, 31].

The amount of data transmitted over the channel is expressed as follows

$$NM + OM \tag{9}$$

And the compression ratio for the amount of data obtained using proposed compression model is computed as follows

$$\mathcal{CR}_{\mathcal{P}} = \frac{N * O}{L_{\text{avg}}[NM + OM]}. \tag{10}$$

where Z depicts, the total sample size. L_{avg} is the average number of bits per Huffman codeword. The L_{avg} of SHCT is 8.6270 bits/sample and the L_{avg} of RHCT is 7.7046 bits/sample [29].

The compressed signals are received at BBU, decompressed and OFDMA demodulator processing is carried to obtain the transmitted binary data. At BBU, the number of Bits in Error (BER), number of Symbols in Error (SER), and throughput of the network are calculated. The proposed model attains superior performance than that of existing SHCT in terms of BER, SER, Compression Ratio (CR), and Throughput (sum rate) which is proved below.

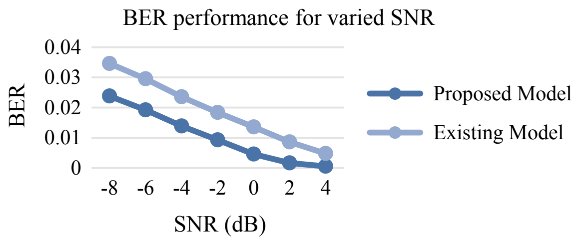
3 Experimental Outcome and Discussion

The experimental outcome attained by proposed refined Huffman compression-based transmission (RHCT) model is compared with the existing standard Huffman compression-based transmission (SHCT) model is described in [30, 32]. The model is implemented using MATLAB 18a framework, using Windows 10 operating system, Intel Pentium-class processor, 12 GB RAM and 4 GB dedicated CUDA-enabled graphic processing unit (GPU). For experimental analysis, researcher considered a LTE channel model with 10 MHz, number of resource blocks are 50, and the channel is AWGN. For constructing antenna correlation matrices, the exponential correlation method [33] is utilized considering uniform linear array. Further, we have considered 32 RA's and 4 AUT's. Then the resource blocks are equally allocated to each user. The performance of wireless network model is evaluated in terms of BER, SER, throughput (Sum rate), and CR.

3.1 Bit Error Rate Performance Evaluation Considering Varied SNR

This section evaluates BER performance of the RHCT model over the SHCT model. For BER experiment analysis, 4 active users and 32 receiving antennas are considered, experiments are conducted by varying SNR from −8 to 4 dB, the BER performance is graphically shown in Fig. 2. The outcome attained shows that the proposed compression model reduces BER over the existing compression model by 31.186, 34.83, 40.96 49.47, 66.09, 80.99, and 87.8% considering −8, −6, −4, −2, 0, 2, 4 dB, respectively. An average BER reduction of 45.09% is achieved by the proposed model over the existing model. From experiment analysis, it is clear that as we increase the SNR, the RHCT model attains better BER performance when compared with that of the SHCT model.

Fig. 2 Performance of BER for varied SNR by considering 32 RA's and 4 AUT's



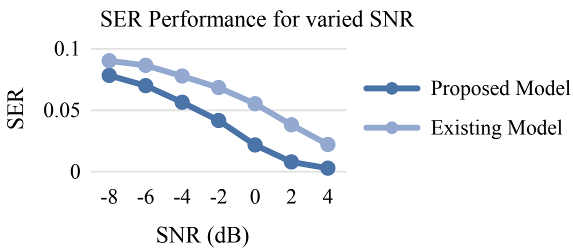


Fig. 3 Performance of SER for varied SNR by considering 32 RA's and 4 AUT's

3.2 Symbol Error Rate Performance Evaluation Considering Varied SNR

This section evaluates SER performance of the RHCT model over the SHCT model. For SER experiment analysis, 4 active users and 32 receiving antennas are considered, experiments are conducted by varying SNR from −8 to 4 dB, the SER performance which is graphically shown in Fig. 3. The outcome attained shows, the proposed compression model reduces symbol error rate over existing compression model by 13.2, 19.19, 27.39, 39.14, 60.78, 79.49, and 87.61% considering −8, −6, −4, −2, 0, 2, 4 dB, respectively. An average SER reduction of 45.09% is attained by the RHCT model over the SHCT model. An average SER reduction of 25.94% is achieved by the proposed model over the existing model considering both the cases. From the experiment analysis, it is clear that as we increase SNR, the proposed RHCT model attains better SER performance when compared to the existing SHCT model. Similarly, the RHCT model shows better SER performance as we increase the number of receiving antennas. The overall result attained shows that SER of RHCT model is better than using SHCT.

3.3 Throughput (Sum Rate) Performance Evaluation Considering Varied SNR

This section evaluates throughput (Sum rate) performance of the proposed RHCT model over the existing SHCT model. For throughput experiment analysis, 4 active users and 32 receiving antennas were used, experiment is conducted by varying SNR from −4 to 4 dB, and the throughput performance is graphically shown in Fig. 4. The outcome attained shows that the proposed RHCTcompression model improves throughput performance over the existing SHCT compression model by 11.41, 9.63, 7.5, 5.389, and 3.4% considering −4, −2, 0, 2, 4 dB, respectively. An average throughput performance improvement of 7.46% is attained by the proposed RHCT model over the existing SHCT model.

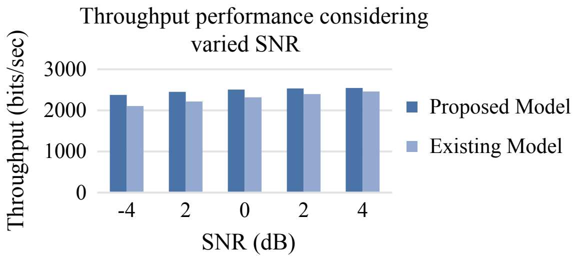


Fig. 4 Performance of Throughput for varied SNR by considering 32 RA's and 4 AUT's

Fig. 5 Flight

3.4 *Compression Ratio Performance Evaluation*

This section evaluates compression ratio performance attained by the proposed Refined Huffman Compression-based Transmission (RHCT) model over the existing Standard Huffman Compression-based Transmission (SHCT) model. The experiment is conducted to evaluate compression ratio performance for flight .tiff input image shown in Fig. 5, by considering fixed number of antennas 32, fixed number of active users 4 and SNR is varied from −4 to 4 dB. The image used for experiment analysis is collected from [34]. The compression ratio performance attained by the proposed RHCT is computed by using Eq. (10). The compression ratio outcome attained for the image in Fig. 5 using RHCT is 6.09 and using SHCT is 5.33. From the overall results attained, it can be seen that the proposed RHCT model attained 14.33% better performance than that of SHCT [30, 32].

4 Conclusion

This manuscript presented a novel Fronthaul compression technique that considered temporal and spatial information of time domain LTE signals. A 10 MHz LTE channel model is simulated which considers higher number of receiving antennas at remote radio heads compared with the number of active user terminals in a particular cell. The RRH's obtains spatial and temporal complex baseband signals and low-rank approximation is applied on the received baseband signal to obtain correlation construction matrices with respect to space and time. Then, the correlated signals are entropy encoded using proposed RHCT and transmitted to BBU. At BBU the decoding of received signals is carried. Experiments are conducted to evaluate the performance of proposed model over the existing SHCT model in the presence of AWGN noise for different SNR by considering 32 receiving antennas and 4 active users. To evaluate the performance, various network parameters such as BER, SER, Throughput (Sum rate) and CR are considered. The experiment outcome shows that the proposed RHCT model reduces BER and SER by 45.09 and 25.94% over the existing SHCT model, respectively. Similarly, the proposed RHCT model improves throughput performance by 7.46% over the existing SHCT model. The proposed RHCT model improves compression performance by 14.33% over the SHCT model. The overall results attained shows that the proposed transmission model with RHCT as a Fronthaul compression technique attains significant performance improvement over the state-of-the art SHCT model in terms of BER, SER, throughput (Sum rate), and compression (CR). The future work would consider performance evaluation by considering large number of AUT's and RA's under different environmental conditions.

References

1. Chen, K., C-RAN, D.R.: The Road Towards Green RAN. China Mobile Research Institute Ver. 2 (2011)
2. Checko, A., Christiansen, H.L., Yan, Y., Scolari, L., Kardaras, G., Berger, M.S., Dittmann, L.: Cloud RAN for mobile networks: a technology overview. IEEE Commun. Surv. Tutor. **17**(1), 405–426 (2012)
3. Zhu, Z., Gupta, P., Wang, Q., Kalyanaraman, S., Lin, Y., Franke, H., Sarangi, S.: Virtual base station pool: towards a wireless network cloud for radio access networks. In: Proceedings of the 8th ACM International Conference on Computing Frontiers, New York, Mar 2011, pp. 34:1–34:10
4. Marsch, P., Fettweis, G.: Uplink CoMP under a constrained backhaul and imperfect channel knowledge. IEEE Trans. Wirel. Commun. **10**(6), 1730–1742 (2011)
5. ETSI GS: 001: Open Radio Equipment Interface (ORI), Requirements for Open Radio Equipment Interface (Release 3) (2014)
6. AB Ericsson, Huawei Technologies, NEC Corporation, Alcatel Lucent, Nokia Siemens Networks: Common Public Radio Interface (CPRI); Interface Specification, CPRI Specification, vol. 5 (2011)
7. CPRI Specification V6.0 (2013-08-30): Common Public Radio Interface (CPRI); Interface Specification (2013)
8. Park, S., Simeone, O., Sahin, O., Shamai Shitz, S.: Fronthaul compression for cloud radio access networks: signal processing advances inspired by network information theory. IEEE Signal Process. Mag. **31**(6), 69–79 (2014)
9. Guo, B., Cao, W., Tao, A., Samardzija, D.: CPRI compression transport for LTE and LTE-A signal in C-RAN. In: Proceedings of International ICST Conference on Communications and Networking in China (CHINACOM), pp. 843–849. IEEE (2012)

10. Samardzija, D., Pastalan, J., MacDonald, M., Walker, S., Valenzuela, R.: Compressed transport of baseband signals in radio access networks. IEEE Trans. Wirel. Commun. **11**(9), 3216–3225 (2012)
11. Vosoughi, A., Wu, M., Cavallaro, J.R.: Baseband signal compression in wireless base stations. In: Proceedings of IEEE Global Communications Conference, pp. 4505–4511 (2012)
12. Vu, T.X., Nguyen, H.D., Quek, T.Q.S., Sun, S.: Fronthaul compression and optimization for cloud radio access networks. In: 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, pp. 1–6 (2016)
13. Heo, E., Simeone, O., Park, H.: Optimal fronthaul compression for synchronization in the uplink of cloud radio access networks. EURASIP J. Wirel. Commun. Netw. 22 (2017)
14. Zhou, L., Yu, W.: Optimized backhaul compression for uplink cloud radio access network. IEEE J. Sel. Areas Commun. **32**(6), 1295–1307 (2014)
15. Ren, Y., Wang, Y., Xu, G., Huang, Q.: A compression method for LTE-A signals transported in radio access networks. In: Proceedings of IEEE International Conference on Telecommunications, pp. 293–297 (2014)
16. Nieman, K.F., Evans, B.L.: Time-domain compression of complex baseband LTE signals for cloud radio access networks. In: Proceedings of IEEE Global Conference on Signal and Information Processing, Dec 2013, pp. 1198–1201
17. Si, H., Ng, B.L., Rahman, M.S., Zhang, J.: A vector quantization based compression algorithm for CPRI link. In: Proceedings of IEEE Global Communications Conference, Dec 2015
18. Sanderovich, A., Somekh, O., Poor, H.V., Shamai, S.: Uplink macro diversity of limited backhaul cellular network. IEEE Trans. Inform. Theory **55**(8), 3457–3478 (2009)
19. Zhou, Y., Xu, Y., Yu, W., Chen, J.: On the optimal fronthaul compression and decoding strategies for uplink cloud radio access networks. IEEE Trans. Inform. Theory (2016). https://doi.org/10.1109/tit.2016.2617862
20. Kang, J., Simeone, O., Kang, J., Shitz, S.S.: Joint signal and channel state information compression for the backhaul of uplink network MIMO systems. IEEE Trans. Wirel. Commun. **13**(3), 1555–1567 (2014)
21. Smys, S., Josemin Bala, G.: Performance analysis of virtual clusters in personal communication networks. Clust. Comput. **15**(3), 211–222 (2012)
22. Coso, A.D., Simoens, S.: Distributed compression for MIMO coordinated networks with a backhaul constraint. IEEE Trans. Wirel. Commun. **8**(9), 4698–4709 (2009)
23. Park, S.-H., Simeone, O., Sahin, O., Shamai, S.: Robust and efficient distributed compression for cloud radio access networks. IEEE Trans. Veh. Technol. **62**(2), 692–703 (2013)
24. Simeone, O., Somekh, O., Erkip, E., Poor, H.V., Shamai, S.: Robust communication via decentralized processing with unreliable backhaul links. IEEE Trans. Inform. Theory **57**(7), 4187–4201 (2011)
25. Lu, L., Li, G.Y., Swindlehurst, A.L., Ashikhmin, A., Zhang, R.: An overview of massive MIMO: benefits and challenges. IEEE J. Sel. Topics Signal Process. **8**(5), 742–758 (2014)
26. Pi, Z., Khan, F.: An introduction to millimeter-wave mobile broadband systems. IEEE Commun. Mag. **49**(6), 101–107 (2011)
27. Yasuda, Y.: Overview of digital facsimile coding techniques in Japan. Proc. IEEE **68**(7) (1980)
28. Arps, R.B., Truong, T.K.: Comparison of international standards for lossless still image compression. Proc. IEEE **82**(6) (1994)
29. Kavitha, T., Jaya Sankar, K.: An efficient compression technique for ITU-T group 3 coded images using variable length codes with reduced average length. In: 2016 IEEE International Conference on India International Conference On Information Processing (IICIP-2016), pp. 1–6 (2016)
30. Ramalho, L., Fonseca, M.N., Klautau, A., Lu, C., Berg, M., Trojer, E., Höst, S.: An LPC-based fronthaul compression scheme. IEEE Commun. Lett. **21**(2), 318–321 (2017)
31. Kavitha, T., Jayasankar, K.: Ideal Huffman code for lossless image compression for ubiquitous access. Indones. J. Electr. Eng. Comput. Sci. **12**(2), 765–774 (2018)

32. Ramalho, L., Freire, I., Lu, C., Berg, M., Klautau, A.: Improved LPC-based fronthaul compression with high rate adaptation resolution. IEEE Commun. Lett. **22**(3), 458–461 (2018)
33. Loyka, S.L.: Channel capacity of MIMO architecture using the exponential correlation matrix. IEEE Commun. Lett. **5**(9), 369–371 (2001)
34. Standard Test Images. Compiled by Mike Waken, University of Michigan. itu.int/net/itu-t/sigdc/genimage/test24.htm

Smart Fleet Monitoring System in Indian Armed Forces Using Internet of Things (IoT)

Mitul Sheth and Pinal Rupani

Abstract Resource management becomes an essential task in our day-to-day lives. Particularly, a few types of assets like fuel which is non-inexhaustible in nature should be overseen in a legitimate path so as to maintain a strategic distance from efficient misfortune. In today's digital world, there should be automatization in Indian forces for fleet management which will greatly help them in maintaining equipment sustenance. The idea is specifically employed for Indian forces by developing three main modules: GPS tracker, weight measurement, and fuel level indicator. We have built up this system to continuously monitor the status of a vehicle by using GPS tracker. Weight sensor is used to measure the weight of the vehicle through which we can get direct information in case of smuggling of weapons. While transmitting fuel to another medium, in order to prevent losses of fuel, we have used fuel level indicator to sense the basic level of medium. The motivation behind developing such a system is to prevent smuggling and losses of fuels and weapons in Indian forces. This solution enhances security as it is accessed remotely over the fleet by an authentic person by using the Internet of Things (IoT) model.

Keywords Fleet · Internet of Things (IoT) · NodeMCU · ESP8266 · GPS · Arduino · Weight sensor · Fuel level indicator · Load sensor

1 Introduction

Internet of Things (IoT), a standout among the most requested word in the field of information technology. IoT focuses to coordinate everything in our reality under a common system, deal with the things, yet it is additionally consistently keeping us refreshed with the current situation of those things. In the present computerized world, IoT is the main stage that spotlights on human comforts [1, 2]. Basically, India is a developing country, where few resources need to be saved such as fuel and

M. Sheth (✉)
Faculty of Engineering, Marwadi Education Foundation, Rajkot, India

P. Rupani
B.H. Gardi College of Engineering and Technology, Rajkot, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_55

electricity that play an important role in the rapid development of India. As India ascends financially and mechanically toward an increasingly prominent position in the locale and the world, it needs to simultaneously expand on its military power, in the advanced era, to impede the threats and difficulties that are likely to be faced along the way. Loss of fuels, weapons, and smuggling are the issues mainly faced by Indian armed forces [3].

The Indian armed forces were smuggled from across the border. In 2018, as per the news, several weapons and guns used by the Indian army and paramilitary forces were smuggled out from a truck that regularly carries garbage and other materials. In 2017, Sambhu Bhattacharjee, a junior works manager of Rifle Factory Ishapore (RFI), was arrested for allegedly smuggling out parts of SLR and INSAS rifles [3]. Keeping these issues in mind, to help Indian armed forces through IoT technology, we developed smart fleet monitoring system [1, 4].

This smart system developed three important modules, GPS tracker which helps in transferring weapons and goods in a secure way by tracking the live location of the particular vehicle. Using weight measurement sensor, we are measuring the weight of the vehicle; in case, if there is loss of fuel or smuggling happens, then we can get direct information. The vehicle fuel tank is fixed with fuel level indicator which gives the data of fuel level at a required moment [5]. These data are controlled in the Arduino which is associated with the sensors and the GPS modem. To track the location of the respective carrier, we need to integrate Node MCU, GPS module with Antenna. Fuel level sensor is fixed in the fuel tank, and it is connected with the GPS system. Special purpose weight sensor is installed with an analog input of tracking device which gives information about changes in the load of the vehicle and the goods [6–8].

2 Literature Review

In today’s fast life, one of the serious issues is the lack of modernization in Indian Armed forces.

This paper furnishes an answer utilizing GPS with Google Map. Despite the fact that this framework gives a speed, driver’s condition, exact situating of the vehicle fundamental the downside of this framework cannot identify the vehicle associated with a mishap or not and finding mishap area and suggesting to vehicle proprietors [1].

This paper built up a vehicle following framework to follow the accurate area of a moving or stationary vehicle progressively. The framework comprises equipment module just as programming module. The framework gives situating and navigational data as far as a number of parameters. Additionally, on and off chance that we need, at that point data with respect to the satellites which are being followed by the framework is likewise shown. The created framework is smaller, minimal effort and dependable based on different highlights systems. For example: fuel level, weight measurement, and so on [1, 8]. Up to this point, the exactness of [9] the fuel level estimation has not been vital. The reason is till now losses of fuel cases were not there in Indian armed forces. The existing system there is an analog fuel meter which indicates either fuel tank is empty, half, or full but we cannot detect the exact level of the fuel tank. Additionally, such a framework cannot shield us from getting cheat at petroleum siphons and these cost more for less measure of fuel so filled. So, it winds up important to grow such a framework which gives a precise (numeric) estimation of fuel in the fuel tank [9].

The major drawback of the existing system is not able to detect the threats facing by Indian armed forces. To overcome this issue, we make an effort to provide a technical solution using IoT at one platform only that addresses all the limitation of the existing system [9].

3 IoT Architecture

IoT is the innovation that manufactures frameworks able to do autonomously detecting and reacting to upgrades from this present reality without human intervention. We thusly need to build up a procedure stream for a positive structure over which an IoT solution is modeled [1, 8]. IoT as an advancement fundamentally contains four guideline parts (sensor, devices, gateway, and cloud). The application and presentation layer are comprised of web protocols [8, 10].

The IoT architecture includes four different phases:

Stage 1: Wireless sensors and actuators. The exceptional characteristic about sensors is their capacity to change over the data got in the external world into data for the survey.
Stage 2: sensor data. Gathering system and analog to digital data transformation. This stage connects with sensor network and fetches the data, whereas Internet gateway works via Wi-fi connection for further process.
Stage 3: Edge system. The edge IT systems perform preprocessing of the data before it continues ahead to the server homestead or cloud.
Stage 4: Cloud analytics. The data needs in-depth process, administered and set away on a standard back-end server (Fig. 1).

4 Related Work

The framework is partitioned into two noteworthy parts: programming and equipment structure. Equipment setup includes microchip, sensors, and actuators, through programming part encases programming that is composed and transmitted in one of the chip. The framework consists of microcontroller linking with sensors and electrical gadgets that are to be observed and controlled. The main purpose of this system is the maintenance of equipment, especially in Indian armed forces. The automation

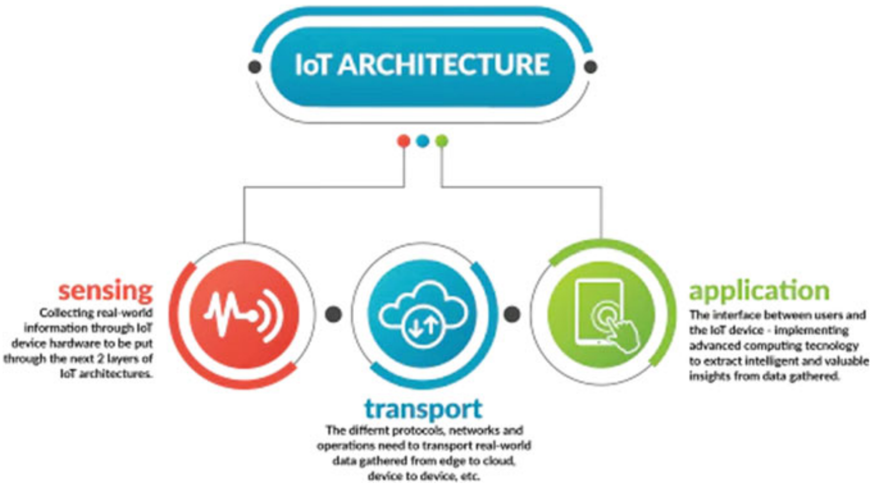


Fig. 1 IoT architecture [10]

designed not only fetching the sensor data but also actuating different parameters, such as fuel sensor will indicate the exact level of fuel present in the tank. The main target is to track a vehicle using a GPS system and also measures the load of a vehicle using a load sensor.

ESP8266 is a Wi-fi system-on-chip (SoC) loaded with an open source and integrated with a 32-bit microcontroller, channels, switches, etc. into a small package. NodeMCU is not a device but it is a firmware that loads onto one board. It provides built-in Wi-fi module and specifically used for IoT applications [6].

A load cell is a transducer that is utilized to change over power into an electrical sign. The most widely recognized utilization of this sensor is in weighing machine. Each machine which shows weight has a load cell as a detecting component [11].

DUT-E fuel level sensors are utilized for exact fuel level and volume estimation in tanks of vehicles and stationary units. This sensor permits to decide current fuel volume and change in volume (refueling or depleting) of fuel in the tank.

It is utilized in GPS tracking and vehicle telematics frameworks as extra fuel sensor or as a substitution of standard implicit (plant) fuel level sensor. At the point when utilized as a piece of telematics framework [12].

5 Working of Fleet Monitoring

The smart fleet monitoring technique comprises three separate sub-modules. In first module, we fetch the location of vehicle using GPS tracker, whereas second module is used to append different sensor such as DUT-E fuel level sensor and weight sensor

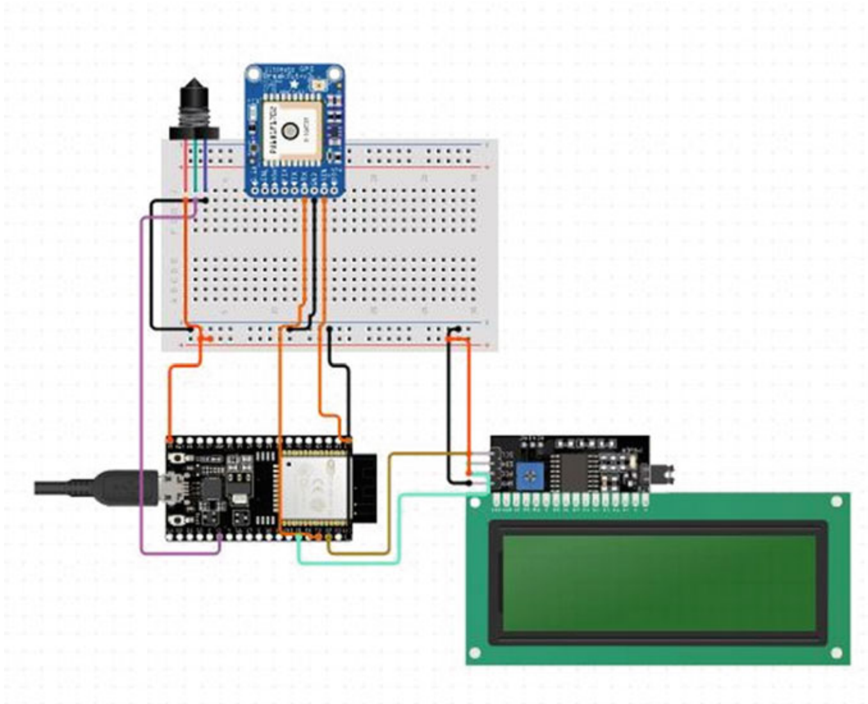


Fig. 2 Hardware setup of fleet monitoring

to quantify weight of vehicle and current level of fuel present in a fuel tank, and lastly, the result will be displayed in LCD screen (Fig. 2).

5.1 GPS Tracker Module

This module is used to track a vehicle. This module consists of a GPS antenna, Node MCU, and a GPS Receiver. The Node MCU is connected to the Internet via a mobile hotspot so that it can work in a range of the hotspot. The GPS Antenna, Node MCU and GPS the receiver is connected to each other. After successful connection, we get the location of a vehicle that is shown in Android or iOS application.

5.2 Load Management Module

This module is used to measure the load of a vehicle whenever a vehicle passes through a road, bridge, etc. This module consists of a load cell, Arduino, hx711 chip, and an LCD display. The load cell is connected with hx711 chip and that chip is connected with the Arduino. The Arduino is connected with the LCD display.

5.3 Fuel Monitoring Module

Fuel monitor sensor is connected with the fuel tank of the vehicle, and buzzer is also connected with the system to notify the person. When someone tries to change in quantity suddenly, automatically alarm starts and authorized person can see on screen as well on the mobile phone.

6 Results and Applications

6.1 Implementation

Implementation of fleet monitoring system is done using NodeMCU ESP8266 microcontroller chip via Wi-fi. For the user interface, we are using Android or iOS version Google Map application through which the user or the authenticated person can monitor the location of the vehicle. Weight measurement and fuel level indication are shown on LCD screen.

6.2 Results

Figure 3 shows the GPS antenna through which we are tracing the location of the vehicle through Google Map. In this technique, we are using a weight sensor and fuel level sensor that can monitor the weight of the vehicle and current fuel present in the tank. Live results clearly viewed in the 16 * 4 LCD screen.

Using DUT-E fuel level sensor, the following values clearly viewed on the LCD screen (Fig. 4):

- Current fuel level (in Liter)
- KMph (Per Hour)

LCD demonstrates the latitude and longitude estimations of the client location, the GPS antenna screens the scope and longitude estimations of the user location, and these fetched data will be sent to the controller (Fig. 5).

Fig. 3 Live project status (weight measurement)

Fig. 4 Live project status (current fuel level)

Fig. 5 Latitude and longitude results

7 Conclusion

The proposed system implemented using Arduino and NodeMCU by integrating varied sensors such as weight sensor, DUT-E fuel level sensor, and GPS tracker. The technique is primarily developed to detect and report regarding smuggling of goods and weapons or fuel theft in Indian armed forces and send an alert regarding theft including the current location of the vehicle. Smuggling and fuel loss are the most challenging issues faced by Indian armed forces. Thus, we used ESP8266 Wi-fi-based module for a better result in terms of utilization, support, and cost. This solution provides security as it is accessed remotely over the fleet by the authentic person only using the Internet of Things.

8 Future Work

This system can be enhanced with a live camera. Using a camera, we can capture the person trying for any unauthenticated activity. We can implement a face detection technique to identify him or her. This will help to improve the security of Indian armed forces.

Compliance with Ethical Standards All authors state that there is no conflict of interest. We used our own data. Humans and animals are not involved in this research work.

References

1. Sheth, M., Vagasia, M., Durani, H., Kotecha, S.: Smart Automated Home Application Using IoT with Blynk App. IEEE (2018)
2. Sheth, M.: Survey on Internet of Things. IJSERT (2018)
3. Saurabh, C.: Smart digital fuel indicator system. Int. Eng. Res. J. (IERJ) (2017)
4. Madakam, S., Ramaswamy, R., Tripathi, S.: Internet of Things (IoT): a literature review. J. Comput. Commun. (2015)
5. Feki, M.A., Kawsar, F., Boussard, M., Trappeniers, L.: The internet of things: the next technological revolution. Computer **46** (2013)
6. Esp8266, NodeMCU. http://www.nodemcu.com/. http://www.nodemcu.com/. [Çevrimiçi]
7. Baraka, K., Ghobril, M., Malek, S., Kanj, R., Kayssi, A.: Low cost Arduino/android-based energy-efficient home automation system with smart task scheduling. In: Fifth International Conference on Computational Intelligence (2013)
8. Rupani, P., Sheth, M.: Smart gardening automation using IoT with Blynk app. IEEE (2019)
9. Penna, M., et al.: Smart fleet monitoring system using Internet of Things (IoT). IEEE (2017)
10. Archicture, IoT. https://www.edureka.co/blog/what-is-iot/. https://www.edureka.co/blog/what-is-iot/. [Çevrimiçi] Edureka
11. Sensor, Load. https://in.omega.com/technical-learning/using-load-cells-to-weigh-trucks-trains-aircraft.html. https://in.omega.com/technical-learning/using-load-cells-to-weigh-trucks-trains-aircraft.html. [Çevrimiçi] Omega (2019)
12. Indicator, Fuel Leval. https://technoton.co.in/. https://technoton.co.in/. [Çevrimiçi] Technoton

Greenhouse Monitoring System Based on Internet of Things

Kantamneni Raviteja and M. Supriya

Abstract In the recent times, IoT is playing a major role in the development of agriculture. Due to the enormous growth of IoT, smart farming is becoming an emerging concept as IoT is capable enough to provide information about the agriculture fields. This work focuses on greenhouses which can be used in growing plants under certain circumstances. Such plant life can be monitored on a regular basis so as to increase the yield with high quality and quantity. This paper aims to provide farmers with an IoT-based Web application for monitoring the agriculture fields and its conditions. With the arrival of open supply Arduino Uno boards beside low-cost wet sensors, it is feasible to make devices that monitor various sensors like temperature/humidity sensor, soil moisture sensor, ultrasonic sensor, PIR sensor, pressure sensor, and light level sensor for consequently irrigating the fields when required.

Keywords IoT · Arduino · Monitoring system · Sensors · Android

1 Introduction

The government of India passed a Digital Republic Campaign to focus on many e-services like e-health, e-education, etc., out of which agriculture is one of the key focus. As per a survey, Food and Agriculture Organization of the United Nations has reported that the global food production should be raised to 70% in order to feed 9.6 billion people and this should be achieved by the end of the year 2050. The agriculture industry is facing challenges in the recent times, such as environment change, shortage of water, restrictions on fertilizing uses, limited availability of agricultural land, and many more [1, 2]. Hence, developing a smart agriculture with IoT technologies is a must to achieve the said goal.

With rising population, there is a desire for increased agricultural production. So as to support larger production in farms, the need of the amount of water utilized and the nutrients and protection of leaves from diseases at earlier stage becomes

K. Raviteja · M. Supriya (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_56

an important concern. Currently, agriculture accounts 83% of the whole water consumption in India, and with the current water scarcity, it becomes a question [3]. The development of monitoring system to monitor the water usage, nutrient need in plants and diseases in leaves is becoming a basic need in day-to-day life, and this is achievable with the use of the IoT technology. Our focus is on greenhouse monitoring system, as it plays a very vital role in the growth of Indian economy. A greenhouse is a building structure inside which plants are grown in a well-monitored environment [4]. Often, this monitoring happens by a worker who keeps visiting the greenhouses at regular intervals to monitor the growth of the plants. But, this system of implementation can be automated by the proposal of a greenhouse monitoring system, which keeps monitoring the temperature, water level, light levels, etc., and the same can be intimated to the user or the owner at regular intervals through GSM technology [5]. Such a system would enable any one to have greenhouses at their building rooftops, and the same can be taken up along with the other regular work.

As per [6], many farmers are shifting to other professions due to insufficient financial protection. Adopting to such greenhouse or hydroponic farming methods, one could earn more profit with less cost. National Horticulture Board of India also provides subsidy for greenhouses that is constructed considering various parameters provided by them.

2 Related Works

A system monitor could be a hardware or software that is generally used to monitor resources and performance in any business. Monitoring systems are accountable for controlling the technology used by the business and help in analyzing their operation and performance so as to detect and alert about possible errors.

Hwang and Yoe proposed a smart greenhouse system using the wireless sensor networks and claimed that the proposed system improves the productivity while maintaining the optimized growth environment [7]. A Zigbee-based monitoring system has been proposed in [8], which claims that the system is completely flexible and less costly when compared to the systems proposed earlier. However, the intimation is not sent through the short messages to the concerned person. Arduino-based monitoring system presented in [9] uses ATMega32 microcontroller designed with pulse-width modulation (PWM) and analog-to-digital converter (ADC). It also uses DHT, LDR, CO_2, and moisture sensors to monitor the greenhouses. A similar system is proposed in [10] and is used to monitor the temperature, humidity, gas, and water levels at the greenhouse. This system uses Freakduino board with Wi-Fi module and also sends the captured data using GSM [11].

Our focus is to provide farmers with an IoT-based Web application for monitoring the agriculture fields and its conditions. With the arrival of open supply Arduino boards, it is possible to make devices that uses sensors like soil moisture sensor, temperature sensor, humidity sensor, light intensity sensor, ultrasonic sensor, PIR sensor, and pressure sensors at the greenhouses to monitor on a regular basis and also as and when required. If the sensed data crosses a predefined threshold range, an alarm may be generated which will alert the resource person to his mobile [12].

3 System Architecture

The architecture diagram of IoT primarily based greenhouse monitoring system is given in Fig. 1. The system has three sensors that are hooked up to the controller, and sensed data from these detectors are sent to an android application. Farmers can start to utilize a variety of monitoring and managed system in order to increase the yield with the help of automation of greenhouse. Parameters like temperature, humidity, soil moisture, LDR, and ultrasonic sensors are monitored and managed in the device, which also can assist the farmers to improve the yield. This system affords uniform and required level of water for the farmhouse. Once, the moisture stage in the soil reaches below threshold value then gadget mechanically switches ON the motor. When the water degree reaches ordinary degree, the motor robotically switches OFF. The sensed parameters and the current status of the motor will be displayed on mobile device. The description of various sensors used is described in the next sub-section.

3.1 DHT11 Sensor

Temperature sensors generally produce standardized digital output and can be interfaced with any microcontrollers like Arduino, Raspberry Pi, etc. It provides high reliability and consists of three predominant components, namely a resistive-kind humidity sensor, an NTC thermistor, and an 8-bit microcontroller. The thermistor is used to measure the temperature, while the microcontroller converts the analog signal from each sensor to a digital signal and sends out the digital signal. The voltage

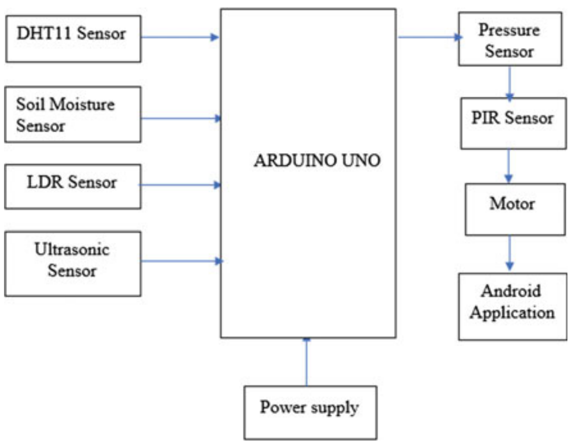


Fig. 1 Block diagram of IoT-based greenhouse monitoring system

for VCC pin generally ranges between 3.5 and 5.5 V. A 5 V supply might be used for high-quality output. The information from the statistics out pin is a serial virtual statistic [13].

3.2 Soil Moisture Sensor

The soil moisture sensor measures the water content in the soil. Once the soil is lacking water below the preset threshold, the module output of this sensor goes high while it is low by default. This tool prompts the user to water their plant life and in addition shows the moist content material of soil on the display screen. It has been loosely applied in agriculture, especially in land irrigation. Specifications of this sensor include, working voltage: 5 V, working current: 20 mA, interface type: analog, and working temperature: 10–30 °C.

3.3 Arduino Uno

Arduino Uno is the heart of this project that is connected to all or any sensors, and a different hardware elements' assembly is needed to realize the bound work. Arduino Uno is the microcontroller board which relays on the ATmega328P. It has six analog inputs, fourteen digital input/output pins among which six pins support PWM outputs, the in-circuit serial programming header, 16 MHz quartz crystal, associated USB attachment, an influence diver, and a reset button.

3.4 Ultrasonic Module HC-SR04

HC-SR04 module contains an ultrasonic transmitter, receiver, and a control circuit. It sends eight 40 kHz to detect the pulse signal back to calculate the test distance. Supersonic sensing elements generate high-frequency sound waves and appraise the echo that is received back by the sensor. It is used for motion or distance sensing with frequency. Supersonic module, HC-SR04, also provides 2–400 cm non-contact activity function with the move accuracy to 3 mm.

3.5 LDR Sensor

Light-dependent resistor (LDR) generally has a resistance and has the capability to change with the fall of light intensity, i.e., these devices depend on the light, when slight light falls on the LDR, the resistance decreases, while the resistance increases in the dark. An LDR has a resistance of 5000 Ω in daylight while it holds a resistance of 20,000,000 Ω in dark.

3.6 Pressure Sensor

These sensors are generally used to measure the fluid/gas flow, speed of the flow, water level in an area, or altitude of a place. It acts as a transducer and generates a signal as an outcome of the pressure forced on it.

3.7 PIR Sensor

A passive infrared (PIR) sensor has two slots made up of a sensitive material and is sensitive to IR. In an idle state, these slots detect same amount of IR, but when an object passes by, due to the warmth of the object, the slots undergo a positive and a negative differential change that causes the output of the sensor. This output can be used to identify the moving object in the field.

3.8 Motor Module

This work uses L293D motor driver module to drive the motor on both the directions with the help of H-bridge. The driver has 16-pins and can control two DC motors in any direction. It also uses a separate motor supply (VSS) to drive the motor, and the supply can vary between 5 and 36 V.

4 Results and Discussion

The hardware setup of the above architecture of greenhouse monitoring system is shown in Fig. 2. This setup has been tested on a plant in a small pot which is converted to a greenhouse. The sensed values of this setup are displayed on an android application which is developed using MIT app inventor. The snapshots of this application are stored on a cloud which can be visualized later for further comparison [14]. The snapshots of various sensed values are discussed in this section.

The graph in Fig. 3 shows temperature values displayed on the users' android application by using ThingSpeak. ThingSpeak is an IoT analytics platform service which allows to consolidate and analyze the data that are streamed in the cloud. This data can also be visualized later and can be used in further processing as can be seen in Fig. 3. This testing has been done between May 14 and 20, 2019, and the details of

Fig. 2 Prototype of greenhouse monitoring system

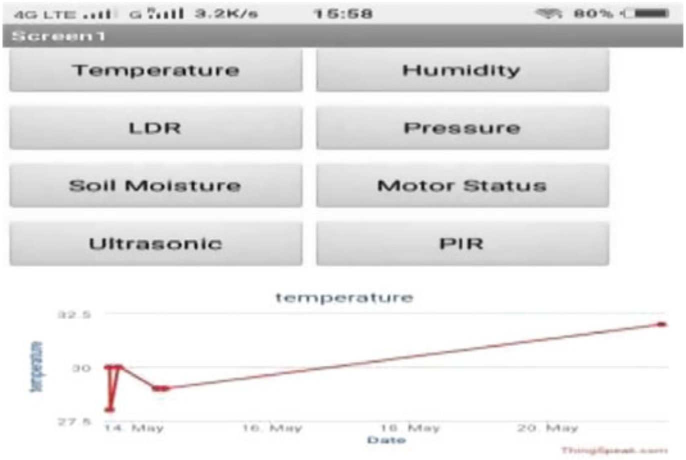


Fig. 3 Temperature values displayed on users' android application

the temperature from this mentioned period are visualized as shown in the figure (i.e., temperature ranges from 27.5 to 31 °C approximately). Based on the temperature values, the greenhouse temperature can be adjusted, so that the plants that are grown are healthy and do not suffer from varying temperature of the outside environment. This data combined with the humidity sensor and moisture sensor values can help to grow the crops healthy in such an environment.

The graph in Fig. 4 shows humidity values between the same mentioned period, and as can be seen, the values lie between 48 and 75%. This data can help the farmer to switch the motor ON and OFF accordingly. Similar to the above, the soil moisture, ultrasonic, BMP, PIR, and LDR values as captured by the proposed system between the mentioned period can be seen in Figs. 5, 6, 7, 8, and 9. The ultrasonic sensor measurements can be used to monitor the growth of the plant at regular intervals, and the result can be used to provide appropriate nutrients to the plants for a healthy growth [15]. The BMP values denote the pressure sensor values which can be used to monitor the pressure in the flow of water. The PIR values can be used to monitor the movement of any unwanted creatures like rats or snakes in the greenhouse. The LDR values help to adjust the light requirements inside the greenhouse for a healthier plant growth.

As per the survey, the ideal levels of the various sensors are as follows:

– Temperature—15–30 °C and not to exceed 35°, exactly 16–30° during day and 13–18° at night.
– Humidity—advisable levels of humidity depend on the temperature at the greenhouses. Temperature of 50 °F needs 83% humidity, 68° needs 91% while 86° needs

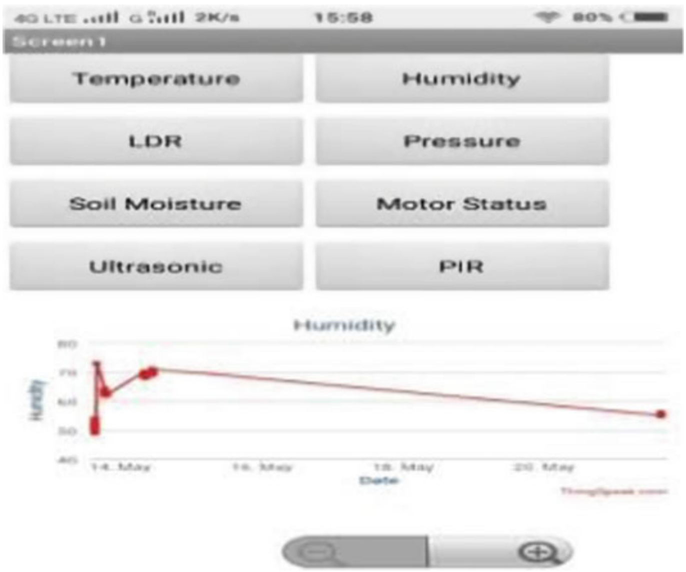


Fig. 4 Humidity values displayed on users' android application

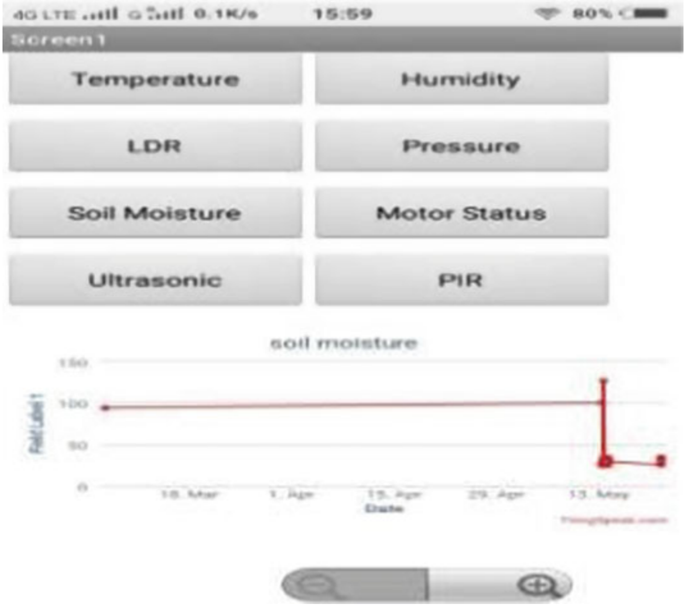


Fig. 5 Soil moisture values

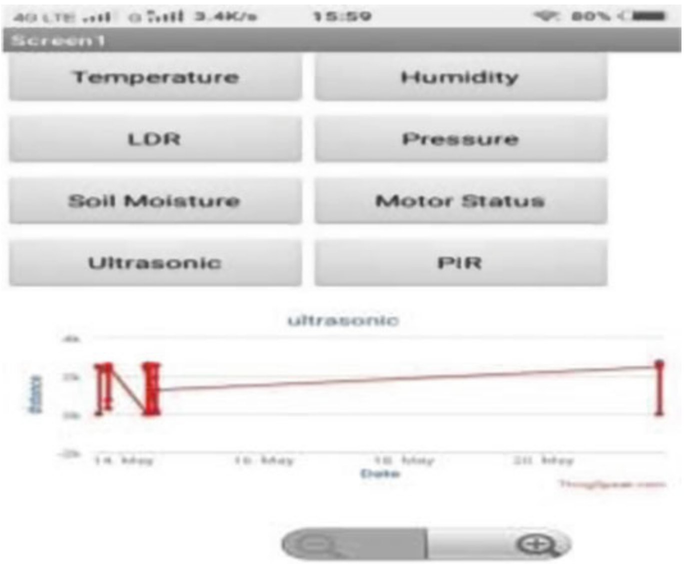


Fig. 6 Ultrasonic values

Fig. 7 BMP values

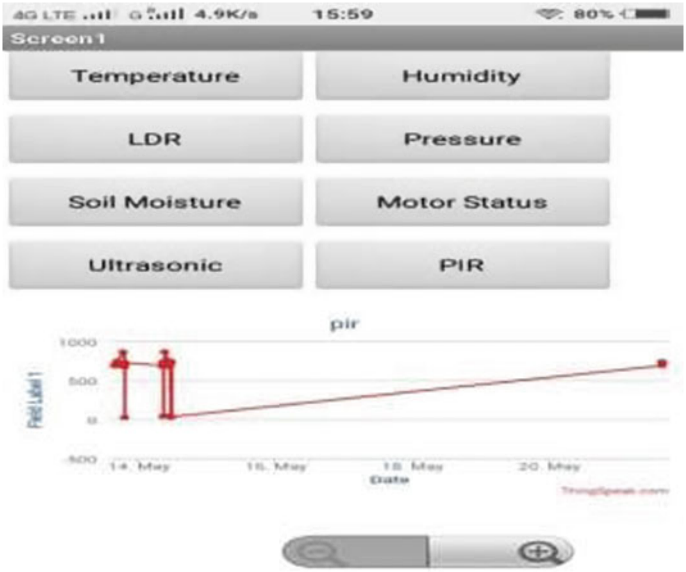


Fig. 8 PIR values

Fig. 9 LDR values

95% humidity levels. Setting humidity relative to the greenhouse temperature prevents the plants from diseases.

- Soil moisture—soil moisture varies based on the soil type and the plant, i.e., about 20% for a silty soil, 15% for a clay soil, and 5% for a sandy soil.
- Ultrasonic—this measures the distance between the sensor and the plant and can be used to monitor the growth. The values depend on the placement of the sensor.
- Pressure—when the pressure values go beyond a threshold, the predefined alarm can inform the farmer.
- PIR—this value is used to find the object movement in the greenhouse.
- LDR—these levels depend on the plants in the greenhouse, and the threshold can be set accordingly.

5 Conclusion

Plants require the proper environmental conditions for optimal growth and health. This work demonstrates the use of sensors that can help to monitor the various parameter levels in the greenhouse. A correct measure and mixture of temperature, humidity, and light levels can increase the production of fruits and vegetable yield in the greenhouse. With additional development of IoT growth in the forthcoming years, these systems may be additionally economical, plenty, quicker, and less costly. In the future, this system can be shaped as a smart system and can be fine-tuned to the need of the user.

References

1. Khanna, A., Kaur, S.: Evolution of internet of things (iot) and its significant impact in the field of precision agriculture. Comput. Electron. Agric. **157**, 218–231 (2019)
2. Kodali, R.K., Jain, V., Karagwal, S.: IoT based smart greenhouse. In: 2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), pp. 1–6, Dec 2016
3. Prabha, R., Sinitambirivoutin, E., Passelaigue, F., Ramesh, M.V.: Design and development of an IoT based smart irrigation and fertilization system for chilli farming. In: 2018 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp. 1–7, Mar 2018
4. Rangan, K., Vigneswaran, T.: An embedded systems approach to monitor green house. In: Recent advances in space technology services and climate change 2010 (RSTS CC-2010), pp. 61–65, Nov 2010
5. Kochar, P., Supriya, M.: Vehicle speed control using zigbee and gps. In: Communications in Computer and Information Science, CCIS, vol. 628, pp. 847–854 (2016)
6. https://agricultureguruji.com/greenhouse-farming/
7. Hwang, J., Yoe, H.: Design of wireless sensor network based smart greenhouse system. In: International Conference on Wireless Networks, pp. 43–48 (2016)
8. Ismail, M.T., Ismail, M.N., Sameon, S.S., Zin, Z.M., Mohd, N.: Wireless sensor network: smart greenhouse prototype with smart design. In: 2016 2nd International Symposium on Agent, Multi-Agent Systems and Robotics (ISAMSR), pp. 57–62, Aug 2016
9. Arif, K.I., Abbas, H.F.: Design and implementation a smart greenhouse. Int. J. Comput. Sci. Mobile Comput. **4**, 335–347 (2015)
10. Mekki, M., Abdallah, O., Amin, M.B.M., Eltayeb, M., Abdalfatah, T., Babiker, A.: Greenhouse monitoring and control system based on wireless sensor network. In: 2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE), pp. 384–387, Sept 2015
11. Pillai, P., Supriya, M.: Real time CO_2 monitoring and alert system based on wireless sensor networks. In: Intelligent Systems Technologies and Applications. Advances in Intelligent Systems and Computing, pp. 91–103, Aug 2015
12. Prasad, S., Peddoju, S.K., Ghosh, D.: Agromobile: a cloud-based framework for agriculturists on mobile platform. Int. J. Adv. Sci. Technol. **59**, 41–52 (2013)
13. http://ag.umass.edu/greenhouse-floriculture/fact-sheets/reducing-humidity-in-greenhouse
14. Gill, S.S., Chana, I., Buyya, R.: IoT based agriculture as a cloud and big data service: the beginning of digital india. J. Organ End User Comput **29**, 1–23 (2017)
15. Akhil, R., Gokul, M.S., Menon, S., Nair, L.S.: Automated soil nutrient monitoring for improved agriculture. In: 2018 International Conference on Communication and Signal Processing (ICCSP), pp. 0688–0692, Apr 2018

Building Personal Marionette (Ritchie) Using Internet of Things for Smarter Living in Homes

Rajkumar Rajasekaran, Ranjan Goyal and Voleti Guru Venkata Mahesh

Abstract The advent of the concepts of Internet of Things has revolutionized and impacted not only the information technology sector but also finance, manufacturing, and the home. The penetration of wireless communication technologies based on mobile networks, bluetooth, and radio-based communication such as Wi-fi, RFID, and NFC has significantly contributed to widespread adoption and acceptance. This research tests the validity of an IoT enabled "marionette" which is armed with an array of sensors, and actuators to achieve easy communications between the user and a home automation system. The system is an integration of a number of IoT modules that provides an automation for several tasks including home utilities, personal/banking information dissemination mechanism, and security. The analysis of the system based on the comparison with traditional one and the power consumption evaluation depicted the need to implement this system for smarter, safer, and easy living in homes.

Keywords Bluetooth · Household mechanism · Information system · Internet of Things · Marionette · Ritchie

1 Introduction

The Internet of Things (IoT) is a network of devices that are connected together for some exchange of data or interact with each other in order to achieve a particular task. There are different types of devices that can be counted in the IoT network such as sensors, actuators, and software. Here, in this paper, the objective is to achieve the task of home automation with the help of an IoT-enabled device that is capable of automating several processes with the help of sensors and actuators. So, in this paper, an IoT-enabled system named "marionette" is proposed which is a personal doll that

R. Rajasekaran (✉) · R. Goyal · V. G. V. Mahesh
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

V. G. V. Mahesh
e-mail: guru.venkatamahesh2017@vitstudent.ac.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_57

can be kept into our homes to perform all the defined tasks in order to provide smarter living in homes.

Keeping in mind the IoT-centric theme of the research, it is understood that the marionette doll must be able to communicate with an array of sensors and possible actuation systems wirelessly. Moreover, since the design of the puppet is similar to a doll, it must be safe to be held or operated by children. Furthermore, it cannot contain any intrinsic moving parts as this would significantly reduce lifetime. Finally, since wireless communication that transcends line of sight is a vital component of this device, there should be minimal internal electromagnetic interference and the body of the doll must be made of and contain suitable non-metallic and non-conductive material to prevent "Faraday Cage" effects. Several benefits of IoT contain smart towns, smart house, smart transport and movement, smart homes and smart groundwork, Shin [1]. Beholden to the above-stated restrictions, the goal of this project is to harness the profits of IoT in smart homes by allowing bank account holders to monitor their finances securely from the comfort of their homes without the persistent difficulties of web and mobile access banking. Various banks have been inadequate in experimenting with different technologies because they appear infeasible for them commercially, though IoT has brought about different changes over the use of low cost yet influential and ground-breaking module.

Therefore, in this paper a personal marionette system is presented that automates the tasks with the help of IoT modules leading to home automation thus providing a smarter living in homes. The rest of the paper is formulated as follows: Section 57.2.1 provides information on the proposed marionette system and its components such as RFID. Section 57.2.2 provides the system architecture of the system and the overview of personal marionette. Section 57.3 describes the home automation provided with the help of this system. Section 57.3.1 describes the structural design of the marionette system and the functional description of the sensors integrated in the system. Section 57.3.2 analyzes the system based on the components in the traditional versus new system and the consumption of power followed by conclusion and future work in Sect. 57.4.

2 Background and Key Issues

2.1 Marionette System

The personal marionette system here, also named as Ritchie, is capable of performing several tasks automatically with the help of an array of sensors and actuators. The marionette system has the following communication set:

- Home utilities systems,
- Banking and finance portal,
- User interface and user experience,
- Warning mechanisms

The planned system Ritchie is a smart inside marionette which uses a variety of wireless sensing knowledge with small and low-priced RFID Schwiegelshohn [2]. Radio-frequency identification (RFID) an important technologies used in the IoT as it can store sensitive data. Automatic identification and data capture (AIDC) methods of automatically identifying objects, collecting data without human involvement. Turcu [3]. Stanford [4] composed in his critique the power of using RFID technology and how it would modification communication as well as our day-to-day process forever. The article plans some details as to why RFID technology remained to be considered an inclined changer. The primary system of user interaction would be via NFC or "Near Field Communication." NFC is a highly specialized subsect of RFID which maintains the speed and convenience of an RFID system while providing an additional benefit of "Security by Proximity." Most phones today now launch equipped with an NRC package. This means that a user's personal information, such as authentication information for banking remains safe from unauthorised, involuntary reading either by the marionette or a third person attack. Moreover, the RFID tags can read multiple codes that are present close to the reader simultaneously which is helpful in retrieving large amounts of data very quickly.

Turcu [3] suggests using radio frequency ID as a form of documentation in an IoT organization as they are comparatively low in cost therefore subsequent in cost-efficient systems. The paper clarifies how the RFID is an enabler in automation through the usage of Automatic credentials and information capture (AIDC) technology that allows various entities to be exclusively identified in the RFID. All RFIDs are branded with the subsequent components, contactless electronic tags (which store exclusive identification data), an RFID reader (to read and write these tags), and submission components used as handling elements. This implies that the RFID system can uniquely identify and read data from a large number of sources wirelessly with a very high rate of success. Lee et al. [5] define a care system as one that safeguards the house beside interruption and uses motion sensors, smoke detectors, investigation cameras, etc. With the use of devices and the microcontrollers, it is probable to implement a safety system that is of less cost and effective because it will be straight connected to the household owner's smartphone done a security application. Many other apparatuses can be built upon the system like reimbursing of utility bills like electricity. There is scope for many sensors and actuators on marionette which can be consumed to the greatest improvement.

2.2 System Architecture

The sections, in our paper, include the banking system anywhere the user checks their bank balance and details; the functions feature where the users monitor. Electricity system spending their bills over months, household privacy system, and early warning systems. Seetharaman [6] proposes that the main motives for which clients have agreed on Internet banking are for their convenience, accessibility, accuracy, and security it gives them. This is true considering how competent with technology people have become. Mishra and Singh [7] conducted a research on how customers perceive the move from traditional banking to Internet banking from the viewpoint of the bank managers. They outlined the factors that need to be considered in order for the adoption of Internet banking to be successful include computer self-efficacy, perceived risk, perceived ease of use, and perceived usefulness. Rui-jin et al. [8] state that Internet banking faces the challenges of low penetration rate in the market, low acceptance, and wasting of company resources while trying to make customers adopt. However, the aforementioned challenges can be mitigated if we provide a way for clients to conduct their banking activities from the comfort of their homes by monitoring their spending, monitoring their utility usage costs, and warning the client of excessive spendings or unnatural changes in usage.

The prototype has four modules which have several advantages. As a bank balance monitor, the system gives the customer convenience in accessing their bank balance without the complexity often brought about by Internet banking. Another advantage of the bank balance monitor is that it is a home assistance device which requires no training to understand. It will help speed up the tasks by giving a simplified bank statement through the use of a GSM module onto their phones or preexisting visual output devices such as monitors or their TV. It can also be used to monitor the electricity balance over a period of time and give an analysis of the spending. When used as an alarm system, it can provide real-time protection to the house through its intrusion detection component which will send warnings of irregularities in standard operations. The Wi-fi pairing with a home automation system provides users with a convenient, albeit indirect method of interfacing and automating with their general electronics to both statistically monitor and alter the states of a majority of their devices truly creating a home of the future

The system is made of a combination of components such as an RFID reader which represents the data acquisition layer where the basic and raw data is collected using RFID reader, sensors, and actuators and the data will be displayed using the LCD or Arduino's serial monitor. The microcontroller and motion sensors represent the data processing layer (Fig. 1).

In this layer, the data acquired from the acquisition layer will be processed using RFID data capture algorithms supported by the Arduino IDE. The stages involved are scanning the tag, verifying tag, granting access, and then updating the card with the new information from the database. An Ethernet cable and Ethernet shield for Arduino are used to provide Ethernet connection for the Internet in the system. The data is sent to and from the database which represents the quality of service layer which ensures that integrity of the user data is maintained, and security is ensured through appropriate identification and authentication of the user. The system will use an LCD display and buzzer which represent the data services layer which shows the desired output from the data processing and in this case if the user has exceeded their monthly spending limit, the system will sound a buzzer and/or give a system notification through the bank monitor component. Ritchie marionette will use motion sensors and actuators; the motion sensors will detect any movement in the house and based on specific recognition algorithms Ritchie will determine whether there is an intruder or not, and react accordingly. Additionally, a camera will be used as a live

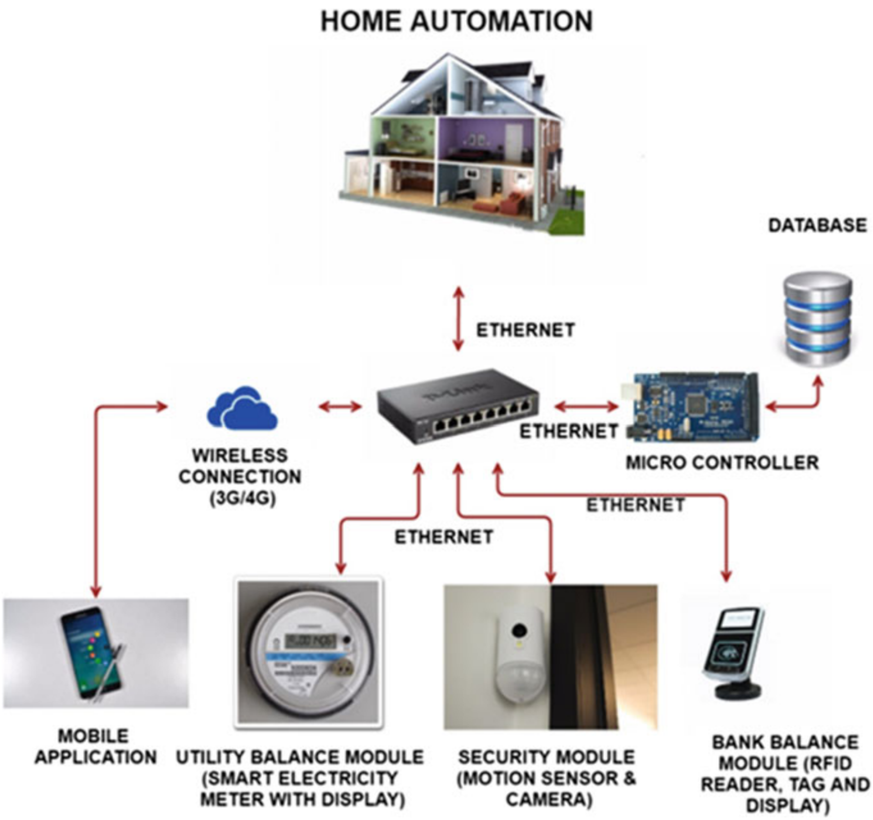


Fig. 1 Architecture of the system

monitoring of the home and also for authentication of users using facial recognition algorithms. Sound sensors will be used to listen and monitor the environment for unusual noise disturbances when the security system is active. It works in conjunction with the other sensors and if the noise persists the sensors will record the sound for a specified interval for later review. Camera will be used for authentication of the user through facial recognition algorithms, video monitoring, and image capture; when the marionette senses intrusion, the camera can send a live feed of the video to the homeowner via a phone application and also take images. RFID reader will be used to detect any RFID-enabled tags and their location in case of lost items, and also for the bank balance monitors which will use the camera to authenticate users. GSM component will be used to record voice messages when the user is not around to take the call (Fig. 2).

The reader will also be used for further home automation for other things. Motion and gesture sensors will be used to detect any movement in the house at specific time intervals and aid the security component of the system to work effectively. Buzzer

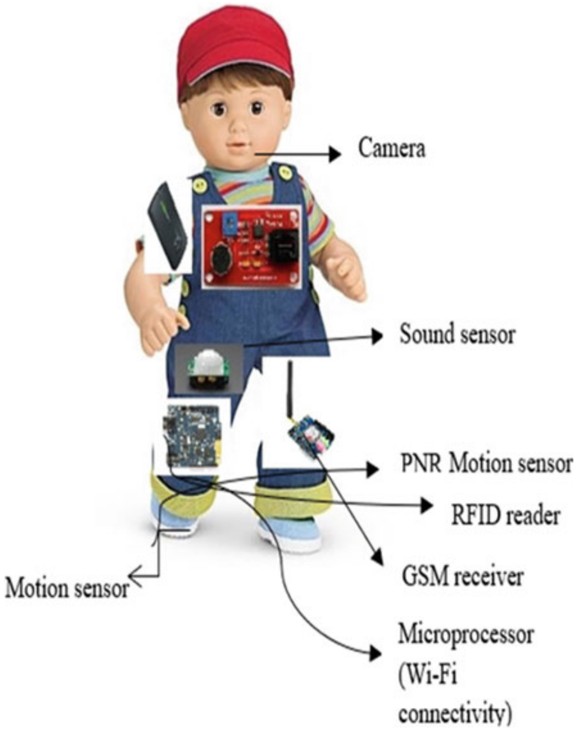


Fig. 2 Personal marionette (Ritchie) [9]

will produce a sound which represents system notifications. An LCD screen can also be added to the marionette.

3 Home Automation

In the home automation system, an array of sensors is placed at strategic or risk positions to monitor light intensity, temperature changes, fire, gas leaks, or physical intrusions. These sensors are connected to a central microcontroller dedicated to monitoring and risk assessment. The data gathered by this microcontroller is sent to the marionette via a 2.4 GHz Wi-fi channel and further to the user from the marionette by a GSM channel. The user can also control the household appliances and electronics via the marionette since all electrical devices would be connected to the aforementioned microcontroller. Table 1 provides the description of the sensors of this system.

Table 1 Sensors description

Sensor	Description
Motion and gesture sensors	Sense any movement of a human being and any gestures done within the radar of the marionette
Sound sensors	Record and filter sounds made in the environment
Actuators	Start the microprocessor
RFID reader	Scan RFID tag and retrieve data from the database
GSM receiver	To send the alert messages
Microcontroller	To control the sensors and to actuate the sensors

3.1 Structural Design of the System

The structure is made of a combination of modules such an Intel Galileo Arduino processor having 32-bit architecture and large number of digital read pins which characterize the communication layer where the simple and raw information is collected using RFID readers, Wi-fi, NFC, and GSM modules, and the information will be processed using the Arduino. The home automation layer contains another secondary microcontroller along with an array of sensors which make up the data acquisition layer. This layer is further complemented by an electricity meter monitor and Wi-fi connection to the bank's authentication webpage. The I/O layer contains a range of actuators that take commands from our secondary microcontroller. Moreover, the GSM shield for our marionette as well as the NFC module that holds the user's banking authentication signature form the primary inputs for our system. We would also receive inputs from other members of the communication layer however; since their primary sensors do not form a part of marionette, they have been excluded from this research. Table 2 provides the functional description of the sensors, actuators, and modules of this system.

3.2 Predictable Results and Analysis

The usage of this system is designed to be intentionally trivial requiring no training or prerequisite knowledge. A user can remotely monitor the conditions of their home from their mobile devices and also take necessary steps to save money and electricity. They can be warned about threats to their home with intrusion detection and hazard sensors. While in proximity to the marionette doll, they can use their NFC tags, often directly on their phones to authenticate the banking details and get that data directly on their phones. Table 3 provides the comparison of the components between traditional and new systems.

To further illustrate how the system will work, we have taken the electricity module and listed the common appliances found in most suburbs. The units consumed for a whole year are shown in Table 4. After getting sample values, there is now

Table 2 Sensors functional description

Sensor/actuator/module	Function
Light dependant resistor, thermometer	Senses ambient light and room temperature and responds with a Boolean value (after threshold conversion) of the lighting and heating requirements
Fire alarm, gas detector, motion sensors	Look for values that fall beyond the range of normal operating conditions and give alerts
Electronic switches	Control household electronics on instruction from user, or by automatic thresholds or timings on cue from the microcontroller
GSM	Primary mode of communication from the user by an exchange of SMS messages
Wi-fi	A significant component of the marionette, Wi-Fi system is responsible for providing communication channels with the user, the secondary microcontroller, the electricity monitors as well as the banking authentication servers
NFC	A proximity triggered sensor that holds the authentication information of the user and sends the banking and utility details on activation

Table 3 Traditional system versus new system

C omponent	Traditional system	New system
Bank details observer	Use Online banking done a mobile phone or computer browser or call the bank for your bank statement. Online banking can be multipart for the less educated people	Requires no user training. The security of a pre-programmed card is user's only responsibility. Easy to use as only requires proximity contact
Electricity bill observer	Hard to keep an analysis of your usage	System gives an investigative report and analysis along with prior warnings of overuse
Security system	Can be turned off automatically and it is not linked to the Internet and mobile presentation	Homeowner can get immediate notifications of any action through the mobile application
Home automation	Exists in isolation from utility and hazard warning	Unified into a single convenient package along with an array of their utilities and uses

a baseline of what units are expected to be consumed by a particular household each month or year. The Ritchie marionette can be set to sound a buzzer when the units consumed have exceeded the normal range and also help to identify which appliance is consuming more electricity. Table 4 provides the power consumption of the respective appliances.

Table 4 Appliances power consumption

Appliance	kWh/year	kWh/day	kW/h
Chest fridge	91–109	0.25–0.3	
New upright fridge/freezer	409	1.12	
Old upright fridge/freezer	500	1.36	
Pool pump (1.5 kW, single speed, average 3 h/day)	1314	3.6	1.2
Pool pump (1.5 kW, variable speed, eco mode, average 5 h/day)	548	1.5	0.3
Waste water treatment aerator	255	0.7	0.065

4 Conclusion and Future Work

This segment deals with the challenges faced by such a device system and the scope to reduce them, as well as expand on its applications. The primary concern is security. An RFID device is easy to unlawfully read since it poses no preventive methods. These can be easily mitigated by the use of Faraday cages as holders for the card. Secondly, banks would like to ensure the security of their servers. Finally, there is a need to build confidence in the abilities of this device. Yet, the scope of this device seems limitless. Its architecture and reference systems allow it to connect to a nearly limitless array of sensors and can be expanded as per individual requirements. The device is easy and safe to use and highly interactive. Based on the analysis of the system, it can be seen, it provides a safe, smart, and easy living in the homes. Building on this base, we can commence with the automation of the doll itself, providing a small personal assistant to our daily needs.

References

1. Shin, D.: A socio-technical framework for Internet-of-Things design: a human-centered design for the Internet of Things. Telemat. Inf. **31**(4), 519–531 (2014)
2. Schwiegelshohn, F., Wehner, P., Rettkowski, J., Gohringer, D., Hubner, M., Keramidas, G., Antonopoulos, C., Voros, N.S.: A holistic approach for advancing robots in ambient assisted living environments. In: 2015 IEEE 13th International Conference on Embedded and Ubiquitous Computing (EUC), pp. 140–147. IEEE (2015)
3. Turcu, Cornel. "The social internet of things and the RFID-based robots." In Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2012 4th International Congress on, pp. 77–83. IEEE, 2012
4. Stanford, V.: Pervasive computing goes the last hundred feet with RFID systems. IEEE Pervasive Comput. **2**(2), 9–14 (2003)
5. Lee, C., Zappaterra, L., Choi, K., Choi, H.A.: Securing smart home: technologies, security challenges, and security requirements. In: 2014 IEEE Conference on Communications and Network Security (CNS), pp. 67–72. IEEE (2014)
6. Seetharaman, A., Singhal, S., Galdhar, P., Rudolph Raj, J., Saravanan, A.S.: Customers' expectations for next generation internet banking. J. Inf. Knowl. Manag. 1650009 (2016)

7. Mishra, V., Singh, V.: Analyzing the gap in the adoption of Internet banking services: managers' perspective. In: 2014 2nd International Conference on Business and Information Management (ICBIM), pp. 42–46. IEEE (2014)
8. Rui-jin, Z., Guo-xin, L., Ze-zhou, S.: Relationship between consumer innovativeness and internet banking acceptance. In: 2014 International Conference on Management Science & Engineering (ICMSE), pp. 308–314. IEEE (2014)
9. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Fut. Gener. Comput. Syst. **29**(7), 1645–1660 (2013)

The Internet of Things (IoT) Routing Security—A Study

M. Durairaj and J. Hirudhaya Mary Asha

Abstract Internet of things (IoT) is the finest metric following technology that turns the attention of the people throughout the world. IoT is a global connecting network that allows people to connect with each other largely. It is a challenging technology due to its complex environment and resource-dependent features. Different surveys depict that there may be several billions of IoT users by 2020. There are numerous companies that offer IoT services. Nowadays, the security of an IoT feature is an issue which is non-measurable in nature. Different research works are being performed to find the optimal solution to provide security of IoT. IoT requires less human commanding and controls which results in vulnerability due to hacking. In this paper, the literature is reviewed based on the types of routing attacks in wireless sensor network interface layer communication and classified the attacks that disturb the communication. These attacks are classified as attacks on topology, resources, and traffic. Based on the classification of attacks, countermeasures are suggested to protect the routing standard for the IoT environment.

Keywords Internet of things · Wireless sensor network · Security attacks · Routing protocols · Resource attacks · Traffic attacks

1 Introduction

Because of the increasing usage of wireless sensor network (WSN) technology and the Internet of things (IoT) application, more sensors and wireless communication technologies are used. The IoT applications are used in military, people's livelihood, industry, education, commerce, and other fields. The IoT devices can be remotely

M. Durairaj · J. Hirudhaya Mary Asha (✉)
Assistant Professor, Department of Computer Science, Bharathidasan University, Tiruchirapalli, Tamilnadu, India
e-mail: hirudhaya_ashaa20@yahoo.co.in

J. Hirudhaya Mary Asha
Research Scholar, Department of Computer Science, Bharathidasan University, Tiruchirapalli, Tamilnadu, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_58

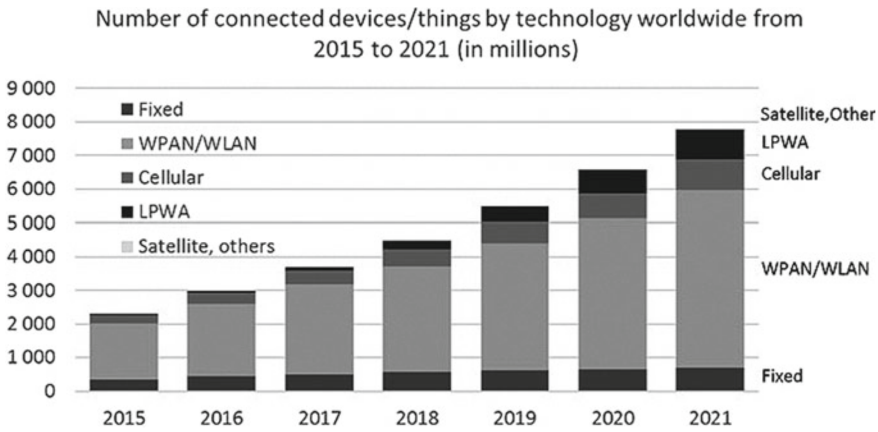


Fig. 1 Number of devices connected

controlled from anywhere to control the home or processes carried out at office. The devices may be wearable accessories or may be larger machines. Each IoT device contains chip like sensor embedded within it. The problems of information security field such as virus attack, data damage, hacker intrusion, malicious code attacks, replication of data attacks, and certain other issues will be very serious for the damage and data losses in the network. The IoT devices followed some standard protocols for sharing information among the devices which are connected in the network. The architecture of the IoT network is constructed based on the communication. Due to some vulnerabilities, the information stored in the network faces great challenges. The defense mechanisms such as encryption and authentication are traditional sources available for ensuring security. Intrusion detection is a new security-enhancing technology which is an active research in the field of network security. Detecting and promoting countermeasure for tackling the drawback of existing traditional security defense technology. The bar chart (see in Fig. 1) shows the number of devices connected by technology among the year 2015–2021 [1].

The research proposed had included the location information in the sensor nodes, which validates the location identity information.

2 Related Works

The connection of the Internet of things tends to increase the income of the society [2]. It is a technology-based economic improvement. The Internet of things (IoT) is called distributed and decentralized systems since this technology built things based on the place which is actually located and service established at the destination network [3]. The correlation between different entities achieves a common goal [3]. There are various attacks which highly affects the nodes in the connected network.

The power consumption levels of the connecting devices are important to distinguish the signals of the distributed devices connected to the gateway [1, 4].

An authentication authorization accounting (AAA) communication system [5] is designed to find how much bandwidth consumed by device during communication throughput. It is difficult to provide data access rights' level to different people based on the bondage of commitment in the privacy policy. Thus, authentication checking is forced by introducing a new framework for reducing the insider attack in the industrial Internet of things (IIoT) [6]. If one or more clients request to modify the same data/file, it must be discovered previously else the data/file would prone to attacks [7, 8]. To avoid critical issues in the connecting devices, the scheduling of task must be done better [9–11]. Security challenges in IoT are as follows.

2.1 Authentication and Integrity

Integrity promises user rights to access or modify data. The data must be accurate and consistent throughout its routing life cycle. Authentication provides the customer an individual or personal secret key as an opener to access the IoT network [5]. Nowadays, various algorithms are followed to make the sensors to respond at real time.

2.2 Access Control for Ensuring Security

Authorization allows the nodes to access different resources based on their access rights. Access control mechanisms check and guarantee the rights of an individual [7]. For verifying access, every IoT node follows some mechanisms different from other nodes for connected in the same network [8, 9]. After the authentication checking, the needed data must be available at the right time in the network routing. Redundancy is the appropriate method to ensure the availability of data [11, 12].

2.3 Privacy and Confidentiality

The private information of the people like health data is recorded by the integration of heterogeneous objects. Intruding is a threat to people's privacy. In the IoT world, the nodes connecting in the network can collect one own's data without their knowledge [10]. In the communicating network, the transmitted data must be accessed and read only by the authorized individuals or parties.

3 IoT Routing Attacks

The IoT devices connected to the network route are prone to various attacks. The attack happens highly which degrades the network performance. Based on the attacking features, the major attacks and the way it affects the routing is described [13–32] and tabulated (see Table 1). A survey was taken on various attacks which takes place in the routing. The types of attacks occurring in the network layer of IoT communication are as follows:

- Sinkhole attack
- Sybil attack
- Clone ID attack
- Selective forwarding attack
- Hello flooding attack
- Local repair attack
- Wormhole attack
- Denial-of-service attack
- Blackhole attack
- Rank attack
- Neighbor attack
- Version number attack.

3.1 Grouping of Attacks

IoT infrastructure or network layer attacks are grouped based on their impact on the network routing, such as attacks against resources, attacks against topology, and attacks against traffic.

Attacks against resources.

In the wireless network routing, the attacker node attempts to destroy the resources that blocks the routing link of this network and also the lifetime availability of the network is reduced. This resource attack will waste the node energy and control overhead. Attacks occur on resources in two ways.

(1) **Direct attack**—attacker node generates overload directly.

   - Flooding attacks
   - Routing table overload attacks in storing mode.

(2) **Indirect attack**—attacker node generates large traffic indirectly. This attack disrupts the route with multiple identities. Some indirect attacks are

   - Increased rank attacks
   - Version number attacks
   - Sybil attack.

Table 1 Techniques available for IoT routing attacks

Techniques for overcoming attacks	Authors	Description of techniques
IDS [33]	Linus Wahlgren et al.	IDS: Intrusive IDS systems capture network traffic and perform analysis by comparing the captured data with the intrusion detection definitions stored by the user
Version number and rank authentication (VeRA in RPL) based solution [28]	Kelpen et al.	Initialization and version number update are the two steps recommended in VeRA. VeRA supports cryptographic security RSA; in this, the nodes use Pk as a public key and Sk as a private key known only to the DODAG root
Trail [34]	Heiner Perrey et al.	Trail is recommended for authentication of the topology. The base node of the tree acts as a true sender where each node linked with base node in tree topology. Trail uses cryptography tests for identifying the topology attacker
Heartbeat protocol [33]	Linus Wallgren et al.	In heartbeat protocol, Internet Control Message Protocol version 6 (ICMPv6) request is a message sent by a node from border router to discover neighbor
Parent fail-over [32]	SukhwinderSharma et al.	The DIO message sent by the node contains the parent node hash value, unheard node set, and the number of hops in the destination path
RPL's global and local repair mechanism [23]	Gill et al.	RPL protocol is an error handling mechanism used for the resource-constrained environment
Rank authentication [32]	Sukhwinder Sharma et al.	According to this method, rank value must be incremented for every hop. For calculating rank value, a hash function is $xr+1 = h(xr)$ function is used

(continued)

Table 1 (continued)

Techniques for overcoming attacks	Authors	Description of techniques
Merkle tree authentication [35]	Li et al.	Merkle tree checks security with cryptographic tests. Each node in the network is accepted with its authentication path information (API)
Svelte [34]	Raza et al.	Svelte is an RPL-based IDS for IoT. It has three modules (1) LoWPAN mapper which traces the RPL network using border router and connects LoWPAN networks to other RPL networks (2) Intrusion detection metric finds the intrusion (3) Firewall for filtering traffic

Attacks on Topology.

The attacker node attempts to intercept the network route, modify the data, and forward the data to the base station [18]. Some topology attacks are

- Routing table falsification attacks
- Sinkhole attacks
- Wormhole attacks
- Routing information replay attacks
- Worst parent attacks
- Blackhole attacks.

Attacks on Traffic.

The traffic attacks consume the resources of the network so that the important requests will not be notified due to the lack of resources for either routing or process the work, such as bandwidth and receiving buffer at the server end [19]. Some traffic attacks are

- Sniffing attacks
- Misappropriation attacks
- Decreased rank attacks
- Identity attacks.

4 Proposed Work Discussing the Attacks and Techniques Which Addresses the IoT Attacks

IoT is an environment where numerous devices communicate with the help of communication technology. After reviewing literature here, a collection of techniques is proposed for addressing the routing attacks. Protocols like RPL's global and local repair mechanism, IDS, heartbeat protocol, parent fail-over, trail, rank authentication techniques, Merkle tree authentication, a Svelte technique for handling network layer routing attacks.

Different IoT routing attacks and the drawbacks of that routing attacks and the overcoming techniques are discussed, and their possible outcomes are also proposed here in (Table 2).

5 Future Research Direction

There is a high demand for structuring an applicable solution to secure essential routing metrics. The Internet of things protocol against attacks like sinkhole attacks, blackhole attacks, selective forwarding attacks and grayhole attacks, version number manipulation attacks should be sorted out and concerned during every IoT evaluation criteria. It is meaningful and useful in investing time to develop various threats overcoming models that faces the routing standard of IoT and its application areas. The standards designed for Internet of things routing do not specify how the node authentication and secured connection mechanisms could be designed. With the increasing feature and usage of IoT projects worldwide, there is a perceptual need for designing secure routing protocols. To design, develop, and implement a secure protocol, which is said to be the undying security goals, all the above-mentioned criteria should be considered.

6 Conclusion

The Internet of things (IoT) focuses more on deploying low power and lossy networks in order to run highly critical communication among the devices and their interconnection to the wireless technology. In this paper, a classification of the attacks is presented and protocol types are also listed. The attacks on resources decrease the network lifetime by generating fake control messages or by building loops of messages. The attacks on topology make the coverage of network to a sub-optimal configuration or node isolation. Finally, attacks on network traffic make a good node to be captured by the malicious node and destroy the larger part of the traffic. From the review of the literatures, comes to know that IDS based system finds the rank attacks and VeRa attack, where TRAIL techniques prevents the Rank attack by minimizing

Table 2 Routing attacks and its survey

Attack types	Drawbacks of attacks	Techniques for the counter-attack	Possible outcome
Rank attack [20] Hashem et al.	• Degrades network performance generates non-optimal loop • Packet delay	IDS [33], VeRA-based solution, trail [34]	Network overhead time is minimized
Selective forwarding [21] Airehrour et al.	• Disturbs routing path and intrudes any protocol	Heartbeat protocol [33]	Creating a dynamic path between parent and children
Sinkhole [22] Tiwari et al.	• Altering the ranking field of the destination information message (DIM)	Parent fail-over method [22, 32], rank authentication technique [22]	The density of the network is increased
Hello flooding [23] Gill et al.	• Creates legitimate traffic	RPL's global repair mechanism	Minimizes communication overhead
Worm hole [24] Ghugar et al.	• Due to this attack, network route topology is disrupted and creates a traffic flow	Merkle authentication [35]	Provides resiliency
Sybil and Clone ID [25] Thomas et al.	• The legitimate nodes are denied access to reserved resources	RPL local repair[36–38]	Reduces traffic and makes resources available
Denial of service [26] Yujie et al.	• Nodes request more resources with fake identity and creates traffic	IDS solution [33]	Reduces traffic
Black hole [27] Singh et al.	• Publish wrong fake route as a trusted route • Dropping the packets instead of forwarding	Svelte [34], parent fail-over [32, 39]	Overhead is small enough with limited energy and memory capacity
Version number [28, 29] Kelpen et al.	• Increases the control overhead • Low packet delivery that drains battery energy	VeRA [28, 35]	Prevents node from increasing the version number and decreased rank values illegitimately
Local repair control overhead [29, 39] Azzuhri et al.	• Affect the network performance by control and routing traffic	Finite state machine-based IDS system solution [40]	Network overhead time is optimized

the network-overhead time. Likewise, parent fail-over technique detects the sinkhole attack, but Svelte technique only can filter the unwanted traffic and helps to avoid the sinkhole attack. As a final prediction, insecure routing plays a major impact in the IoT network and the entire network functioning must be safe by finding a universal solution applicable to all the present and future routing attacks.

References

1. Akpakwu, G.A., et al.: A survey on 5G networks for the Internet of Things: communication technologies and challenges. IEEE Access **6**: 3619–3647 (2018)
2. Vermesan, O., et al.: Internet of things strategic research roadmap. In: Internet of Things-Global Technological and Societal Trends, vol. 1.2011, pp. 9–52 (2011)
3. Roman, R., Zhou, J., Lopez, J.: On the features and challenges of security and privacy in the distributed internet of things. Comput. Netw. **57**(10), 2266–2279 (2013)
4. Balevi, E., Al Rabee, F.T., Gitlin, R.D.: ALOHA-NOMA for massive machine-to-machine IoT communication. arXiv preprint arXiv:1803.09323 (2018)
5. Zhou, R., et al.: File-centric multi-key aggregate keyword searchable encryption for industrial internet of things. IEEE Trans. Ind. Inf. (2018)
6. Silva, E.F., Muchaluat-Saade, D.C., Fernandes, N.C.: ACROSS: a generic framework for attribute-based access control with distributed policies for virtual organizations. Fut. Gener. Comput. Syst. **78**: 1–17 (2018)
7. Huotari, S., Rothstein, K.M.: Mechanism for executing server discovery. U.S. Patent No. 9,871,872, 16 Jan 2018
8. Kogias, D.G., et al.: Realizing the wireless technology in the Internet of Things (IoT). In: Emerging Wireless Communication and Network Technologies. Springer, Singapore, pp. 173–192 (20118)
9. Jiang, Y., Huang, Z., Tsang, D.H.K.: Challenges and solutions in fog computing orchestration. IEEE Netw. **32**(3): 122–129 (2018)
10. Rao, A.P.: Adaptive control strategies for task scheduler using the Internet of Things. In: Exploring the Convergence of Big Data and the Internet of Things. IGI Global, pp. 129–140
11. Morabito, R., et al.: Consolidate IoT edge computing with lightweight virtualization. IEEE Net. **32**(1): 102–111 (2018)
12. Silva, B.N., Khan, M., Han, K.: Internet of things: a comprehensive review of enabling technologies, architecture, and challenges. IETE Techn. Rev. **35**(2), 205–220 (2018)
13. Restuccia, F., D'Oro, S., Melodia, T.: Securing the Internet of Things: new perspectives and research challenges. arXiv preprint arXiv:1803.05022 (2018)
14. Wang, H., Zhang, Z., Taleb, T.: Special issue on security and privacy of IoT. World Wide Web **21**(1), 1–6 (2018)
15. Osanaiye, O., Alfa, A.S., Hancke, G.P.: A statistical approach to detect jamming attacks in wireless sensor network. Sensors **18**(6), 1691 (2018)
16. Weekly, K., Pister, K.: Evaluating sinkhole defense techniques in RPL networks. In: 2012 20th IEEE International Conference on Network Protocols (ICNP), pp. 1092–1648, 14 Feb 2013
17. Gao, S., et al.: Security threats in the data plane of software-defined networks. IEEE Netw. (2018)
18. Feng, Y., et al.: Vulnerability of traffic control system under cyber-attacks using falsified data. Transportation Research Board 2018 Annual Meeting (TRB) (2018)
19. Valanarasu, M.R.: Smart and secure IoT and ai integration framework for hospital environment. J. ISMAC **1**(03):172–179 (2019)
20. Airehrour, D., Gutierrez, J., Ray, S.K.: A trust-based defense scheme for mitigating blackhole and selective forwarding attacks in the RPL routing protocol. Aust. J. Telecommun. Dig. Econ. **6**(1), 41 (2018)

21. Tiwari, R., Saxena, T.: A review on Sybil and sinkhole of service attack in VANET. Recent Trends Electron. Commun. Syst. **5**(1) 7–11 (2018)
22. Gill, R.K., Sachdeva, M.: Detection of hello flood attack on LEACH in wireless sensor networks. In: Next-Generation Networks, pp. 377–387 (2018)
23. Ghugar, U., Pradhan, J.: Intrusion detection system in wireless sensor networks for wormhole attack using trust-based system. In: Handbook of Research on Information Security in Biomedical Signal Processing. IGI Global, pp. 198–209 (2018)
24. Thomas, A., Gireesh Kumar, T., Mohan, A.K.: Neighbor attack detection in the internet of things. In: Advanced computational and communication paradigms. Springer, Singapore, pp. 87–196 (2018)
25. Yujie, Z.H.A.O., et al.: Techniques for automatically mitigating denial of service attacks via attack pattern matching. U.S. Patent No. 9,912,678. 6 Mar 2018
26. Singh, M.M., Mandal, J.K.: Impact of black hole attack on the reliability of mobile ad hoc network under DSDV routing protocol. Int. J. Syst. Control Commun. **9**(1), 20–30 (2018)
27. Kelpen, K., Simo, H.: Privacy and data protection in the domain name system. Privatheit und selbstbestimmtes Leben in der digitalen Welt. Springer Vieweg, Wiesbaden, pp. 253–302 (2018)
28. Azzuhri, S.R., et al.: Towards a better approach for link breaks detection and route repairs strategy in AODV protocol. Wireless Commun. Mobile Comput. (2018)
29. Devibala, K., et al.: Neighbor constraint traffic centric distributed sinkhole detection and mitigation approach for quality of service improvement in wireless sensor networks. In: Industry Interactive Innovations in Science, Engineering, and Technology. Springer, Singapore, pp. 357–366 (2018)
30. Cho, J.-H., Chen, R.: PROVEST: provenance-based trust model for delay tolerant networks. IEEE Trans. Dependable Secure Comput. **15**(1), 151–165 (2018)
31. Sharma, S., Bansal, R.K., Bansal, S.: Issues and challenges in wireless sensor networks. In: 2013 International Conference on Machine Intelligence Research and Advancement (2013)
32. Le, A., Loo, J., Luo, Y., Lasebae, A.: Specification-based IDS for securing RPL from topology attacks (2011)
33. www.postscapes.com/internet-of-things-protocols/
34. Perrey, H., Landsmann, M., Ugus, O., Wahlisch, M., Schmidt, TC: TRAIL: topology authentication in RPL. ACM 978-1-4503-1169-4 (2016)
35. Raza, S., Wallgren, L., Voigt, T.: SVELTE: real-time intrusion detection in the Internet of Things. Ad Hoc Netw. (2013)
36. Patil, M., Biradar, R.C.: A survey on routing protocols in wireless sensor networks. 978–1-4673-4523-1/12/$31.00 ©2012. IEEE
37. Li, H., Lu, R., Zhou, L., et al.: An efficient Merkle-tree-based authentication scheme for smart grid (2013)
38. Wikipedia: Wikipedia: the free encyclopedia. Retrieved from https://www.wikipedia.org/ (2017)
39. Thubert, P., et al.: Co-existence of a distributed routing protocol and centralized path computation for deterministic wireless networks. U.S. Patent No. 9,882,804, 30 Jan 2018
40. Mihovska, A., Sarkar, M.: Smart connectivity for the Internet of Things (IoT) applications. In: New Advances in the Internet of Things, pp. 105–118. Springer, Cham (2018)

Efficient Hybrid Method for Intrinsic Security Over Wireless Sensor Network

G. Sangeetha and K. Kalaiselvi

Abstract Secrecy communication is promising particularly for wireless systems due to the transmission environment of the radio path, which is simply interrupted. Wireless security methods have classically improved for conventional wireline applications, and these techniques are not assumed substantial properties of the wireless channels. To overcome these problems, in this research work a foundation was developed to introduce and examine the wireless networks inhibiting confidentiality presented via node spatial distribution, wireless propagation medium and combined network interference. This work proposed an approach, called as blowfish algorithm and secure hash algorithm (SHA), for the security which are inner qualities. Blowfish, a 64-bit symmetric block cipher which utilizes a key having the variable length from 32 bits to 448 bits, includes 16 rounds and produces the key-dependent S-boxes. It is faster speed for the procedure of encryption/decryption of group communication information from the given network. A hybrid method blowfish with SHA is proposed in this work to improve the security for ensuring secrecy from the eavesdroppers. It is used to ensure the higher security in terms of reliability and security in the given network by using efficient cryptography algorithm. The result shows that the formulated system produces higher efficiency in terms of better security rather than the existing system. The proposed hybrid BF + SHA algorithm provides higher ratio of packet delivery, average delay and throughput than the existing system.

Keywords Intrinsic security · Blowfish algorithm and secure hash algorithm (SHA) · Communication · WSN

G. Sangeetha (✉) · K. Kalaiselvi
Department of Computer Science, School of Computing Sciences, Vels Institute of Science Technology and Advanced Studies (VISTAS), Formerly Vels University, Chennai 600117, India

K. Kalaiselvi
e-mail: kalairaghu.scs@velsuniv.ac.in

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_59

1 Introduction

WSN which is expanded to a wireless sensor network is a set of specific transducers combined with an interacted networks planned to check and confirm constraints in several positions. Usually, supervised metrics are such as heat, moisture, stress, wind direction and speed, lighting strength, vibration strength, resonance strength, electrical energy, substance meditation, impurity stages and essential body operations. A sensor network also included various recognition positions named sensor nodes, every of which is tiny, insubstantial and transportable. With a transducer, microcomputer, transceiver and energy foundation are capable of each sensor. Depending upon sensed physical causes and phenomena, transducer produces electrical signals [1, 2]. Through microcomputer, it records the result of the sensor. A hard-wired or wireless, which receive orders from a middle system, is a transceiver and transfer information to the same system. From the electric utility or from a battery, the energy for every sensor node was extracted.

The receivers in the WSN must be able to notice if the swapped content among the interacting members of the WSN is modified. In addition, the reliability service must guarantee that the swapped content is not deleted, duplication of old information, fake, or old in the network [3]. Reliability of the swapped content is typically given with the grasp of the content added to the content itself. While the receiver sensor node gets the information, it verifies to distinguish if the digest of the content that it calculates and the digest received equals each other. If they are, then it accepts it as legitimate information [4].

With access control, illegal usage of supplies is prohibited in WSNs. It indicates which member of the system attains which content or facility. For example, sensor nodes must not be authorized to have the privileges of sinks like varying network-wide performance metrics of the WSN protocols. Therefore, restrictive services or operations based on the member would be suitable.

Every node is restricted in its estimating and interaction capabilities over WSN. But, interesting in-system information aggregation and examination are executed via set of nodes. For instance, a set of nodes might be answerable for mutually tracing a vehicle via the system. The genuine nodes include the set which may vary constantly and rapidly [5]. Various key services are implemented through groups over WSN. Subsequently, secure protocols for group organization are necessary, firmly admitting novel group participants and helping secure group interaction. The result of the group's calculation is usually sent to a base station. The result should be legitimate to guarantee it comes from a legal group. Any solution should also be competent by means of time and power (or engage lower computation and communication costs), excluding several standard group management resolutions.

Modern security networks are depending on cryptographic primitives, who often refuse to take two key factors for wireless sensor network: (a) the corporeal possessions of the wireless medium and (b) the spatial organization of the genuine and dangerous nodes [6]. In these two, such aspects play a significant because they involved in the interaction channels among the nodes, which produces rise to primary secrecy restrictions of a WSN. The intrinsic arbitrariness of the wireless medium and the spatial position of such nodes are lender inherent secrecy through interactions infrastructure in the physical layer level.

In this network scenario, nodes' spatial distribution, the stage of a significant role and the Poisson point process (PPP) are utilized to examine wireless networks along with security. Advocate the effective use of wireless system intrinsic possessions. When the output is usually taken as causing damage for the purpose of communication, expected such interference be able to become helpful for network secrecy. Thus, it is significant to typify the network inferences' consequence at together receivers and eavesdroppers that are permissible. From this, reasonable techniques are able to be formulated designed for uplifting the privacy of the network to an innovative state.

In this research work also establishes the foundation for providing designs and investigation of wireless networks with intrinsic secrecy. The major objective of this research work is to propose a new schema BF + SHA approach to resolve security problem in wireless systems regardless of the transmit originality of the network, which make easy the interception of two-way radio communications. The proposed approach is based on encryption and decryption security methods.

2 Literature Review

Liu et al. [7] present an alternative depiction of the security facility of the several antenna wiretap outlets under various common matrix conditions on the channel input via a channel development spat. This classification is through natural history information-theoretic and openly constructed on the instinct concerning to the best broadcast scheme in this communication set-up.

Pinto et al. [8] introduced the essentially secured interaction graph (iS-graph), which is an arbitrary graph provides the links that is found along with well-built secrecy on huge network, in the occurrence of eavesdroppers. It was studious on the iS-graph's restricted link, and hence, this method is used to improve it. The next part also shows among the node along with neighbour, the maximum secrecy rate. It is assumed the set-up where the eavesdroppers are certified to swap and unite all data. It measures precisely how eavesdropper collusion decreases with the security properties of the given system, in comparison to a non-colluding set-up.

Geraci et al. [9] consider the transmission channel along with secret information and outside eavesdroppers in which parallel multi-antenna base station interacts with numerous potentially fake users, in the occurrence of randomly positioned outside eavesdroppers. Through this method, it examines the secrecy rates attainable along with normalized channel inversion pre-coding via implementing a huge network examination that unites outcomes through random matrix and stochastic geometry theory, where the amount of customers and the amount of broadcast antennas are denoted as K and N, respectively, in a fixed ratio both yields result as infinity. It demonstrated both such quantities as scale and the density of an external eavesdropper. In addition to this, it also defines a constructive rule for the option of the parameter to be normalized, which is sceptic of channel condition data.

Krishnamurthy et al. [10] focused on the key agreement issues and introduced a novel cross-layer secure coding design protocol over the number of block-fading channels. This method is in need of statistical information about the data of eavesdropper channel level. The issue of key agreement to make the capability to prove the secure coding problem per block in the channel has been decreased by using a secrecy amplification method. Focusing on this issue, for the binary input despoiled wiretap channel, it is exposed that polar codes tend to attain nonzero ideal secrecy rates, while enjoying an extraordinarily low encoding and decoding difficulty. It also demonstrates that the secrecy ability was ensured by the special case of symmetric main and number of eavesdropper channels. This method is very much extensive to the several accessible channels with a despoiled eavesdropper through which a nontrivial attainable privacy area has been recognized. This polar coding scheme is then exploited in the key agreement protocol, where the secure coding per block in the channel is helps to produce an advantage for the legitimate nodes over the number of channels in eavesdropper, through privacy amplification module which is then turned into a secured private key.

Dworkin [11] built an endeavour to demonstrate that this development will not break the secrecy while evaluated to that of previous blowfish approach. For this reason, it utilizes avalanche effect [3] as the basis of privacy study.

Sudha and Monica [12] intimate the secure hash algorithm 3 relation of operations on binary data. Hash functions are mechanism for various significant data privacy purposes which include (1) the production and confirmation of digital signatures in it, (2) key root and (3) production of pseudorandom bit. Extendable result operations are diverse from hash operations, although it is probable to use them in many comparable ways, along with the flexibility to be modified openly to the needs of any entity applications, which is issue to other safety valedictions.

3 Proposed Methodology

In this work, blowfish algorithm and secure hash algorithm (SHA) are proposed for ensuring better intrinsic secrecy over WSN significantly.

3.1 Network Module

An undirected graph $G(V, E)$ where mobile nodes are marked as V the set of vertices in the network and the physical/logical links in the middle of the mobile nodes denoted as E the set of edges. At the same level, sensor nodes are also placed. Edge can connect the two nodes within a network that can communicate directly in the graph. Let N be denoted as network of m mobile nodes, $N_1, N_2 \ldots N_m$ and D denoted by group of n data objects $d_{1,}\ d_{2,}\ \ldots, d_n$ distributed within the system. Figure 1 shows the network scenario.

(1) The genuine network is the collection of nodes which targets to swap secret data. This system is defined through the point procedure π_l with spatial density λ_l. π_l is set of point procedure and along with spatial concentration and equivalent towards the genuine transmitters π_{tx} and π_{rx} the genuine receivers, correspondingly. Therefore, and $\lambda_l = \lambda_{tx} + \lambda_{rx}$ and

$$\alpha \triangleq \frac{\gamma_{tx}}{\gamma_{tx} + \gamma_{rx}} \tag{1}$$

are defined such that $\lambda_{tx} = \alpha\lambda_l$ and $\lambda_{rx} = (1 - \alpha)\lambda_l$ with $\alpha \in (0, 1)$.

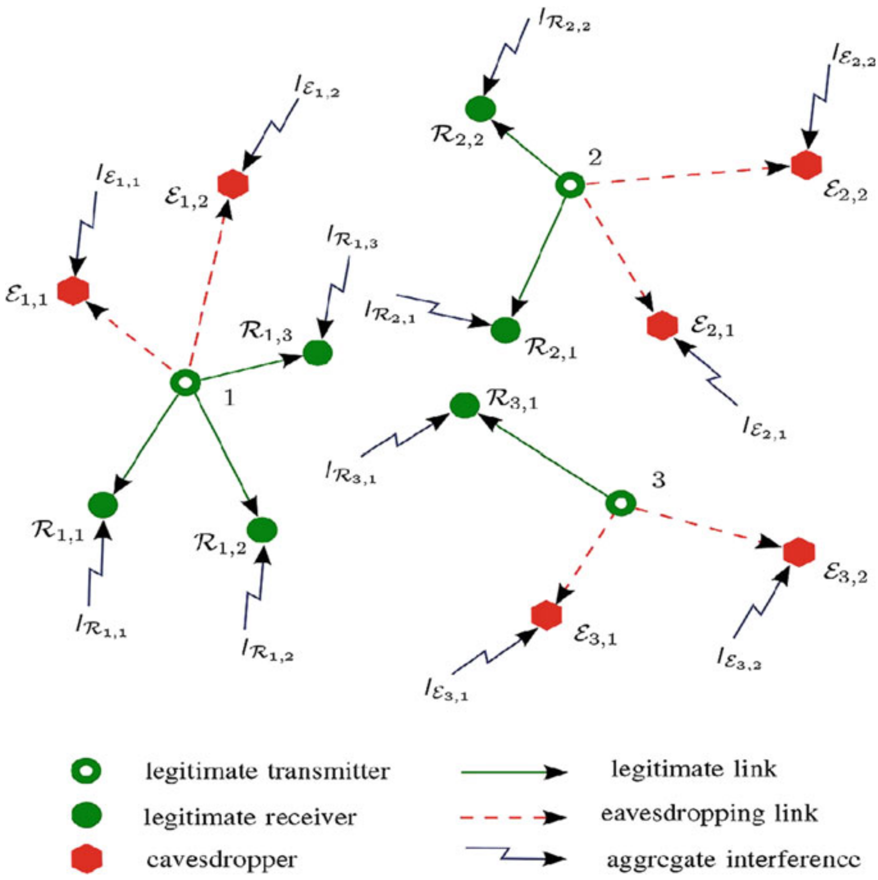


Fig. 1 Network scenario

(2) The eavesdropping network is bound of nodes which make an effort in the direction of interrupt the confidential data passing by the genuine network. It is computed by point process π_e with spatial density λ_e.
(3) Nodes in the interfering network are interfering with both legitimate receivers and eavesdroppers [13, 14]. Every eavesdropper is concerned by unintended and intentional interference produced by the genuine transmitters and the intended interferers, correspondingly. It is computed by point process π_{jx} with spatial density λ_{jx}. The quantities $\lambda_{ir} = \lambda_{tx}$ and $\lambda_{ie} = \lambda_{tx} + \lambda_{jx}$ represent the sum spatial density of nodes in the network which interferes the legitimate network and the eavesdropping network, respectively.

From the homogeneous spatial PPPs $\pi_{tx}, \pi_{rx}, \pi_e,$ and π_{jx}, legitimate receivers, legitimate transmitters, intentional interferers and eavesdroppers are geographically distributed in an n-dimensional Euclidian space R^n. Let T denotes the index sets of genuine transmitters and J of planned interferers. Let us assume that the bounded set $A \subset R^n$, for the jth legitimate transmitter in A.

For the jth transmitter, R_j and ε_j are represented as the index set of possible genuine receivers and index set of eavesdroppers attempting to intercept the secret data from the jth transmitter correspondingly.

3.2 Wireless Tap Channel in Network Setting

This section introduces wireless tap channel which is composed of a genuine transmitter with indexes as R_j, kR_j, a genuine receiver with index and an eavesdropper with index j, i attempting to intercept the transmission of secret data [15]. At a given time, the inward signal at a node with index v as of the uth transmitter is computed by

$$\gamma_{u,v}^{(t)} = \sqrt{P^T}\frac{H_{u,v}}{D_{u,v}^b}S_u + \widetilde{W}_v \quad (2)$$

where for the jth legitimate transmitter

$$v \in \begin{cases} R_j & \text{for } t = l \\ \varepsilon_j & \text{for } t = e \end{cases} \quad (3)$$

where l or e denoting the genuine transmitter or eavesdropping link. In (2), from the transmitter, P_{T} the signal power at the referred distance $d_0 H_{u,v} \in C$ denotes the quasi-static channel gain; $D_{u,v} = \|X_u - X_v\|/d_0$ is the normalized Euclidian distance present among the transmitter X_u and X_v the receiver at the random positions X_u and X_v, correspondingly; S_u specified as the transmitted symbol; b denoted as the amplitude for path loss exponent. $\widetilde{W}_v$ is the disorder which together the network interfering with receiver noise. Pointedly

$$\widetilde{W}_v = \sqrt{P^T} \sum_{q \in I_v} \frac{H_{q,v}}{D_{u,v}^b} S_q + W_v \tag{4}$$

where I_v is the index set of nodes affecting the intrusion to the receiver, i.e.,

$$I_v = \begin{cases} \frac{T}{\{j\}} & \text{for } v \in R_j \\ JUT/\{j\} & \text{for } v \in \varepsilon_j \end{cases} \tag{5}$$

In the proposed system, intrinsic secrecy I achieved by using secure hash algorithm along with blowfish algorithm [16]. It issued for encryption and decryption was as SHA employed for computing the hash value of file which is uploaded by the user on the network side. The proposed SHA considers data file of less than 264 bits in length and creates a 160-bit information digest which is developed; thus, it must be expensive to discover a data which relates a known hash value, i.e., $H(A)$; it is issue to discover a file B which also contains same hash, and even more hard to coordinate with the purpose of file B says what you want it in the direction of say.

Proposed BF + SHA algorithm

1. Start the process
2. x = obtain_file_information(), where x is the input of file with 64 bit
3. x will be further split into two halves $x1$ and $x2$.
4. $x1 = x1{:}32$
5. $x2 = x33{:}64$
6. For $i = 1$ to 16 do
7. $x_1 = x1$ XOR P_i X-OR K_1
8. $x_2 = F(x1)$ XOR x_2 X-OR K_2
9. Swap x_1 and x_2
10. After the 16th round, swap x_1 and x_2 again to undo the latest swap.
11. $x_2 = x_2$XOR P_{17}
12. $x_1 = x_1$ XOR P_{18}.

4 Simulation Results

This segment provides the secrecy presentation of a huge wireless network. In this section, the simulation results of packet delivery ratio, computation cost, detection accuracy are measured to compare the existing method and the proposed insider attacker detection method, in the simulation.

4.1 Packet Delivery Ratio (PDR)

The ratio of the amount of effectively delivered information messages to the sink over the sum amount of packets produced via each and every source. High percentage of the delivery of packets ratio gives rise to the network consistency and also satisfies the QoS.

4.2 End-to-End Delay

The average time taken by a packet to transmit from source to destination in the network is named as end-to-end delay

$$\text{End-to-end delay} = \frac{\sum_{i=1}^{n}(t_{ri} - t_{si})}{n} \quad (6)$$

where t_{ri} is the receiving time of ith packet, t_{si} is the sending time of ith packet, and n represents the amount of packets.

4.3 Throughput

The speed in which the information data is sent across the system is known as throughput and is measured in bits per second (bit/s).

$$\text{Throughput} = \text{packet received amount of packet forward} \quad (7)$$

Figure 2 also shows the simulated results of the proposed and existing methods in terms of inference range.

Figure 3 illustrates the comparison between the interference dynamics and proposed BF + SHA technique for the packet delivery ratio performance. The nodes in the network are ranged from 2 to 12, and the PDR is plotted graphically for these nodes. The experimentation result proves that the proposed BF + SHA algorithm yields higher PDR results of 85.00% for 12 numbers of nodes, whereas interference dynamics technique provides only 78.00% of PDR. It concludes that the proposed system has improved PDR in terms of optimized better secrecy using BF + SHA scheme over a given network.

Figure 4 illustrates the comparison between the interference dynamics and proposed BF + SHA technique for the throughput performance. The nodes are ranged from 2 to 12, and the throughput is plotted graphically for these nodes. The result shows that the proposed BF + SHA algorithm which yields a higher throughput of 482 kbps of than the interference dynamics technique gives only 418 kbps. Thus, the

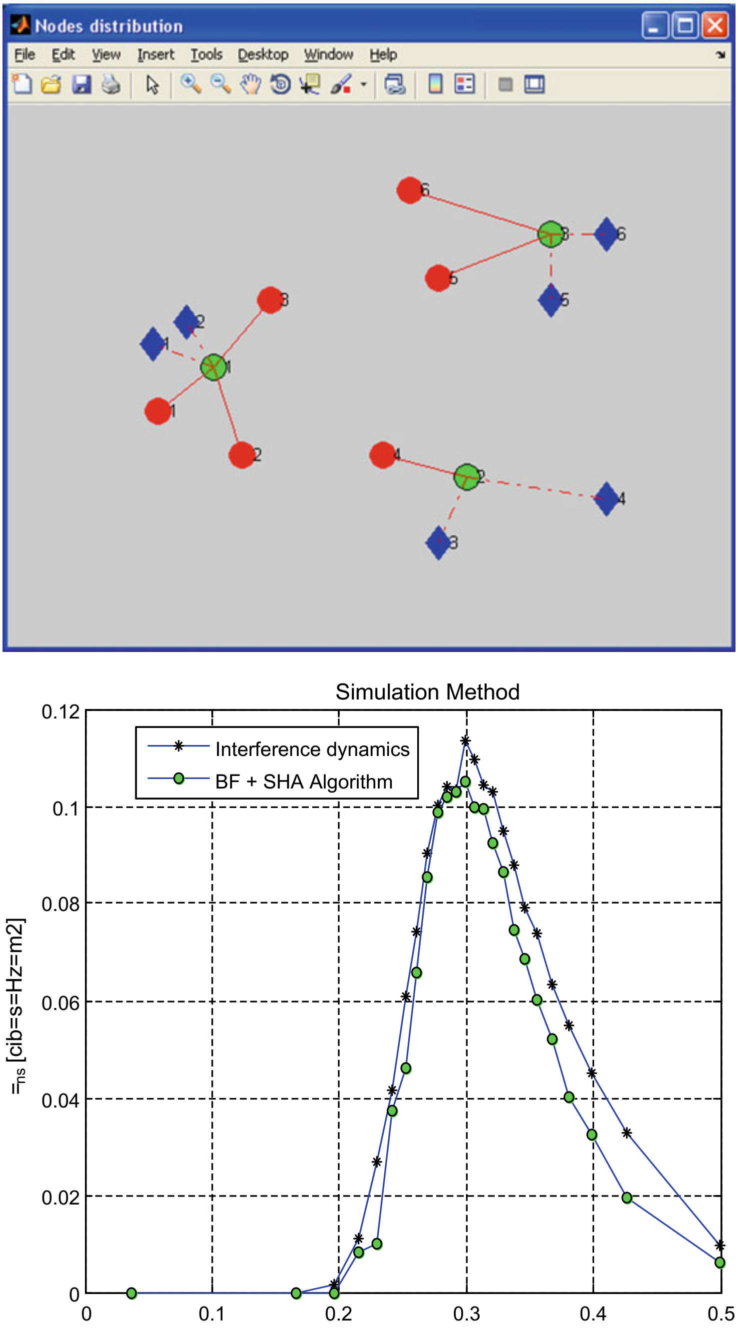


Fig. 2 Simulation model

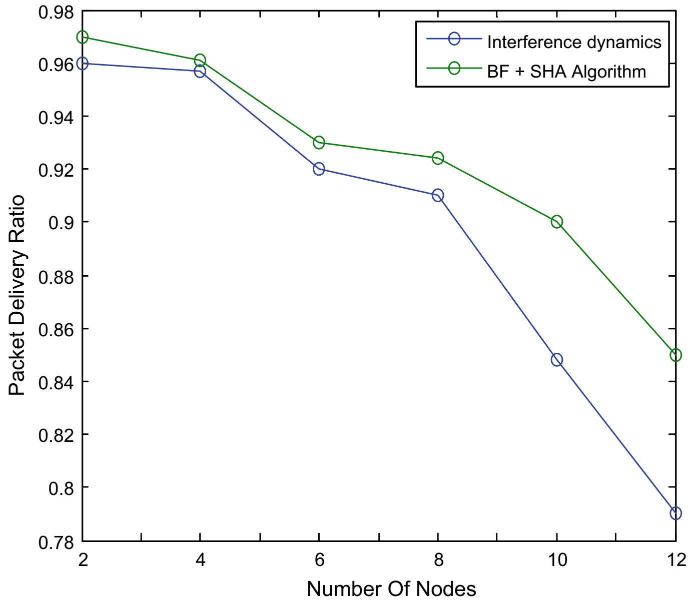


Fig. 3 Packet delivery ratio versus number of nodes

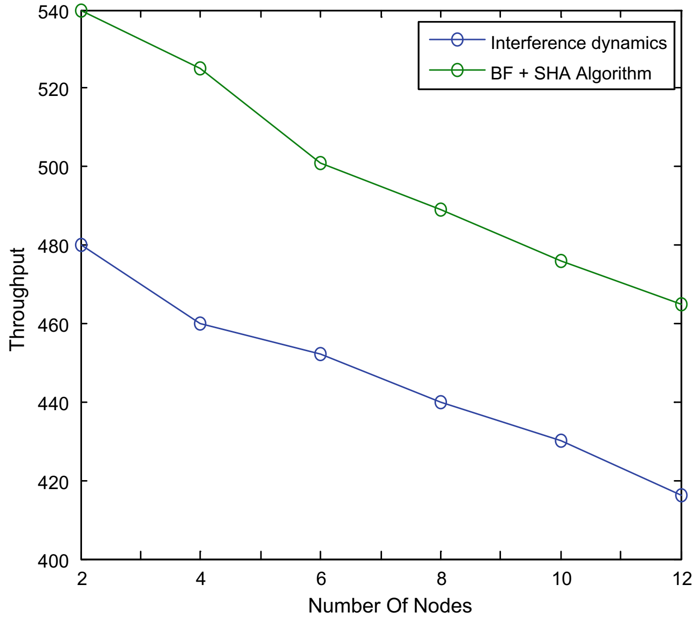


Fig. 4 Throughput versus number of nodes

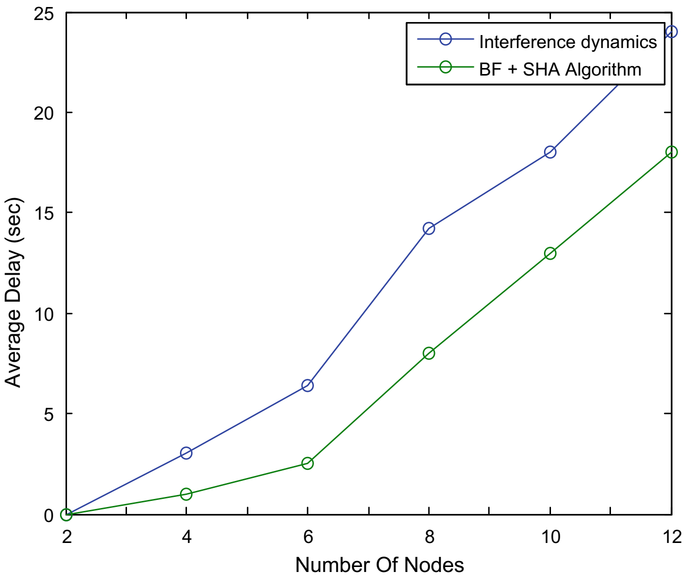


Fig. 5 End-to-end delay versus number of nodes

result concludes that the proposed system has performed in terms of secrecy using BF + SHA scheme.

Figure 5 provides the comparison between the interference dynamics and proposed BF + SHA technique for the delay performance end to end. The nodes in the network are varied from 2 to 12, and delays between the nodes are marked graphically for these nodes. Experimentation result provides the proposed BF + SHA algorithm has lesser end-to-end delay of 17 s, whereas the existing interference dynamics technique has higher delay of 24 s. The proposed system has superior presentation in conditions of end-to-end delay because of BF + SHA scheme over given network.

5 Conclusion

This work introduced network secrecy measurements and distinguishes the result of the network intrusion on network secrecy in a large-scale wireless network. In this research, a new approach of hybrid BF + SHA algorithm is proposed along with 256-bit key size to encrypt the plaintext and SHA algorithm to calculate hash value for achieving that the secret key which is transmitted safely. It is used to ignore the requirement of third-occasion auditors. The resultant hash from this utility gets stored at secure regional hash repository. The simulation result concludes that the proposed system provides higher performance in terms of better security rather than the existing system. The proposed hybrid BF + SHA algorithm achieves higher packet delivery ratio and average delay throughput than the existing system.

References

1. Li, J., et al.: An intelligent wireless sensor networks system with multiple servers communication. Int. J. Distrib. Sens. Netw. **11**(8), 960173 (2015)
2. Rezvani, M., et al.: Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks. IEEE Trans. Dependable Secure Comput. **12**(1), 98–110 (2014)
3. Barros, J., Bloch, M.: Strong secrecy for wireless channels. In: Proceedings of International Conference on Information Theoretic Security, Calgary, Canada, Aug 2008
4. Koyluoglu, O.O., Koksal, C.E., Gamal, H.E.: On secrecy capacity scaling in wireless networks. Information Theory and Applications Workshop, San Diego, CA, pp. 1–4, Feb 2010
5. Ghasemi, A., Sousa, E.S.: Interference aggregation in spectrum-sensing cognitive wireless networks. IEEE J. Sel. Top. Signal Process **2**(1):41–56 (2008)
6. Shen, J., et al.: Enhanced secure sensor association and key management in wireless body area networks. J. Commun. Netw. **17**(5), 453–462 (2015)
7. Liu, T., Shamai, S.: A note on the secrecy capacity of the multiple-antenna wiretap channel. IEEE Trans. Inf. Theory **55**(6), 2547–2553 (2009)
8. Pinto, P.C., Barros, J., Win, M.Z.: Secure communication in stochastic wireless networks—Part II: Maximum rate and collusion. IEEE Trans. Inf. Forensics Secur. **7**(1), 139–147 (2011)
9. Geraci, G., et al.: Secrecy rates in broadcast channels with confidential messages and external eavesdroppers. IEEE Trans. Wirel. Commun. **13**(5), 2931–2943 (2014)
10. Krishnamurthy, G.N., et al.: Performance enhancement of Blowfish and CAST-128 algorithms and security analysis of improved Blowfish algorithm using Avalanche effect. IJCSNS **8**(3), 244 (2008)
11. Dworkin, M.J.: SHA-3 standard: permutation-based hash and extendable-output functions. Federal Information Processing Standards (NIST FIPS)-202 (2015)
12. Sudha, M., Monica, M.: Enhanced security framework to ensure data security in cloud computing using cryptography. Adv. Comput. Sci. Its Appl. **1**(1), 32–37 (2012)
13. Oliviero, F., Romano, S.P.: A reputation-based metric for secure routing in wireless mesh networks. IEEE GLOBECOM 2008–2008 IEEE Global Telecommunications Conference. IEEE (2008)
14. Bloch, M., Barros, J., Rodrigues, M.R.D., McLaughlin, S.W.: Wireless information-theoretic security. IEEE Trans. Inf. Theory **54**(6), 2515–2534 (2008)
15. Bhalaji, N.: Performance evaluation of flying wireless network with Vanet routing protocol. J ISMAC **1**(01), 56–71 (2019)
16. Rabbachin, A., Conti, A., Win, M.Z.: Intentional network interference for denial of wireless eavesdropping. In: 2011 IEEE Global Telecommunications Conference-GLOBECOM 2011. IEEE (2011)

Cloud-Based Healthcare Portal in Virtual Private Cloud

R. Mahaveerakannan, C. Suresh Gnana Dhas and R. Rama Devi

Abstract Healthcare system providing cloud-based storage has the potential to store the patient's therapeutic records to the remote server than maintaining the files and radiological images on a hard drive or local storage device that enables the patient to access their medical records at any place by using the Internet through a Web-based application. The structure of cloud application guarantees the privacy and security of health-related data to preserve the sensitive health information. The architectural design of cloud computing aims to alleviate the privacy concern and to fulfill the confidence and trust of the cloud-based healthcare organization. The requirements, architecture design, software components, and validation methods of cloud-based healthcare system are discussed.

Keywords Healthcare system · Public cloud · VPC · VPN · NAT · Client system

1 Introduction

The organization requires different types of software and hardware to run the organization. They need a different expert to maintain the software and hardware [1]. These are the challenging task before cloud computing. After cloud computing, an organization can get rid of those problems. If the organization has not had enough resources to invest in infrastructure and platforms to deploy their applications, they can take advantage of the cloud services to suit their specific needs. Since the cloud provides a scalable infrastructure to handle the load effectively, the organization can pay for what they used based on a pay-as-you-go model. Since they can deploy and run the different applications in the cloud, they do not have to buy and maintain the

R. Mahaveerakannan (✉)
Information Technology St. Peter's University, Chennai, India

C. Suresh Gnana Dhas
Vivekananda College of Engineering for Women, Namakkal, India

R. Rama Devi
Janson Institute of Technology, Coimbatore, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_60

infrastructure like hardware, software, networking, security, firewall, etc. The cloud service provider should produce their data from the negative impact on business. So, the cloud service provider provides a high level of services as the expectation of the organization. Google and Microsoft provide Web-based email service. The emails are stored in a Google server and Microsoft server instead of storing in a client computer. Microsoft also offers a Web-based office online app, Web-based apps, social networking sites, and media services. Healthcare sensitive data must be treated with high-level security of current legislation [2, 3]. Public cloud computing provides the virtual private cloud (VPC) which allows the organization to isolate their cloud instance from other organizations in order to meet the security issues and provides great control over the cloud environment, and also, it creates a hardware virtual private network (VPN) to connect the own data centers with public cloud.

2 Requirements

Cloud-based healthcare system should maintain the security and privacy of data within the organization in order to protect the sensitive data. The cryptographic and non-cryptographic methods are used to preserve the privacy and security concern [4]. The taxonomy of security and privacy requirements [5] is categorized in Fig. 1.

2.1 Confidentiality

The sensitive healthcare data not only protected from cloud service provider and also protect from unauthorized insiders.

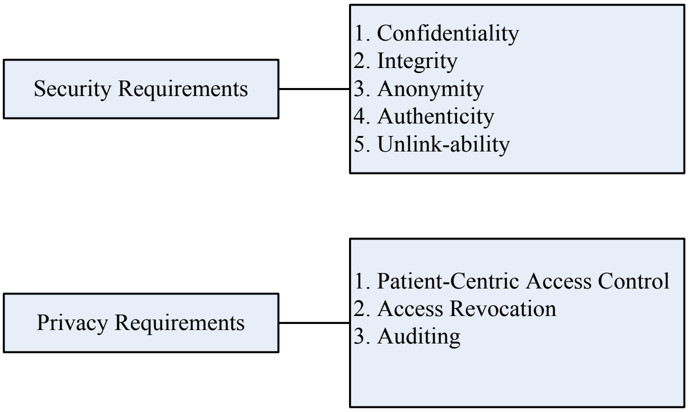


Fig. 1 Security and privacy requirements

2.2 Integrity

The healthcare data should not be modified through any illegal action of authorized or unauthorized persons. Identifying attributes such as name, date of birth, and address is encrypted with cryptographic method and access control mechanism enforced on the attribute of healthcare data file stored in the cloud.

2.3 Collision Resistance

Collision resistance mechanisms provide resistance among unauthorized users or authorized users to prevent the illegitimate actions among authorized or unauthorized users. It ensures the privacy of health data not only from the unauthorized users, but also from the authorized users.

2.4 Anonymity

The methodology of pseudonyms is identifiers that are used to identify the data owner instead of real name. The methodology used to prevent revealing the data owners' identities such as name, security number form CSP, authorized and unauthorized users.

3 Architecture Design

3.1 Healthcare System (HCS) in the Public Cloud Service

The following architecture is proposed in the figure. It comprises the three systems are:

1. The healthcare system (server)
2. Amazon Web Service (public cloud)
3. Client system (authorized clients).

The public cloud supports sensitive data of different departments in HCS. The client application runs under the HTTP and interacts with HCS. The architecture of HCS in public cloud is illustrated in Fig. 2.

HCS can create the different applications, Web services, and database storage in elastic cloud computing (EC2) instance in the public cloud. Moreover, other organizations can create the instance in the EC2 server in the public cloud and the instances are shared by flat network. Since no organization has own network or own IP address,

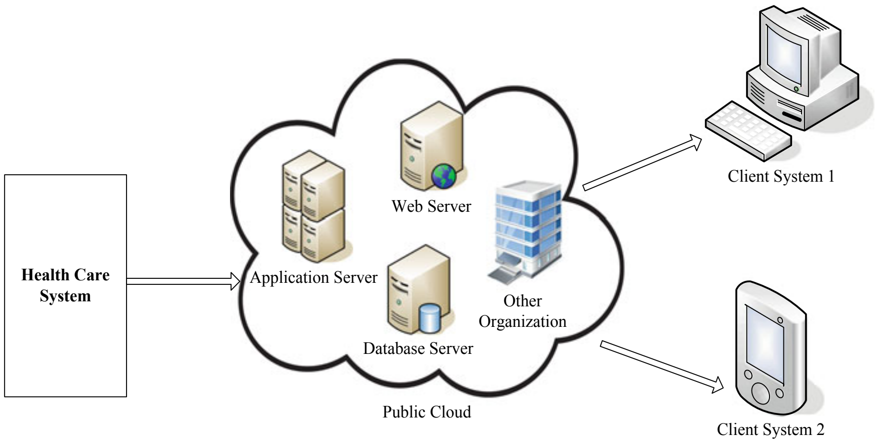


Fig. 2 Healthcare system in flat network

it is not possible to prevent the others to try to get accessing own instance in the cloud. Anyone can directly launch EC2 anywhere in the cloud. Each instance can be directly accessed by public IP address and based on access control mechanism, so each entity should be individually managed. Since the hackers are easier to intercept data on the flat network, it is creating network security issues on the network [5].

3.2 *Proposed Architecture*

Healthcare System (HCS) in Virtual Private Cloud (VPC).

The work introduces the virtual private cloud (VPC) in the public cloud to allocate the part in the public cloud. So, the existing network and VPC become as a single network. The own defined virtual network allows the organization to control the network environment. When the existing customer gateway (CG) in HCS connected to virtual private gateway (VPG) in VPC using VPN or direct connect to establish the VPN connection, it allows internal system (IS) in HCS connected to the VPC through the private IP address without Internet or public IP address. It can also allow VPC peering in different regions in public cloud that can connect the multiple VPC within public cloud to talk with each other without the Internet or direct connect which is illustrated in Fig. 3.

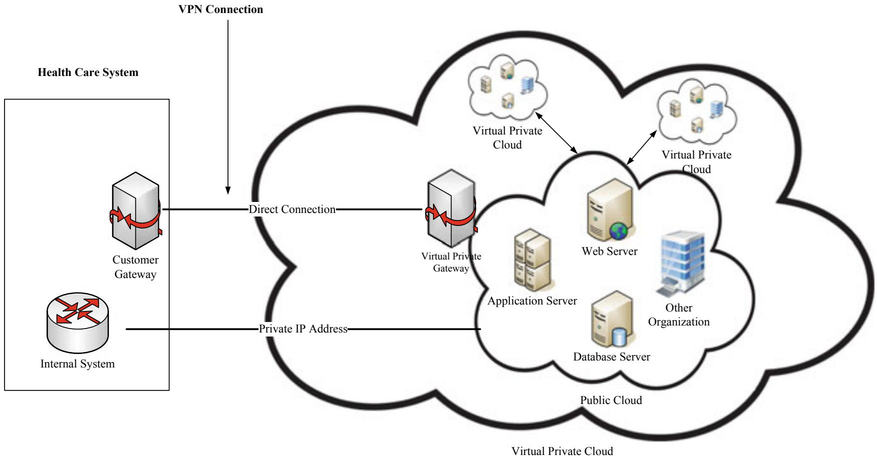


Fig. 3 Virtual private cloud for security

4 Establish Flexibility of Access in VPC

HCS has different servers, such as application server, Web server, and database server. It can improve the security in the EC2 instance, in public cloud by providing appropriate access control to a different server and each server can be individually controlled by providing appropriate subnet; for example, connect the Internet gateway (IGW) to the Web server to establish the external access to the Web server [6]. The database server and application server are protected from external server. The large network is segmented by subnet to provide public and private subnet to increase the flexibility of access. Customer gateway in HCS can connect privately with private IP address into the VPC in public cloud through virtual private Gate (VPG) without using Internet gateway. It can protect the sensitive data by providing private IP address if properly secures the network, which is illustrated in Fig. 4.

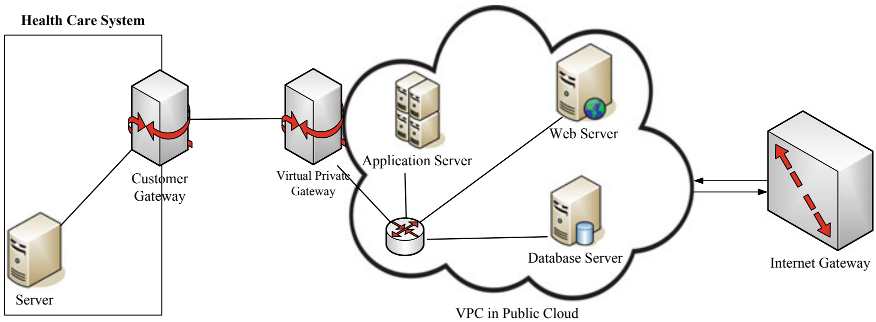


Fig. 4 Create private cloud for healthcare system

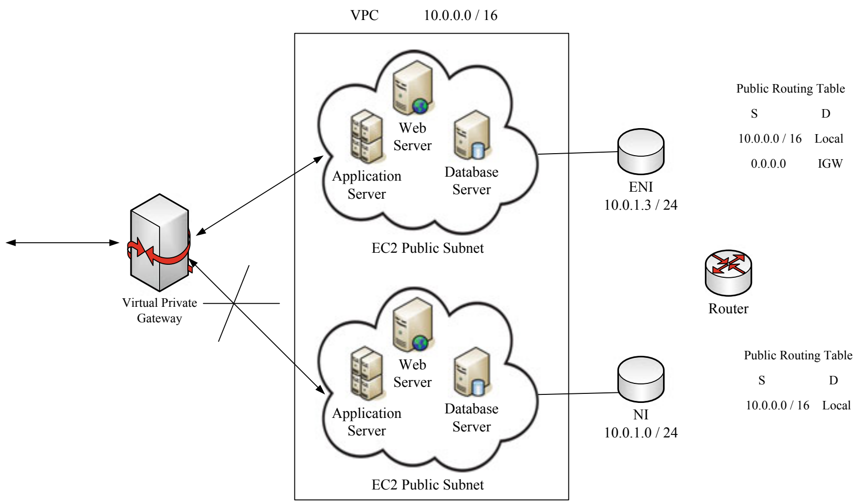


Fig. 5 Public and private subnets for servers

4.1 Create the Public and Private Subnet for VPC IP Address 10.0.0.0/16

RFC 1918 standard of IP Range: 10.0.0.0/16
Availability Zone 1
Public subnet1A: 10.0.1.0/24 for Web servers
Private subnet1B: 10.0.2.0/24 for database servers

Access control lists (ACL) act as a firewall for control the traffic into and out of the subnet [7]. All Web servers are managed in public subnet to allow inbound and outbound access. The public subnet 10.0.1.3/24 wants to communicate to the private subnet 10.0.3.48/24 which sends the request to the router to search for the routing information. It checks the destination IP address is local to the VPC that communicate locally; otherwise, it communicates through an Internet gateway [8]. Since public routing table associated with the public subnet, it can be possible to communicate both locally and outside. All the database servers have managed in the private subnet to prevent the inbound and outbound access [9]. Since private routing table associated with private subnet, it has only the possibility to communicate local subnet which is illustrated in Fig. 5.

4.2 Network Address Translation (NAT)

Proposed works introduced network address translation (NAT) gateway used instead of developing software to perform IP filtering and intrusion detection and prevention

(IDP) to enable connection in the private subnet to the Internet or other services in public cloud and also block the inbound public access through the Internet [10]. For example, private subnet can communicate with the external IP address 52.19.1.4 through a NAT gateway which is illustrated in Fig. 6.

1. Database server wants to communicate with external IP address 52.19.1.4; it sends the request to router.
2. Router search in the private routing table.
3. Get the NAT gateway.
4. NAT gateway connects the router.
5. Communicate with external IP address through IGW.

The healthcare system's Web server can handle the patients' request and deliver the medical information anywhere through the NAT gateway with the corresponding security policy of authorization and authentication [11]. Then, it is responsible for information is wrapped with their authentication keys for converting the data into an encrypted object under the security and privacy requirements [12].

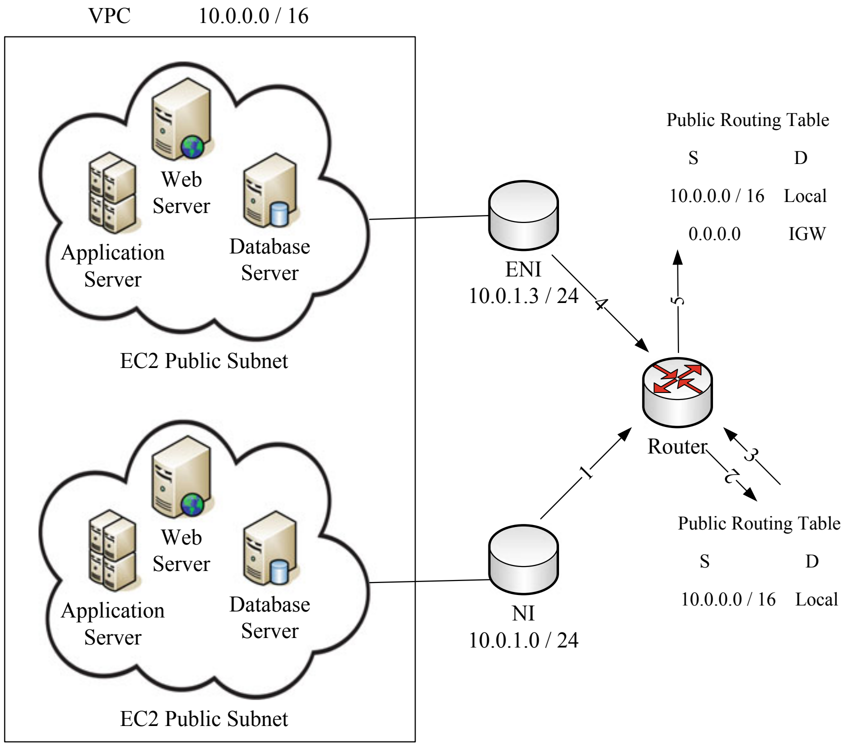


Fig. 6 Secured database server with NAT gateway

Table 1 Evaluation parameter

Goal	Metric
Patient can access medical images and reports	Patient's name, identification number, and number of tries
Patient's privacy protected	VPC and access control list
Patient's data protected	Private subnet

5 Evaluation Framework

The goal of evaluation parameter of patient's authentication and privacy requirements is defined in Table 1.

6 Conclusion and Proposed Work

The healthcare system keeps up and protects the health-related data in VPC architecture. They can keep the advantages of public cloud with respect to flexibility, scalability, elasticity, performance, availability, and the pay-as-you-use pricing model for both small- and large-scale health organizations. The healthcare sensitive data in VPC is intended to meet the prerequisite of security, privacy, and confidentiality. Future work is to extend to the analysis of cryptographic solutions along with a digital signature and build up the robust methodologies of SSH-based layered encryption approach to ensure the integrity for transferring data from one point to another. The cloud-based healthcare system can gain the overall acceptance by providing strong security mechanisms to protect the patients' records.

References

1. Abbas, A., Khan, S.U.: A review on the state-of-the-art privacy preserving approaches in e-health clouds. IEEE J. Biomed. Health Inf. **18**(4), 1431–1441 (2014)
2. Zhang, R., Liu, L.: Security models and requirements for healthcare application clouds. In: 3rd IEEE International Conference on Cloud Computing, Miami, FL, USA, July 2010, pp. 268–275
3. Li, J.: Electronic personal health records and the question of privacy. Computers (2013). https://doi.org/10.1109/MC.2013.225
4. Leng, C., Yu, H., Wang, J., Huang, J.: Securing personal health records in the cloud by enforcing sticky policies. TELKOMNIKA Indonesian J. Electr. Eng. **11**(4), 2200–2208 (2013)
5. Subashini, S., Kavitha, V.: A survey on security issues in service delivery models of cloud computing. J. Netw. Comput. Appl. **34**(1), 1–11 (2011)
6. Metri, P., Sarote, G.: Privacy issues and challenges in cloud computing. Int. J. Adv. Eng. Sci. Technol. **5**(1), 001–006 (2011)
7. Li, Z.R., Chang, E.C., Huang, K.H., Lai, F.: A secure electronic medical record sharing mechanism in the cloud computing platform. In: 15th IEEE International Symposium on Consumer Electronics (ISCE '2011), pp. 98–103, June 2011

8. Ramgovind, S., Eloff, M.M., Smith, E.: The management of security in cloud computing. In: Information Security for South Africa (ISSA). IEEE, 2010, pp. 1–7
9. Fabian, B., Ermakova, T., Junghanns, P.: Collaborative and secure sharing of healthcare data in multi-clouds. Inf. Syst. **48**, 132–150 (2015)
10. Google Online Documentation: Google's approach to IT security. Google White Paper (2010)
11. Duraipandian, M., Vinothkanna, R.: Cloud based internet of things for smart connected objects. J. ISMAC **1**(02), 111–119 (2019)
12. Winkler, V.: Securing the Cloud Computer Security Techniques and Tactics. Elsevier, Amsterdam. ISBN: 978–1-59749-592-9 (2011)

Interference Aware Cluster Formation in Cognitive Radio Sensor Networks

Jayashree Agarkhed and Veeranna Gatate

Abstract Cognitive radio sensor networks (CRSNs) is a branch of wireless sensor networks (WSNs) where cognitive intelligence is used in order to utilize the underutilized spectrum. Imbibing cognitive intelligence in WSNs can overcome various drawbacks faced by traditional WSNs as cognitive technology has an ability to adapt to surrounding environment, thereby effectively utilizing the electromagnetic spectrum. In order to have an effective communication between nodes of CRSN, an energy-efficient and adaptive medium access control (MAC) layer protocol is necessary. In this work, a novel clustering algorithm called interference-aware cluster formation in CRSNs (IACFC) is proposed. This mechanism selects vacant channels for data transmission based on optimal number of clusters formed. We identify the vacant channels based on channel availability and channel overlapping probability. Binomial distribution is used to allocate the idle channels to cognitive users. Simulation results show that IACFC has a better performance in comparison with existing clustering algorithms in CRSNs.

Keywords Cognitive radio · Channel interference · Cluster-head · Channel overlapping · Energy efficiency

1 Introduction

WSNs are controlled by governmental policies to work in a fixed radio spectrum policy. In the recent times, there has been wide range of applications developed in various fields (sensor networks, body area networks, mesh networks, personal area networks, etc.) [1] as more technologies are growing everyday there is a huge scarcity of the available radio spectrum, and since there is a rapidly growth in demand, cognitive radios can be used in the form of a solution to improvise utilization of radio spectrum [2].

J. Agarkhed · V. Gatate (✉)
Computer Science and Engineering Department, Poojya Doddappa Appa College of Engineering, Kalaburagi, Karnataka, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_61

The current static allocation of channels is not been efficiently utilized as studies made by Federal Communication Commission (FCC) say that spectrum allocated to wireless technologies are underutilized. The fixed spectrum allocation policy has led to scarcity of spectrum and also leads to interference between technologies [3]. According to the authors in [4], FCC report says that there is very less percentage of spectrum occupied by licensed users and major part of the spectrum is left unutilized. To address this problem, we can use cognitive technology to overcome the issue of inefficient spectrum usage.

Minimizing the interference between the PUs and SUs is discussed in [5] where the authors proclaim that the SUs have to vacate the channels whenever PUs try to occupy the same channels. The CRSN has the capability of reducing the power consumption, reduction in packet loss, better communication quality and high degree of buffer management.

However, the SUs can communicate with other SUs only when there is spectrum availability. Therefore, CRSN imposes spectrum awareness constraint [4], i.e., nodes communicate only when there is spectrum availability. By spectrum, awareness nodes can avoid communicating over the crowded licensed band and thereby utilizing the radio spectrum in a more efficient manner. As with any other wireless network, nodes in CRSN are also battery-powered with limited energy. Energy in CRSN is consumed in different ways, i.e., to perform sensing of radio spectrum, to detect the presence of PU, to discover the route and thereby to route the data packets and channel negotiation. In Sect. 2, we illustrate some similar works on clustering in CRSN with their potential drawbacks. In Sect. 3, the proposed protocol is explained in detail. Performance analysis of the proposed protocol is performed by comparing with existing clustering protocols such as CogLEACH [6] and DSAC [7] in Sect. 6 and finally Sect. 7 derives the conclusion.

2 Related Work

In this section, study of some of the existing approaches in clustering protocols in CRSN is made. In [8], clusters are formed to minimize the energy consumed by sensor nodes taken during communicating the information to base station. Authors describe hierarchical clustering algorithm suitable for large number of wireless sensor nodes. But, due to the fact that CRSN nodes are constrained with radio spectrum availability and CR constraints, such algorithm is not applicable for cognitive environment.

In [6], authors propose a extension of LEACH protocol called CogLEACH suitable for cognitive environment. Each sensor node sends channel state to the base station. The expected number of clusters is determined by checking the number of nodes required to cover the area and the availability of channels at that node. Cluster-head (CH) is chosen based on probability value of having high availability of channels. Nodes with high probability value are chosen as cluster-head. Upon formation of clusters, CH nodes generate TDMA schedules and all the cluster members start data transmission in their allotted timeslots. CogLEACH achieves better throughput and lifetime when compared to traditional LEACH protocol.

In [7], authors propose a spectrum-aware clustering algorithm called distributed spectrum-aware clustering (DSAC) algorithm for CRSN. The algorithm performs clustering by sensing the vacant channels and if there is change in PU activity. DSAC uses group-wise constrained clustering to minimize the energy of CRSN nodes. As the number of active PU nodes increases, more spectrum-aware constraints are imposed on the process of clustering. Hence, DSAC has poor performance in terms of energy consumption while clustering.

In [9], authors describe about the control channel allocation in cognitive radio networks called spectrum opportunity-based clusters (SOCs). This mechanism allocates channels dynamically in opportunistic CRNs. Here, the authors develop cluster-based methods for control channel allocation in CRNs. Using this approach, the SUs of the network are partitioned into clusters on the basis of location and time-dependent spectrum availability. The control channel among the list of available channels is rotated, so that cluster–cluster communication can take place.

In [10], authors propose event-driven spectrum-aware clustering algorithm (ESAC) which forms clusters based on the detection of an event. The algorithm achieves energy efficiency by avoiding re-clustering as the cluster formed is maintained till the end of event. Clusters are formed by exchange of control messages such as EFC_REQ, EFC_REP, C_REQ and C_REP, thus generating control overhead. In dynamic networks such as CRSN, exchange of control messages causes delay; therefore, the protocol suffers from delay in cluster formation.

Authors in [11] propose a cluster-based cognitive radio network for mesh networks (CogMesh). CogMesh is a multi-channel and multi-access network in which the channels available undergo changes during lifetime of a node. By sharing local common channels, clusters are constructed, and these clusters together form the network. Each cluster is controlled and coordinated by a cluster-head.

A clustering protocol called energy-efficient routing in wireless sensor networks through balanced clustering (ECHERP) is discussed in [12] which focus on energy conservation through balanced clustering. This novel algorithm selects a random node with higher energy as cluster-head. The protocol considers the current and estimated future residual energy of nodes. The protocol is scalable as the network topology adjusts based on signal interference-to-noise ratio (SINR).

Energy-efficient heterogeneous clustered (EEHC) scheme is discussed in [13] for WSN. Here, clustering process is achieved by taking heterogeneity of wireless sensor nodes. Cluster-heads are selected based on the residual energy of nodes. Authors assume that placing heterogeneous nodes is an effective way to increase network lifetime and network reliability. The proposed algorithm maximizes the network lifetime by electing cluster-heads in a distributed fashion.

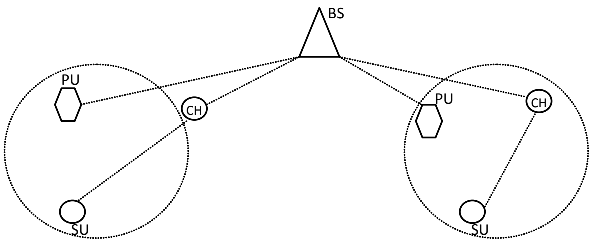
An energy-efficient hierarchical clustering algorithm is proposed in [14]. In this paper, the authors propose a fast, distributed algorithm in order to organize the sensors of a WSN in a hierarchy of clusters with an objective to minimize the energy consumed by the sensor nodes to communicate the information to the base station.

The authors assume a contention-free and error-free environment which is not the case in CRSN.

3 Proposed Protocol

In this section, we emphasize on describing a mechanism in CRSN that focuses on efficient energy utilization by forming clusters, and the channels are allocated to secondary users on the basis of binomial distribution. In a typical CRSN, primary user activity is dynamic, i.e., PU can start its activity and disappear at any point of time. The SU gets an opportunity to access the spectrum only in the absence of PU. Therefore, on the basis of channel availability, the SU nodes communicate with each other. Since the PU activity is dynamic and detection of channel availability should be continuously updated, we present here a cluster-based channel allocation algorithm which permits the SUs to utilize the spectrum on the basis of channel availability. Clusters are formed in a way, so that communication takes place between a cluster member (CM) to cluster-head (CH) and further down the hierarchy, CH communicates with the neighbor CH till information reaches base station (BS).

We consider a CRSN consisting of PUs and SUs defined as $S = \{N, C\}$, where N indicates the total number of sensor nodes in the network area and $C = \{c_{12} \ldots c_{ij}\}$ indicates the connections between peer nodes from C. All the sensor nodes are deployed randomly, and channel states are updated periodically for effective communication. The network is constructed with the multiple primary users PUs; secondary users, SUs, with the sensing ability with the centralized base station B_S are shown in Fig. 1. The SUs in the network are imbibed with channel sensing intelligence which helps in determining spectrum availability. Each node maintains the state timer to update the channel state information and the state probability. The SUs find the channel occupancy and channel availability during the channel sensing process. Each node maintains a routing table containing list of neighbor nodes which are immediate one-hop neighbors. The sensed data from sensor nodes is sent to CH and routing list containing upstream nodes is forwarded to CH.



PU-Primary User, SU-Secondary User, BS- Base station, CH- Cluster Head

Fig. 1 Secondary user communicating with Base station using Primary channel

4 Channel Allotment Scheme

Network communication is source initiated where BS intimates the nodes to perform channel sensing to check for available vacant channels. The channel list is maintained at BS and is updated always. The base station B_S broadcasts the announcement message periodically. All SUs in network on receiving the announcement information get initialized along with timer used by each sensor to update its location and the channel occupancy state. All the nodes forward location information to the B_S which stores the collected information. Once the location information is sent from all SUs in the network, B_S initiates the clustering process. Using binomial distribution, the idle channels are distributed to SUs by computing the probability of channel availability. The probability of channel available is computed by taking ratio of number of idle channels at a node to total number of nodes available at all nodes in the network. Neighbor distance from a node is calculated by knowing their location indicated by X, Y coordinates. Let two nodes i and j have the site points of $(X1, Y1)$ and $(X2, Y2)$ in sequence. Distance D_t between node i and node j is computed as

$$(D_t) = \sqrt{|X1 - X2|^2 * |Y1 - Y2|^2}$$

Clustering process is initiated by computing approximated sum in the inverse form ratio of the total number of nodes with the product of the computed probability using mean value evaluated by manipulating the binomial distribution. The number of overlapped channels is computed along with channel interference whose values are used in updating the channel availability.

Within the cluster, active PUs can use accessible channel and the remaining channels can be accessed by the SUs. One method to avoid channel interference between SUs within a cluster is to avoid the current reachable channel that is selected within private group. By computing binary co-channel conflict list from the binary matrix, we divide the channel between ith and jth SUs. Channel interference is where two or more sensors share the same band and in further words, when there are at least two CRSN nodes employing the same channel and if they can listen to each channel at the same time. By adjusting mobility and transmission power, we can minimize the channel interference between the sensor nodes.

We divide the spectrum into several channels on basis of bandwidth and frequency with some interspaces between the channels. Some channels may not be available for use because of the regularity of the domain. We may get some overlapping channels within the frequency band and sometimes we may not get. Initially, all nodes find channel time, and this is converted into a binary matrix. In the binary matrix of the channel availability, if the channel is accessible, then the value will be one else zero as shown in the expression.

We compute incentive cost for each possible channel, based on the availability to the PU. The incentive cost is computed on the basis of throughput, bandwidth and the connectivity distance to the PU. Based on the connection between two nodes, the channel bandwidth C_B can be estimated as follows

$$C_{\mathrm{B}} = \frac{\mathrm{bytes}X8}{\text{Current Time/Bandwidth}}$$

Throughput is defined based on the number of receiving bytes per second at the receiver end as

$$\text{Throughput} = \frac{\mathrm{bytes}X8}{\text{Current Time}}$$

Each node updates these parameters in its information table. If there are *n* number of overlapped channels and the channel interference range is more than one sensor node, then we compute the channel communication state. We then update number of occupied channels from the channel list C_{L}. Find the channel information C_{I} based on the number of channels m and occupied channel list as $O_{\mathrm{c.}}$ Determine the cluster and cluster-head at each round R of the communication range C_{R} Channel information is updated using the channels' current scenarios

$$C_{\mathrm{I\_}} = (\mathrm{m} - C_{\mathrm{I}} + 1) \quad \text{and} \quad C = C_{\mathrm{I\_}}$$

We compute the channel value by calculating as $\mathrm{cv} = \frac{m}{C}$.

Compute the updated channels using the expression

$$C_{\mathrm{R\_}} = \frac{KcvC_I}{m, 1}$$

If there is no change in the number of updated channels, then new $C_{\mathrm{R_}}$is computed as

$$C_{\mathrm{R\_}} = \min \frac{KC_I}{\sum C}$$

Now, effective channel selection is determined based on the transmission range of $C_{\mathrm{R\_}}$. We then determine the probability of nodes covered by the PU and nodes that are not within coverage area of PU. Using these two factors, we determine the final probability of the node to become the cluster-head. Static channel probability C_{P} for two sensors is computed and $s2 = s(1 - C_{\mathrm{P}})$.

If $C_{\mathrm{P}} > 0$ and $C_{\mathrm{P}} < 1$, then channel selection C_{S} is computed as shown in the expression:

$$C_{\mathrm{S}} = \min \frac{\mathrm{KC}_I}{((s1m + s2m), 1)}$$

Select the channel as available channel CA, and then, we compute the value (CA). If the interference range is >0 and the channel overlapping is >0, then the final channel value is computed as

$$\frac{\text{value}}{(O_{\text{L}} + C_{\text{I}})}$$

Finally, the incentive is computed which is used to assign the channel to each sensor.

5 Clustering Process

As similar to work in DSAC [5], expected power P_{E} is computed to achieve communication process by minimizing the sum of squared distances between SUs and their group centers $P_{\text{E}} = (I_{\text{N}} + I_{\text{G}})$. Using channel bandwidth, channel value as computed using expression () throughput and distance, the stable cluster-head is selected. Based on this final probability value, the node executes the cluster-head decision process to belong to a particular cluster. The sensor node with highest probability becomes a tentative cluster-head and broadcasts to all nodes within coverage area a cluster-head announcement message.

The selected CH nodes start the final cluster-head timer to receive the join message from the group members. Upon hearing the tentative CH announcement, the non-CH nodes store the id, location and the probability value of the CH. The non-CH nodes select the best available cluster-head with the greatest probability and least distance. The non-CH nodes send a join message to the selected group-head node. CH keeps the member list to accept the join replies from all members.

The node with the highest probability of channel availability and least interference value will become the cluster-head in the next iteration. To avoid reassignment of channel to same node, round-robin policy is implemented where it assigns all channels in a round-robin manner. If the non-head nodes belong to more than one cluster region, then it selects the group-head with the least distance. During the group formation, if no beacon message is received from the head, then the non-head nodes become the group-head. Once the clustering process is completed, then the head node generates and broadcasts the time slots for each node for cluster communication. We execute channel nomination in terms of both fairness-based allocation and priority-based allocation.

The fairness index is computed by calculating the number of channel allocations to each node during the current iteration. The fair value is calculated by applying the fairness index which is computed by the addition of incentive cost and channel factors. For each node, the required data rate is computed by checking all incoming and outgoing traffic. By comparing the required data rate of each node, priority of the data transmission for a particular node is computed. During the channel allocation process for each group, non-overlapping channels and with least interference get allocated to the high fairness value with the high-priority data groups. The same process gets repeated for all formed clusters, and it performs the data transmission in the allocated channel. If there are no available clusters in the coverage area, then the node itself becomes the CH and transmits the CH announcement message. Once the cluster is formed, the CH nodes synthesize the time-division multiple access (TDMA) schedule for all non-CH nodes. The generated TDMA slots are broadcasted to all the member nodes. Based on the received TDMA slot, the cognitive sensor nodes perform the data transmission. And in remaining time period, the sensor node goes to the sleep state in the specific timeslots allocated by CH.

6 Performance Evaluation

In this section, we analyze our proposed work in terms of various performance parameters such as energy consumption, node's residual energy, throughput, packet received and packet dropped. The implementation of proposed protocol is carried over in three scenarios-distributed, centralized and co-operative. We compare our simulation results with earlier clustering algorithms, CogLEACH and DSAC.

a. *Energy Consumption*: In CogLEACH protocol, the node which has more idle channels becomes the CH. If the spectrum is busy, the protocol does not allow any packets to send. This approach in a dynamic environment like CRSN is not scalable, whereas CH rotation is made in DSAC protocol and hence energy is saved. The protocol avoids overlapped channels but does not take into consideration the channel interference. IACFC has more energy efficiency than these two since both channel interference and channel overlapping are considered. In terms of residual energy also, IACFC has better efficiency when compared to cogLEACH and DSAC (Fig. 2).
b. *Throughput*: IACFC implements a separate channel for transferring control messages which is not done by DSAC and CogLEACH. To select best channel, priority- and fairness-based channel selection is not considered. However, IACFC has a separate channel for control messages and for data transmission, and priority-based channel selection is also considered which leads to better throughput in comparison with CogLEACH and DSAC. Figure 3 shows the throughput comparison.

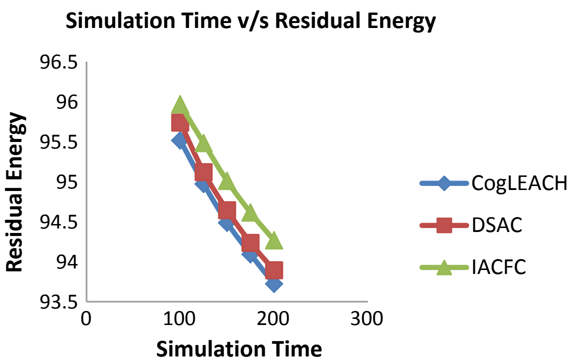


Fig. 2 Node's average residual energy

Fig. 3 Throughput

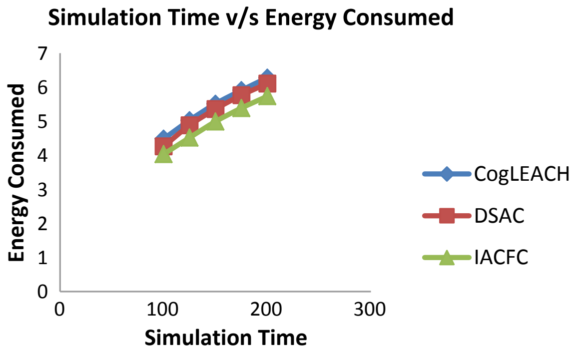
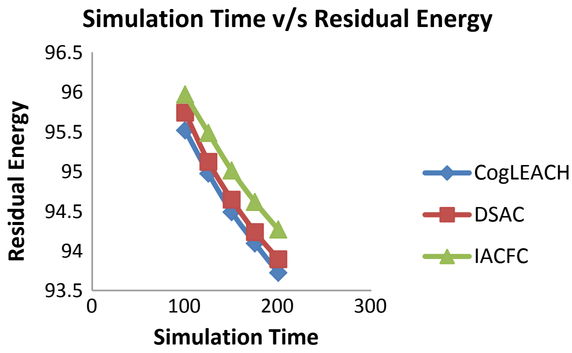


Fig. 4 Packet delivery ratio



c. Figure 4 shows the packet delivery ratio comparison. Since IACFC uses channel allocation based on clusters which is allocating a dedicated channel only after cluster is formed, the proposed work has improved performance in packet delivery ratio as compared to DSAC and CogLEACH.

7 Conclusion

Cognitive intelligence-based networks can play a vital role in efficient utilization of unutilized vacant channels as they have ability to adapt to the changes in environment and detect the vacant channels in a dynamic spectrum. In this work, we put forth a novel method to have effective communication between various cognitive users by the process of clustering. Existing clustering protocols in CRSNs only focus on cluster formation, giving importance to energy preservation. Here, along with energy conservation, focus is given on fairness- and priority-based channel allocation, taking channel interference and channel overlapping into consideration. Channel fairness is computed based on channel bandwidth, throughput and distance to the PU, whereas priority is determined by taking the data transmission rate of channel into consideration. These two criteria make IACFC a better energy-efficient fairness-based clustering protocol in CRSNs. Simulation results show that IACFC has better performance in terms of various Quality of Service parameters such as throughput, energy consumption and packet delivery ratio when compared to other existing clustering algorithms in CRSNs like DSAC and CogLEACH.

References

1. Cesana, M., Cuomo, F., Ekici, E.: Routing in cognitive radio networks: challenges and solutions. Ad Hoc Netw. **9**(3), 228–248 (2011)
2. Zhang, H., Zhang, Z., Dai, H., Yin, R., Chen, X.: Distributed spectrum aware clustering in cognitive radio sensor networks. In: Proceedings of the IEEE GLOBECOM 2011, Houston, TX, pp. 1–6 (2011)
3. Sarma, H.K.D., Bhuyan, B., Dutta, N.: An energy balanced routing protocol for cognitive wireless sensor networks. World Congress on Engineering & Computer Science, San Francisco, USA (2013)
4. Wang, B., Ray Liu, K.J.: Advances in cognitive radio networks: a survey. IEEE J. Sel. Top. Signal Process. **5**(1), 5–23 (2010)
5. Askari, M., et al.: A channel assignment algorithm for cognitive radio wireless sensor networks, pp. 12–12 (2012)
6. Eletreby, R.M., Elsayed, H.M., Khairy, M.M.: CogLEACH: a spectrum aware clustering protocol for cognitive radio sensor networks. In: Proceedings of IEEE CROWNCOM 2014, pp. 179–184 (2014)
7. Zhang, H., et al.: Distributed spectrum-aware clustering in cognitive radio sensor networks. In: 2011 IEEE Global Telecommunications Conference-GLOBECOM 2011. IEEE (2011)
8. Song, Y., He, X., Binsack, R.V.: Energy aware routing protocol for cognitive radio networks. J. Wirel. Sens. Netw. **09**(03), 103 (2017)
9. Liu, S., Lazos, L., Krunz, M.: Cluster-based control channel allocation in opportunistic cognitive radio networks. IEEE Trans. Mobile Comput. **11**(10), 1436–1449 (2012)
10. Ozger, M., Akan, O.: Event-driven spectrum-aware clustering in cognitive radio sensor networks. In: Proceedings of IEEE INFOCOM, pp. 1483–1491 (2013)
11. Chen, T., Zhang, H., Maggio, G.M., Chlamtac, I.: CogMesh: a clusterbased cognitive radio network. In: Proceedings of the IEEE DySPAN, Apr 2007, pp. 168–178
12. Nikolidakis, S.A., Kandris, D., Vergados, D.D., Douligeris, C.: Energy efficient routing in wireless sensor networks through balanced clustering. Algorithms **6**(1), 29–42 (2013)
13. Kumar, D., Aseri, T.C., Patel, R.B.: EEHC: energy efficient heterogeneous clustered scheme for wireless sensor networks. Comput. Commun. **32**(4), 662–667 (2009)
14. Bandyopadhyay, S., Coyle, E.J.: An energy efficient hierarchical clustering algorithm for wireless sensor networks. In: IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428), vol. 3. IEEE (2003)

Efficient Utilization of Resources of Virtual Machines Through Monitoring the Cloud Data Center

H. Priyanka and Mary Cherian

Abstract In cloud computing paradigm, the management and utilization of infrastructural resources are a challenging process to service providers as well as the cloud users. A cloud user accesses services based on the service provider's pay and use strategy. In this case, the users fail to utilize the computational resources effectively; it does not only lead to paying more, but also degrade the performance of the cloud data center. To overcome these inopportune situations, we propose a conceptual framework with a heterogeneous environment in a cloud data center. This model is proposed to enhance the management of infrastructural resources by provisioning virtual machines effectively and to improve the efficiency of resource utilization of applications running inside virtual machines (VMs). This paper presents the initial framework for data center setup with the results.

Keywords Cloud data center · Virtual machine · Resource utilization · Computer and control systems

1 Introduction

Cloud computing is a computing model that relies on using resources on a remote server located on the Internet. With this user can store or access information as needed, instead of maintaining data on one or individual computer.

Cloud computing has developed into a revolutionary computing model in which users can retrieve a set of organized computing resources (services, servers, apps, databases, and networks) from an Internet browser. The cloud framework is made up of virtualized data centers with thousands of huge computing servers. The data centers can use numerous megawatts of energy and since the servers are almost all inactive, most of the energy generated by these data centers is dissipated. The software that maintains the VMs on the physical machines (PMs) is known as VM kernel or hypervisor [1, 13]. As cloud computing is becoming increasingly popular, data centers are also increasing rapidly. These result in an enormous rise in the

H. Priyanka (✉) · M. Cherian
Dr. AIT, Bangalore, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_62

quantity of energy consumed. Therefore, it is essential to minimize execution and maximize resource usage to maintain cloud data centers efficiently [2, 12].

Load balancing [3, 15] has become one of the main cloud computing crises. It is indeed a technique of assigning loads to multiple virtual machines in cloud data centers to boost execution and resource usage, respectively, by commonly checking overloaded and underloaded virtual machines. There are two kinds of algorithms for load balance, static and dynamic. In static, when it starts allocating inputs, it does not reflect on a node's previous state. On the other hand, in dynamic [4], it preserves and verifies a node's prior state while allocating inputs, such as CPU, RAM, storage, bandwidth, and network. Dynamic load balancing strategy is implemented in our proposed algorithm.

This paper introduces a heterogeneous environment for cloud data center that is capable of managing a cloud infrastructure's physical resource. This will boost the cloud provider's gain by decreasing the physical infrastructure's energy consumption. Our algorithm utilizes both genetic algorithm [5, 6] and shuffled frog leaping algorithm [7] in computing minimum execution time of all incoming applications and maximizing the resource usage for cloud data centers. We validate the results using the CloudSim simulation tool. Our results demonstrate the only initial framework such as setting up of data center and monitoring incoming applications.

2 Related Work

A study on load balancing and resource usage problems in the cloud data center is discussed in this section.

Yu and Gao [8] suggested an energy-aware load balancing algorithm for the heterogeneous data center in the cloud. The proposed algorithm will minimize the energy usage of physical infrastructure and the resources are managed efficiently. This maximizes the income of the cloud vendor. The suggested algorithm uses both dynamic voltage/frequency scaling (DVFS) and server stabilization to reduce the energy usage of cloud data centers. This algorithm cannot study the impact of other system assets such as bandwidth and other network metrics on both electricity consumption and average wait time (Table 1). He-Fan Chen et al. [1] proposed an algorithm called dynamic power-saving resource allocation (DPRA) mechanismto minimize the energy rate and increase resource usage.The DPRA mechanism based on particle swarm optimization (PSO) [10, 14] is originally proposed to assign more kinds of resources and to reinforce VMs across different data centers with a view to minimize the energy usage of Physical Machines (PMs) and air conditioning units and the electrical bill of the data center. The total energy consumption of the DPRA is lower than that of the PSO. The limitation of the proposed system is to enhance the effectiveness of the PSO in the DPRA system.

Table 1 Limitations of algorithms

Author	Proposed method	Disadvantages
Ghribi et al. [9]	Usage of dynamic and static power-saving methods to achieve power-efficiency while allocating resources and migrating	It considers only non-heterogeneous environment. The SLA violations occur
Ramezani et al. [10]	The technique called task-based system load balancing method using particle swarm optimization (TBSLB-PSO) transfers only extra tasks from an overloaded VM	It considers only the homogeneous virtual machines. It migrates VMs from overloaded hosts alone, it does not consider under loaded VMs
Khaoula Hamdi, Yu et al. [8]	Energy-aware load balancing algorithm named as PLBA	It focused on the effect of system resource only on energy consumption and but not on average response time
Sharma et al. [11]	A combination of algorithms known as HGAPSO based on genetic algorithm (GA), particle swarm optimization (PSO), and Euclidean distance is suggested to achieve an optimum energy efficiency and resource usage point	Due to the space limitations, they considered only four types of VMs but the proposed algorithm can be extended for large number of VMs combinations

3 Proposed System

In our proposed model, four distinct states are possible for a specified host, namely Balanced (B), High (H), Imbalanced (I), and Low (L). Based on the state, it decides whether the migration of VM's is really essential.

The genetic algorithm–shuffled frog leaping algorithm (GASFLA) model is proposed for VM distribution and migration. In this section, the architecture of the proposed model is shown in Fig. 1. This paper is mainly concentrated on setting up of data center with heterogeneous environments and monitoring the data center dynamically based on the status of the host.

A. System Architecture

A large data center, represented by the many hosts, is an infrastructure as a service (IaaS) setting. The data center is made up of heterogeneous hosts with various VMs in each host. Virtual Machine Monitor (VMM) can be used to allocate multiple VMs to each host. In addition, the CPU performance metrics defined in terms of the millions of instructions per second (MIPS), the amount of RAM and network bandwidth is characterized by each host and VM. Our proposed model interacts as shown in Fig. 1. It consists of the following setup.

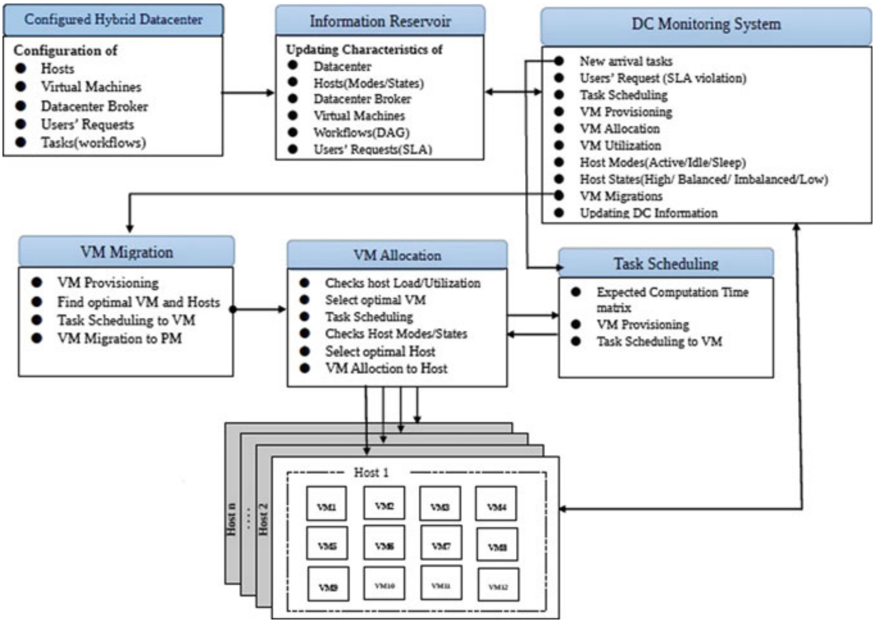


Fig. 1 Architecture of GASFLA framework

1. **Setup of the Hybrid Data Center**

To construct an infrastructure of cloud data center, the servers and virtual machines have to configure by assigning their basic entities of MIPS, the size of RAM, bandwidth, storage, and number of cores, hosts, and VMs. The CloudSim is initialized and constructed the data center for cloud user's utilizations.

2. **Information Reservoir**

All the characteristics of resources of data center, service level agreement, states of hosts (High/Balanced/Imbalanced/Low), and modes of hosts (Active/Idle/Sleep) are updated into it. Initially, the mode of newly created hosts is set to "Idle," and it is considered as switched "ON" mode and consumes the energy which is lesser than 15% and greater than 5%. In starting, the new arrival of tasks is scheduled only to the host which is in "Idle" mode and then it will be set to "Active" mode. If a host is in "Sleep" mode, a task cannot be scheduled into it and it is considered as switched "OFF" and consumes 0% energy.

Algorithm 1: Setup of data center
Input:

1. *HostConfiguration* = "*MIPS, Memory, PEs, Storage, Bandwidth, No of hosts*"
2. *VM Configuration* = "*MIPS, Memory, PEs, Storage, Bandwidth, No of VMs*"
3. *Cloudlets (Tasks)* = "*Length, PEs (CPU), input file, output file and no of cloudlets*"

Output: ***Do some operation depending on the status of the host***

4. if *datacenter setup* submitted values are *true*
5. *Go to DC Monitoring system and initialize values true*
6. *Information Reservoir → Heterogeneous environment is created*
7. *else*
8. *datacenter setup false*
9. *return.*

3. **Data Center (DC) Monitoring**

It acts upon the results, which is received from information reservoir. It checks and observes all activities of data center, and it ignites and initializes the processes. If there is new arrival of tasks, it will be scheduled to the virtual machines and assigned to the data centers. The users' service level agreement of host utilization and resource management is maintained not to exceed the agreed limit. By considering upper and lower limits of the threshold and also with the host state, we decide the VM migration. Based on the information which is collected by monitoring system, VM migration takes place. Also, after the migration, it updates the states and modes of hosts and all the activities of data center dynamically. If we have "m" number of physical machines (PM) and "n" number virtual machines (VM) to execute the tasks in a data center, the utilization of physical machine (PM_u) is calculated on the basis of the allocated number of virtual machines on it.

$$\mathrm{PM}_u = \sum_{i=1}^{n} \mathrm{CVM}_{ij}/\mathrm{CPM}_j \quad i = 1 \text{ to } n, \quad j = 1 \text{ to } m$$

where

CVM_{ij} is the processing capacity (MIPS * Pes) of CPU of ith virtual machine (VM_{ij}),
CPM_j is the processing capacity (MIPS * Pes) of CPU of jth physical machine (PM_j).
Pes = no. of processors

The utilization of PM's is considered as the threshold limits of hosts in "active" mode to identify the states of the computational resources.

Algorithm 2: Data Center (DC) Monitoring system

```
Input: High (H), Low (L), Balance (B), and Imbalanced (I).
Output: ϒ-(current host state).
if (hostCPUUtilization > 0.75)
ϒ – H
else if (hostCPUUtilization <= 0.75 && hostCPUUtilization > 0.60)
ϒ-B
else if (hostCPUUtilization <= 0.60 && hostCPUUtilization > 0.40)
ϒ-I
```

else if (hostCPUUtilization <= 0.40 && hostCPUUtilization > 0.0)
ϒ-L
else
ϒ-*Sleep*
Return ϒ

4. **Task Scheduling**

This process is performed in two stages: (1) It starts submitting of tasks to virtual machines when the new arrival of workflows and (2) whenever the need for migration of virtual machines. In this scenario, we find the expected computation time matrix to the received tasks based on the capacity of the memory and CPU of virtual machines. Also, this matrix is considered as the initial population for task scheduling to genetic algorithm (GA).

5. **VM Allocation**

In this section, it acts in the process of selection of optimal virtual machine, tasks scheduling, determining the optimal hosts based on their modes and states and allocation of virtual machines to the optimal hosts. To do all these processes we introduce genetic and meta-heuristics GA-shuffled frog leaping algorithm.

4 Experimental Results

The suggested GASFLA approach employs an optimization method within the cloud data centers to improve performance parameters. The suggested GASFLA model is scalable to a big heterogeneous cloud setting and we used an open-source cloud framework called CloudSim to perform the simulations. The proficiency of the proposed model has been displayed with host configuration, VM configuration, and cloudlet configuration.

Table 2 gives the details of creation of hosts and VMs. Here, we have considered 5 hosts, 5 VMs, and 5 cloudlets. Each host and VM will be having different host id and VM id and it will be assigned with million per instruction (MIPS). The values

Table 2 Creation of hosts and VMs

Sl. no.	Host		VM	
	Host ID	MIPS	VM ID	MIPS
1	Host Id 1 MIPS = 19676		VM Id 1 MIPS = 1961	
2	Host Id 2 MIPS = 18747		VM Id 2 MIPS = 1531	
3	Host Id 3 MIPS = 19796		VM Id 3 MIPS = 986	
4	Host Id 4 MIPS = 18,574		VM Id 4 MIPS = 1730	
5	Host Id 5 MIPS = 16853		VM Id 5 MIPS = 1506	

Table 3 Assignment of cloudlets to VMs

Sl. no.	Cloudlet Id	VM Id	Computation time (s)
1	Cloudlet 0	VM Id 0, 1, 2, 3, 4	109.33, 106.99, 106.99, 137.03, 121.2
2	Cloudlet 1	VM Id 0, 1, 2,3, 4	104.22, 107.33, 105.67, 212.78, 131.78
3	Cloudlet 2	VM Id 0, 1, 2, 3, 4	107.16, 104.33, 109.55, 213.89, 137.54
4	Cloudlet 3	VM Id 0, 1, 2, 3, 4	154.08, 139.97, 143.78, 115.89, 142.89
5	Cloudlet 4	VM Id 0, 1, 2, 3, 4	123.78, 175.07, 147.98, 225.08, 186.08

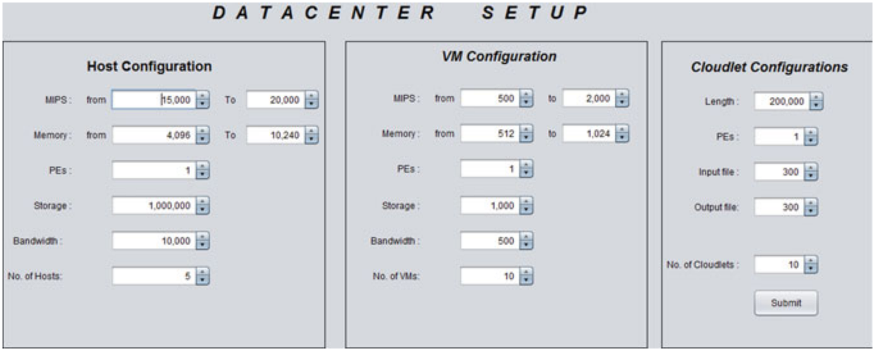


Fig. 2 Setting up of data center

for each host and VM are different as tabulated below which shows heterogeneous environment of data center. Table 3 shows the assignment of cloudlets to VMs. Each cloudlet is assigned to all 5 created VMs, each of them have different computation time for assignment as tabulated in the table.

Figure 2 gives the configuration of the data center where user can input the values. This creates the heterogeneous environment of the cloud data center. The parameters used in configuration are indicated in Tables 2 and 3.

Figure 3 gives the details of creating hosts, VMs, and cloudlets. Each host and VM will be having different host id and VM id so that it can be identified separately with their resource configuration. Each cloudlet (tasks) will be having different cloudlet lengths.

5 Conclusion

We present the monitoring of data center based on the host states. The data center set up in the heterogeneous environment is discussed, where host configuration, VM configuration, and cloudlet (Task) are given as input by the user. The suggested approach maintains the load on all the hosts by setting up the threshold values, by this server overhead can be eliminated. The experimental findings indicate that data center features such as host configuration, VM configuration, and cloudlet configuration can

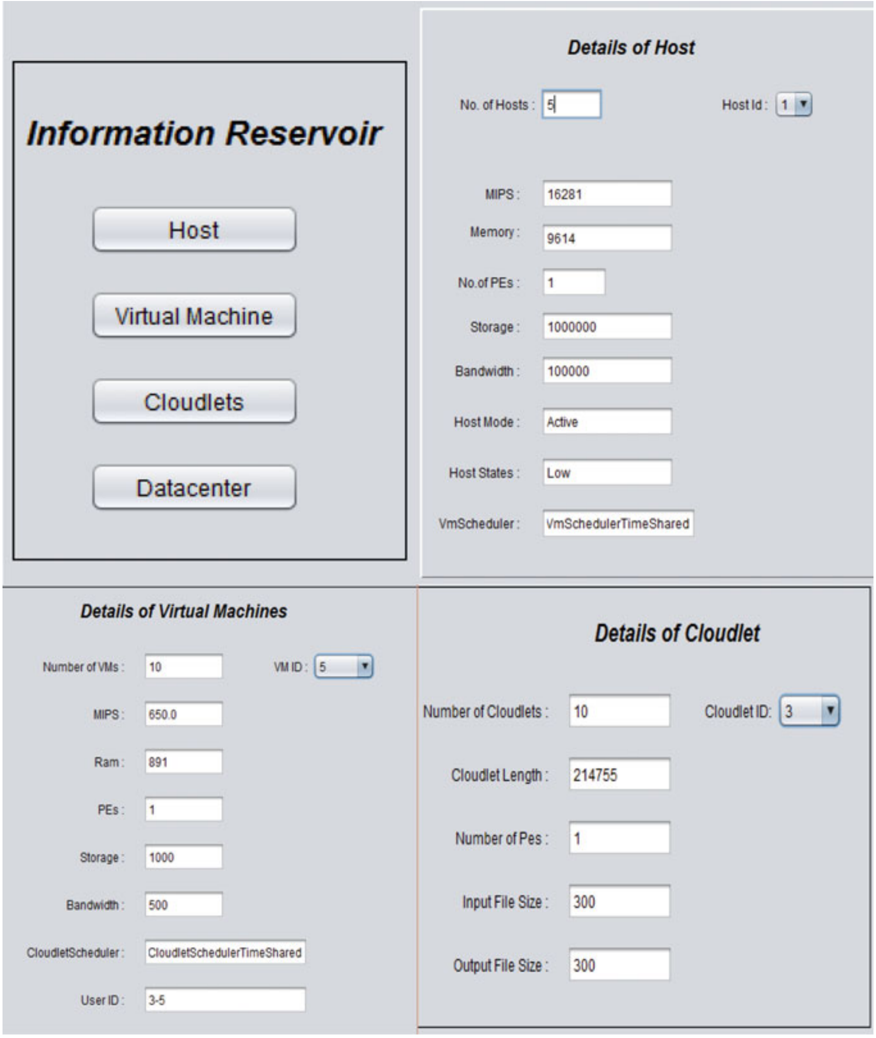


Fig. 3 Heterogeneous environment of data center

be set up with distinct input values. The actual implementation and results of VM migration and allocation using GASFLA framework will be carried out as further work of the research study. The VM allocation will be further extended using new VM allocation policy based on Algorithm 2.

References

1. Nashaat, H., Ashry, N., Rizk, R.: Smart elastic scheduling algorithm for virtual machine migration in cloud computing. J. Supercomput. 1–24, Jan 2019
2. Smys, S., Josemin Bala, G.: Performance analysis of virtual clusters in personal communication networks. Cluster Comput. **15**(3), 211–222
3. Zin, T.T., Lin, J.C.W., Pan, J.S., Tin, P., Yokota, M.: Genetic and evolutionary computing. In: Proceedings of 9th International Conference on Genetic and Evolutionary Computation (ICGEC), vol. 1. Springer, Yangon (2015)
4. Priyanka, H., Cherian, M.: A review on virtual machine migration, resource management and challenges. In: Third International IEEE Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), pp. 697–703, Dec 2018
5. Sharkh, M.A., Shami, A., Ouda, A.: Optimal and suboptimal resource allocation techniques in cloud computing data centers. J. Cloud Comput. Adv. Syst. Appl., 1–17
6. Hamad, S.A., Omara, F.A.: Genetic-based task scheduling algorithm in cloud computing environment. (IJACSA) Int. J. Adv. Comput. Sci. Appl. **7**(4), 550–556 (2016)
7. Hu, B., Dai, Y., Su, Y., Xu, L.: Feature selection for optimized high-dimensional biomedical data using an improved shuffled frog leaping algorithm. IEEE/ACM Trans. Comput. Bioinf. **15**(6), 1765–1773 (2018)
8. Yu, L., Gao, Y.: Energy-aware load balancing in heterogeneous cloud data centers. In: ICMSS '17, Wuhan, China, ACM, 14–16 Jan 2017
9. Ghribi, C., Hadji, M., Zeghlache, D.: Energy efficient VM scheduling for cloud data centers: exact allocation and migration algorithms. In: Proceedings of 13th IEEE/ACM International Symposium Cluster, Cloud, and Grid Computing, pp. 671–678 (2013)
10. Ramezani, F., Jie, L., Khadeer Hussain, F.: Task-based system load balancing in cloud computing using particle swarm optimization. Int. J. Parallel Prog. **19**, 739–754 (2013)
11. Priyanka, H.: Analytics of application resource utilization within the virtual machine allocation at the cloud data center. IJSR **5**(4), 1690–1694 (2016)
12. Chou, L.-D., Chen, H.-F., Tseng, F.-H., Chao, H.-C., Chang, Y.-J.: DPRA: dynamic power-saving resource allocation for cloud data center using particle swarm optimization. IEEE Syst. J. **12**(2) (2018)
13. Sharma, N.K., Ram Mohana Reddy, G.: Multi-objective energy efficient virtual machines. IEEE Trans. Serv. Comput. **12**(1), 158–171 (2019)
14. Cho, K.-M., Tsai, P.-W., Tsai, C.-W., Yang, C.-S.: A hybrid meta-heuristic algorithm for VM scheduling with load balancing in cloud computing. IEEE Trans. Neural Comput. Appl. **6**, 1297–1309 (2015)
15. Gamal, M., Rizk, R., Mahdi, H., Elhady, B.: Bio-inspired load balancing algorithm in cloud computing. In: Proceedings of AISI, pp. 579–589 (2017)

A Study of Energy Management Techniques for Smart City Applications on Educational Campus

Mohammad Zeeshan and Majid Jamil

Abstract Energy management in educational institutions is a much needed requirement due to the random energy needs and changing occupational behavior. A smart energy management system will be able to counter the different energy needs of the university building so that the energy needs are reduced to a minimum. Hence, a smart energy framework is proposed by comprising several layers of information transfer, and algorithms for smart light and consumption control based on the data are discussed.

Keywords Energy management system · Campus · Occupancy · Control · Sensor · Data acquisition

1 Introduction

There has been an unparalleled increase in the consumption of energy resources leading to their rapid depletion until new resources emerge which can tackle the current energy demands [1–3]. The energy consumed in buildings is now becoming a major source of usage in modern-day infrastructure. A precarious problem is the energy consumed in an academic campus which can affect the contentment of students. This imparts an unnecessarily heavy burden on the university accounts owing to wastage of energy [4–6]. An energy management system for the building needs to be assimilated to enable it to reduce the building energy requirement, the costs involved and the emissions which increase pollution. An energy management system is required for proper coordination and monitoring of energy needs of a building [7, 8]. This energy management setup comprises various energy solutions, management systems and software solutions. The energy consumption at the campus can be controlled by using control devices like sensors, microcontrollers and wireless networks [9]. Cost reduction and revenue enhancement can be obtained for saving on energy expenditure. These systems contribute to provide a detailed assessment of the

M. Zeeshan (✉) · M. Jamil
Department of Electrical Engineering, Jamia Millia Islamia, Maulana Mohammad Ali Jauhar Marg, Jamia Nagar, New Delhi 110025, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_63

energy consumption to residents and the building supervisors of campus building. These systems influence the way in which consumers behave; however, majorly, it is unto the consumer to control the energy expenditure. The equipment which is not desirable may be turned off or operated in a way such that it reduces energy consumption. The aspiration is to create a green university campus which encompasses effective energy usage in university campus. This will in turn promote environmental awareness among staff and the students [10]. The proposed methodology aims to build a green institution with sensor control-based technique for energy optimization [11]. The method attempts to study the energy consumption for all conditions. The motivation is to give a feedback of the energy consumption methods [12, 13] to the students and staff so that they are well aware of the energy consumed by them and make them better citizens by availing them the facility to improve their behavioral habits. The sensors will be deployed at various checkpoints so that the complete information regarding the data pattern is studied.

2 Data

The campus selected for the study and consumption data collection is National Institute of Technology (NIT), Trichy [14]. NIT campus consists of 777.77 acres of land which is segregated into three zones, namely the academic block, hostels and quarters. The academic block consists of 15 departments and one administrative building. The hostel count is 24 with 3 computer labs. The monthly energy consumption for the campus has been discussed in Table 1. The electrical equipment installed at the institute campus along with their rating and quantity has been discussed in Table 2.

Table 1 Monthly energy consumption for campus

S. no.	Month	Energy consumption (units)
1	January	416,840
2	February	515,500
3	March	485,380
4	April	614,380
5	May	532,560
6	June	450,240
7	July	534,380
8	August	660,860
9	September	599,860
10	October	472,040
11	November	415,280
12	December	315,600

Table 2 Details of equipment installed at the institute campus

S. no.	Equipment	Rating	Count	EMS for equipment control
1	Air conditioner 1.5/2 ton	20 A/30 A	89	Occupancy-based climate control and time table-based ON/OFF
2	Lights	40 W	1400	High quality fluorescent lamps with clustering-based control
3	Fan	65 W	571	IoT-based fan speed control and occupancy-based power modes
4	Exhaust fan	45 W	8	Humidity sensor controlled exhaust control
5	Water cooler	160 W	1	Water temperature sensor and thermostat controlled EMS
6	Water filter	40 W	2	Water quality management system and auto-turn off
7	EPBX	20 W	1	N.A
8	Computer systems B/W or color	60 W/80 W	219	Server controlled auto-turn off with data loss protection
9	Printer	120 W	4	Server checked pending print job detection and ON-OFF control
10	Microprocessor kit	10 W	18	Manual turn off. Instructions to students
11	UPS, stabilizer	5 kVA	5	Central MCB controlled switchboard
12	Others (plugs etc.)			Manual, awareness

3 Results and Discussion

The different segments for developing a smart university are as follows:

3.1 Data Monitoring and Supervision

The energy consumption of the various devices being assimilated in the different departments and corridors of the university campus needs to be monitored continuously and controlled for proper control mechanisms. The environmental parameters too need to be monitored and sensed for accurate control of different processes. This would provide vital information regarding the consumption so that this data may be forwarded to the data management devices. Sensors are installed at various corners of the department to measure various parameters such as consumption. This data is forwarded at regular intervals of time by utilizing a data collection device. The data is then forwarded to the servers for proper machine learning procedures. The data from various departments and laboratories is shown in Table 3. This includes the equipment installed, energy consumption of respective departments, storage options and logged data from the servers. Some sensors are included such as temperature, motion, occupancy, moisture, carbon monoxide and infrared sensors. These sensors are equipped to detect the presence or absence of students and measure the room temperature as well as moisture levels. The similarity between different appliances can be analyzed by the real-time data and comprehended by the statistical comparison and commands. The residents can monitor and control the various processes and improve feedback with devices. The various sensors installed across the smart campus are discussed in Table 4.

3.2 Data Storage

A data profile is maintained by various inputs from various sensing ports across the campus which reflects various parameters as discussed in Sect. 1. Different types of data are collected from different parts of the building like the temperatures of heaters, etc., which compose the data profile. The data collected from the rooms is collected in a data collection center and then sent to the server. This data increases rapidly with time as the data from different sensors gets accumulated. The data is filtered, and important information is extracted so as to act on the appliances which contain energy consumption data at regular intervals of time. To speed up the data execution time, some significant information is extricated. This information is then evaluated to extract useful information and stored in a database. The data extracted is categorized into various parts depending upon the type of usage like the equipment type, time, level and the room style. The data is then stored for a longer term and analyzed on a historical basis, and the data patterns are studied to detect aberration in the energy consumption data based on factors such as climate and usage. The equipment state is monitored to study the various states of devices and minimize the energy utilized and maximize the energy savings.

3.3 Server

The server manages the collected information data, and the smart mechanism measures the degree of aberration in the data and feedback from the customer. This energy mechanism program is stored in the server.

Table 3 Departmental consumption data

S. no.	Department	Equipments	Energy consumption	Energy storage	Data collected
1	Computer Science	32 AC's, 287 tubelights, 129 fans, 2 exhaust, 1 water cooler, 2 filters, 1 EPBX, 105 CPU, 4 printers, 18 microprocessors	38,015 W, 860 A	Battery storage system, UPS	Data from occupancy, temperature, humidity, CPU status, optical data
2	Chemical	5 AC's, 181 lights, 22 fans, 1 xerox machine	7240 W, 200 A	Battery storage system, UPS	Data from temperature and occupancy
3	CEESAT Lab	4 AC's, 84 lights, 26 fans, 2 UPS, 2 stabilizer	3856 W, 120 A	UPS	Data from temperature and occupancy
4	Chemistry	Vacuum pumps, air oven, heating metal, hot plates, gas plant, water distillation plant, spectrophotometer, lights, fans, exhaust	16,122 W	Battery storage system	Data from temperature, smoke and humidity sensor
5	Architecture	5 AC's, 114 fans, 313 lights, 20 CPU	18,290 W, 130 A	Battery storage system, UPS	Data from temperature and occupancy sensor
6	P.M.D	3 AC's, 60 lights, 17 fans, 2 CPU, 4 oven heater, 1 motor	9265 W, 80 A	Battery storage system, UPS	Data from temperature, smoke and occupancy sensor
7	MBA	18 AC's, 130 lights, 55 fans, 40 CPU, 1 oven heater, 1 motor, 1 xerox machine	15,775 W, 460 A	Battery storage system, UPS	Data from temperature, smoke and occupancy
8	ICE	7 AC's, 128 lights, 32 fans, 37 CPU, 5 oven heater, 1 xerox machine	12,480 W, 190 A	Battery storage system, UPS	Data from temperature, smoke and occupancy sensor

(continued)

Table 3 (continued)

S. no.	Department	Equipments	Energy consumption	Energy storage	Data collected
9	Library	18 AC's, 130 lights, 55 fans, 40 CPU, 1 oven heater, 1 xerox machine	11,275 W, 460 A	Battery storage system, UPS	Data from temperature, smoke and occupancy sensor
10	Mechanical	AC, motors, grinding machine, EMCO, Dekal milling machine, VMC, HMT-spindle, compressor motor, lights and fans	141.7 kW	Battery storage system, UPS	Data from temperature and occupancy sensor

Table 4 Sensors installed in smart campus

Category	Occupancy	Humidity	Temperature	Motion
Function	Sensing room occupancy	Relative humidity measurement	Room temperature measurement	Room motion detection
Sensing technique	Microwave	Capacitive	R.T.D	Infrared
Range	16 m	0–100% RH	0–50 °C	6 m
Operating voltage	1–10 V	3.5–5.5 V	0–10 V	4.5–20 V
Accuracy	10–90%	±5% RH	±0.3 °C	±0.2%

3.4 Information System

The guidelines prepared from the data collected from the energy consumption comprise the smart energy management system. This data is composed of observed energy expenditure and saving data from the educational campus. The rule administrator defines rules and set parameters so as to efficiently control the system. This collected information is analyzed and used to establish and collect rules to the database. Whenever there is an aberration in the setup, then a new rule is established. If the system detects any aberration, then the rule base is updated according to the classification algorithm, and the cause of the aberration is sorted out. Accordingly, the proper procedure is carried out to rectify the aberration. The data collected from the department such as temperature, energy consumption etc. is sent to the data collection center. From there, the routing device collects the data, sorts it according to location, level, type, etc., and routes it to the data server where the data is processed for machine learning algorithms as shown in Fig. 1.

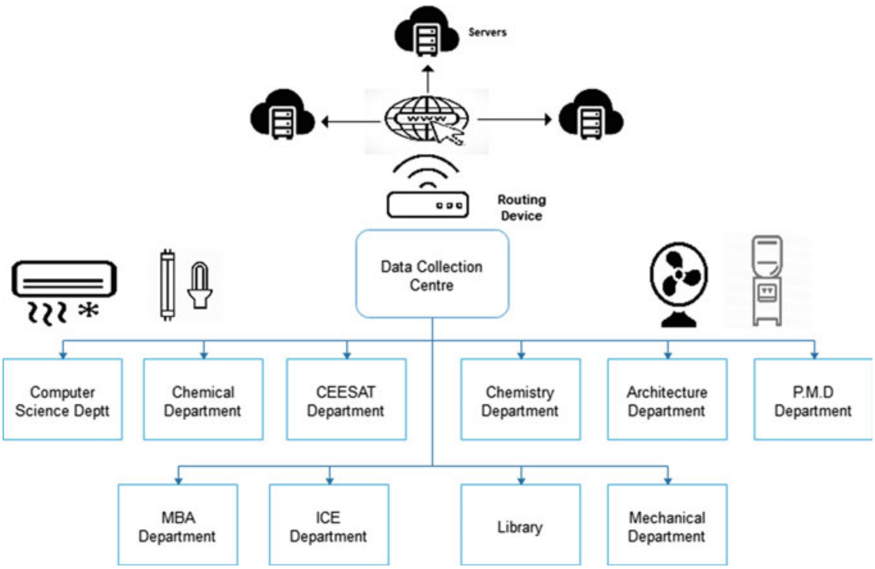


Fig. 1 Departmental categorization

An occupancy-based light control bundled with a cluster-based light density control [15] is installed at the building premises as shown in Fig. 2. The light will turn ON if any occupancy is detected within the building, and the light density will increase or decrease based on the intensity of the student cluster focused around the same.

The motion detector will detect any motion inside the lecture room, laboratories and corridors. If any motion is detected inside the room, then the occupancy counter increments to 1 and the static counter switches back to zero. The clusters of students are then detected and the light intensity increases at the cluster center as shown in

Fig. 2 Clustering-based light control

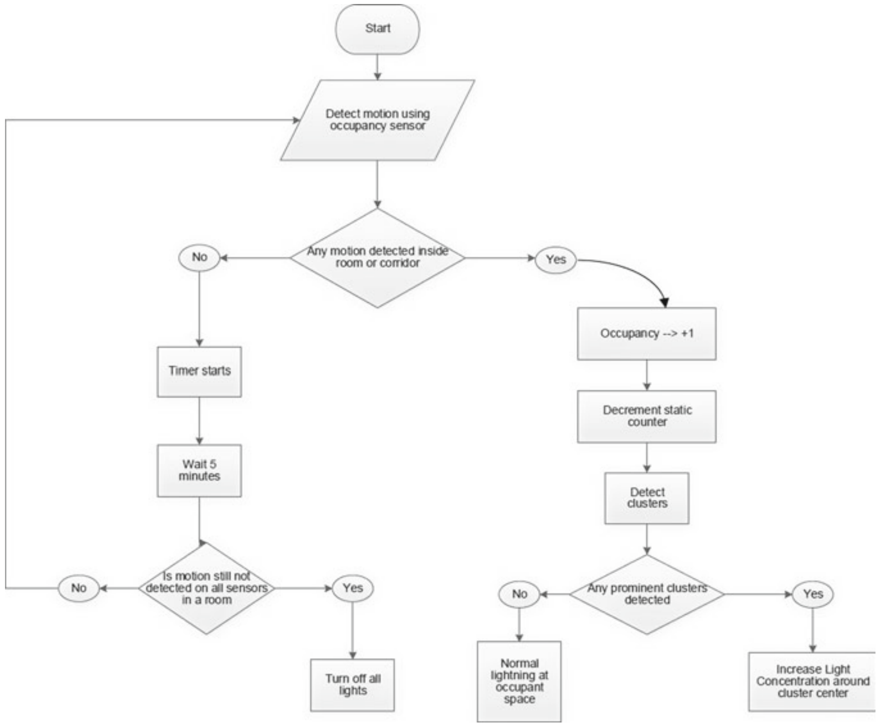


Fig. 3 Flowchart for light control

Fig. 3. If no motion is detected, then a timer starts which stays for 5 min, and then, the lights are turned off.

To reduce the energy consumption of the building, the consumption data is read from the server, and the data is categorized according to peak time and off-peak timings. If off-peak timing is detected, then the appliance turns ON without any delay else the data is further analyzed. If the demand exceeds the supply, then renewable storage options are analyzed. If the storage is able to fulfill the demand, then the demand is fulfilled else the nonessential appliances are switched off. This is discussed in detail in Fig. 4.

Assuming daily room activity of 8 hours, the smart system was installed and the room was lit based on the smart algorithm based on occupancy by students which is depicted in Table 5.

We assume that on an average, nine appliances remain ON during an 8-h period without the smart setup. The details of ON equipment utilizing the smart setup have been assumed and discussed in Table 5. For an 8-h period, we can then say that the energy consumption without the smart setup is found to be 4.32 kWh, while the energy consumption utilizing the smart setup was found to be 3.24 kWh leading to savings of approximately 33%. For an average supply of 700 units/h, the consumption by the institute has been assumed as shown in Table 6. For off-peak hours 1, 2, 3,

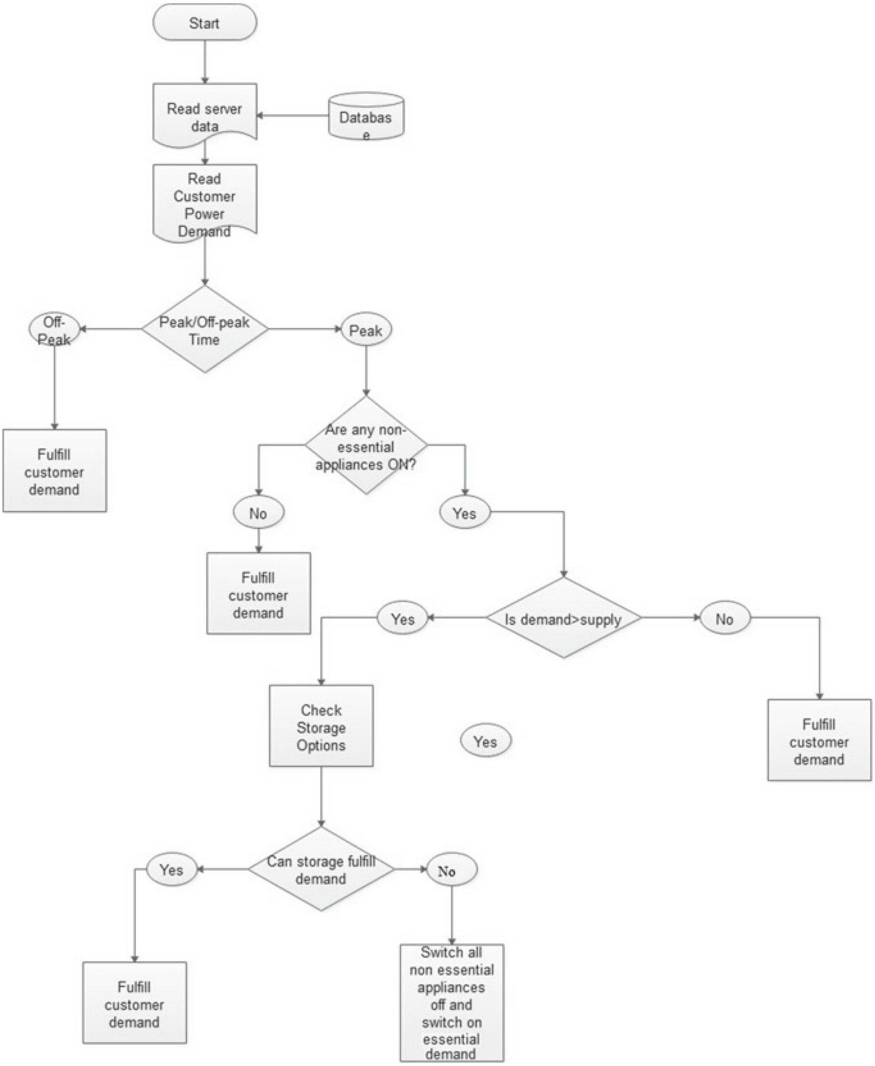


Fig. 4 Flowchart for energy consumption

9 and 10, the appliance is turned ON if there is some consumer demand. Storage options are analyzed by the data processing network if any demand arises in peak hours 4 to 8. If storage options are unable to fulfill the demand in case of shortage of energy, non-essential loads are turned off as discussed in flowchart.

Table 5 Occupancy-based equipment control

Time	Student occupancy	Equipments ON	Equipments OFF	Average ON period without smart lights
Hour 1	5	1	11	9
Hour 2	15	3	9	9
Hour 3	40	8	4	9
Hour 4	52	12	0	9
Hour 5	48	12	0	9
Hour 6	35	9	3	9
Hour 7	25	7	3	9
Hour 8	8	2	10	9

Table 6 Appliance unit consumption during peak and off-peak hours

Time	Average units consumed per hour	Actual units consumed	Peak/off-peak	Storage required
Hour 1	700	580	Off-peak	No
Hour 2	700	650	Off-peak	No
Hour 3	700	780	Off-peak	Yes
Hour 4	700	800	Peak	Yes
Hour 5	700	820	Peak	Yes
Hour 6	700	890	Peak	Yes
Hour 7	700	750	Peak	Yes
Hour 8	700	700	Peak	No
Hour 9	700	640	Off-Peak	No
Hour 10	700	320	Off-peak	No

4 Conclusions

The study was made to analyze the possibility of a smart energy management infrastructure which can be utilized to minimize the energy consumption at the selected site. Some advanced occupancy-based light ON-OFF and dimming control methods were analyzed for their feasibility. Energy saving infrastructure was also analyzed where renewable energy storage was used to overcome outages in electricity and prevents energy losses and revenue to the institute. Future prospects will deal with advanced environmental control and energy saving infrastructure which will bring down the energy expenditure of the institute.

References

1. Qureshi, M.I., Rasli, A.M., Zaman, K.: Energy crisis, greenhouse gas emissions and sectoral growth reforms: Repairing the fabricated mosaic. J. Clean. Prod. **112**, 3657–3666 (2016)
2. Tripathi, L., Mishra, A.K., Dubey, A.K., Tripathi, C.B., Baredar, P.: Renewable energy: an overview on its contribution in current energy scenario of India. Renew. Sustain. Energy Rev. **60**, 226–233 (2016)
3. Zhang, D., Wang, J., Lin, Y., Si, Y., Huang, C., Yang, J., Huang, B., Li, W.: Present situation and future prospect of renewable energy in China. Renew. Sustain. Energy Rev. **76**, 865–871 (2017)
4. Aman, S., Simmhan, Y., Prasanna, V.K.: Improving energy use forecast for campus micro-grids using indirect indicators. In: 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW), pp. 389–397. IEEE (2011)
5. Brinkhurst, M., Rose, P., Maurice, G., Ackerman, J.D.: Achieving campus sustainability: top-down, bottom-up, or neither? Int. J. Sustain. High. Educ. **12**(4), 338–354 (2011)
6. Gul, M.S., Patidar, S.: Understanding the energy consumption and occupancy of a multipurpose academic building. Energy Build. **87**, 155–165 (2015)
7. Missaoui, R., Joumaa, H., Ploix, S., Bacha, S.: Managing energy smart homes according to energy prices: analysis of a building energy management system. Energy Build. **71**, 155–167 (2014)
8. Ruusu, R., Cao, S., Hasan, A., Kortelainen, J., Karhela, T.: Developing an energy management system for optimizing the interaction of a residential building with the electrical and thermal grids. In: 12th REHVA World Congress, CLIMA 2016, p. 129 (2016)
9. Steinberg, J.D.: System and method for using a wireless device as a sensor for an energy management system (2016)
10. Faruqui, A., Sergici, S., Sharif, A.: The impact of informational feedback on energy consumption—a survey of the experimental evidence. Energy **35**(4), 1598–1608 (2010)
11. Agarwal, Y., Balaji, B., Gupta, R., Lyles, J., Wei, M., Weng, T.: Occupancy-driven energy management for smart building automation. In: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, pp. 1–6 (2010)
12. Buchanan, K., Russo, R., Anderson, B.: The question of energy reduction: the problem(s) with feedback. Energy Policy **77**, 89–96 (2015)
13. Zhou, K., Yang, S.: Understanding household energy consumption behavior: the contribution of energy big data analytics. Renew. Sustain. Energy Rev. **56**, 810–819 (2016)
14. https://aditiword.files.wordpress.com/2015/11/blaze_audit-report-for-nitt_3140474.pdf
15. Lu, J., Birru, D., Whitehouse, K.: Using simple light sensors to achieve smart daylight harvesting. In: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, pp. 73–78 (2010)

Low-Noise Amplifier for Wireless Local Area Network Applications

Malti Bansal and Jyoti

Abstract Wireless local area network is a data transmission technique developed mainly to allocate location-independent network system linked between devices by utilizing radio frequency waves rather than cable infrastructure. The standard protocol used for wireless local area network is IEEE 802.11. The wireless technology offers the capability to allow communication between two or more bodies over distances without the use of wires or cables. LNA is the active block in radio transceiver systems. The design specifications for LNA are generally dependent upon the value of S parameters, power consumption, noise figure, linearity, and gain of transistor. This paper will review the design parameters as well as provide comprehensive review of the existing topologies used for LNA in WLAN applications.

Keywords Low-noise amplifier · Gain · Noise figure · WLAN · S parameters · IIP3

1 Introduction

The wireless transmission was established in 1870. In the twentieth century, the analog communication system was taken over by digital communication and transmission of data started over the radio frequency signals. The organization of radio frequency (RF) spectrum and the FCC international agreement permitted various parts of radio frequency bands to utilize licensed free for International, Scientific and Medical (ISM) radio bands. The classification of radio bands is licensed band and unlicensed band. The licensed bands are controlled bands, i.e., paid licensed for the use of spectrum segment. The unlicensed bands are ISM bands which are free, but the main disadvantage is interference from other networks. In the year 1997, the initiations of first IEEE 802.11 standard; it defined that signals propagate through 2.4 GHz ISM radio bands to transfer the digital data. The protocols and topologies were described in wireless networks in the year 1997. The fundamentals

M. Bansal (✉) · Jyoti
Department of Electronics and Communication Engineering, Delhi Technological University, Delhi 110042, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_64

and terminology conception of wireless field network were well established [1]. The classifications of wireless network are categorized into different technology which depends upon the range to obtain, various devices to link as well as quantity of data to be transmitted.

WPAN denoted as a wireless personal area network, having a shorter range network which lies up to the range of 20–30 ft. or 7–10 m. The commonly use 802.15 standard family specification is to associate with two or more connected devices having low power consumption. The examples of WPAN protocol are bluetooth, bluetooth low energy, etc. The WLAN has range 300 m ft. It consumes more power. The word WMAN denoted as a wireless metropolitan area network extends spectrum to specific geographic area like metro city, metropolis and town, etc. Basically, WMANs use licensed frequency band, and WiMax is an example of WPAN. Lastly, WWAN is wireless wide area network. It provides the connectivity or range over a wide geographical area. It typically uses licensed frequency bands. These networks are utilized for data services, mobile phones, and operated by carriers. The basic principle of WWAN is mobile terminal and base transceiver station. The examples of WWAN are GSM and UMTS. Structured network of base stations of wireless wide area networks is mandatory to cover a country. Every base station occupies a specific area for the network types called as cellular networks. Although the network cells for wireless wide area network also cover small portion networks, i.e., WLAN and wireless personal area networks (WPAN), the WLAN covers small area; it emphasizes that these network types are well suited for services such as consumer portal services. On the other hand, WWANs support full-scale services such as safety, fleet management and telemetric services, which cover all regions of a country [1, 2].

In the domain of wireless communication systems, a main module in transceiver system is the low-noise amplifier (LNA), which is the first amplifier in a succession of amplifiers that are used to demodulate the signal as well as process a low-level signal as received by an antenna. A proficient LNA amplifies very low-power signals without adding up excessive noise, due to this maintaining the essential signal-to-noise ratio (SNR) at tremendously low-power stages. The LNA has a great impact on the quality as well as capabilities of a RF receiver due to the effective noise of all the consequent stages which is reduced by the gain of the LNA [3–5].

Due to the development of wireless communications, various standards like WiMax, wireless local area network (WLAN), bluetooth, global positioning system (GPS), etc., have developed. Hence, an exceptional methodology developed in the designer's brain was to construct the design with various multi-standard RF circuits that are known as "reconfigurable circuits." From this method, the evaluation of standard circuitry design is recreated in such a manner that advanced constructions will favor additional frequency bands. Firstly, focusing on reconfigurable circuit receivers, several structures were recommended for gratifying these approaches [6–10]. The review of various topologies of LNA at 2.4 GHz band and other applications of low-noise amplifier is illustrated in [11–16].

2 Wireless Local Area Network

Wireless communication has evolved into worldwide multi services with working frequencies of 2.4 GHz/5.2 GHz/5.7 GHz bands for WLAN, 1.5 GHz band for global positioning system, 900 MHz/1.8 GHz/1.9 GHz bands for GSM, etc. Hence, the requirement of wireless communication system is the combination of two or more standards in one mobile unit. A WLAN is also known as local area wireless network. Basically, it is wireless distribution methodology for various devices that utilizes radio bands which have higher frequency and also provides an access point to Internet. However, wireless local area network permits end users to proceed throughout the area which has been covered, a home town or offices with balancing a network connection. In the twentieth century, wireless local area network (WLAN) proved to be very cost-effective. The purpose of WLAN was where wired connection was strategically impossible. Nowadays, the proprietary protocols and solutions of WLAN are removed due to wireless distribution method for two or more devices that uses high-frequency radio waves and often IEEE 802.11 standards in several versions. The cost of WLAN also began to reduce considerably. On the other hand, wireless fidelity (Wi-Fi) is recognized as a super-sequence of IEEE 802.11 standard. Currently, the wireless fidelity devices should not need a Wi-Fi certification because the Wi-Fi devices are also utilized by hot spots in mobile appliances. According to the survey, approximately 700 million customers proceed through about 750,000 mobile data hot spots. In case of wireless local area network, the component is appraised as a station and falls further which have categorized into two parts such as access points as well as clients. Furthermore, the function of access points (APs) is transmitting and receiving the RF signals with devices, Clients incorporate a number of devices like IP phones, smartphones, other cell phones, workstations, laptop computers and desktop, etc. The communication between APs and clients among each other is known as BSSs, i.e., basic service sets. These sets are categorized as an independent basic service sets and infrastructure basic service sets. In case of independent BSSs, the communication between two clients is done without utilizing access points. It cannot connect to any another BSS. Hence, it is also known as peer-to-peer or ad hoc WLANs. In infrastructure based BSSs, the communication with other stations uses other BSSs with the utilization of access points [17] (Fig. 1).

3 Low-Noise Amplifier Topologies Used for WLAN Applications

This section comprises the popular LNA topologies for WLAN applications. We first consider the bipolar junction transistor (BJT)-based LNA circuit topologies for WLAN (Fig. 2).

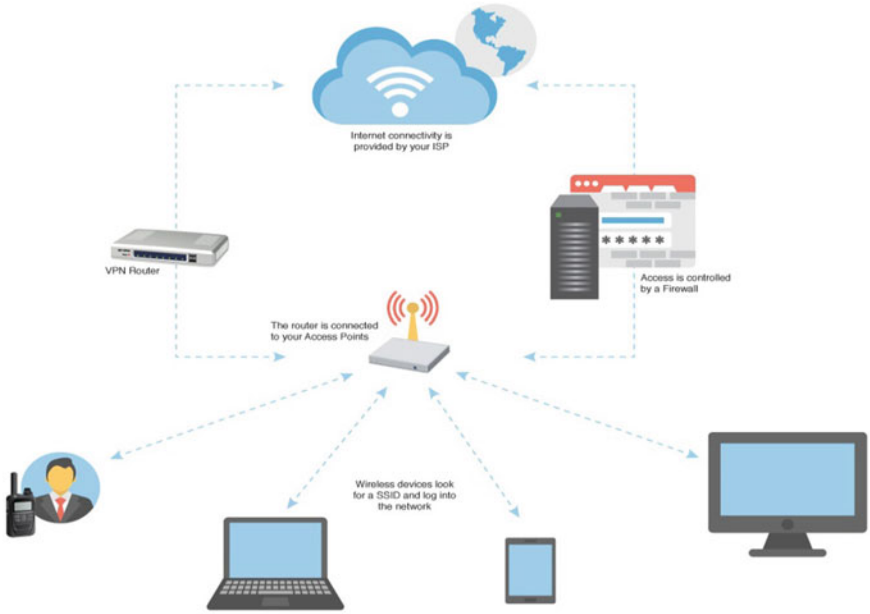


Fig. 1 Architecture of WLAN [38]

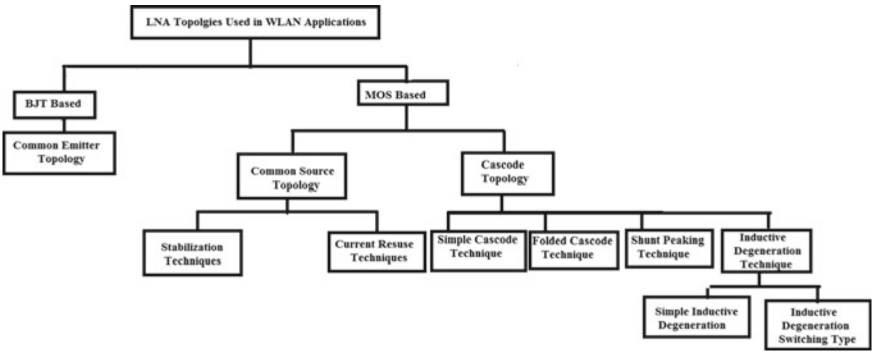


Fig. 2 LNA topologies used for WLAN applications

Common Emitter Topology

The comparative study of designing of LNA methodologies is discussed in [18, 19]. The architecture of LNA using common emitter topology is as shown in Fig. 3. The author Lee et al. [20] proposed a LNA using common emitter topology. Basically, it was composed of two stages of common emitter amplifier. The input portion was comprised of transistor preceded with an inductor providing impedance matching for the transistor. The other transistor gives nominal gain and output matching. The shunt feedback resistor had been employed to reduce the die area (Fig. 4).

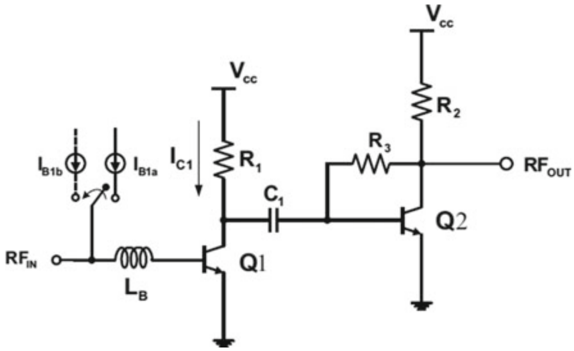


Fig. 3 Common emitter topology [20]

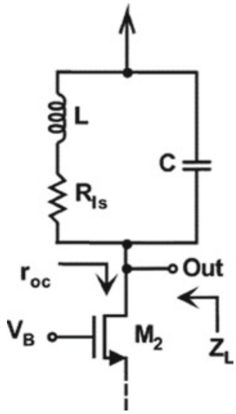


Fig. 4 Gain stabilization technique [23]

We now consider MOSFET-based circuit topologies for WLAN.

Common Source Topology

(i) Stabilization Techniques

The author Makesh Iyer et al. proposed a distinct technique to enhance the stability of the amplifier in Advanced Design System (ADS). Firstly, the author presented the designing of low-noise amplifier with series gate resistance. So, it improved the stability of amplifier while increasing the thermal noise. The other technique was designing of LNA using series drain resistance; it would reduce the gain of amplifier but it also gave optimum noise of the amplifier [21, 22]. The schematic of LNA using gain stabilization technique was as shown in Fig. 3 [23].

(ii) Current Reuse Topology

The LNA used in biomedical applications using different topology is illustrated in [24, 25]. Hossein Khosravi et al. present a low power with high gain of LNA which is employed in 2.4/5.2 GHz for WLAN applications. In this paper, LNA is comprised of two stages of common source amplifier with notch filters. Basically, foremost stage of common source amplifier obtained characteristics, i.e., higher gain of LNA operated at low frequencies. It also provided wide input matching by employing supplementary inductances, i.e., Ls and L_1 with capacitance (C1). The passive elements were connected with MOSFET M1 in series and parallel connections. Furthermore, the next portion of LNA utilized a current reuse methodology to obtain nominal gain with low power consumption. The transistor M_3 released a DC current which could be reused by transistor M_2, and hence, there was no requirement of driving current for transistor M_1. To obtain input matching with wide bandwidth, the series inductor connected to the source of transistor and resistive feedback technique was applied in the circuit. The frequency response of concurrent dual-band low-noise amplifier employed a notch filter circuit in the architecture. Therefore, the simulation results gave power gain of 15.9 dB, input reflection coefficient of -14 dB with low noise figure 1.8 dB at 2.4 frequency radio band. Additionally, for 5.2 GHz frequency spectrum, the parameters like noise figure lied at 2.7 dB, power gain lies at 14.3 dB with input reflection coefficient −12.8 dB [26] (Figs. 5 and 6).

Cascode Topology

(i) Simple Cascode Topology

Cascode is two-stage amplifier consisting of common emitter stage with common gate amplifier. Advantages of the cascode amplifier are sufficient gain, low noise,

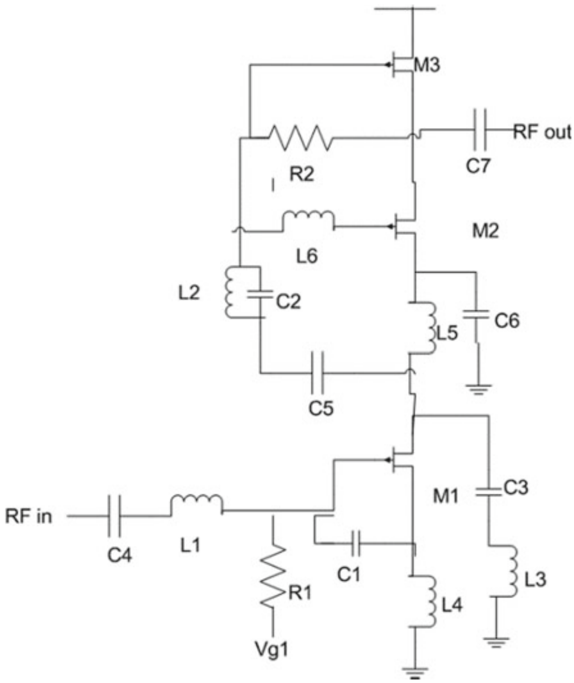


Fig. 5 Dual-band LNA [26]

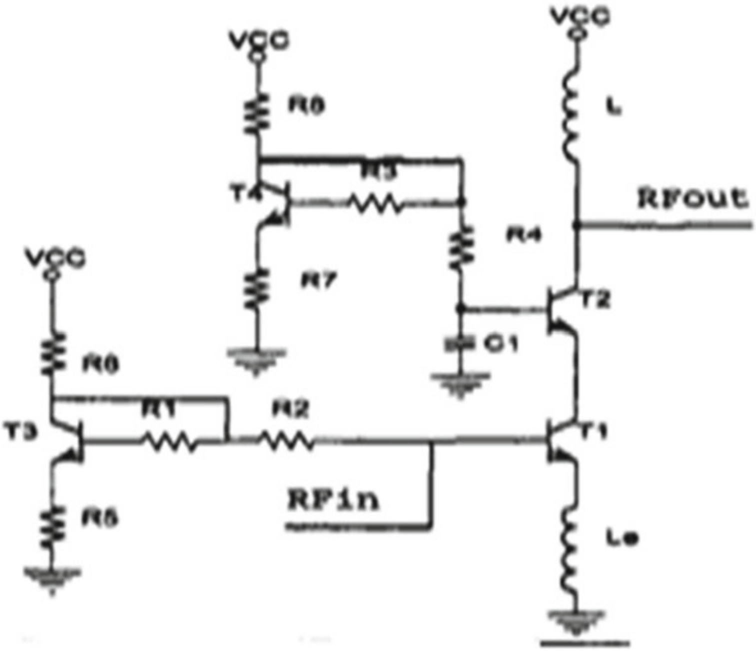


Fig. 6 Cascode amplifier [28]

etc. Cascode amplifier provides a good reverse isolation and high-frequency bandwidth. It also provides better circuit stability. J. Sadowy et al. presented a LNA for WLAN applications. The configuration of LNA consisted of transistor T_1 providing the sufficient gain of the amplifier. On the other hand, the T_2 transistor gives the reverse isolation. It also removes the Miller effect of the base-collector capacitance of transistor T_1. The proposed LNA was implemented using SiGe BiCMOS 350 nm technology. The results of the LNA demonstrated 17 dB of power gain, NF less than 2.58 dB having supply voltage 3.3 V in the 6 GHz range [27–29].

(ii) Folded Cascode Topology

In the year 2003, the author Tommy K. K. Tsang et al. presented a dual-band CMOS LNA for 2.4 and 5 GHz frequency bands and implemented them using 180 nm technology. The folded cascode topology was applied for 1 V low-noise amplifier applications. The architecture is comprised of NMOS with PMOS transistors to decrease the voltage supply as shown in Fig. 7. The transistor connected at input terminal denoted as M_1 provided the transconductance of the low-noise amplifier, and the other transistor M_2 was utilized as buffer circuit to reduce the miller effect and also enhanced the performance of frequency response at higher frequencies. The analysis of result obtained in paper was suitable for higher-frequency, low-voltage, and low-power applications [30] (Fig. 8).

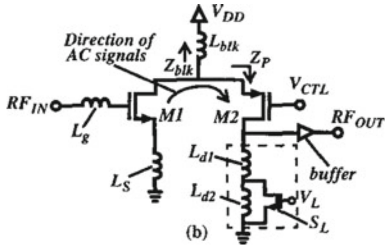


Fig. 7 Dual-band 1 V LNA [30]

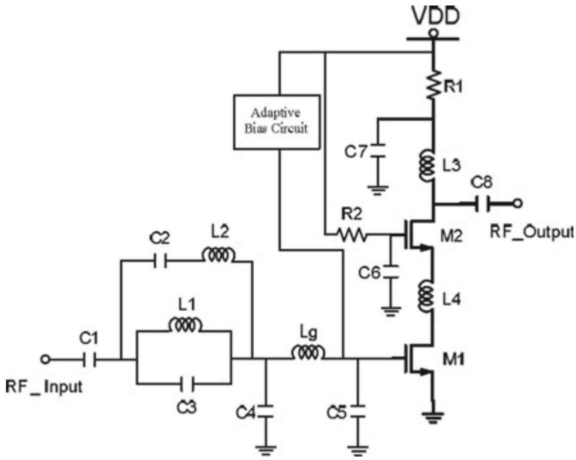


Fig. 8 Concurrent multi-band LNA [31]

(iii) Shunt-Peaking Technique

The author Chih-Yuan Kao et al. were described the fully integrated multi-band LNA are employed different frequency bands, i.e., 2.4, 3.5, and 5.2 GHz for WLAN applications. The shunt-peaking technique is integrated with inter-stage matching was described in the paper. The low-noise amplifier was implemented using 180 nm technology CMOS process. The amplifier was composed of three modules, firstly input matching network for three resonant frequencies, amplifier section consisting of cascode topology, and lastly inter-stage matching architecture of output matching network [29, 31].

(iv) Inductive Degeneration Technique

(a) Simple Inductive degeneration Technique

The author Bayoung chi et al. proposed a LNA which functions via two modes of operation: low gain mode and high gain mode. The LNA was implemented using 180 nm technology having 1.8 V power supply. The common source with common gate using inductive degeneration topology had been implemented in this paper. The architecture of LNA is depicted in [32]. Enable input was used to select gain mode to select the LNA gain. The enable input was selected to the ground, and the LNA was operated in high mode. If the enable input was selected to the V_{dd}, the LNA operated in low gain mode. The input transistor provided the complete consequence on the noise parameters. The analysis of results was obtained in both the modes: Low gain and high gain operations were done. In 2.4 GHz frequency band, the high gain mode provided input reflection coefficient as −10 dB, gain 12 dB, and noise figure 4 dB [32, 33]. The author Omid Eslamifar et al. utilized a cascode source degenerated topology that had various benefits like nominal gain, good isolation, and decreased Miller effect. The switches were composed of n-channel MOSFET on the chip. The sizes of Msw1 and Msw2 were governed by contemplating the parasitic capacitance of the switches with on resistance. The inverter was comprised to switch ON or OFF the transistors. Likewise, the dimensions of inverter were directed to be the least size of transistor, which are not degrading the performance of LNA [34, 35] (Fig. 9).

(b) Inductive Switching-Type Circuit Topology

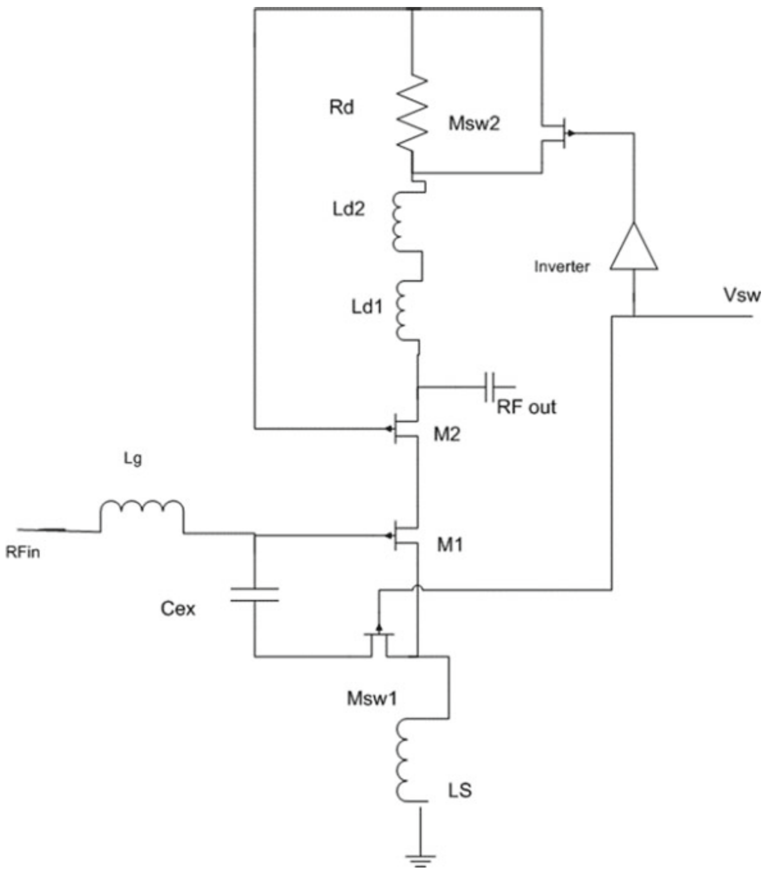


Fig. 9 Dual-band LNA with switching transistors [34]

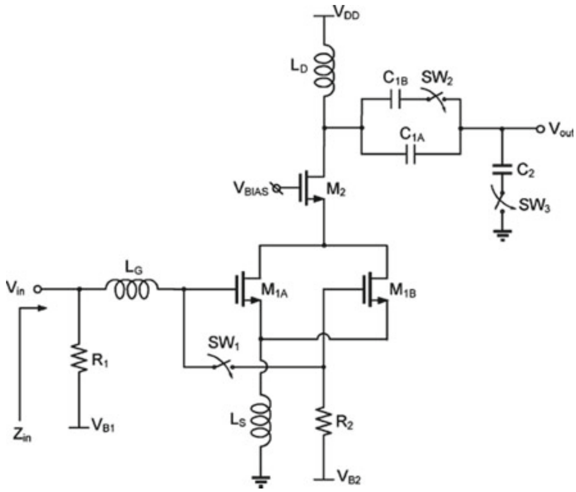


Fig. 10 Switching-type circuit topology [36]

In the year 2005, the author Liang-Hung Lu presented a dual-band LNA operated in 2.4 GHz/5.2 GHz frequency band for WLAN applications. The switching-type circuit topology was implemented in this paper. The author presented dual-band LNA along with all on-chip components. The architecture of low-noise amplifier is depicted in Fig. 10. The LNA employed a cascode source degeneration amplifier with switched capacitors with transistors. The results were obtained in 180 nm process technology, having nominal gain of 10.1 and 1 dB compression point as −7 dBm [36, 37, 39] (Table 1).

4 Conclusion and Future Scope

The review of various LNA topologies used in WLAN applications has been presented in this paper. The LNA utilized in WLAN applications requires enhancing the performance of different parameters like input matching stage, nominal gain, output matching stage, with low noise figure and good intercept point by using different topologies. The survey of different topologies of LNA used in WLAN applications has also been done. The distinctive characteristics of different topologies can simultaneously obtain input matching, nominal gain, output matching, and a low noise figure with good intercept point for WLAN LNA. Hence, we discussed different parameters and topologies and concluded as to which topology is best suitable for WLAN application. We conclude that the stabilization technique discussed is the most superior technique for WLAN applications. Distinct tradeoffs among all parameters must be determined while designing WLAN LNAs.

Table 1 LNA topologies used in wireless local area network applications

Sl. no.	Type	Topology	Technique	Technology	BW (GHz)	NF (dB)	S_{11} (dB)	S_{21} (dB)	S_{12} (dB)	S_{22} (dB)	Power (mW)	IIP3 (dBm)	Vdd (V)	Characteristics of LNA circuit used	Ref.
1.	BJT based	Common emitter topology	–	SiGe	2.4	3.18	−27.6	13.8	–	–	8.7	–	1.5	The LNA obtained low reflection coefficient at all desired frequency	[20]
					5.2	3.42	−47.1	14.4	–	–	7.5	–			
					5.7	3.21	−24.7	13.3	–	–	7.5	–			
2.	MOSFET based	Common source topology	Stabilization techniques	HEMT	5	0.308	−45.106	11.256	−20.094	−37.011	–	–	–	The LNA achieved good reflection coefficient, high gain and low noise figure	[22]
			Current reuse technique	180 nm	2.4	1.8	−14	15.9	–	–	2.25	–	–	The LNA obtained low reflection coefficient and high gain	[26]
					5.2	2.7	−12.8	14.3	–	–	2.25	–			
		Cascode topology	Simple cascode technique	180 nm	1.4	2.3	–	17	–	–	13	−18	3.3	The LNA exhibits high gain and low reflection coefficient	[28]
					2.4	2.26	−24	18.3	–	−11	6	–	2	The LNA exhibits low noise figure and low input reflection coefficient	[19]
					5.2	3.2	−28	15.2	–	−7.2	–	–			

(continued)

Table 1 (continued)

Sl. no.	Type	Topology	Technique	Technology	BW (GHz)	NF (dB)	S_{11} (dB)	S_{21} (dB)	S_{12} (dB)	S_{22} (dB)	Power (mW)	IIP3 (dBm)	Vdd (V)	Characteristics of LNA circuit used	Ref.
			Folded cascode technique	180 nm	2.4	2.3	−5.1	11.6	–	–	14.2	−7.9	1	The LNA exhibits high gain and low reflection coefficient	[30]
					5	2.9	−26.3	10.8	–	–	–	−7.1			
			Shunt-peaking technique	180 nm	2.4	3.89	−10.37	11.79		−12.47	13.5	−3.0	1.4	The LNA exhibits low input and output return loss at all desired frequency band	[31]
					3.5	4.03	−10.41	11.70		−13.17	13.5	−2.1			
					5.2	3.73	−13.56	10.06		−19.42	13.5	0.4			
			Inductive degeneration technique 1. Simple inductive degeneration topology	180 nm	3.1–10.6	<3	<−11	13.2–14.8	<−33	<−10.5	23.7	−3.1– −8.6	1.8	The LNA achieved high gain and good reflection coefficient.	[22]
				180 nm	2.4	4	−10	12	–	−12	7.94	−2	1.8	The LNA exhibits very high gain and low noise figure.	[34]
			2. Inductive degeneration switching type	180 nm	2.4	2.9	−10.1	10.1	–	−10.5	11.7	4	1.8	The LNA exhibits high power disspiation and achieves good linearity	[36]
					5.2	3.7	−11	10.9	–	−17	5.7	−5			
					3.5	4.03	−10.41	11.70		−13.17	13.5	−2.1			
					5.2	3.73	−13.56	10.06		−19.42	13.5	0.4			
					5.2	3.2	−28	15.2		−7.2					

References

1. http://www.ciscopress.com/articles/article.asp?p=1876001
2. http://www.ecartouche.ch/content_reg/cartouche/LBSbasics/en/html/LBSBasicsU4_learningObject3.html
3. Ooi, B. Z. M., Wong, C. L., Hoh, C. W.: Low-Noise Amplifier with High Linearity. US 8,803,612 B1
4. Azevedo, F., Fortes, F.: A 2.4 GHz Monolithic Single-ended-Input/Differential-Output Low-Noise Amplifier, pp. 1–4. (2007)
5. Srivastava, G., Bansal, M.: A high linearity shunt capacitive feedback LNA for wireless applications. In: Smys, S., Bestak, R., Rocha, Á. (eds.) Inventive Computation Technologies. ICICIT 2019. Lecture Notes in Networks and Systems, vol. 98, Springer, Cham (2020)
6. Gyamlani, S., Zafar, S., Sureja, J., Chaudhari, J.: Comparative study of various LNA topologies used for CMOS LNA design. Int. J Comp Sci. Emerging Tech. 3, 41–49 (2012)
7. Aditi, Bansal, M.: A high linearity and moderate gain LNA for receiver front-end applications in 2.4 GHz ISM band. In: 2017 International Conference on Innovations in Control, Communication and Information Systems (ICICCI), pp. 1–5, Greater Noida, India (2017). https://doi.org/10.1109/ICICCIS.2017.8660854
8. Srivastava, V. M., Kumar, R.: Low noise amplifier for 2.45 GHz frequency band at 0.18 μm CMOS technology for IEEE standard 802.11 b/g WLAN, Int. J. Intell. Syst. Appl., 11, 68–74 (2012)
9. Aditi, Bansal, M.: High linearity and low noise shunt resistive feedback CMOSLNA in 2.4 GHz ISM band. In: 2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE), pp. 95–99, Noida (2017). https://doi.org/10.1109/RDCAPE.2017.8358247
10. Shankar, S. U., Dhas, M. D. K.: Design and performance measure of 5.4 GHZ CMOS low noise amplifier using current reuse technique in 0.18 μm technology. Procedia Comput. Sci. 47, 135–143 (2015)
11. Miguel, J., Machado, H.: LNA for a 2.4 GHz ISM receiver. Dissertation submitted for obtaining the degree of Master in Electrical and Computer Engineering Jury, July (2010)
12. Bansal, M., Jyoti.: A review of low noise amplifier for 2.4 GHz frequency band. In: 2017 International Conference on Innovations in Control, Communication and Information Systems (ICICCI), pp. 1–6, Greater Noida, India (2017). https://doi.org/10.1109/ICICCIS.2017.8660895
13. Bansal, M., Jyoti.: A review of various applications of low noise amplifier. In: 2017 International Conference on Innovations in Control, Communication and Information Systems (ICICCI), pp. 1–4, Greater Noida, India (2017). https://doi.org/10.1109/ICICCIS.2017.8660954
14. Bansal, M., Jyoti: CMOS LNA for BLE applications. Int. J. Eng. Technol. Sci. Res. **4**(11), 324–331 (2017). ISSN 2394-3386
15. Bansal, M., Jyoti.: Utilizing CMOS low-noise amplifier for bluetooth low energy applications. In: Malik, H., Srivastava, S., Sood, Y., Ahmad, A. (eds.) Applications of Artificial Intelligence Techniques in Engineering. Advances in Intelligent Systems and Computing, vol. 697, Springer, Singapore (2019)
16. Bansal, M., Jyoti.: Low noise amplifier in smart healthcare applications. In: 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 1002–1007, Noida, India (2019). https://doi.org/10.1109/SPIN.2019.8711705
17. https://www.techopedia.com/definition/5107/wireless-local-area-network-wlan
18. Aditi, Bansal, M.: Design, analysis, and comparison of LNA topologies for IEEE 802.15.4 Zigbee standard. In: 2017 Conference on Information and Communication Technology (CICT), pp. 1–7, Gwalior (2017). https://doi.org/10.1109/INFOCOMTECH.2017.8340597
19. Aditi, Bansal, M.: High linearity and high input impedance matching common gate CMOS LNA in 2.4 GHz ISM band. In: 2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE), pp. 90–94, Noida (2017). https://doi.org/10.1109/RDCAPE.2017.8358246

20. Lee, P. W., Chiu, H. W., Hsieh, T. L., Shen, C. H., Huang, G. W., Lu, S. S.: A SiGe low noise amplifier for 2.4/5.2/5.7 GHz WLAN applications. In: IEEE International Solid-State Circuits Conference, pp. 364–365 (2003)
21. Bansal, M., Aditi.: A high linearity and low noise shunt resistive feedback UWB LNA. In: 2017 Conference on Information and Communication Technology (CICT), pp. 1–5, Gwalior (2017). https://doi.org/10.1109/INFOCOMTECH.2017.8340596
22. Iyer, M., Shanmuganantham, T.: LNA design for WLAN applications. In: IEEE International Conference on Circuits and Systems, pp. 319–323 (2017)
23. Sivonen, P., Vilander, A., Pärssinen, A.: A gain stabilization technique for tuned RF low-noise amplifiers. IEEE Trans. Circuits Syst. **15**, 1702–1705 (2004)
24. Bansal, M., Singh, D.: LNA for neural applications. Int. J. Comput. Math. Sci. **6**(11), 74–81 (2017). ISSN 2347-8527
25. Bansal, M., Singh, D,: Low noise amplifier in bluetooth and bluetooth low energy (BLE) application. In: National Conference on Emerging Trends in Electronics and Communication (ETEC-2019), ISSN 0975-9514, pp. 110–113 (2019)
26. Khosravi, H., Zandian, S., Bijari, A.: A low power, high gain 2.4/5.2 GHz concurrent dual-band low noise amplifier. In: IEEE Annual computing and communication workshop and conference, pp. 788–792 (2019)
27. Bansal, M., Singh, D.: Design and implementation of low noise amplifier in neural signal analysis. In: Gani, A., Das, P., Kharb, L., Chahal, D. (eds.) Information, Communication and Computing Technology. ICICCT 2019. Communications in Computer and Information Science, vol. 1025. Springer, Singapore (2019)
28. Sadowy, J., Tellie, I., GraffeUil, J., Tournier, E., Escotte, L., Plana, R.: Low noise, high linearity, wide bandwidth amplifier using a 0.35 pm SiGe BiCMOS for WLAN applications. In: IEEE Radio frequency Integrated Circuits Symposium, pp. 217–220 (2002)
29. Bansal, M., Singh, D.: Cascode common source LNA with inductive degeneration topology utilizing different output matching circuits in 45 nm CMOS technology. In: 4th International Conference on Communication and Electronics Systems (ICCES-2019), pp. 594–598 (2019)
30. Tsang, T. K. K., El-Gamal, M. N.: Dual-band sub-1 V CMOS LNA for 802.11a/B WLAN applications. In: International Symposium on Circuits and Systems, pp. 217–220 (2003)
31. Kao, C. H., Chiang, Y. T., Yang, J. R.: A concurrent multi-band low-noise amplifier for WLAN/WiMAX applications, In: IEEE International Conference on Electro/Information Technology, pp. 514–517 (2008)
32. Chi, B., Shi, B.: A 1.8 V 2.4 GHZ CMOS on-chip impedance matching low noise amplifier for WLAN applications. In: International Symposium on Circuits and Systems, pp. 189–191 (2003)
33. Bansal, M., Srivastava, G.: High linearity and low power cascode CMOS LNA for RF front-end applications. In: Proceedings of Third International Conference on Intelligent Computing and Control Systems (ICICCS-2019), pp. 1702–1705 (2019)
34. Eslamifar, O., Shirazi, R. F.: Design a dual-band low-power CMOS low noise amplifier for use in WLAN applications. IN: 22nd Iranian Conference on Electrical Engineering, pp. 101–105 (2017)
35. Bansal, M., Srivastava, G.: Design and implementation of LNA for biomedical applications. In: Pandian, A., Ntalianis, K., Palanisamy, R. (eds.) Intelligent Computing, Information and Control Systems. ICICCS 2019. Advances in Intelligent Systems and Computing, vol. 1039. Springer, Cham (2020)
36. Lu, L. H., Wang, Y. S.: A compact 2.4/5.2 GHz CMOS dual-band low-noise amplifier. IEEE Microw. Wirel. Compon. Lett. 15, 685–687 (2005)
37. Bansal, M., Srivastava, G.: A high linearity & high stability cascode CMOS LNA for RF front-end applications. Int. J. Adv. Res. Comput. Commun. Eng. **8**(7), 49–54 (2019)
38. http://www.icomamerica.com/en/products/network/wlan/wlan/wlan-network.aspx
39. Bansal, M., Singh, D.: Different input impedance matching circuits for cascode common source LNA with inductive degeneration topology in 45 nm CMOS technology. In: 4th International Conference on Communication And Electronics Systems (ICCES-2019), pp. 589–593 (2019)

Indoor Mobile Robot Path Planning Using QR Code

Bhusapalli Dhamodar Reddy and A. A. Nippun Kumaar

Abstract Mobile robots that addresses the automation needs are extensively used in service-based environments like warehouses, hotels, hospitals, restaurants etc. Path planning to reach destination from a source is a crucial task for any mobile robot. An idle path planning algorithm given a source and destination, should plan a shortest path with less computation time autonomously. In this work Quick Response (QR) code is used as via point/landmark in the environment. QR code is a 2D barcode which stores more data than a traditional barcode system. Path matrix which contains navigation information to reach nearby via points or destinations is stored in QR code. These QR codes are placed on the floor strategically, such that given a destination, a mobile robot can navigate in a shortest route possible without any overall environment information. The proposed system is implemented and tested using an in-house built omnidirectional mobile robot powered by BeagleBone Black. The obtained results proves that the mobile robot is able to reach its destination point accurately.

Keywords Mobile robot · Path planning · Navigation · QR code · BeagleBone black

1 Introduction

Today, mobile robots are widely used in indoor environments. A robot is mobile when it can move in the environment without a fixed structure. Mobile robots have exceptions when comes to the payload like sensors they carry, computational capacity and the limitation of energy they carry to drive the robot. The essential component of mobile robot in indoor environments is path planning.

Path planning algorithm in mobile robot given a source and destination, should be able to plan a path in that environment to make the robot reach its desired destination point. Simple and effective navigation system is essential for any indoor mobile robot. An idle path planning algorithm for a mobile robot should plan a shortest path

---

B. D. Reddy (✉) · A. A. N. Kumaar
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bengaluru, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_65

with less computation time autonomously. In order to plan a path, the path planning algorithm requires environment information.

QR code is a 2D barcode which stores more data than a traditional barcode system. The advantages of using QR codes over the other alternative technologies such as RFID tags are easy localization, economical and minimal maintenance.

In this work, an idle path planning algorithm for indoor mobile robots using QR codes is proposed. Path matrix which contains navigation information to reach nearby via point or destination is stored in QR codes. These QR codes acts a via point/landmark and these are placed on the floor strategically. Mobile robot is equipped with a camera to capture the QR codes present in floor, the decoded data from the QR codes is used by path planning algorithm to get the environment information. Thus, path planning algorithm will be able to plan and reach the desired destination point.

The rest of paper is organized in sections as follows.

Section 2 presents a brief background of current technologies and summary of related work, Sect. 3 explains about the implementation of the system, Sect. 4 discuss about the results followed by conclusion in Sect. 5.

2 Literature Survey

To navigate in indoor environment mobile robot needs a Path planning algorithm. Path planning algorithm requires environment information for planning a path, this information for a robot is provided using RFID and QR code [1]. RFID is used to obtain the position of destination point and QR code [2] is used to align the robot so as it reaches the desired destination point. To obtain and navigate through the shortest path A* algorithm is used [3]. A* algorithm is typically used in all the path planning algorithms. A* algorithm is a heuristic search algorithm searches all the paths in the surroundings [4]. After every search, algorithm evaluates the searches and picks up the best path from the search results which has the shortest path. This search and picking up process is repeated till the path to the destination point is obtained. One-way channel mechanism is implemented hence the robot doesn't need any collision avoidance mechanism. Since all the robots will be navigating in a single channel.

Usage of QR codes as landmarks for mobile robot localization and navigation is proposed in [5]. QR codes contains the data of coordinates of the position, with the help of that localization and navigation is achieved [6]. Dijkstra algorithm is used for path planning based on available paths [7]. Mobile robot is equipped with laser ranger to find the obstacles in the line of sight of the robot so that robots doesn't collide with the other static objects present in the indoor environment and with the other mobile robots operating in that area. Mapping of the obstacles detected by laser ranger is done and those data is stored so that any other robots can be used. This approach using QR codes made the mobile robots to read the QR codes faster and accurately thus resulting in reading the QR codes without reducing the speed of the mobile robot. There exists a critical need of cost-effective, reliable and accurate solutions which helps the mobile robots to navigate in an indoor environment [8]. Artificial labels like QR codes address those challenges.

Two-dimensional codes were used for navigation of Autonomous Ground Vehicle [9]. In this work QR codes are used to achieve autonomous navigation system for AGV. The QR codes are positioned on the floors and the positioning information is stored in those. The data obtained from QR code is used by the AGV to navigate [10]. The QR codes are also consist of the obstacle's location near to that QR code which will help AGV to navigate without any collisions with the obstacles.

Usage of artificial labels to help the mobile robots to navigate in the environments [11, 12]. When mobile robots not only need structural information of the environment to navigate but also, they need to have deeper knowledge in order to have high degrees of autonomy and intelligence. These artificial labels like QR code and barcodes have tendency to store the data required by the mobile robot to possess deeper knowledge [13] about the environment thus helping the mobile robots in its navigation.

Localization of the robot in the environment is needed for any mobile robot to obtain its position relative to its environment. It is obtained with the help of artificial labels, particularly "QR code" because it is suitable for mobile robot applications and its robustness against the noise [14]. QR code is encoded not only its position in its state space coordinate system, but also direction (i.e. normal vector) and physical size [15]. Using those data from the QR code the robot can compute relative posture and position of the 2D code without any additional data.

3 Implementation

3.1 Architecture of Proposed System

Figure 1 shows the overview of the architecture of proposed system.

A USB camera module which is used to get the QR code is connected to beagle bone black. The captured images by the camera module are interpolated to 25 Mega Pixels for image quality. The controllable pins of the motors present in the motor driver board are connected to the GPIO pins of the BeagleBone black. BeagleBone black is well suited for embedded applications due to its high performance and low power consumption. 12 V DC power supply is connected to the motor driver board to provide the required power to the 4 DC motors which are connected to motor driver board. BeagleBone black is powered by a 5 V power supply.

3.2 Flow Diagram of the System

Figure 2 refers to flow diagram of the implemented system.

The system initializes with all the required packages and gets the user provided destination point. The robot moves forward till it finds a QR code. If a QR code

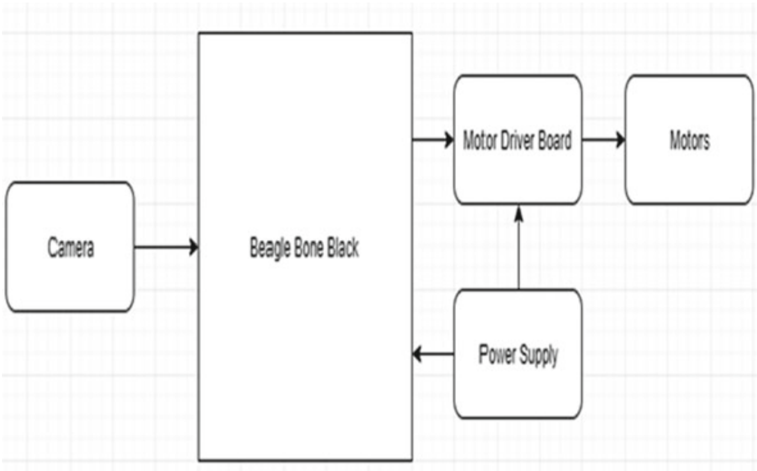


Fig. 1 Architecture of proposed system

is detected mobile robot will align its position with the QR code and obtains the QR code data. If the destination point information is present in the obtained data, the mobile robot will navigate to destination point based on that data. If destination point information is not present, it will pick the shortest path details and navigates to the nearest via point. This process is repeated until it reaches its destination point.

QR code contains navigation information of the nearby via points or destinations. This is stored as a matrix, every row in the matrix contains path information to reach a particular point. Equation (1) shows the matrix format.

$$Vqrno, \text{destinations} = \begin{bmatrix} V1/D1 & FD1 & BD1 & LD1 & RD1 & D1, D2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Vn/Dn & FDn & BDn & LDn & RDn & Di \end{bmatrix} \quad (1)$$

Generic naming convention is followed for naming the QR code, where V stands for Vertex followed by the unique number of that vertex and nearest destination points separated by commas.

Here,

Vi = Nearest QR code
Di = Nearest destination point
FDi = Forward Distance in centimeter
BDi = Backward Distance in centimeter
LDi = Left Distance in centimeter
RDi = Right Distance in centimeter
$i = 1$ to n
n = Number of near points

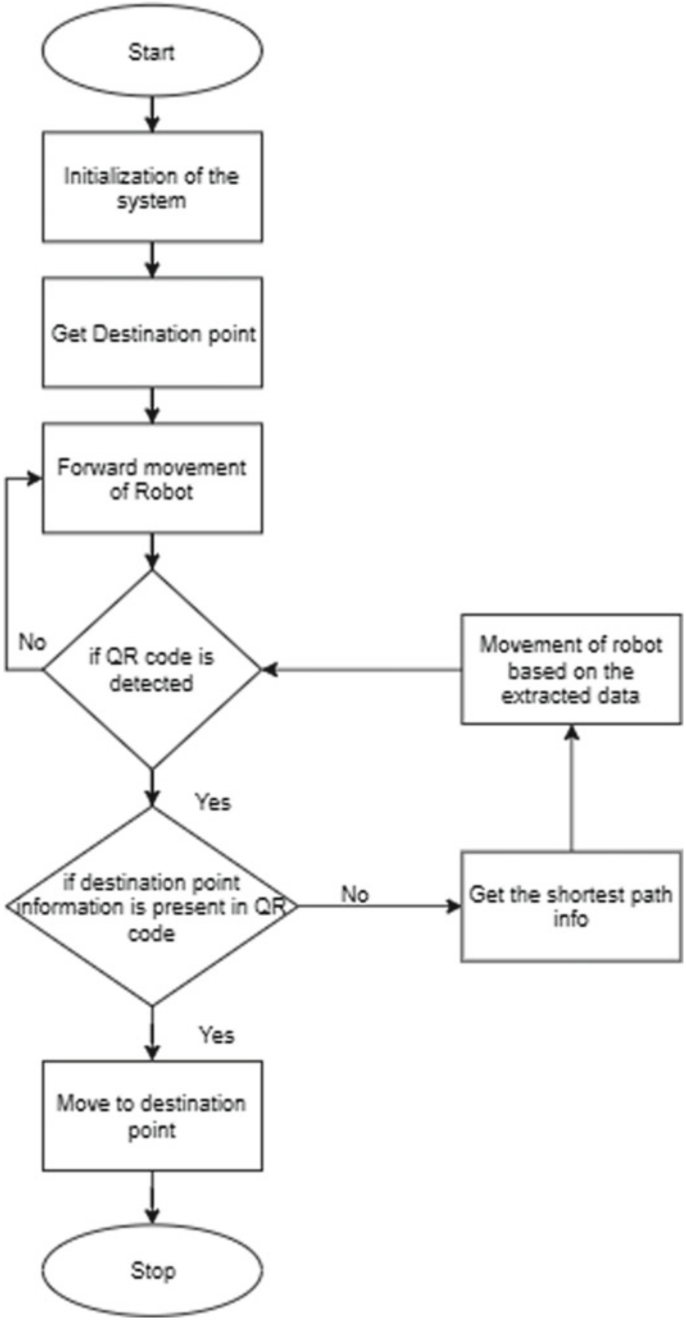


Fig. 2 Flow diagram of the system

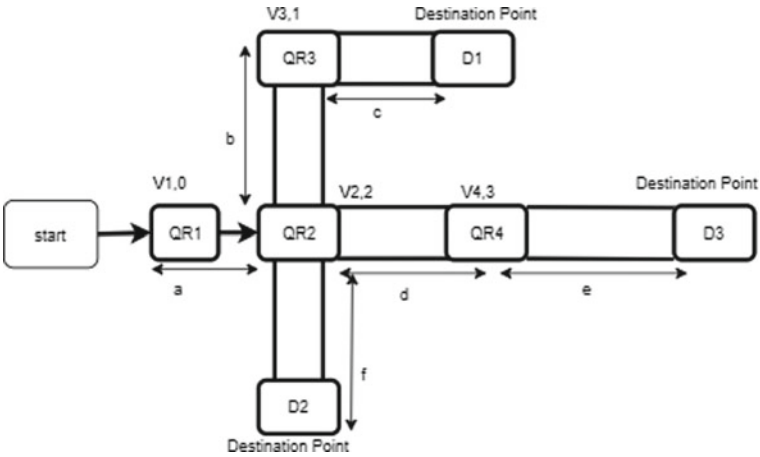


Fig. 3 Example scenario

3.3 Proposed Path Planning Algorithm

In the proposed path planning algorithm, the mobile robot starts from its initial point. The destination point is given by the user. The mobile robot starts to move till it finds a QR code in forward motion once it detects the QR code the robot will align its position and with the obtained path matrix data present in QR code, the path planning algorithm will plan a path to navigate.

Figure 3 shows an environment with three destinations and four QR codes as via points. Data in the QR code is given in Eqs. (2–5)

$$V1,0 = \begin{bmatrix} V2 & a & 0 & 0 & 0 & D2 \end{bmatrix} \tag{2}$$

$$V2,2 = \begin{bmatrix} D2 & 0 & 0 & f & 0 & 0 \\ V4 & d & 0 & 0 & 0 & D3 \\ V3 & 0 & 0 & b & 0 & D1 \\ V1 & 0 & a & 0 & 0 & 0 \end{bmatrix} \tag{3}$$

$$V3,1 = \begin{bmatrix} D1 & c & 0 & 0 & 0 & 0 \\ V2 & 0 & 0 & 0 & b & D2 \end{bmatrix} \tag{4}$$

$$V4,3 = \begin{bmatrix} D3 & e & 0 & 0 & 0 & 0 \\ V2 & 0 & d & 0 & 0 & D2 \end{bmatrix} \tag{5}$$

Case 1: Robot needs to reach destination point $D1$ from start point

In this case the robot starts to move from its initial stating position in forward motion until it obtains the first QR code $V1$,0. Once it detects $V1$,0 it decodes the data to obtain the data as shown in Eq. (2). Since the there is no info related to destination point $D1$ in the captured QR code $V1,0$, the robot picks the only available path to $V2$. The robot travels 'a' distance in forward motion to reach the next QR $V2,2$. Then it obtains the encoded data in $V2,2$ as shown in Eq. (3). The destination point $D1$ can be reached from $V3,1$ hence the robot travels to $V3,1$ by moving 'b' distance left. Once it reaches $V3,1$ it obtains the data as shown in Eq. (4), then the destination point $D1$ is reached by moving 'c' distance forward.

Case 2: Robot needs to reach destination point $D3$ from start point

In this case the robot starts to move from its initial stating position in forward motion until it obtains the first QR code $V1,0$. Once it detects $V1,0$ it decodes the data to obtain the data as shown in Eq. (2). Since the there is no info related to destination point $D3$ in the captured QR code $V2,2$, the robot picks the only available path to $V2$. The robot travels 'a' distance in forward motion to reach the next QR $V2,2$. Then it obtains the encoded data in $V2,2$ as shown in Eq. (3). The destination point $D3$ can be reached from $V4,3$ hence the robot travels to $V4,3$ by moving 'd' distance forward. Once it reaches $V4,3$ it obtains the data as shown in Eq. (4), then the destination point $D3$ is reached by moving 'e' distance forward.

Case 3: Robot needs to reach destination point $D2$ from start point

In this case the robot starts to move from its initial stating position in forward motion until it obtains the first QR code $V1,0$. Once it detects $V1,0$ it decodes the data to obtain the data as shown in Eq. (2). Since the there is no info related to destination point $D2$ in the captured QR code $V2,2$, the robot picks the only available path to $V2$. The robot travels 'a' distance in forward motion to reach the next QR $V2,2$. Then it obtains the encoded data in $V2,2$ as shown in Eq. (3). The destination point $D2$ can be reached by moving 'f' distance right.

4 Results

Acrylic sheet is used for building the mobile robot chassis. A USB camera is mounted on the robot chassis. The four Dc motors are coupled with the Omni wheels which helps with forward and sideways mobility. The motors are driven by an H-Bridge module for precise control. BeagleBone black is mounted on top of chassis which controls the movement of mobile robot. A 12 and 5 V power supply is given to motors and BeagleBone black respectively (Fig. 4).

The advantage of using QR code is, it can be used for position alignment of mobile robot since robots tends to move out from actual path due to hardware manufacture issues. QR codes acts as a reference point (Fig. 5).

The robot is able to align its position with respect to the QR code by obtaining the QR code coordinates. The four coordinates of the captured QR code are calculated with the help of OpenCV then the mismatch distance of QR code in the frame from its intended position is calculated in pixels and it is then converted to centimeter so as to align the robot with help of obtained distance data in centimeter.

Figure 6 shows an environment with four destination points and nine QR codes as via points.

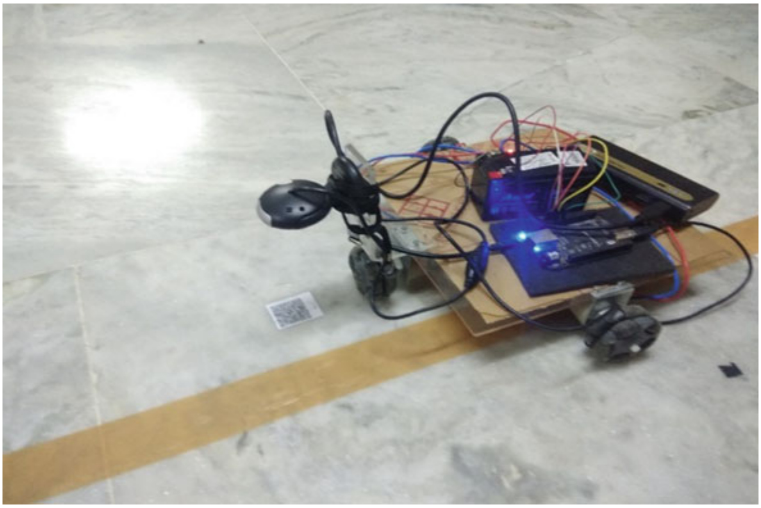


Fig. 4 Built mobile robot powered by BeagleBone black

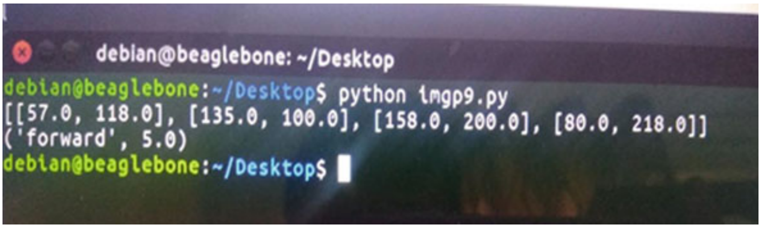


Fig. 5 Misplacement distance forward of 5 pixels

In Fig. 7, the robot started from its starting position. It moves forward until it reaches the QR $V1,0$ then align its position with respect to the captured QR code $V1,0$. QR $V1,0$ contains only the one via point QR $V2,0$ details hence the robot picks the only available path to reach the via point $V2,0$. Once it reaches $V2,0$ it adjusts its position with respect to the QR code and obtains the destination point details from $V2,0$ QR code, since its nearest QR $V3,1$ have the destination point $D1$ details, the robot navigates to $V3,1$ with the help of data present in $V2,0$. Once it reaches the robot is able to reach its destination point $D1$ with the help of data present in $V3,1$. In this way the robot is able to successfully navigate to the destination point $D1$ from its initial starting position.

In Fig. 8, the robot starts to navigate from the QR $V1,0$ to reach its destination point $D5$. QR $V1,0$ contains only the one via point QR $V2,0$ details hence the robot picks the only available path to reach the via point $V2,0$. Once it reaches the via point $V2,0$ it adjusts its position and obtains the data of nearest via points and destination can be reached from that via point. Destination point $D5$ can be reached from the

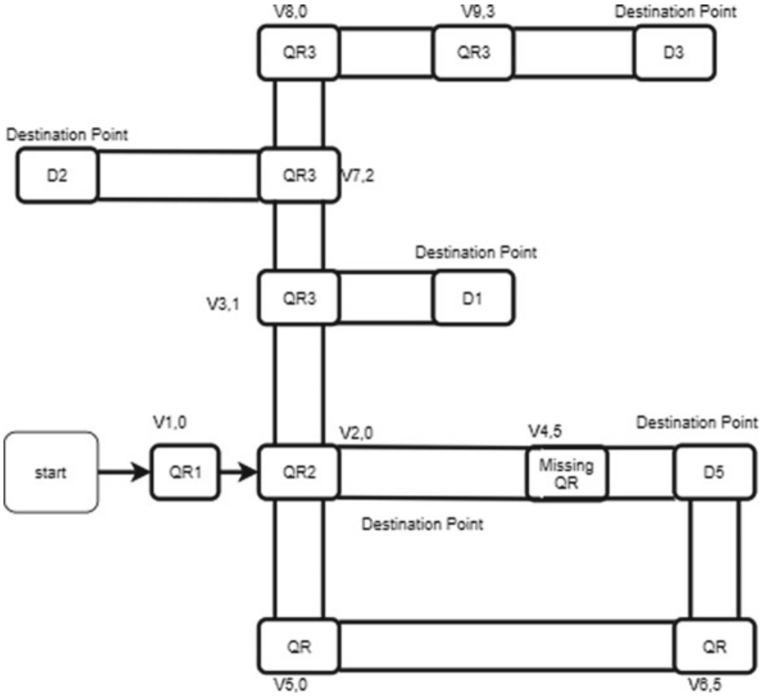


Fig. 6 Test environment

nearest via point $V4,5$, hence the robot picks that path to reach $V4,5$. Robot travels to reach QR $V4,5$ but the via point $V4,5$ QR code is missing in that point, hence to have a path planning algorithm robust against these kind of scenarios the algorithm is designed to handle such scenarios to move back to the previous via point. In this case the robot travels back to via point $V2,0$ then picks the other path to reach via point $V5,0$ from which the robot travels to the next via point $V6,5$. Destination point $D5$ is reached using the path details present in $V6,5$. This proves the proposed path planning is robust to the missing QR codes or damaged codes.

In Fig. 9 robot starts to navigate until it finds the QR $V1,0$. The destination points for the robot to navigate in this scenario are $D1$, $D2$, $D3$. The first destination it needs to reach is destination point $D1$, thus when it finds the QR $V1,0$ it picks the only available path to via point $V2,0$. $D1$ can be reached from via point $V3,1$ hence the robot picks the path to reach via point $V3,1$. Once it reaches via point $V3,1$ it gets the destination point $D1$ path details and travels to reach the destination point $D1$. Once it reaches $D1$ then the next destination point it need to reach is $D2$ so in order to navigate to $D2$ point, the robot travels back to previous QR $V3,1$. In $V3,1$ the robot obtains the path details for the destination point $D2$ which can be reached from via point $V7,2$. Once it reaches QR $V7,2$ it obtains the path details to navigate to $D2$ point and navigates. Once it reaches $D2$ the next destination point it need to travel is $D3$, so the robot travels back to QR $V7,2$ to obtain the destination point $D3$ path

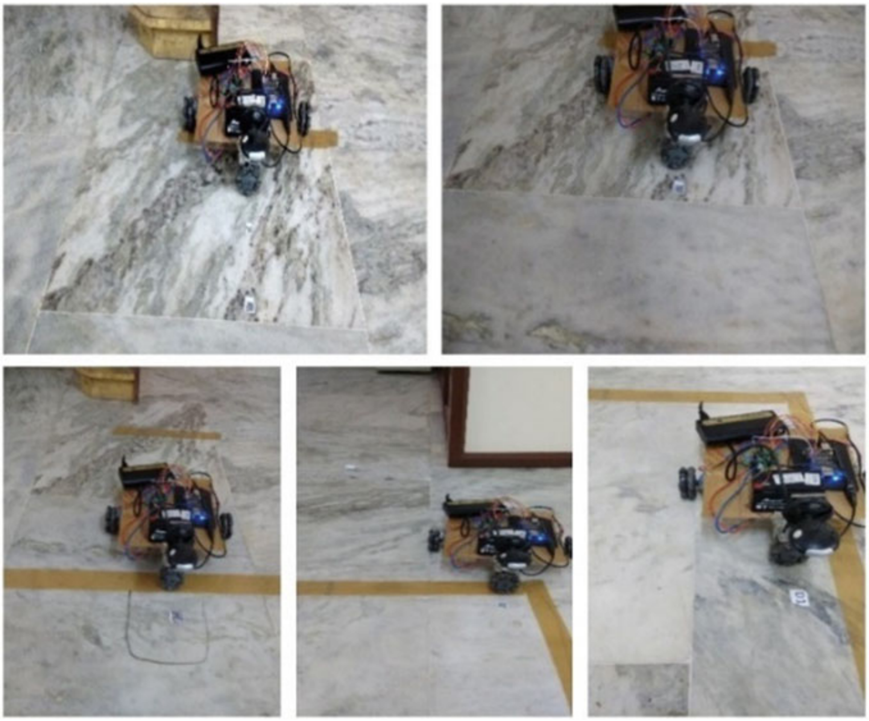


Fig. 7 Robot navigation to destination point $D1$ from start point

details. QR $V7,2$ doesn't contains the $D3$ path details, hence the robot navigates to nearest via point 8,0. In QR $V8,0$ the robot obtains the path details of QR $V9,3$ from which the destination point $D3$ can be reached. The robot travels to QR code $V9,3$, adjusts its alignment with respect to QR code and travels to reach its final destination point $D3$.

5 Conclusion

A path planning algorithm using QR code as via points is proposed for the navigation of mobile robots in an indoor environment. The proposed algorithm uses QR codes to get the environment information for the navigation. Proposed path planning is implemented successfully and tested using an inhouse built mobile omnidirectional robot powered by BeagleBone black. The algorithm is tested for multiple scenarios and the mobile robot is able to navigate to single or multiple destination points accurately. The algorithm is robust to missing QR codes or damaged codes at the

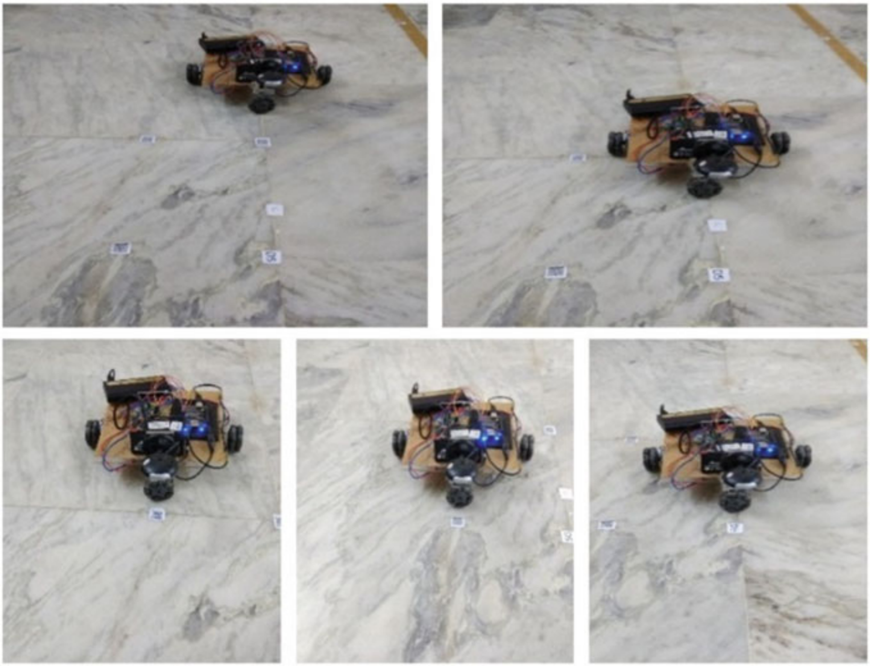


Fig. 8 Robot navigation to destination point with a missing via point in its path

desired points in the system. The work can be enhanced by equipping the built mobile robot with wheel encoders for accurate position control and enhancing the positional alignment of the robot with respect to the QR code in both hardware and software areas.

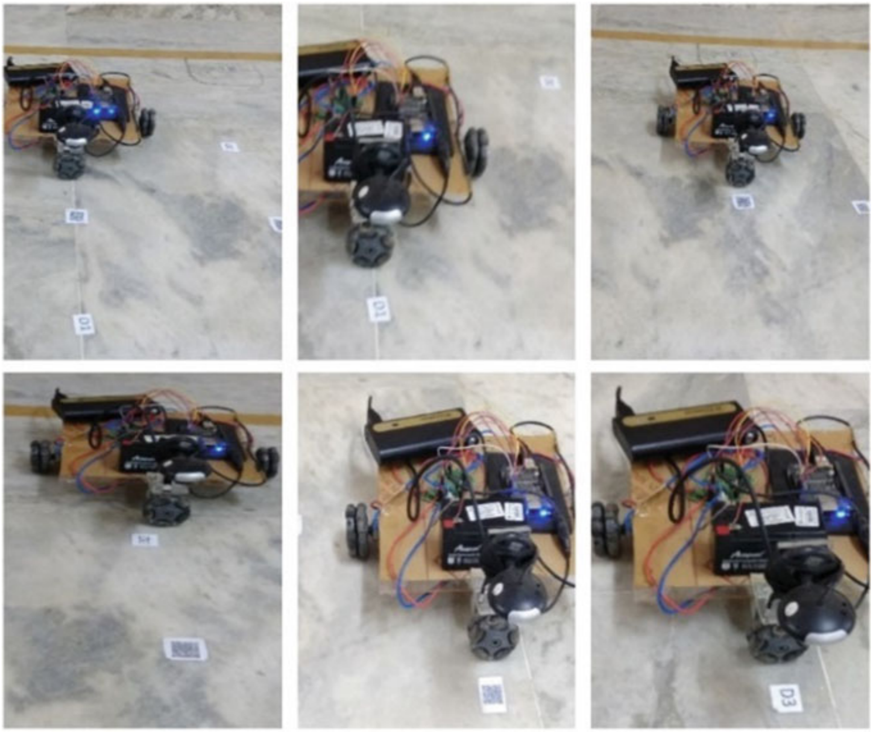


Fig. 9 Robot navigation to multiple destination points

References

1. Xiao-Long, W., Chun-Fu, W., Guo-Dong, L., Qing-Xie, C.: A robot navigation method based on RFID and QR code in the warehouse. In: 2017 Chinese Automation Congress (CAC), Jinan, pp. 7837–7840 (2017)
2. Zhang, H., Zhang, C., Yang, W., Chen, C. Y.: Localization and navigation using QR code for mobile robot in indoor environment. In: 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, pp. 2501–2506 (2015)
3. Teja, P. R., Kumaar, A. N.: QR Code based Path Planning for Warehouse Management Robot. ICICCS, (2018)
4. Lin, M., Yuan, K., Shi, C., Wang, Y.: Path planning of mobile robot based on improved A∗ algorithm. In: 2017 29th Chinese Control and Decision Conference (CCDC), Chongqing, pp. 3570–3576 (2017)
5. Joshy, P., Supriya, P.: Implementation of robotic path planning using ant colony optimization algorithm. In: International Conference on Inventive Computation Technologies (ICICT), vol. 1, IEEE (2016)
6. Li, Z., Huang, J.: Study on the use of Q-R codes as landmarks for indoor positioning: Preliminary results. In: 2018 IEEE/ION Position, Location and Navigation Symposium (PLANS), Monterey, CA, pp. 1270–1276 (2018)
7. Narayan, S., Prasannakumar, T. V., Prasanth, M., Velavan, N. P., Kumar, N. P., Pillai, A. S.: A priority based exploration algorithm for path planning of an unmanned ground vehicle. In: 2014 International Conference on Embedded Systems (ICES), Coimbatore, pp. 275–280 (2014)

8. Quilez, R., Zeeman, A., Mitton, N., Vandaele, J.:. Docking autonomous robots in passive docks with Infrared sensors and QR codes. In: International Conference on Testbeds and Research Infrastructures for the Development of Networks Communities (TridentCOM), Vancouver, Canada Jun 2015
9. Dutta, V.: Mobile robot applied to QR landmark localization based on the keystone effect. In: Zhang, D., Wei, B. (eds.) Mechatronics and Robotics Engineering for Advanced and Intelligent Manufacturing. Lecture Notes in Mechanical Engineering, Springer, Cham (2017)
10. Cherni, F., Boutereaa, Y., Rekik, C., Derbel, N.: Autonomous mobile robot navigation algorithm for planning collision-free path designed in dynamic environments. In: 2015 9th Jordanian International Electrical and Electronics Engineering Conference (JIEEEC), Amman, pp. 1–6 (2015)
11. Zhou, C., Shuai, P., Dai, C.: The application of QR codes and WIFI technology in the autonomous navigation system for AGV. Advances in Engineering Research, pp. 2352–5401 Jan (2018)
12. Hao, W., Guohui, T., Chunwei, P., Xin, W., Jinshan, Y.: Three hierarchy map building based on artificial label. Open Cybern. Syst. J. **9**, 2834–2841 (2015)
13. Lee, S. J., Lim, J., Tewolde, G., Kwon, J.: Autonomous tour guide robot by using ultra-sonic range sensors and QR code recognition in indoor environment. In: IEEE International Conference on Electro/Information Technology, Milwaukee, WI, pp. 410–415 (2014)
14. Kobayashi, H.: A new proposal for self-localization of mobile robot by self-contained 2D barcode landmark. In: 2012 Proceedings of SICE Annual Conference (SICE), Akita, pp. 2080–2083 (2012)
15. Tian, W.: The research into methods of map building and path planning on mobile robots. In: 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, pp. 1087–1090 (2017)

A Novel Privacy Preservation Scheme for Internet of Things Using Blockchain Strategy

Dolagobinda Samal and Rajakumar Arul

Abstract Internet of Things (IoT), the modern approach of science and technology, is currently experiencing a rapid evolution in the area of research and industry, but it still suffers from various security issues [1]. In an IoT environment, privacy is the prime concern, and the information is being handled mainly through the decentralized topology [2]. The commonly utilized methods for communicating and fetching data are carried out through the Internet by utilizing different types of smart devices that have the ability to go around the Internet into an unsafe platform. It is due to the receptiveness and heterogeneousness humor of IoT terminals, where they are uncovered to attackers by making its user's identity easily traceable and for the most part, is not secured-by-design [3]. It is essential for an upgradation, to fix their vulnerabilities and to anticipate attackers from enlisting them into botnets. This paper is based on the relevant security issues and confidential challenges upon various vulnerabilities of Internet services for IoT devices by presenting some typical attack cases with an outlook on possible solutions that are carried out in a systematic literature review. We have focused on the privacy issue that prevails in the IoT environment and thereby proposed an appropriate security model through Blockchain strategy for establishing a secure and reliable communication among peers.

Keywords Attacks · Blockchain · Botnet · Data integrity · Denial of service · IoT · Security

D. Samal (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru, India

R. Arul
Amrita Vishwa Vidyapeetham, Bengaluru, India
e-mail: rajakumararul@ieee.org

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_66

1 Introduction

1.1 IoT and Attacks

IoT has been around us for a while; It is the modern-edge innovation having the capacity to associate communicate and controlling gadgets remotely through the Internet [4]. The word "Internet of Things" is an organization of objects connected to the Internet which can collect data from the environment and also exchange information, mentioned as "IoT." Nowadays, from smartphones, watches and cars to even smart-homes and many more connected technologies are gradually transforming our lives certainly for the better [5]. At present, Internet of Things (IoT) is rapidly evolving in our lives and our businesses in a countless ways having the potential to become mainstream for everyday tasks to make people's lives easier, and the numbers prove it: in Fig. 1, from [5] while in 2005, there were over 3 billion people and 600 million devices associated with the Internet and later is likely to grow to 80 billion connected devices by the year 2022 was estimated.

The IoT space is evolving and so the attack surface of IoT devices [6]. It is been more than two years, there was a threat of cyber attacks in 2016 which targeted the IoT devices moved from theoretical to actual. In the same attack, tens of thousands of poorly secured IoT devices were manipulated to cause several distributed denial of service (DDoS) by crippling high volumes of traffic to targeted Web sites. In the attack of September 2016, the attacker simulated a botnet with enlisted around 150,000 devices, programmed to send 1 TB of data per second at the victim's servers. Generally, routers and connected cameras make up 80 percent of infected devices, and researchers suggest that every IoT end device is vulnerable, from smart bulbs, smart home products, IoT-enabled vehicles to voice assistants. Mostly, the majority of IoT attacks have occurred in the DDoS realm. According to Symantec's ISTR reports here are three forms of DDoS-attack—LightAidra, Kaiten and Mirai—collectively reasoned for nearly 80% of the IoT attacks in the year 2018 [7]. The cyber researchers from F-secure have issued a warning that cyber attacks on IoT nowadays increasing

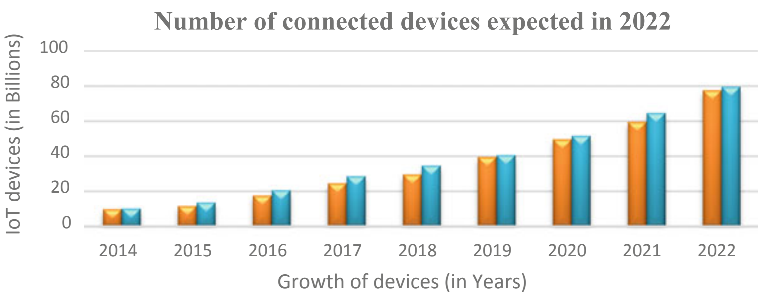


Fig. 1 Internet of Things (IoT): Number of connected devices worldwide from 2014 to 2022(in billions)

rapidly. The company's "Attack Landscape H1 2019" measured a three-fold increase in attack traffic to more than 2.9 billion events [8]. In Imperva's recent report in a duration of 13 days, more than 400,000 infected devices utilized for botnets. The Mirai botnet, appeared in 2016 produced 292,000 requests per minute to the victim's server associated DDoS with it [9]. IoT devices need to be patched and updated regularly, which itself a challenge [6]. As of now, IoT is in its earliest stages, and subsequently, more consideration ought to be paid on issues such as security and privacy of the IoT information. The vulnerabilities and loopholes in the IoT systems require seriously attention from the research and advancement community [10]. Because of the receptiveness, universality and multi-source heterogeneity of the IoT devices, moreover, it faces genuine security issues and that should be fathomed earnestly. So, there should be step-by-step directions to exploit rising innovations to take care of the issues of vulnerability, reliability and data protection for the advancement of the IoT [3]. In this paper, the key commitment of our work is to provide an appropriate security framework based on the leading-edge technologies which is doable for the IoT nodes carrying valuable information and to safeguard and verifying based on tracking of the information, by removing of a central, assured middle man within two or more IoT end devices.

The paper is cataloged as follows. The related work regarding different security issues in the IoT framework in Sect. 2.1 and some recent researches toward the security solution are discussed in detail in Sect. 2.2. An overview of the Blockchain technology briefed in Sect. 2.3, and a clear problem statement defined in Sect. 2.4. Section 3 describes our proposed work with a novel security framework in Sect. 3.1 and implementation details in Sect. 3.2. Section 4 analyzes our results. The concluding statements from our research with a future outlook are summarized in Sect. 5.

2 Related Work

2.1 Review Stage 1: Security Issues in Various Layers of IoT

IoT is based on four simple building blocks, one of the parts is sensors and actuators which measure a physical amount from the environment and convert them into electrical data for the framework to understand and act accordingly. Then come the framework and the gateway which interact with the Internet and act as a medium between the sensor nodes with the Internet. Information mostly collected from the sensor nodes and transmitted over the Internet through the gateway. Next comes the Server which acts as a backbone for an IoT network. Routers, back-end processing node, bridges and different sort of applications are acting like components of a server which manages things like data acquisition, device management and configuration, processing and storing the information inside the cloud servers where the data can be utilized in various business analytics. Lastly comes the user interface which offers assistance to clients for controlling their IoT devices [11]. The open three-tier architecture of the Internet of Things analyzes most of the security questions confronting each layer. The perception layer is the establishment of IoT architecture. The perceived nodes exhibit multi-source heterogeneous data and generally lack of effective monitoring which makes most of the terminals exposed over the Internet and more vulnerable to security threats. There is the possibility that the attackers can use the security flaws of the IoT terminals to obtain the original identity and sensitive information of the user, and using those fake identities, the attackers communicate with other nodes to perform illegal actions or malicious attacks. The network interconnected layer gives access to the sensor layer for transmission of data to the IoT service platform. A small amount of data is transmitted from the IoT devices, and generally, complex algorithms are not used for the protection of data which results in various attacks, tempering and destruction of data. When data is transferred between IoT devices, there is a possibility of occurring of vulnerabilities such as issue of authentication, key mismatch, confidentiality of information, data protection issues and increase security hazards for the network layer. The application layer provides an interface between the user and the system. The application layer is directly facing the outside world, and it is the most sensitive and risky part. The commonly utilized authentication method adopted by the application layer of IoT is by message verification code when the sender and the receiver determine for the communication. However, during the communication process, the attackers try for a faulty verification by hacking the devices, which leads to some common risks such as spear phishing, social engineering, spoofing and Web application attacks to steal a person's identity. These attacks are often successful by making an assumption of being in a trusting organization [12]. Due to the heterogeneity nature of the IoT structures, there is a need for a proper security plan for current IoT devices. The current operating environment of IoT devices is a traditional centralized system that requires an assured middle man to safeguard the data retrieved from all gadgets. The whole IoT system requires a decentralized operating environment, and a trustful new mechanism needs to be established to maintain consensus among devices and ensure the overall operation of the system.

Attacks on IoT surface. Here, in Table 1, different types of security issues and vulnerabilities are addressed to aware mankind about possible threats in the IoT environment, and some open research challenges for security and some key points are addressed in the Sect. 2.2 are like authenticity, data integrity, lightweight mechanisms and upgradation of security software for the IoT devices to avoid getting hacked and attacks like denial of service, gateway or portal attacks, Ransom-ware attacks and to resolve unauthorized access in the network [13].

Table 1 Taxonomy of different security attacks possible within every layer of the Internet of Things (IoT) framework

IoT layers	Components	Possible attacks
Application layer	Personalized information service, intelligent transportation service, environmental monitoring	Cookie injection, software attacks, DoS attacks, cyber attacks, man-in-the-middle attacks
Support layer	Cloud computing, intelligent computing	Ransom-ware attacks, side-channel attacks, Sybil attack
Network layer	Internet, mobile communication, network infrastructure and communication protocols	Man-in-the-middle attacks, DoS attack, IP spoofing, routing information attacks
Perceptual layer	The RFID reader, sensor, GPS	Malicious attacks, physical attacks, DoS attacks

2.2 Review Stage 2: Security Solutions for IoT

Sicari et al. in 2016 represented a layer named NOS as a distributive middleware, which manages humorous data and also evaluates quality and security levels associated with each data, additionally proposed an algorithm for security by measuring the trust for the data of registered devices. For the high-security perspective, results were also showing good signs for future success where a key administration framework has to be presented for future work [14]. The authors S. Raza, et al. in 2016 focused on empowering secure communication among the gadgets. From the case study of the paper, a prototype is displayed which grants the devices of IoT for utilizing Datagram Transport Layer Security (DTLS). And also presented an Internet security service provider (IoTSSP) acts as a third party which offers an appointment for the handshaking and exchange of sessions to another module [15]. H. Bostani et al. in 2017 proposed a framework for the detection of the sinkhole in the IoT environment. The proposed architecture was modeled for continuous observation of the behavior of the hosting node in the network, but in the proposed solution, there is a voting method for the detection of unusual behavior of nodes. The model outruns from the existing solution, and in the future work, the performance may improve by the techniques of data mining and AI technologies [16]. The authors B. Daghighi, et al. in 2017 mentioned that protection of the device is an important factor, and there are lots of ways to prevent the device mainly giving restrictions from external unauthorized devices. By disabling, the loosely configured devices from the access of the Internet by ensuring the specified services are enabled, booting should be done securely, and in each connection, there should be establishment of applying device authentication, updating the latest security patches on the device and also while building connection implementing secure key for the security establishment [17]. From our literature work, we concluded that there is a requirement for lightweight security innovation and a system to address the security issues of the IoT. The Blockchain is leading-edge innovation that employments cryptography and an agreement to confirm exchanges,

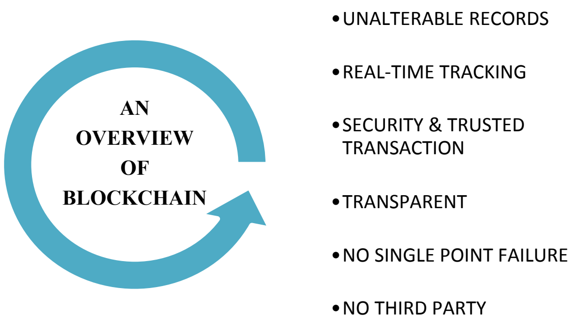


Fig. 2 An outline of Blockchain technology

which guarantees the authenticity of an exchange, avoids twofold investing and permits for high-value exchanges in a trustless environment that make this technology secure and unique and one of the unexplored regions for the IoT security. In the next section, we have discussed an overview, advantages and disadvantages of the Blockchain technology and how it could be a proper security answer for the IoT architecture for the prevention of malicious attacks that might happen in the IoT environment.

2.3 Overview of Blockchain Technology

The Blockchain technology is the leading-edge technology devised originally from digital cryptocurrency where information is available in a distributed manner could not be copied and controlled by any single entity of the block. Technically, a Blockchain is digitized, decentralized public ledger having no single point failure [18]. In the model development of the Blockchain at the point when information is transmitted starting with a single node then to another node, the starting one initially confirms the data coming into one specific node, and if the information is verified, it transmits the verified data all through the network. The decentralized feature can enhance the centralized state of the current IoT and keep the whole IoT framework from being harmed because of the obliteration of the central node [19]. For our work, Blockchain is addressed as the priority for the security of IoT devices (Fig. 2).

2.4 Problem Statement

The colossal number of uncertain Internet of Things (IoT) devices with high computation power makes them a straightforward and appealing center for the attackers. Taking advantage of heterogeneity of the IoT terminals, the attackers deal with the less secured devices to use them to create a huge scale of botnets inside the organization [20]. Once a machine is compromised, a small program (trojan, malicious

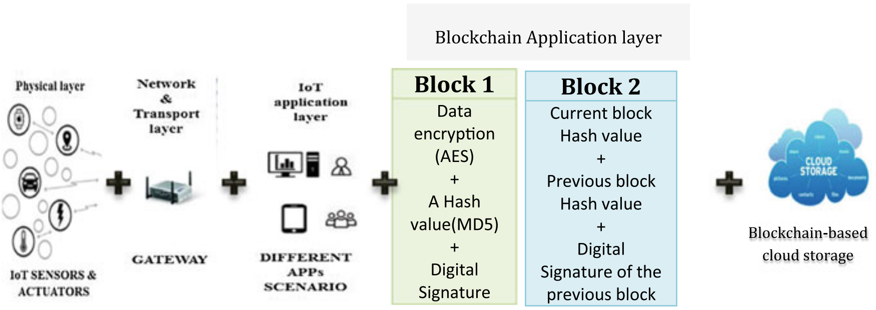


Fig. 3 Proposed architectural framework for IoT: Introducing the Blockchain layer with the basic IoT infrastructure

program or virus) is initialized by the attacker, who can instruct the bots over the network at a certain time to execute actions such as denial of service for the existing users which may cause much delay in the devices where cost and managing of the device will be more, and moreover, user satisfaction level will go down.

3 Proposed Work

3.1 Security Framework for IoT

To provide security to the IoT devices from attackers thereby looking after the malicious devices IP address over the Network by monitoring the data coming from the devices. For the malicious IP addresses, the administrator should able to block the device from accessing the whole network. In our proposed model, we are proposing a novel security concept sponsored by Blockchain-based security system with the base IoT Architecture, as shown in Fig. 3. We acquaint a Blockchain application layer with an IoT framework. Utilizing the Blockchain strategy, we can control and arrange the information generating from IoT terminals where data tempering will not possible. And also, we can able to track the malicious IP addresses which may avoid the attacker. With the implementation of Blockchain strategy, we can able to prevent the bot users from using the service.

3.2 Implementation

For the execution of our proposed plan, we have illustrated any sort of attack circumstance for IoT. To neutralize the attack and to provide security to the IoT terminals, we adopted our proposed Blockchain app layer with the basic IoT architecture. Basically, in our implementation, we have demonstrated the DoS attack case scenario by creating botnets in the IoT environment. A smartphone as an IoT end device, a real cloud for storing the real data and a local server, we have used for the uploading and retrieving data (for our case, we have taken images). In our first module, to show the attack case scenario, the IoT device is compromised to upload many images at once to the cloud through the application which causes network congestion, if the image transfer rate crosses the threshold limit, the user will be detected as a botnet in the system, and then, the device IP is added into bot list. The botnet detection system is placed in the Web server for monitoring from which device within a stipulated time interval exceeds the number of images uploaded. In the Web server, the encryption and decryption process were running. While an image is uploaded into the server, it is encrypted and stored into the cloud based on Blockchain strategy. We have created the hash value and digital signature of the data and stored it as a genesis block. Based on Blockchain strategy, the next blocks are designed to store the hash code and digital signature of the previous block, observed in Table 2. For data encryption, we have used the Advanced Encryption Standard (AES) algorithm. For the generation of the hash code of the data, we have used the MD5 mechanism.

4 Result Analysis

Our prime goal was to provide secure and reliable communication among peers. In our model where we are trying to secure the data uploaded into our cloud server. It is observed that if any bot user resides in the network and if any temper did to the F_hash_code or from Genesis_block values, the botnet users will not able to retrieve the data (images) from the server, and the user will be logged out from the application which neutralizes the attack and enhances the security. We have neutralized the botnets through the Blockchain strategy in a novel way by encrypting the data and creating its hash value or digital signature and storing them in the cloud server. From our implementation, clearly, we have proved that the Blockchain strategy could be a proper security concept for the IoT devices by implementing it with the cloud server. And it is also demonstrated that implementation of this technology can prevent the data from the botnets, so that the system will be secured.

5 Conclusion and Future Work

Here, in our work, we have addressed the privacy problems that prevail in IoT through our literature work and made a clear survey on all the current research solutions that exist for the IoT environment and found that the Blockchain is a moderately unexplored region for IoT security. In cloud computing, privacy plays a dynamic role in protecting user's data, but in multi-domain environments and service-oriented

Table 2 Database table structure: Generated hash code stored in the database based on Blockchain strategy

File-code	File-name	File-type	File-upload date	User-id	Genesis block	F-hash code	Block-Id
1001	63_40_download.jpg	.jpg	16-06-2019/ 10:35:14	1	68b7ab2cae159ca054c30f61az27a00	68b7ab2cae159ca054c30f61az27a00	BLK-100163_40_download.jpg
1002	14_save_20190622_122626.jpg	.jpg	26-06-2019/ 17:04:59	2	68b7ab2cae159ca054c30f61az27a00	d2u47shd59udf7aa38q6sa6oha3aq54t	BLK_100214_save_20190622_122626.jpg
1003	74_save_2019068_125446.jpg	.jpg	26-06-2019/ 23:16:36	1	d2u47shd59udf7aa38q6sa6oha3aq54t	Eec87s34ma97bfr3uan2kg5ga28fcg25a7	BLK_100374_save_2019068_125446.jpg

architecture like IoT, it is essential to implement a secured service [21]. We have uploaded only image files, and enhancement can be done on this project by uploading all types of files like video, audio, text, etc. In future work, we can approach a module for the idea of home automation, where the arrangement of smart home devices will run at a reasonable time restricting the power costs. Another case of an important area for the implementation is social insurance IoT applications, where the altering of information could be tragic. At the point when the applications and their security prerequisites have been built up, the subsequent stage is to assess for implementation of Blockchain innovation.

Acknowledgements D. Samal gratefully acknowledges support from Amrita School of Engineering, Amrita University for research fellowship.

Rajakumar Arul and D. Samal gratefully acknowledge support from the Department of Computer Science, Amrita Viswa Vidyapeetham, Bengaluru.

References

1. Dorri, A., Kanhere, S. S., Jurdak, R.: Blockchain in the internet-of-things: Challenges and solutions. Aug (2016)
2. Conoscenti, M., Vetrò, A., De Martin, J. C.: Blockchain for the internet of things: A systematic literature review. In: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), Agadir, pp. 1–6 (2016)
3. Boudguiga, A., et al.: Towards better availability and accountability for IoT updates by means of a blockchain. In: 2017 IEEE European Symposium on Security and Privacy Workshops (Euros & PW), Paris, pp. 50–58 (2017)
4. Rakesh, N.: Performance analysis of anomaly detection of different IoT datasets using cloud microservices. In: 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, pp. 1–5 (2016)
5. greenstechnologys.com/Source: Internet
6. zscaler.com/blogs/research/sneak-peek-recent-iot-attacks//Source: Internet
7. Davis, D. B.: www.symantec.com/blogs/expert-perspectives/istr-2019-internet-things-cyber-attacks-grow-more-diverse/Source: Internet
8. forbes.com/sites/zakdoffman/2019/09/14/dangerous-cyberattacks-on-iot-devices-up-300-in-2019-now-rampant-report-claims/#5e8bc4295892/Source: Internet
9. bankinfosecurity.com/massive-botnet-attack-used-more-than-400000-iot-devices-a-12841/Source: Internet
10. Arul, R., Raja, G., Bashir, A.K., Chaudry, J., Ali, A.: A console GRID leveraged authentication and key agreement mechanism for LTE/SAE. IEEE Trans. Industr. Inf. **14**(6), 2677–2689 (2018)
11. IoTdunia https://iotdunia.com/ Source: Internet
12. Praveena, A., Smys, S.: Anonymization in social networks: A survey on the issues of data privacy in social network sites. J. Int. J. Eng. Comput. Sci. **5**(3), 15912–15918 (2016)
13. Yaqoob, I., Ahmed, E., ur Rehman, M. H., Ahmed, A. I. A., Al-Garadi, M. A., Imran, M., Guizani, M.: The rise of ransomware and emerging security challenges in the Internet of Things. Comput. Netw. **129**, Part 2 (2017)
14. Sicari, S., Rizzardi, A., Miorandi, D., Cappiello, C., Coen-Porisini, A.: A secure and quality-aware prototypical architecture for the internet of things. Inf. Syst. **58**, 43–55 (2016)
15. Raza, S., Seitz, L., Sitenkov, D., Selander, G.: Scalable security with the symmetric keys (DTLS) key establishment for the internet of things. S3k IEEE Trans. Autom. Sci. Eng. **13**(3), 1270–1280 (2016)

16. Bostani, H., Sheikhan, M.: Hybrid of anomaly-based and specification-based ids for the internet of things using unsupervised of based on mapreduce approach. Comput. Commun. **98**, 52–71 (2017)
17. Huh, S., Cho, S., Kim, S.: Internet of things architecture: recent advances, taxonomy, requirements, and open challenges, managing IoT devices using block-chain platform ICACT (2017)
18. Investopedia.com/ Source: Internet
19. Lei, W.: Analysis of using blockchain technology to solve the security problem of the internet of things. In: 2018 2nd International Conference on Systems, Computing, and Applications (SYSTCA) (2018)
20. Bertino, E., Islam, N.: Botnets and internet of things security. Computer (2017)
21. Karthiban, K, Smys, S.: Privacy preserving approaches in cloud computing. In: 2018 2nd International Conference on Inventive Systems and Control (ICISC), IEEE, pp. 462–467 Jan 19 (2018)

Dolagobinda Samal currently perusing his Master's at the Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru. He pursued his Bachelor in Electrical and Electronics Engineering from Biju Patanaik University, Odisha. His research interests include IoT, artificial intelligence, WSN and mobile communication networks.

Rajakumar Arul currently serves as an Assistant Professor at the Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru. He pursued his Bachelor's and Master's in Computer Science and Engineering from Anna University, Chennai. He completed his Doctorate of Philosophy requirements in Full Time under the Faculty of Information and Communication, Department of Computer Technology, Anna University—MIT Campus. His research interests include Security in Broadband Wireless Networks, WiMAX, LTE, Robust resource allocation schemes in mobile communication networks. He is a member of IEEE and a Professional Member of ACM.

Logically Locked I2C Protocol for Improved Security

S. Rekha, B. Reshma, N. P. Dilipkumar, A. Ajai Crocier and N. Mohankumar

Abstract The inter-integrated circuit is a serial, half-duplex protocol. It is a synchronous device that can implement multi-masters and multi-slaves. The masters are the controlling blocks that initiate all activities of the slaves. It is important to ensure that this communication is secured. This paper aims at a simplistic approach to lock the clock line of the I2C using logic gates so as to protect the information passing through the data lines. It makes use of a four-digit hexadecimal password. This allows up to 65,536 combinations enhancing the security and to gain additional control over I2C. This proposal uses a concept of logic locking to gain security.

Keywords Logic locking · Hardware security and trust · I2C protocol

1 Introduction

Communication interfacing plays an important role in today's world as there is a necessity for communication medium between a source and destination. I2C is one such protocol which is being used as a communication media. These interfaces are of two types parallel and serial. Serial interface transfers the data in bit-bit fashion with a clock signal to synchronize it, thus making it a simple protocol. But parallel protocol transfers data in byte-byte fashion; it is more complex than serial. I^2C is serial protocol. I^2C is very flexible because it supports multi-master and multi-slave communication that can be used in the circuit. It requires only two wires for communication and it minimizes the interconnection among IC's and hence it is less expensive than PCB's using serial peripheral interface (SPI) protocol. Chip addressing and addition of any other extra devices is very simple as we do not need any chip select (CS) lines to add extra devices like SPI. It requires only two bidirectional signal lines for communication among multiple devices. I^2C protocol has better error handling mechanism as it has the ACK and NACK feature, where ACK stands

S. Rekha · B. Reshma · N. P. Dilipkumar · A. A. Crocier · N. Mohankumar (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: n_mohankumar@cb.amrita.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_67

for acknowledgement and latter for no acknowledgement. Logic locking provides an efficient way to protect digital circuits. A logic lock requires a key input and a previously set password that is stored in the memory. It is secure and cost effective. It helps to defend the circuit against piracy and reverse engineering attacks. It hides the functionality of the entire circuit or can also be used to hide particular blocks of the circuit, as in the case of I2C. In I2C, it helps defend the clock line from the master to slave.

1.1 Problems with I2C

In most architectures, a single clock line controls the activity of the slaves. So, an attacker can easily monitor the slaves. I2C is not designed for messages that are long enough for encryption to make sense (no built-in encryption). This makes it easy to modify the data that is received by the slaves. Since the buses are bidirectional, miscommunication can happen between master and slave in case of an attack. The paper aims at a simplistic approach for protocol security. The design makes use of logic gates that are easily available. It is easy to implement and effective. Also, the power consumed is negligible in comparison with the security offered. The lock is also efficient in terms of area. The password can be changed according to the convenience of the user. I2C allows multiple slaves to communicate with multiple masters. It provides a simple way to communicate between IC's using minimum number of pins. Locking ensures that other devices access the masters. Risk of hardware Trojan altering the global behaviour is reduced. Relay attacks on master i.e., unauthorized blocks suppressing authorized blocks, will fail. It also helps prevent an attacker from disrupting the acknowledge bits.

2 Literature Survey

Inter-integrated circuit protocol is a serial communication protocol in which either the master or multiple masters communicate with the slave components. In the research paper published by Kumari et al., [1] FPGA has been used as a master and MEMS motion sensor has been used as a slave. The data of the MEMS motion sensor is passed through the I2C protocol to the FPGA. In this, I2C protocol is given preference to interface the FPGA and the MEMS sensor because it gives more simplicity as it uses only two wires (SDA and SCL) and a smaller number of pin connections. In the research paper published by Yadav et al. [2], two masters have been interfaced in the protocol. Here, the two masters are the FPGA ICs and the slave is the DC motor. To develop a dual master set-up, first an I2C protocol bus with single master has been setup then using arbitration technique another master is interfaced in the protocol. When the protocol is implemented the two FPGAs gain the control over the DC motor. In the above-mentioned research paper, only two masters were used in the entire I2C protocol but what if really more masters are used. Reference [3] shows the working of the I2C protocol for multiple masters-multiple slaves. To know which master has got the bus at a particular instant of time, arbitration technique is used satisfying the DO-254 standards. Arbitration refers to the cancelling the communication of one master if one master is in the communication. This occurs during the start condition of the data transfer. As the master and the slave components of the I2C protocol bus works at different voltage levels there is a risk that the slave component might get a high voltage and thus gets damaged. The applications of I2C protocol bus are very wide and it can be used in our day-to-day life. One such application is illustrated in Ref. [4]. It proposes an idea that an I2C protocol bus can be used as waiter system in the restaurants. The customer can give command in speech mode to a smart phone which converts speech to text and sends to the master node in the kitchen using I2C protocol architecture. Here, the customer's smart phone is the slave device and the device in the kitchen is the master. Hence, the waiter system is easy and convenient. In Ref. [5], it is mentioned that malicious hardware is not always detected at the time of testing. Mostly, the presence is identified only after the Trojan begins its attack. To solve this, a method to encode the authentication keys with the help of a pseudo random number generator has been proposed. Reference [6] talks about trouble shooting the I2C protocol and has provided solutions to various conflicts that I2C faces. These include unexpected SCL pulses, missing bytes, incomplete clocks, missing acknowledgements, address conflicts and various others. In protecting integrated circuits from piracy with test-aware logic locking, a paper by Plaza and Markov [7], it has been discussed that remote foundries forsake security to make IC manufacturing cheaper. These are sold in the black markets. EXOR-based chip-locking, cryptography and active metering are used to safeguard unlocked circuits. Multiplexer-based locking method has been used to demonstrate attacks on locked circuits that do not preserve test response. The paper published by Ranjani et al. [8] have devised a method to prevent hardware encryption. They have presented a highly secured mechanism which counters many key guessing methods and proved with experimental results.

2.1 Logic Locking

Reference [9] contains detailed explanation about the various threats like IP piracy, reverse engineering, hardware Trojans, etc. in their paper Evolution of Logic Locking. They have dealt with different logic locking techniques like random logic locking, strong logic locking. Thangam et al. [10] proposed different hardware techniques to secure IP piracy in their paper Novel Logic Locking Technique for Hardware Security on. This is done by implementation of a design by inserting key gates into the original design. Functional Analysis Attacks on Logic Locking, a paper by Sirone et al. [11], proposed logic locking algorithms like FALL which use structural analyses of locked circuit to identify the key. It can defeat locking without oracle access by just analysing the net list. These experiments have succeeded against 65 upon 80 (81%)

using Secure Function Logic Locking. Reference [12] found a new approach for protection of digital circuits against IC piracy. They have used multiple key inputs to control key gates unlike the conventional method. This technique is immune to key-sensitization attacks and execution time for locking circuits is reduced drastically. Sengupta et al. [13] have discussed various hardware security threats in their research titled: ATPG-based cost-effective, secure logic locking. They have used the recent locking technique SFLL. It mainly focuses on cost effectiveness. They have devised a method named SFLL-fault which uses VLSI testing and tools which automates critical steps in SFLL and to minimize the cost. Their results have reduced the cost by 35% compared to SFLL without any compromises in security.

3 Securing I2C Protocol

Inter-integrated circuit is a serial protocol. It has two wires: serial clock and serial data; these wires are used to interface peripherals. In this protocol, usually a master communicates with many slaves, however, there can be multi-masters communicating with multi-slaves. The functional diagram shown in Fig. 1a has two pull up resistors (*Rp*) and a power supply (+*Vdd*). The communication between master and slave happens when the SDA line is low and SCL line is high as shown in Fig. 1b. So, to attain maximum security the lock is inserted in the clock line which initiates the communication. Then output of the logic lock and serial clock line is given through a two-input AND gate. The output of the AND gate is 1 when the password is correct and 0 when the password is wrong.

The block diagram of the logic lock circuit is shown in Fig. 2. The input is given through hexadecimal keypad (four bits) and is converted to binary number system using a hexadecimal to binary decoder. The actual password is stored in the memory. Both these hexadecimal codes are compared using a comparator (built from 16 Ex-Nor gate which is the more efficient than using 16 full adders). The final output of the lock is either logic 1 or logic 0. If it is logic 0, then the password is incorrect and lock cannot be opened, and for logic 1, the lock is opened. There is a need for locking this protocol to secure it from all the threats.

The password is stored in memory using 16 D flip-flops. The input condition is copied to the output only when the clock input is high. If the clock input is in low level, the flip-flop will not store the input condition. The logic lock has been simulated using LabVIEW; Fig. 3 is the simulation for the digital lock in I2C protocol. The results of simulation have been plotted in a waveform.

The circuit below shows the lock built from various gates and memory elements. It uses D flip-flop to store the password and is followed by Ex-Nor gates used as comparators.

The correct password has been set as ‘EDAD’ in the result, which converts to 1110110110101101 in binary. This has plotted in the waveform shown in Fig. 4a. Since this is the correct input, the logic state from the lock output is 1 as shown in Fig. 4c. Any password other that the one stored in the memory produces logic state

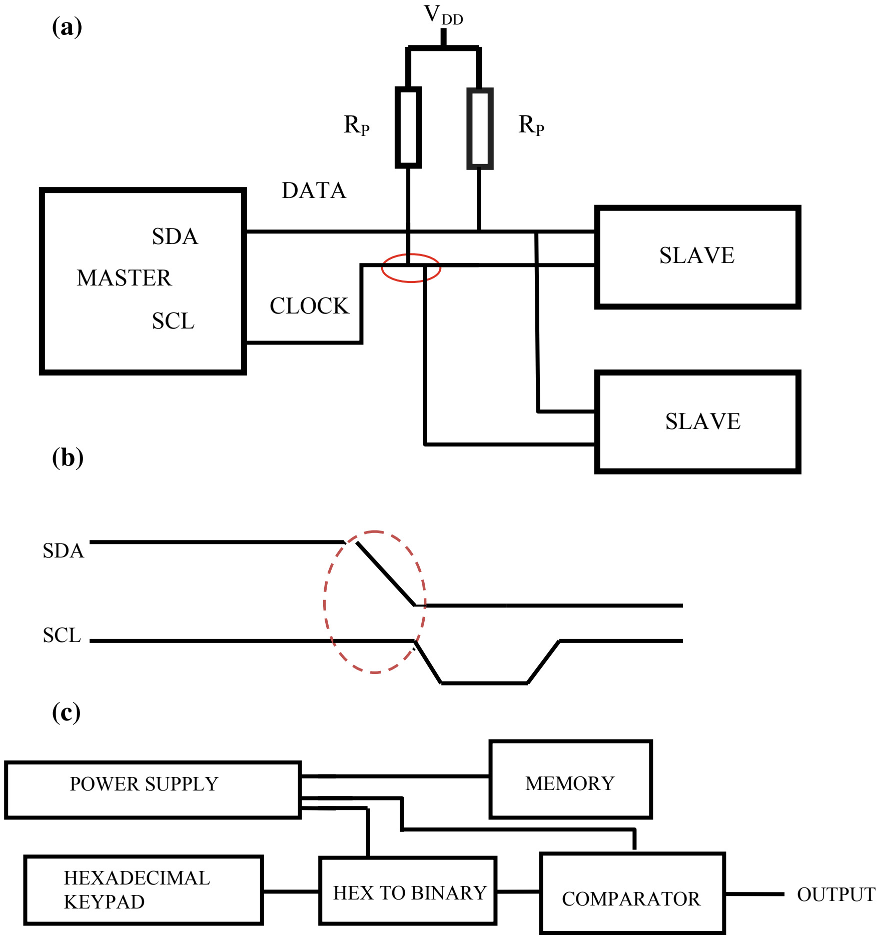


Fig. 1 **a** Block diagram of I2C protocol, **b** serial data line and serial clock line conditions for communication, **c** functional block diagram of logic lock

0 as in Fig. 4d. For example, ‘ABCD’, has been considered as the user input. This gives the waveform 1010101111001101 (Fig. 4b) (Table 1).

The Hamming distance for correct password it is 0.

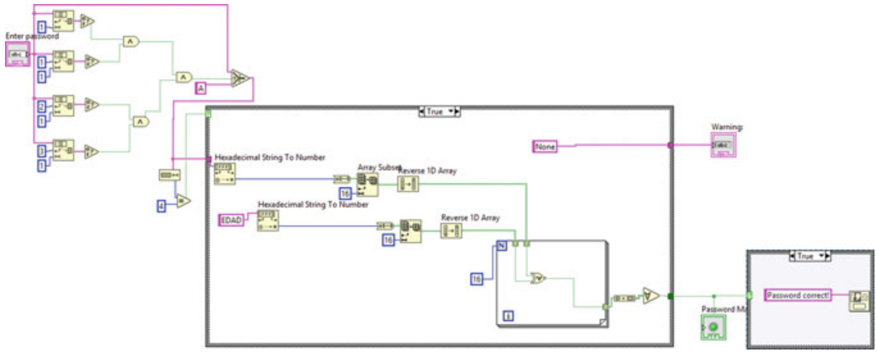


Fig. 2 LabVIEW simulation

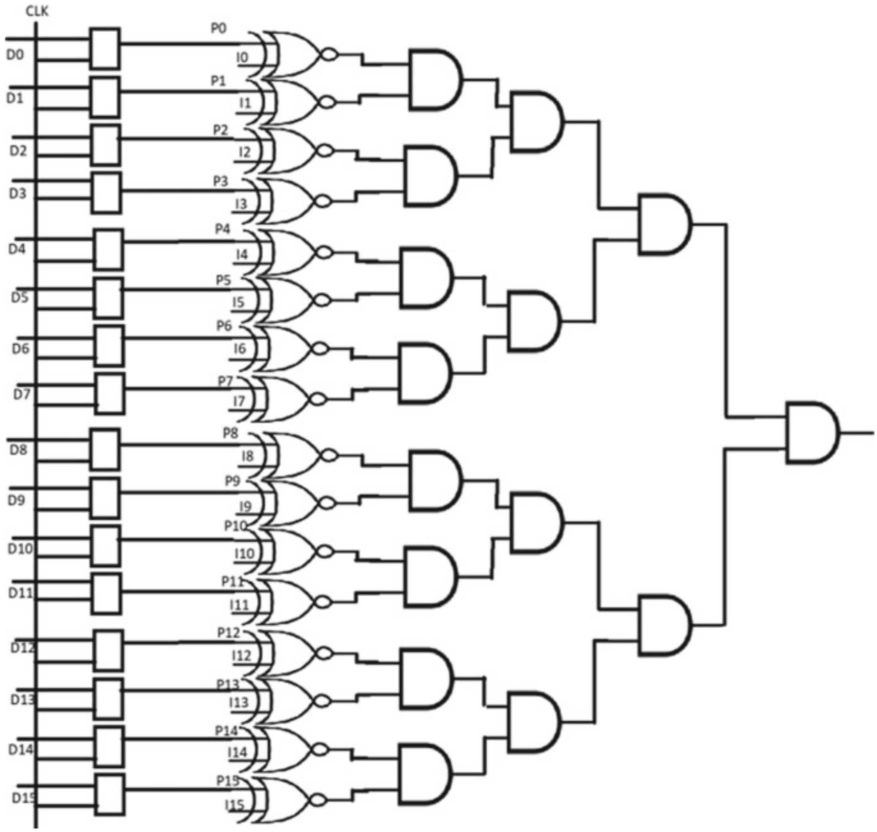
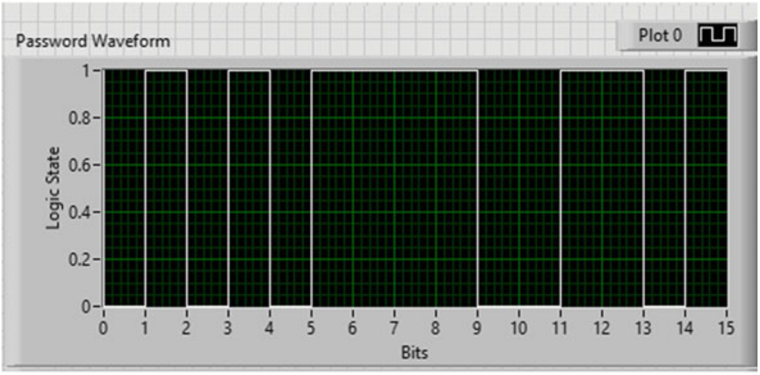


Fig. 3 Combinational circuit for memory

(a)



(b)

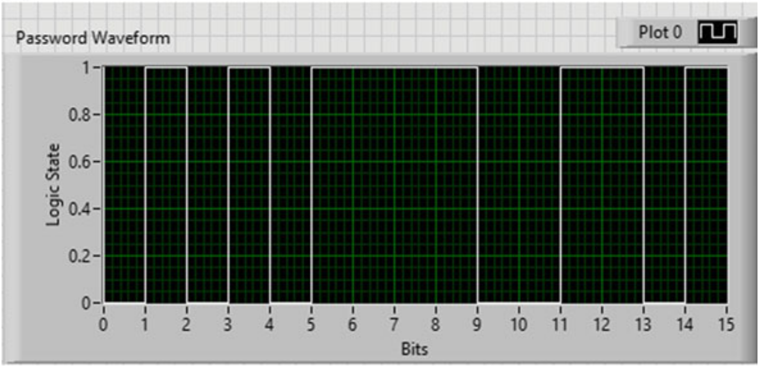
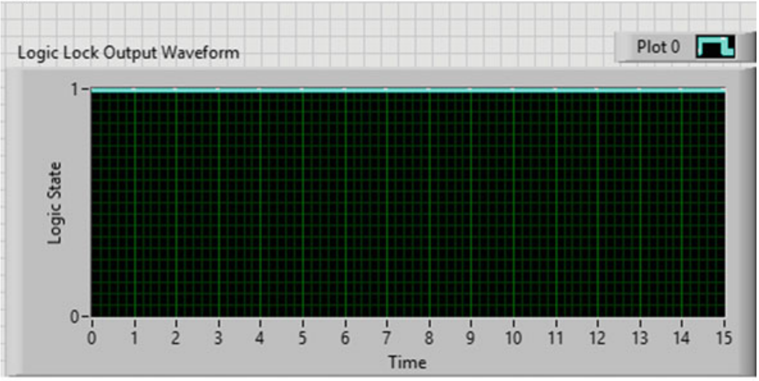


Fig. 4 **a** Waveform of correct password, **b** waveform of wrong password, **c** logic lock output for correct password, **d** logic lock output for wrong password

(c)



(d)

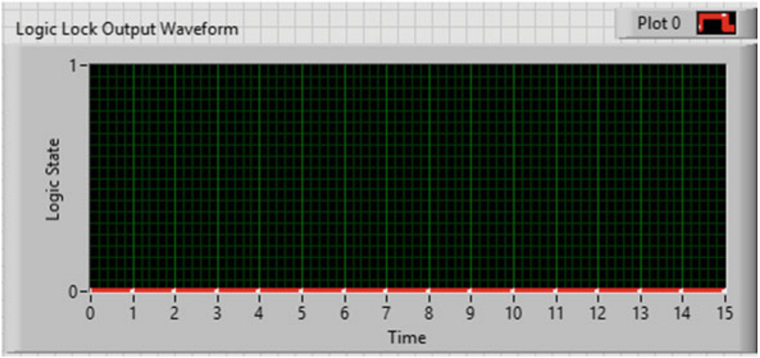


Fig. 4 (continued)

4 Conclusion

I2C protocol is secured by locking the serial clock line which enhances the security by mitigating other hardware threats like Trojans. Since the hexadecimal keypad is used, the number of combinations is around 65 k. The logic is not developed using other complex combinational circuits instead using simple Ex-Nor gate which does not occupy much space, and is aimed at cost effectiveness which has been verified. Besides, the lock also has a lot of future scope since it can easily be extended to more bits hence increasing security level. The first two blocks can be replaced with an ASCII keypad and ASCII to binary decoder. This increases the number of combinations drastically.

Table 1 Hamming distance for various passwords

S. no.	Input I	Input II	Hamming distance
1	ABCD	ABCD	0
2	456A	456E	1
3	56CD	56CE	2
4	4567	456C	3
5	FFFF	DDDD	4
6	CFEE	ABCC	5
7	ABCD	CDEF	6
8	12AB	34CD	7
9	DAC4	EDED	8
10	78AB	89CD	9
11	FEFE	4567	10
12	BCBC	1245	11
13	BBBB	CCCC	12
14	7654	89A5	13
15	AF6E	D093	14
16	765B	89A5	15
17	AF6E	5091	16

References

1. Kumari, R. S. S., Gayathri, C.: Interfacing MEMS motion sensor with FPGP using I^2C protocol, 21–22 Oct (2016)
2. Deepika, D., Yadav, N.: Design of dual master I^2C bus controller and interfacing it with DC motor. In: 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 12–13 Oct (2018)
3. Bharath, K. B., Kumaraswamy, K. V., Swamy, R. K.: Design of arbitrated I^2C protocol with DO-254 compliance. In: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 17–18 March (2017)
4. Vishakh, B. V., Khwaja, M. K., Vidhyapathi, C. M.: Comprehensive automated device for hotel management using I2C protocol. In: 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 10–12 Dec 2015
5. Reddy, D. M., et al.: BHARKS: Built-in hardware authentication using random key sequence. In: 2017 4th International Conference on Signal Processing Computing and Control (ISPCC), pp. 200–204 (2017)
6. Ferrando, M. P.: Troubleshooting I 2C bus protocol, Application Report SCAA106–October (2009)
7. Plaza, S. M., Markov, I. L.: Protecting integrated circuits from piracy with test-aware logic locking. In: 2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2–6 Nov (2014)
8. Ranjani, R. S., Devi, M. N.: Enhanced logical locking for a secured hardware IP against key-guessing attacks. In: Proceedings of International Symposium on VLSI Design and Test, pp. 186–197 June (2018)
9. Yasin, M., Sinanoglu, O.: Evolution of logic locking. In: 2017 IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC), 23–25 Oct (2017)

10. Thangam, T., Gayathri, G., Madhubala, T.: A novel logic locking technique for hardware security. In: 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE), 27–28 April (2017)
11. Sirone, D., Subramanyan, P.: Functional analysis attacks on logic locking. In: 2019 Design, Automation & Test in Europe Conference & Exhibition (DATE), 25–29 March (2019)
12. Karousos, N., Pexaras, K., Karybali, I. G.: Weighted logic locking. In: 2017 IEEE 23rd International Symposium on On-Line Testing and Robust System Design (IOLTS), 3–5 July (2017)
13. Senugupta, A., Nabbel, M., Yasin, M., Sinanoglu, O.: ATPG-based cost-effective, secure logic locking. In: 2018 IEEE 36th VLSI Test Symposium (VTS), 22–25 April (2018)

Broadband Circularly Polarized Microstrip Patch Antenna with Fractal Defected Ground Structure

B. Naveen Reddy and V. Mekaladevi

Abstract A new technique to design single-fed circularly polarized (CP) microstrip patch antenna is presented in this paper. The estimation of bandwidth enhancement is based on dimension adjustment of the fractal defected ground structure (FDGS). Then, by inserting circular slot on the patch, circular polarization (CP) is obtained. The parametric optimization is done in ANSYS HFSS simulator in order to fix the dimensions of the microstrip patch antenna. The microstrip patch antenna with third iterative FDGS is fabricated and measured. The bandwidth of the measured microstrip antenna with CP radiation is about 120 MHz (2.13–2.25 GHz) and axial ratio of 2.3 dB to ensure circular polarization.

1 Introduction

For many years now, the microstrip antennas with defected ground structure (DGS) are commonly used in the bandwidth enhancement, because of low profile and easy fabrication. Compared with the conventional microstrip patch antenna (MSA) with DGS, FDGS has more advantage and better prospects in enhancing the bandwidth. Single-fed microstrip antenna is preferred because of its simple structure to achieve CP radiation. This antenna can also be designed using improvised particle swarm optimization technique as it was used in many applications like microwave absorbers design [1] to reduce the time. In [2], rectangular microstrip patch antenna with stub and multiple feeding points [3] increased the bandwidth up to 6%.

Different techniques are proposed to overcome the narrow bandwidth limitations such as by implementing different feeding techniques, frequency selective surface (FSS), electromagnetic band gap (EBG), photonic band gap (PBG), meta material and so on. DGS is being used due to its simple design compared with other techniques.

B. Naveen Reddy (✉) · V. Mekaladevi
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

V. Mekaladevi
e-mail: v_mekaladevi@cb.amrita.edu

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_68

Slot etched on the ground plane is considered as DGS. Circular patch antenna with two circular slots [4] etched on the ground plane suppressed cross-polarization (XP) level ranging from 5 to 8 dB, bandwidth of 160 MHz. In L-shaped asymmetric DGS [5], XP level is suppressed up to −15 dB. V slot etched on square patch in [6] gave bandwidth of 2.13 GHz. In all the reported designs of microstrip patch antenna with DGS, the enhancement of bandwidth is considerable but not up to the required level.

The basic way of achieving CP radiation is to increase the XP level of the microstrip antenna. Asymmetric U slot structure is used to generate CP radiation in [7]. Square ring is used for the CP operation in [8]. But the conventional DGS's like dumbbell shape or U shape could not increase the XP as required. Different techniques have been presented to increase the XP level like using fractal defected ground structure (FDGS), using double feed maintaining 90° phase shift, truncating the corners of patch or inserting a slot on the patch. Y-shaped FDGS was designed to increase XP level to achieve CP radiation in [9]. Self-iterated square-shaped lattices are etched in the ground plane of the microstrip patch antenna in [10]. The corners of the patch are truncated to achieve the CP radiation in [11]. A slot inserted in the patch has been presented in [12].

To overcome the drawbacks, a new technique is proposed in this paper. The design of the broadband single-fed CP square microstrip antenna is done by etching FDGS on the ground plane and inserting a circular slot on the patch with coaxial feed.

2 Proposed Antenna Design

The length and width of the microstrip patch antenna with substrate ($\varepsilon_r = 4.4$, $h = 1.6$ mm) is calculated using the conventional formulas at 2.2 GHz. In this design, square patch is used with $W_p = L_p = 32$ mm. Ground dimension is set by keeping 6 h distance on all sides and the dimensions are $W_s = L_s = 51.2$ mm.

The CP radiation is achieved by exciting two orthogonal modes with equal magnitude. The original resonant mode produces a field in one direction and the increased XP resonant mode produced by etching the FDGS with a circular slot produces another field orthogonally and a slight difference in the magnitude is maintained. The design of FDGS etched on ground plane with circular slot on patch is shown in Fig. 1a, b.

3 Parametric Analysis

The dimensions obtained from the formulas are merely an idea of around what values of dimensions to be considered. But practically, optimal value can be greater or less than the values obtained from the standard formula. First, we design the antenna with the standard formula values and then we vary the width and length parameters for the optimum values. Coaxial feed position is also important to have good matching

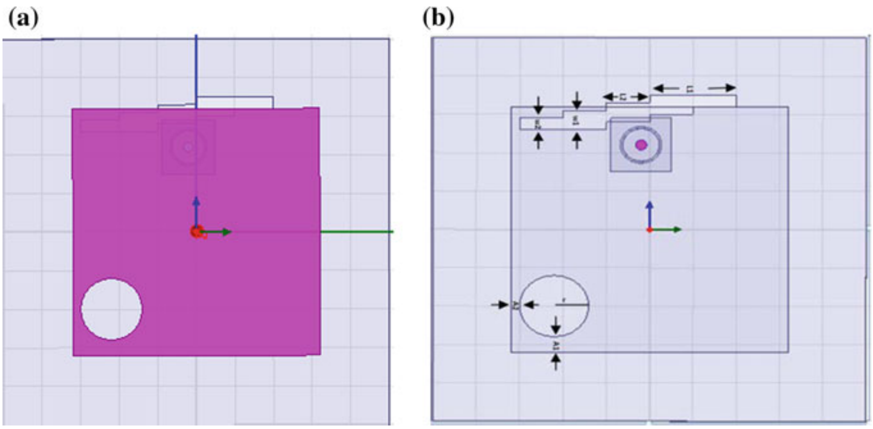


Fig. 1 **a** Proposed FDGS design with circular slot, **b** dimensions of proposed design

with antenna. Unless there is a perfect impedance match, the antenna will not radiate much. Feeding position plays important role in the impedance matching, and it is important to find out the feed location for the desired line impedance Z_o [13].

3.1 *Varying the Parameters*

The length of the patch and substrate are optimized using high-frequency structure simulator (HFSS). When the parameters are varied with a set of values, the optimum value obtained for patch length of MSA is $L_p = 32$ mm and optimum value of substrate length obtained is $L_s = 50$ mm and optimum distance to be maintained between the center of patch and feed point is 8 mm. Based on these results we fix the dimensions of the microstrip patch antenna. The corresponding plots for the variation in parameters are shown in Fig. 2a, b, c.

The optimized parametric values of the dimensions used in the proposed design are shown in Table 1.

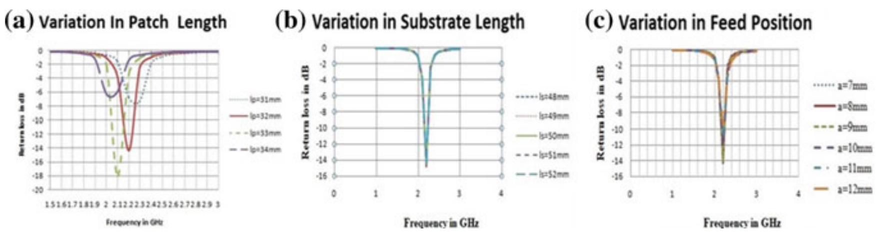


Fig. 2 **a** Variation of L_p, **b** variation of L_s, **c** variation of ‘a’

Table 1 Dimensions of proposed design

	Simple MSA (mm)	MSA with FDGS (mm)	Proposed antenna (mm)
W_p	32	32	32
L_p	32	32	32
W_s	50	50	50
L_s	50	50	50
L_1	–	10	10
L_2	–	5	5
W_1	–	2.5	2.5
W_2	–	1.5	1.5
r	–	–	4
A_1	–	–	2
A_2	–	–	1

4 Results and Discussion

Comparison of return loss for different designs has been shown in Fig. 3. The proposed antenna has the highest return loss of −25.27 dB followed by the design of MSA with FDGS with return loss of −15.12 dB and then followed by the design of simple MSA with return loss of −14.3 dB. Accordingly, the bandwidth obtained for the proposed antenna is 150 MHz, for MSA with FDGS the bandwidth is 86 MHz

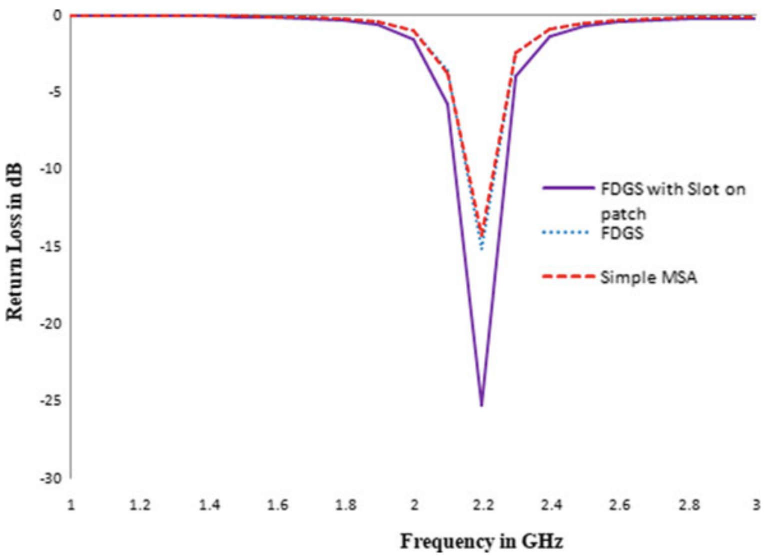


Fig. 3 S_{11} plot for different configurations

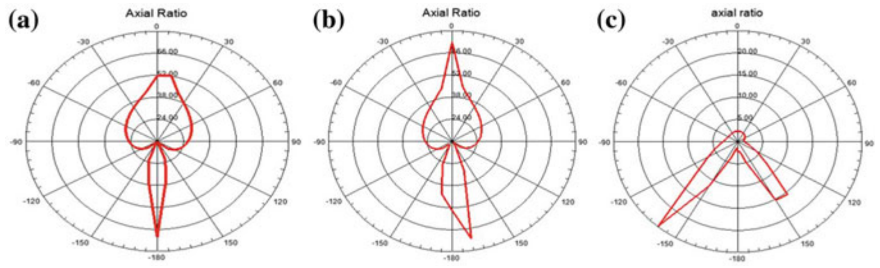


Fig. 4 **a** Axial ratio for simple MSA, **b** axial ratio for MSA with FDGS, **c** axial ratio of proposed antenna

and for the design of simple MSA bandwidth is 76 MHz. The proposed antenna has very much improved bandwidth compared to other designs and our requirement of wider bandwidth is also satisfied.

From the axial ratios plotted in Fig. 4a, b, c, both the designs simple microstrip patch antenna and FDGS design in Fig. 4a, b have the axial ratio more than 3 dB which indicates that they are linearly polarized. The bandwidth enhancement in both the designs is very less. But proposed FDGS design in Fig. 4c has the axial ratio less than 3 dB that is 2.3 dB which assures that the antenna is circularly polarized and eventually the bandwidth enhancement is high in the design.

Gain and directivity plot has been shown in Fig. 5a with gain of 0.01 dB and directivity of 2.46 dB. As the circular polarization is achieved, the right-hand circular polarization (RHCP) and left-hand circular polarization (LHCP) has been plotted in Fig. 5b. In Fig. 6a, axial ratio versus frequency has been plotted and it was clear that at 2.2 GHz the axial ratio is less than 3 dB and started increasing after that. Axial ratio versus theta has been plotted in Fig. 6b and it was clear from the plot that axial ratio is less than 3 dB from theta $= -100^0$ to theta $= 100^0$ while other part of plot has more than 3 dB.

The comparison between simulated values and measured values has been shown in Table 2. And the polarization of the antenna for the first two designs is linear polar-

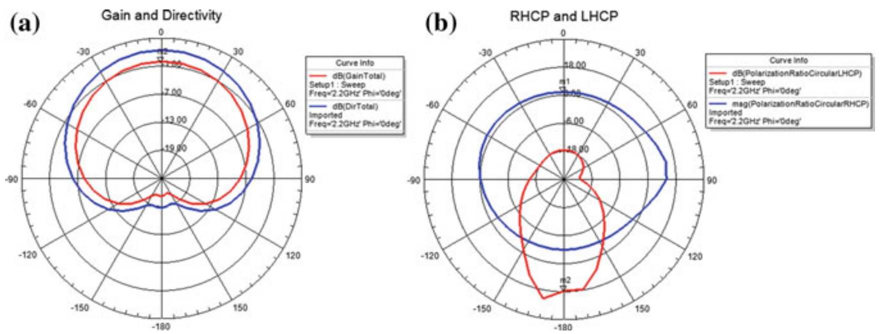


Fig. 5 **a** Gain and directivity plot of proposed antenna, **b** RHCP and LHCP plot of proposed antenna

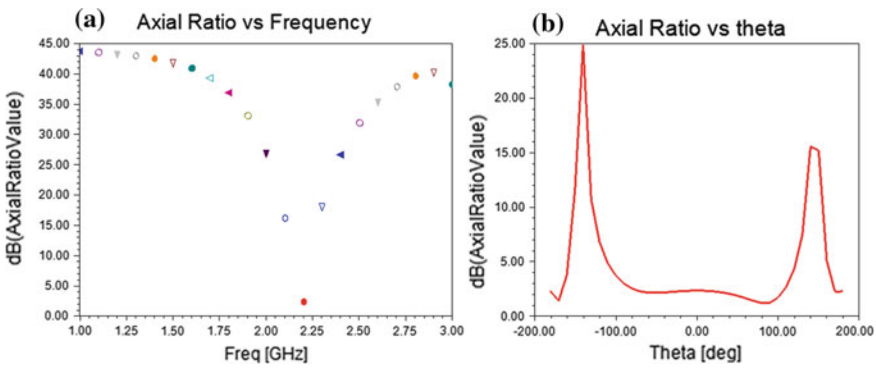


Fig. 6 **a** Axial ratio versus frequency, **b** axial ratio versus theta

Table 2 Comparison between simulated and measured results

	S_{11}		Gain (dB)	Directivity (dB)	Axial ratio (dB)	Bandwidth	
	Simulated (dB)	Measured (dB)				Simulated (MHz)	Measured (MHz)
Simple MSA	−14.30	–	−0.02	2.46	51.6	76	–
With FDGS	−15.12	–	0.01	2.46	72	86	–
FDGS with circular slot	−25.27	−17.2	0.01	2.46	2.3	150	120

ization and the last design was a circularly polarized design because the axial ratio was less than 3 dB. And comparatively proposed antenna has maximum bandwidth. Top view and a bottom view photographs of fabricated antenna are shown in Fig. 7a, b and comparison between measured and simulated results is shown in Fig. 7c.

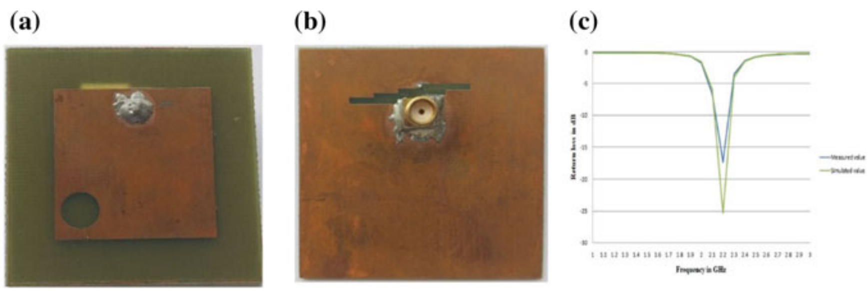


Fig. 7 **a** Top view, **b** bottom view, **c** comparison between measured and simulated results

5 Conclusion

The study of the FDGS has made more progress in recent years and has more advantages and better prospects. With this motivation, the proposed antenna has been fabricated with FR-4 substrate of dimensions (50 mm × 50 mm × 1.6 mm). By implementing this design, we achieved broadband bandwidth up to 150 MHz which was quite satisfactory. The value obtained for gain and directivity is 0.02 and 2.46 dB. The FDGS design with a slot on the patch which is simulated in HFSS has been fabricated and measured and the bandwidth obtained in the band 2.13–2.25 GHz is 120 MHz while its axial ratio is about 2.3 dB.

References

1. Mouna, H., Mekaladevi, V., Devi M. N.: Design of microwave absorbers using improvised particle swarm optimization algorithm. J. Microw. Optoelectron. Electromagn. Appl. **17**(2), São Caetano do Sul, Apr/June (2018)
2. Feng, M., Zhang, J., Wu, W.: A compact microstrip patch antenna with bandwidth enhancement. In: 2017 7th IEEE International Symposium on Microwave, Antenna, Propagation, and EMC Technologies
3. Koutinos, A. G., Ioannopoulos, G. A., Chryssomallis, M. T., Kyriacou, G. A.: Bandwidth enhancement of rectangular patch antennas using multiple feeding points: A review. In: 2018 7th International Conference on Modern Circuits and Systems Technologies, MOCAST 2018, pp. 1–4 (2018)
4. Kumar, C.: Nature of cross-polarized radiations from probe-fed circular microstrip antennas and their suppression using different geometries of defected ground structure (DGS). IEEE Trans. Antennas Propag. **60**(1), 92–101 (2011)
5. Fields, C.: Communications asymmetric geometry of defected ground structure for rectangular microstrip: A new approach. **64**(6), 2503–2506 (2016)
6. Esa, M., Jamaluddin, U., Awang, M. S.: Antenna with DGS for improved performance. In: 2010 IEEE Asia-Pacific Conference on Applied Electromagnetics, APACE 2010-Proceedings, no. Apace, pp. 1–4 (2010)
7. Tong, K., Wong, T.: Circularly polarized u-slot antenna. IEEE Trans. Antennas Propag. **55**(8), 2382–2385 (2007)
8. Chen, W.S., Wu, C.K.: Square-ring microstrip antenna with a cross strip for compact circular polarization operation. IEEE Trans. Antennas Propag. **47**(10), 1566–1568 (1999)
9. Wei, K., Li, J. Y., Wang, L., Xu, R., Xing, Z. J.: Communication a new technique to design circularly polarized microstrip antenna by fractal defected ground structure. **65**(7), 3721–3725 (2017)
10. Verma, S., Rano, D., Hashmi, M. S.: A novel miniaturized band stop filter using fractal type defected ground structure (DGS). In: Asia-Pacific Microwave Conference Proceedings, APMC, pp. 799–802 (2018)
11. Alam, S., Surjati, I., Ningsih, Y. K.: Patch modification and slot arrangement of microstrip antenna for improving the axial ratio. In: 2017 International Conference on Broadband Communication, Wireless Sensors and Powering, BCWSP 2017, vol. 2018–January, no. August 1945, pp. 1–5, 2018
12. Malekabadi, S. A., Attari, A. R., Mirsalehi, M. M.: Compact and broadband circular polarized microstrip antenna with wideband axial-ratio bandwidth. In: 2008 International Symposium on Telecommunications, IST 2008, pp. 106–109 (2008)

13. Sudhakaran, G., Kandipati, B., Surya, G. B., Shree, V. K., Sivaprasad, M., Jayakumar, M.: Evolutionary algorithm based structural optimization for patch antenna design and its performance analysis. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, pp. 2189–2192 (2017)

A Novel Technique for Vehicle Theft Detection System Using MQTT on IoT

K. Aishwarya and R. Manjesh

Abstract Automobile theft is a worldwide immense problem. A vehicle of top-notch security features is usually higher in the cost, and it cannot be afforded by middle-class people. By considering all these parameters, we aimed to design a low-cost, real-time, robust security system for vehicles. The main purpose of this project is to notify the vehicle owner, when the vehicle is moved/theft from the parking area and to monitor the movement of the vehicle in real time. Raspberry Pi 3 board connected with SIM908 GPS module will be placed in the vehicle. GPS module receives the signal and provides the data in NMEA format. Out of several messages, GPRMC message will be sent from the board to the cloud. GPRMC message type has much information like latitude, longitude, speed, time, etc. Raspberry Pi 3 which is connected to mobile hotspot will communicate with the cloud using MQTT protocol. MQTT protocol is a lightweight protocol which comes on top of TCP/IP. The web application will be connected to the cloud by providing configuration details. Once the web application is subscribed to the GPS topic, it starts getting the GPS data which will be displayed on the web page.

Keywords Raspberry Pi · NMEA · Hotspot · GPS module · GPRMC

1 Introduction

In recent years, owning a vehicle is a passion as well as a requirement. Every home has one vehicle for transportation; it may be a low-cost bike to high-cost car depending on their source of income. As vehicle production increases, vehicle theft also increased drastically [1–5]. Presently there are few types of anti-theft equipment, one is steering locking system, second is auto guard system to close and open vehicle doors, third is electronic alarm system. Sometimes vehicles are not even secured under CCTV surveillance. Watchman or security also fails sometimes to guard the vehicle all the time [2–4]. Available access control systems are failed to provide information

K. Aishwarya (✉) · R. Manjesh
Department of Computer Science and Engineering, Srinivas Institute of Technology, Mangalore, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_69

to the right user [4–6]. Traditionally, vehicle theft is being investigated using the vehicle number plate, but most of the time criminals used to change number plate successfully [7–9]. And also there is a rapid growth in the number of vehicles since 2011, so the parking place is the biggest problem and no parking problem is revealed [9–12]. We proposed a vehicle tracking and theft detection system which is designed to track and monitor the status of the vehicle, so that the owner can park anywhere regardless. We adopted technology like global positioning system to get information like location and time. GPS data can be visualized in Google Map technology [12–15]. The main purpose of this project is to notify the vehicle owner if the vehicle is moved/theft from the parking area and also to monitor the vehicle movement in real time. This will accurately detect the vehicle position and also inform the owner on time by playing siren at the application end, so that alert messages can be sent to reduce time in a short time.

2 Literature Review

Hu et al. [1] proposed a model for detecting vehicle theft using GPS module and GSM. In this design, the remote monitoring of the vehicle is applied and the user can set vibration and alarm to alert when a theft happens. Through the tracker, location information will be sent to the user's phone every once in a while. Here, MCU handles the whole process where different GPS modules and wireless modules are attached to it. This proposed method does not show more information of hardware connection and real-time working module. GSM module creates a connection between the user and the system. Shammi et al. [2] proposed a method using a machine learning technique to find vehicle theft happening from the parking. Here, they introduced the edge detection technique that is a canny method. The canny method observes the changes happening in the vehicle position and then the movement of the object gets alerted to the owner. The monitoring of the vehicle is by CCTV camera of the parking area and that will be controlled by security panel. Recording of video and converting that to frames to achieve edge detection and movement detection is done by implementing algorithms. Sreedevi and Nair [3] introduced a new technique using image processing to prevent vehicle theft detection and prevention. In this proposed model, vehicle owner's photographs were stored in the database and this will be compared with photographs taken through the webcam; if both the pictures get matched, then only the vehicle can be accessed. If photographs did not get matched, automatically the alarm gets activated to alert the owner. Here, the technology used is image processing, real-time face detection and recognition. The algorithm used is PCA and DCT normalization gives higher accuracy. Babu and Raghunand [4] proposed a technique to find traffic violence. Here, photographs of the suspected vehicle are taken through digital camera concentrating number plate. Boundary box method with image segmentation is carried out to recognize the image of the vehicle's number plate. After the segmentation, the character recognition is done by template matching. Once the processing of these images is done, then the required data like vehicle number is accessed easily and accurately. Li et al. [5] presented an article based on ZigBee technology which is used to detect the vehicle.

Vehicle scan is done through infrared rays keeping transmitter and receiver tubes on both sides of the road. Once the vehicle comes to contact with the scanner or infrared rays, vehicle information is collected and will be recorded such as size, shape, and color. Collected data is processed and analyzed using tools and software programming. ZigBee networks provide efficient results and also cost effective.

3 System Design

This module includes a Raspberry Pi which is used to interconnect different hardware components. The vehicle tracking system consists of GPS module and Raspberry Pi. The GPS module receives the information from the GPS antenna and generates a data string in the form of the NMEA message format. This data string is given to the Raspberry Pi which acts as an interface between the GPS module and the antenna. It then sends this conditioned data string to the cloud. This data string can be forwarded either by using GPRMC format to cloud. The data is processed further by Raspberry Pi and finally given to PC or Laptop with local/Google Maps for display. In this project, we are using Microsoft Visual Studio software and Linux-based operating system for mobile devices such as smartphones and tablet computers. A C program is used for our project to detect the GPS data from the tracker. HTML and JavaScript are used for designing the front end web page (Fig. 1).

Tracker installed inside the car with Raspberry Pi will give the accurate position of the vehicle. The message broker on the publisher side keeps publishing the data to the cloud broker. A user can able to access the data by subscribing based on the topic. There are many MQTT brokers available to use freely. There is no direct connection between the publisher and broker. All the data must be transferred through the broker. This protocol uses three qualities of service parameters. So the failure rate in message delivery is very less (Fig. 2).

Global positioning system usually carried in phone or high-end vehicles to track the vehicle movement and to detect the location. The GPS navigation system is a device that accurately shows the location of the vehicle or any other equipment (Fig. 3). The data recorded by the tracker is stored in the file or it will be transferred to the device connected to it through GPRS. Satellite-based GPS tracking systems work anywhere in the world. GPS is most used by the military system, civil and commercial purposes. This is freely available around the world. So there is no need to pay for a subscription or any other service charges. The position of the vehicle is determined by taking longitude and latitude of that area, which can also include the speed of the moving vehicle, distance traveled, etc.

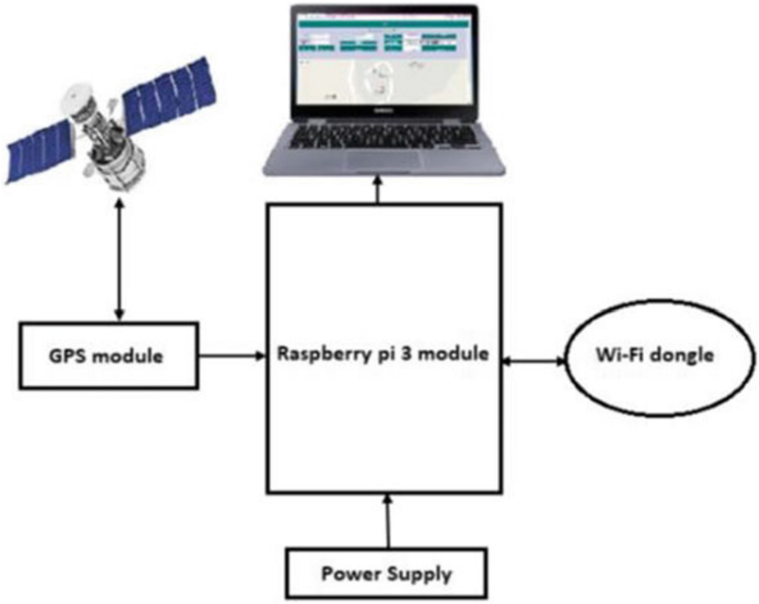


Fig. 1 Block diagram of vehicle tracking system

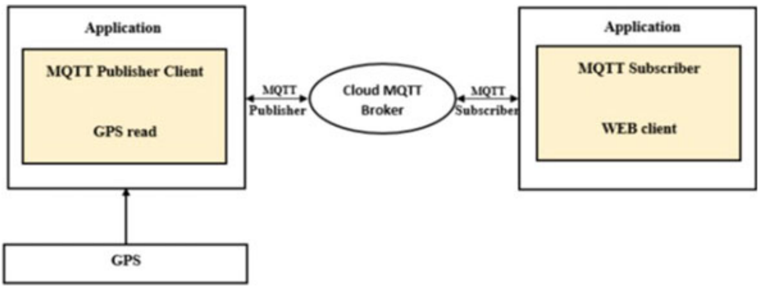


Fig. 2 MQTT protocol communication

4 System Implementation

It consists of a Raspberry Pi kit which includes GPS modules and antenna and a front end web page to display, Remote car, and power supply.

- GPS is used to get information like the location of the vehicles.
- Raspberry Pi is used to connect the tracker.
- The web page is used to display data (Figs. 4 and 5).

Fig. 3 GPS module with antenna

Fig. 4 Experimental set-up

This activity template provides input fields and a sample implementation of an activity that asks users to connect to the cloud with their credentials. This template includes:

- Setting Connection—which provides login to the cloud.
- Entering device ID—which provides subscription (Fig. 6).

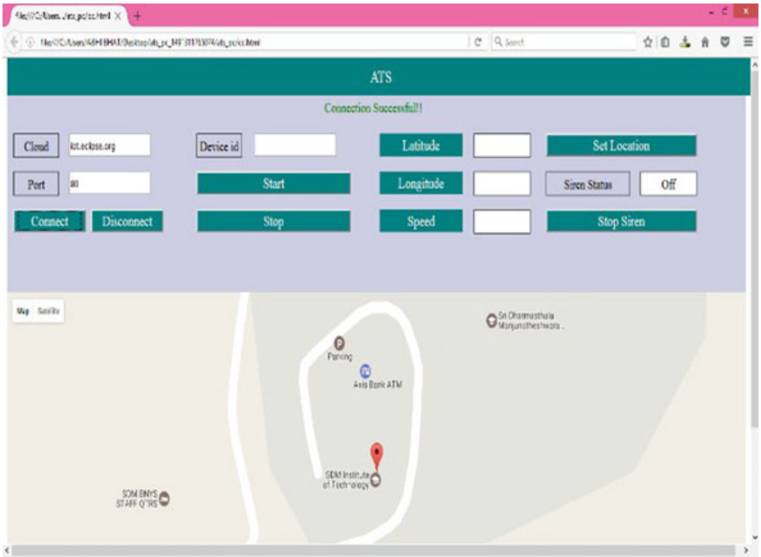


Fig. 5 Setting up connection to cloud, and displaying the connection successful

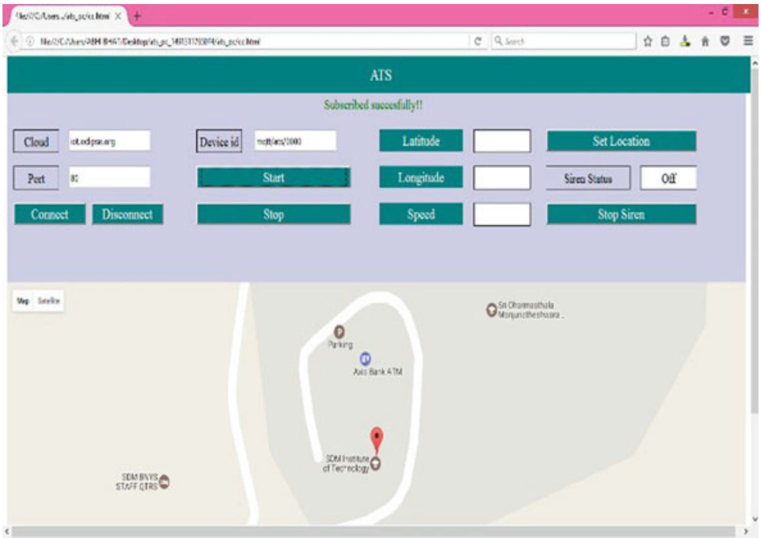


Fig. 6 Entering the device ID

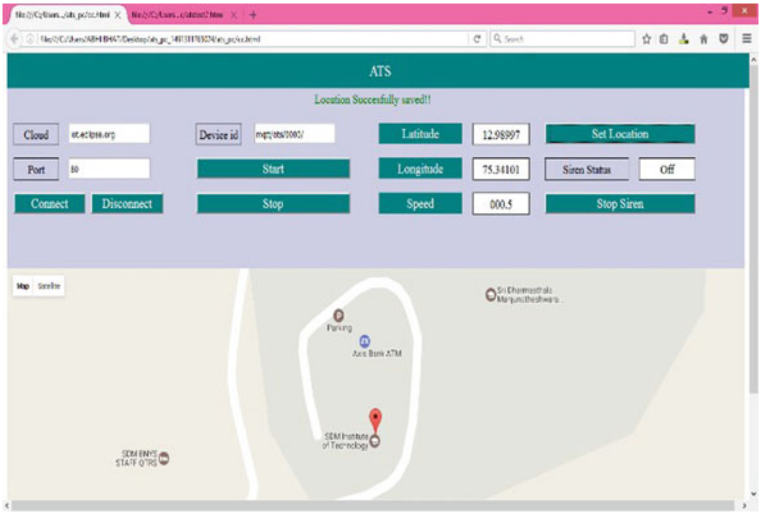


Fig. 7 Setting up set location, and displaying location where the vehicle is parked

This activity template provides input fields and a sample implementation of activities that asks users to enter their device ID with their credentials. This template includes:

- Entering device ID which provides subscription.
- START button to allow to display the location values.
- STOP button to stop the subscription (Fig. 7).

This activity template provides input fields and a sample implementation of activities that asks users to press the set location button. This template includes:

- Latitude button—which displays the latitude value of the parked vehicle.
- Longitude button—which displays the longitude value of the parked vehicle.
- Speed button—which displays the value at which speed the location values are received from the tracker (Fig. 8).

This activity template provides input fields and a sample implementation of activities that asks users to press the set location button. This template includes:

- Latitude button—which displays the change in latitude value of the parked vehicle.
- Longitude button—which displays the change in longitude value of the parked vehicle.
- Speed button—which displays the value at which speed the location values are received from the tracker.

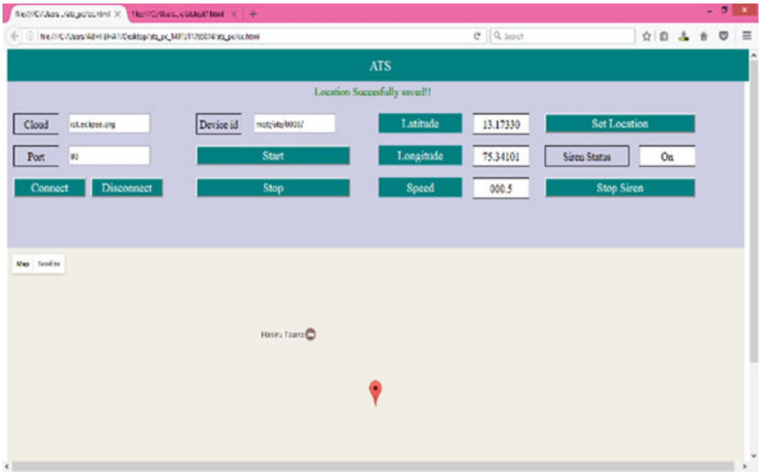


Fig. 8 Displaying the change in the location values of the vehicle that is parked

5 Conclusion

In this paper, the proposed model plays a vital role in real-time monitoring and tracking of vehicles. The system is designed in such a way that the user need not worry about parking the vehicle in public places. In the system set-up, the vehicle is composed of Raspberry Pi, Wi-Fi connection, a GPS tracker to get location information. The GPS navigator continually tracks the vehicle from the parking place. When the theft occurs immediately, users get alerted through siren happening in the application end. Longitude and latitude of the vehicle get recorded in the time of parking, so if this gets changed, a bit user gets the siren. Small vehicles and small scale vehicles will not be having the facility to track the theft. This system is very easy to configure and cost effective. This model works in real-time and gives accurate results.

6 Future Work

If the vehicle speed is beyond the "speed limit" in sensitive areas like schools, hospitals, etc. The driver must get worned through alert to get control over the accelerator. Providing an accurate details like phone number and distance of the nearest hospital, police station from the current location. Designing an android application for convenient usage. Auto alert system when disaster happen. Mobile application can be developed using android tool.

References

1. Hu, J. M., Li, J., Li, G. H.: Automobile anti-theft system based on GSM and GPS module. In: 2012 Fifth International Conference on Intelligent Networks and Intelligent Systems (ICINIS), pp. 199, 201, 1–3 Nov 2012
2. Shammi, S., Islam, S., Rahman, H. A., Zaman, H. U.: An automated way of vehicle theft detection in parking facilitated by identifying moving vehicles in CCTV video stream. In: 2018 International Conference on Communication, Computing and Internet of Things (IC3IoT), pp. 36–41, Sept 2018
3. Sreedevi, A. P., Nair, B. S. S.: Image processing based real time vehicle theft detection and prevention system. In: Process Automation, Control and Computing, pp. 1–6, Jan 2011
4. Babu, K. M., Raghunand, M. V.: Vehicle number plate detection and recognition using bounding box method. In: 2016 IEEE International Symposium on Communication Control and computing Technologies (ICACCCT), pp. 106–110
5. Li, L., Zhang, Y., Lv, S., Chen, Y., Liu, T., Xu, X.: Design of intelligent infrared vehicles detect system based on ZigBee. In: Chinese Automation Congress (CAC), p. 1
6. Stallings, W.: Cryptography and Network Security‖, 4th edn, Prentice-Hall, Englewood Cliffs, NJ, USA
7. Almomani, I. M., Alkhalil, N. Y., Ahmad, E. M., Jodeh, R. M.: Ubiquitous GPS vehicle tracking and management system. In: IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT) 2011
8. Lita, I., Cioc, I. B., Visan, D. A.: A new approach of automobile localization system using GPS and GSM/GPRS transmission. In: 29th International Spring Seminar on Electronics Technology, 2006. ISSE'06, pp. 115, 119, 10–14 May 2006
9. Fleischer, P. B., Nelson, A. Y., Sowah, R. A., Bremang, A.: Design and development of GPS/GSM based vehicle tracking and alert system for commercial intercity buses. In: 2012 IEEE 4th International Conference on Adaptive Science & Technology (ICAST), pp. 1, 6, 25–27 Oct 2012
10. Nagaraja, B. G., Rayappa, R., Mahesh, M., Patil, C. M., Manjunath, T. C.: Design & development of a GSM based vehicle theft control system. In: 2009 International Conference on Advanced Computer Control, ICACC'09, pp. 148, 152, 22–24 Jan 2009
11. Qian, M., Gao, H., Liu, W.: Android based vehicle anti-theft alarm and tracking system in handheld communication terminal. In: IEEE International Conference on Consumer-Electronics-Taiwan (ICCE-TW), 2018
12. Mukhopadhyay, D., Gupta, M., Attar, T., Chavan, P., Patel, V.: An attempt to develop on IOT based vehicle security system. In: IEEE International System on Smart Electronic Systems (iSES) (Formerly iNis), 2018
13. Mohanasundaram, S., Krishnan, V., Madhubala, V.: Vehicle theft tracking, detection and locking system using open CV. In: 5th International Conference on Advanced Computing and Communication System (ICACCS), 2019
14. Sathyanarayana, M., Mahendra, S., Vasu, R. B.: Smart security system for vehicle using IOT. In: 2nd International Conference on Green Computing and IOT (ICGCIOT), 2018
15. Geetha, M., Priyadarshini, T., Sangeetha, B., Sanjana, S.: Anti-theft and tracking mechanism for vehicle using GSM & GPS. In: Third International Conference on Science Technology Engineering & Management (ICONSTEM), 2017

Secure Wireless Internet of Things Communication Using Virtual Private Networks

Ishaan Lodha, Lakshana Kolur, K. Sree Hari and Prasad Honnavalli

Abstract The Internet of Things (IoT) is an exploding market as well as an important focus area for research. Security is a major issue for IoT products and solutions, with several massive problems that are still commonplace in the field. In this paper, we have successfully minimized the risk of data eavesdropping and tampering over the network by securing these communications using the concept of tunneling. We have implemented this by connecting a router to the Internet via a virtual private network while using PPTP and L2TP as the underlying protocols for the VPN and exploring their cost benefits, compatibility and most importantly, their feasibility. The main purpose of our paper is to try to secure IoT networks without adversely affecting the selling point of IoT.

Keywords IoT · Networks · Security · Sensors · Wireless communication · Sensor networks

1 Introduction

IoT is the connection of everyday mundane devices and things to a network like the Internet. IoT has gained immense popularity in the recent years due to the low cost of and easily deployed sensors and actuators which can easily be controlled by micro-controllers and availability of IPv6 addresses.

IoT is also one of the largest sources of data in the world. It is estimated to have generated more than 500 ZB by the end of 2019 [1]. The data will be mainly sensor data from the manufacturing and healthcare industries as they account for over 70% of IoT devices. Only 0.06% of total devices that can be connected to the Internet are actually connected, which shows the potential for a huge IoT data explosion.

I. Lodha (✉) · L. Kolur · K. S. Hari · P. Honnavalli
PES University, Outer Ring Road, Bangalore, India

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_70

2 USP of IoT—Challenges for Security

The main selling points for the widespread and easy adoption of IoT are the very reasons that hinder its security and do not allow easy implementation of conventional security procedures on them. Thus, any security procedure that is designed for IoT devices should be in tandem with these USPs and complement them in order for these procedures to be adopted by the industry.

- Small Size: IoT components are typically very small with minimal processing power. They have the minimum processing capability to serve their purpose, which allows them to remain cheap and small. They are not capable of executing conventional forms of heavy encryption and other security procedures.
- Energy Efficiency: IoT devices are predominantly battery powered and consume minuscule amounts of power. In adding security features, we cannot drastically increase the power consumption of the devices as it would render them incompetent in the market [2]. The small battery also accounts for its small size and low cost which enable their large-scale deployment.
- Usage Lifetime and Accessibility: IoT devices are deployed for considerably long periods of time of up to 5 years without almost any direct user contact or maintenance intervention. Thus the methods implemented should not require users to access the IoT device for anything in this duration. They are also deployed in remote inaccessible locations, need for repeated intervention would be a hindrance to this cause.
- Cost: IoT devices are generally cheap and thus if implementation of security solutions makes them considerably more expensive, then it would not perform well in the market, thus the cost effect of security features has to be minimal.

Thus, several constraints are in place in making IoT networks secure without disrupting the present status quo in the industry. It has been attempted to secure these networks with minimum impact on the status quo and thus device a solution acceptable to and implementable by the industry.

3 Approaches and Their Feasibility

The most common approaches to IoT security that are being extensively explored in ongoing research are as follows.

3.1 Hardware-Implemented Security

Implement on chip encryption for secure end to end communication by keeping the encryption mechanism separate from the rest of the system, making it harder to break or intercept. These security chips essentially give you a trusted environment that can be used for what is called a 'hardware root of trust.' These have a very specialized operating system, a specialized environment, that is built into that chip, all designed from scratch with security as a top priority. The first layer of security is on-device mitigation and hardware-implemented security is one way of achieving that. It is for security over the application layer and despite the network being secure if the end devices are not, attacks and threats are possible.

3.2 Device Authentication

There is a conflict in the implementation of device authentication, that is between how secure the authentication technique is against how practical it is to implement it onto a small, low-cost IoT device. We need to apply the Principle of Adequate Protection where we need to choose a method that is a balance of both, that is secure enough but also cost-effective. The RSA key used for communication like secure shell (SSH) between hosts can be used for authentication of IoT devices and the recipient host. The drawback of this method is that the RSA key is large in size and its generation is computationally intensive because it has to generate a large number of bits. This is not feasible to be implemented on minute IoT devices since they need to be energy effective [3].

3.3 Lightweight Encryption

There are various requirements to decide what sort of cipher to use: size, power consumption and processing speed (throughput, delay). Once we judge the edge device based on these parameters, we decide on using either a symmetric or an asymmetric encryption algorithm [4]. Symmetric encryption is where only one key is used to encrypt as well as decrypt information and asymmetric encryption is where there are two cryptographic keys, namely a public key as well as a private key where the public key can be used to encrypt information and data but only the private key of the user can be used to decrypt it [5].

3.4 Securing the Network

A plethora of companies use this process in allowing employees remote access to their business systems over the public Internet. Virtual private networks give a virtual direct connection between systems over the public Internet. They extend personal networks over the Internet. By connecting the IoT devices to the edge devices over a VPN we can ensure that the data is not eavesdropped on or altered en route. This is a tried and tested method of transferring sensitive data over an unsecured network and does not add computational or temporal overhead.

4 Tunneling

The mechanism of the tunneling protocol is as follows: It takes the part of the packet that contains information and uses it to store the packets that actually assist in solving the problem. The protocols used by tunneling are layered, which consist of protocols such as OSI or TCP/IP. The packets that assist in service are known as payloads, that operate at a layer which is one below that of the parent packets in which they are encapsulated [6]. Proposed work uses tunneling protocol as a way to create a virtual private network. It hides all communication between the hosts connected to it by way of encryption and hiding the hosts, thus creating an illusion of the absence of any devices or communication.

The most commonly used protocols for tunneling are:

4.1 Point-to-Point Tunneling Protocol

Point-to-point tunneling protocol also known as PPTP. The main concept of this protocol is a client–server design. This runs on the data-link layer in the OSI hierarchy. This protocol is widely used to tunnel data over the Internet. In order to initiate the connection, the user launches a point-to-point tunneling protocol client that in turn establishes a link to their Internet provider. After this, there is a TCP connection established between the client and the server of the virtual private network. It makes use of port 1723 and the concept of General Routing Encapsulation (GRE) to create the tunnel. This setup can also be done over a local network. The next step is the information flow [7].

4.2 Layer 2 Tunneling Protocol

The objective of this protocol is to use an encryption mechanism in conjunction with it, that is passed through the tunnel in order to achieve security of data. This is also an improvement over the point-to-point tunneling protocol. The protocol it uses to send the data and the header of L2TP is UDP. The reason for choosing UDP over TCP is so that we can avoid the ‘TCP meltdown problem.’ The point-to-point sessions are established within an L2TP tunnel. The protocol does not provide any sort of confidentiality by itself. Protocols such as IPsec are used to secure the L2TP tunnel. Together they are known as L2TP/IPsec. There are two ends of any L2TP tunnel. They are known as the LNS and the LAC. The LNS or the L2TP network server keeps searching for new tunnels. Once a connection is set up the transfer of data is full duplex. There are upper-level protocols which are used in order to use this protocol for networking; this is done using an L2TP session. Either the L2TP network server or the L2TP Access Concentrator can send a request for a session. Since each session is independent from the other, we can set up a large number of virtual networks through one tunnel. There may be a constraint since each packet data should not be greater than the maximum transmission unit of that datagram. The types of packets are similar to the ones in PPTP: control and data. Here, there is a guarantee that the protocol provides security when transmitting control packets. If we need reliability for data packets, it must be done by the protocols that run inside L2TP.

5 Campus Building Network

In an attempt to implement our VPN-based security concepts, we tried to tackle another real-life problem around us, that of physical security of the university campus. The system we have implemented is a secure IoT system with wireless sensor networks which serve as a tool for the estate management for remotely checking and controlling infrastructure. For the purpose of this tool, we have divided the rooms of the campus into two categories, one comprising classrooms, seminar halls, washrooms and so on which are considered non-sensitive and should be accessible to all students while the second category is the one with sensitive rooms and to which only select persons should have access which include the likes of offices of faculty members, storerooms, examination offices, etc. For the scope of this paper, we will concentrate on non-sensitive type of rooms only. With respect to these rooms, the tool should be able to detect occupancy and at the end of the college day should lock a vacant room and notify the estate management about occupied rooms post college hours. It should also automatically switch off all electrical appliances of unoccupied classrooms. The estate management should also be able to give overriding commands for each and every room from a central management system. The communication between sensors/actuators and the decision-making server must be wireless and secure.

The IoT network comprised of NodeMCUs with on chip ESP8266 Wi-Fi module and Arduino UNO with external ESP8266 Wi-Fi module as the micro-controllers. We have used a variety of micro-controllers to demonstrate scalability in that respect. Each room has a set of sensors comprising humidity sensor, passive IR motion sensor and temperature sensor. These sensors combined together give a stock of the situation in the classroom and about its occupancy. Each room also has a set of actuators, which include a door lock (electromagnetic or servo based), and AC relays to control the electrical appliances. The sensors log their reading in the server every second along with the state of the actuators, i.e., if the room is locked or not and if the appliances are on or off. The server reports the state of the rooms to user and saves the incoming data in a log file along with a room ID for each room. According to some rules, the server sends back instructions to each room's micro-controllers such as to switch on the appliances or to lock the room. The micro-controllers are programmed to automatically switch off the appliances if the room is empty and if the server sends a command to lock the room but if the room is occupied, then the micro-controllers will not lock that room but send an error back to the server for the security to go check the room.

The architecture of this system has all sensors and actuators connected to micro-controllers and the Wi-Fi module of both micro-controllers connected to a router. Multiple adjacent rooms share a common secure router and this router is connected to the public Internet and over a VPN on the Internet to the central server. For the sake of the prototype, the secure router was connected to another unsecured open router which mimicked the public Internet. We tried both point-to-point tunneling protocol (PPTP) and Layer 2 tunneling protocol (L2TP) as the underlying protocol of the VPN. For PPTP we used the pated tool on Ubuntu machines and for L2TP we user the macOS Server application which created a L2TP VPN server. The secure router connects to the VPN server and allows secure tunneled communication with the server. We connected another computer to the unsecured router or the Internet and used Wireshark to sniff for packets being transmitted through that router. In both tunneling protocols, we could not sniff any packets while without the VPN, all packets could be sniffed, which meant we could intercept them and even use ARP spoofing to mount a MITM attack. Thus, the use of tunneling returned positive results.

The server is written in Java and uses swing GUI toolkit and several threads. Each secure router has port forwarding enabled to allow each ESP8266 module to have full-duplex communication with the central server. The central server has a Hash Table of room IDs mapped to the IP address of the secure router for that room and port numbers for the port forwarding from the router to that room's micro-controllers. The GUI has a text box for room ID and buttons to send commands like Lock/Unlock the room and turn on/off the appliances, and it returns the status of execution of these commands and errors faced if any. It also has an option to get the status of a room which is retrieved from the most recent log and printed in proper format, giving status of the lock, appliances, occupancy and temperature. The server has also been configured to automatically sound the fire alarms if the temperature of the room exceeds a certain threshold. The GUI also allows the user to sound a buzzer in the entered room ID which can be used as a signal to empty the room. The format of the log is the room number hyphen flag 0 or 1 for motion sensed or not, followed by flag for if appliances are on or not, then it has the angle of the servo which gives the status of the door lock, ending with the temperature. IT further has an administrator panel to add, change and delete room IDs and the corresponding IP address and port numbers.

6 Major IoT Cybersecurity Concerns

6.1 Network Attacks

- MITM: In network security, a man-in-the-middle attack (MITM) is an unauthorized attack where the attacker tries to intercept communication between two parties and gain access to their sensitive information. Active eavesdropping is a common type where an attacker intercepts the incoming messages and sends new messages of his own to the receiver. These types of attacks are more prevalent in IoT devices due to the lack of attention given to changing passwords, and also because they do not have any check on security certificates' validity.
- Botnets: Botnets are a myriad of devices connected by the Internet, handled by the attacker who then manipulates these devices, sometimes without their knowledge, to deploy DDOS attacks. IoT devices are more susceptible because they lack an antivirus protection layer which would probably detect the malware, that is in contrast, not detected when dormant, hence greatly easing the process of deploying DDOS attacks because of the reduced security complexity aspect [8].

6.2 Other Attacks

All the software attacks like viruses, worms, trojans, etc., that can be used on a computer system can also be used to attack IoT devices. In IoT devices, this threat is amplified due to their minimalistic nature and size. They do not have any protection in form of malware and cannot support them as well. IoT devices can also be subjected to side channel attacks where the attacker uses data like computational power used in encryption to statistically guess the algorithm and key used.

7 Future Work

A defense in-depth approach with multiple security mechanisms in place would be an ideal setup. Lightweight encryption of data before it is sent to the router to be transmitted via the tunnel covers all bases and makes the network completely secure. Lightweight encryption is also the most suitable combination as a multitude of research has been undertaken in the same and is comparatively more feasible. It could also be device authentication on connecting to the VPN and before communication can commence via the tunnel.

8 Conclusion

Wireless networks of IoT devices have been secured by hiding the devices on a virtual private network. The proposed work connects IoT devices to the main servers through a VPN server which hides the communication between them from the public Internet by way of VPN's underlying encryption. L2TP and PPTP have been used, tested and discussed for the given scenario. The proposed security works on the principle of concealment. It has a classroom network to control locks and appliances which are controlled by a central admin server, the communication between them is over an open Wi-Fi, to mimic the public Internet, and has been secured by the use of a VPN server that facilitates the communication.

References

1. Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper, Nov 2018
2. Haroon, A., et al.: Constraints in the IoT: The World in 2020 and Beyond, pp. 1–2 (2016)
3. Zhang, J., et al.: Proximity based IoT device authentication. In: IEEE INFOCOM 2017-IEEE Conference on Computer Communications, IEEE, 2017
4. Sridhar, S., Smys, S.: Intelligent security framework for IoT devices cryptography based end-to-end security architecture. In: 2017 International Conference on Inventive Systems and Control (ICISC), IEEE, pp. 1–5, Jan 19 2017
5. Katagi, M., Moriai, S.: Lightweight Cryptography for the Internet of Things, pp. 2–4. Sony Corporation, 2008
6. Riahi, A., Challal, Y., Natalizio, E., Chtourou, Z., Bouabdallah, A.: A systemic approach for IoT security. In: Proceedings—IEEE International Conference on Distributed Computing in Sensor Systems, DCoSS 2013. pp. 351–355 (2013) https://doi.org/10.1109/dcoss.2013.78
7. Miklós, G., Harmatos, J.: Mobile layer 2 virtual private network over internet protocol networks. U.S. Patent No. 9,173,153. 27 Oct 2015
8. Nawir, M., et al.: Internet of things (IoT): Taxonomy of security attacks. In: 2016 3rd International Conference on Electronic Design (ICED), IEEE, 2016

A Contingent Exploration on Big Data Tools

Latika Kharb, Lakshita Aggarwal and Deepak Chahal

Abstract In past few years, size of the data is growing exponentially by extreme fast rates, for instance, the size of data growth was ten times faster in the growth due to various means such as the data from mobile devices, remote sensing, sensing aerial devices, recording frequency of radio waves. Until most recently, most of the data was never analysed and most of the time it was discarded. The data stored requires lots of storage space whereas later due to lack of storage space the data is either ignored or deleted due to lack of storage space to process the data. Sometimes, we are even capable of storing the data but until that data is not processed, it is raw useless data to us because that will not be able to fetch with new insights. In the analysis, we face two types of challenges, first is the lack of storage space and second a suitable software to process this data. In this paper, we have discussed about the evolution of 4 V's of big data, levels of big data tools, various data tools along with a comparative analysis of those tools on the basis of distinguished features like mode of software, data processing, language support, data flow security, latency and fault tolerance is also generalized for better understanding.

Keywords Data growth · Data flow security · Big data · Data tools

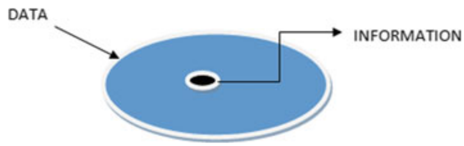
1 Introduction

The process of building applications has been a journey, and it varies depending on one's application requirements and purpose [1]. In today's scenario, where the world is budding at an massive scale from all source that has led to the development of big data; big data is just a compilation of huge quantity of data. These are the data sets that are so big which cannot be processed from the traditional tools or traditional application software. These data includes those data sets that are extremely large in size that are beyond the ability of mankind and even the commonly used software tools cannot even process them (Figs. 1 and 2).

L. Kharb (✉) · L. Aggarwal · D. Chahal
Jagan Institute of Management Studies, Sector-5, Rohini 110085, Delhi, India
e-mail: latika.kharb@jimsindia.org

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1_71

Fig. 1 Data versus information



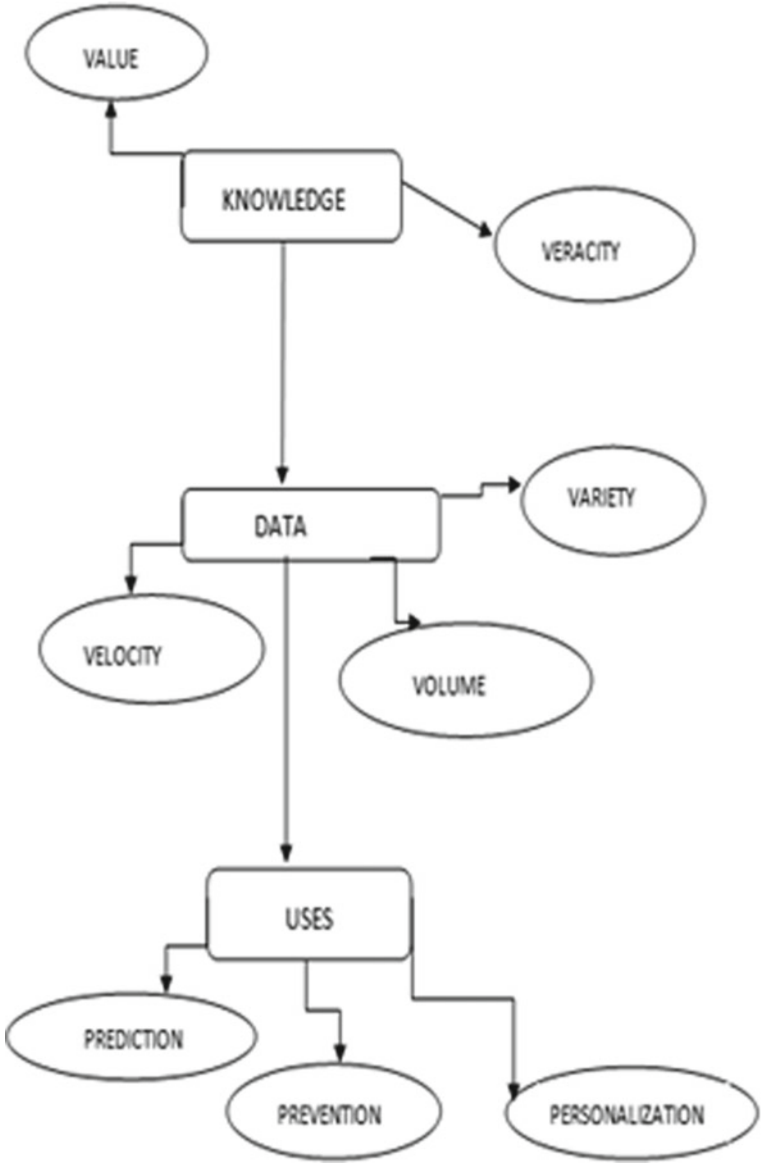


Fig. 2 Information and knowledge

Big data faces numerous challenges such as data capturing, data storage, data analysis, searching, sharing, transferrable data, data visualization, throwing queries on it, updation, security. The data included in this is usually unstructured data, or sometimes it is semi-structured data like the data that does not include completely defined data models and are even not organized properly to be executed or processed. Big data has lots and lots of text which is even difficult to find out the important information out of it. The size of the data is increasing per day per second. The size of the data varies from bytes to terabytes and to exabytes of data which is not even processed to get information out of it. The urgent need of an hour is to use those new tools and techniques that are able to process and execute large data sets and reduce complexity even at a massive scale [2].

2 Characteristic Features of Big Data

Big data is known from 3 V's to 4 V's. The 3 V's were variety, velocity and volume. As the digitization increased its pace one more V was added, then it came to be known as the 4 V's, i.e. veracity [3, 4].

1. **Variety**: it reflects the nature and type of the data, i.e. whether the structured, semi-structured or unstructured and in any form: text, image, audio or video.
2. **Velocity**: velocity is the rate at which the data is generated and the speed at which the data generated is used to process the data sets in order to overcome the growing demands of the industry. As the digital world is growing with the corresponding rate at which the data is on the rise.
3. **Volume**: volume is the amount of data through which it decides whether the size of the data is small, big or massive. So, the task of big data is to analyse those data sets and extract useful information out of it.
4. **Veracity:** veracity tells about the captured data is accurate or not. It also tells about the accountability of the data.

Shift from 4 V's to 5 V's

1. **Value**: it is to assess the quality assurance of a data before analysing the data sources. It is necessary to find whether the data sources that are likely to generate information are accountable or not.

3 Beyond 5 V's, 3 P's

1. **Prediction:** it helps to enquire knowledge forecast demand, predict problems and behaviours of the changing influence of the big business houses.
2. **Prevention:** it identifies risks dangers and prevents them. Security is the major concern nowadays this application is associated more with the health sector.

3. **Personalization:** it shows the amount of information necessary for the individual sitting in real time at the front of the screen to analyse the data at an accountable time producing quick responses.

4 Levels of Big Data Tools

Data privacy incorporates not only the security aspects but it is the data that you are collecting and people do have an expectation of privacy. For this, big data tools could be divided into three levels on the basis of complexity (Fig. 3).

1. **Layer one**: it is the largest and wide area to process "open source tools" and includes information regarding servers, infrastructure and storage space, e.g. cloud players use this layer.
2. **Layer two**: business players build proprietary apps in it, e.g. Hadoop
3. **Layer three**: it includes vertical-specific apps, and it includes all the effective big data tools. It includes the applications such as: PWC, Accenture.

Now, the broad division comes into play is big data tools are divided broadly into two categories:

- **Proprietary apps**: these are those apps which are the private property of an individual no other can access it without seeking the permissions. In this, only the owner can make changes in it. Some examples of proprietary tools are: Pentaho Business Analytics, Splunk and Jaspersoft.

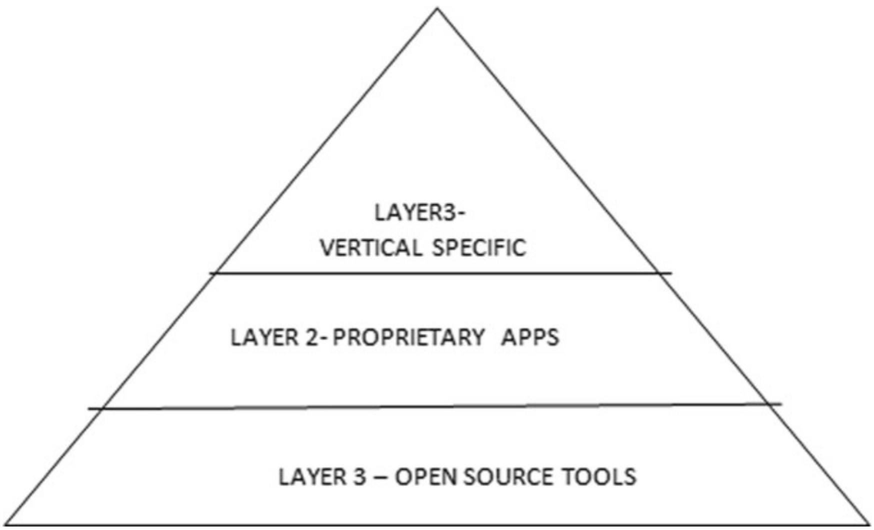


Fig. 3 Level wise classification of big data tools

- **Open source tools**: these tools have the source code that is openly available to anyone. People freely use these open source tools. The software is openly distributed to all that any can use for their personal use. Some examples of open source big data tools are: Mongo DB [5], Hadoop, Apache Cassandra, Apache SAMOA, Elastic search, and High-Performance Computing Cluster (HPCC).

5 Exploration of Big Data Tools

A huge repository of terabytes of data is generated each day from modern information systems and digital technologies such as Internet of Things and cloud computing [6]. The data collected from different sources of applications all over the world from various fields is expected to double in recent two years and so. The data stored has no purpose until it is able to provide with useful information. Large number of tools whether open source or proprietary are available to process big data. These big data tools prove to be useful for analysing the raw data to useful code of information.

- **Tool 1: Apache Cassandra**

It is an open source big data tool that is used due to its effective management of large amount of data sets. Cassandra offers varied amount of scalable data in data analysis. Cassandra has numerous advantages such as improved performance, fault tolerance and scalable data. It offers "NO SQL" database and high-speed transactional data. Cassandra offers huge workloads of big data across many nodes and does not fail even at a single point.

- **Tool 2: Hadoop**

Hadoop is one of the open source big data tool which is used for storage of huge amount of data. In 2006, it was developed by computer scientists Doug Cutting and Mike Cafarella. Hadoop has advanced software library that contains number of superior libraries for large volume of data sets. Developers of Hadoop are working more even to upgrade the product more so that it could deal with great processing capabilities. It is extremely useful for the applications where fault tolerance is an issue, storing large amount of data and providing high throughput after processing the data. It is used in batch processing. It works on the concept of MapReduce for applications. Its architecture includes kernel Hadoop, MapReduce, Apache Hives and Hadoop Distributed File System (HDFS). The techniques of reducing maps are used for large data sets programming that is internally based on the technique that divides the large data sets into different corresponding data sets and conquers them all together to process and further analyse the data.

In Hadoop, we part the data into subsequent steps and then add up those methods in two steps first map step known as master node which is followed by the second step known as reduce step. The first step divides the data into small sub-problems which are then distributed further to second step. After all the master nodes combine, output of smaller problem is produced by second step. Both the steps, i.e. map and reduce steps, work synchronously for problem-solving.

- **Tool 3: Plotly**

It is an open source software that is used for Web-based data applications and analytical applications. It is an online tool that makes information graphics easy. It also allows to share interactive graphs and apps by data scientists throughout the world. Many companies use this tool for graphs and charts designing, e.g. Google, New York University.

- **Tool 4: Mongo DB**

It is an open source free software. Mongo DB is developed by Apache licence and GNU Affero General Public License. It contains schema less documented database. It is used mainly on those data sets that vary or change frequently and contains unstructured or semi-structured data. It is adaptable in nature that fits in almost every use case. It offers varied uses such as: storage of data from many sources, catalogues and many more. It also handles even the complex queries fired on database. The architecture is a single database split into numerous databases which are further split into number of collections, collections into many documents and lastly documents into key-value pairs.

- **Tool 5: Drill**

It is a distributed open source framework that works efficiently on huge data sets. It is another product produced by Apache to process large data sets, i.e. even petabytes and exabytes of data sets in seconds and can process millions of records. Drill supports large tons of file system with many data formats, varied query languages and data sources. Drill divides data sets in seconds analysing large data sets even in a blink of an eye. It uses Hadoop Distributed File System (HDFS) for storing the data. Drill also performs MapReduce jobs and analysis of batch processing of the big data.

- **Tool 6: Storm**

It is a distributed open source big data tool designed for processing unstructured data even in real time. It is easy to configure, reliable technology, handles fault tolerance and provides competitive performance. Storm cluster is same as the cluster of Hadoop. The tasks are all performed in a storm cluster in different topologies. It is a great competent to any programming language. It is a real-time framework, e.g. twitter uses storm. A storm senses channels for messages all time until user terminates to end the sensing of the channel. A cluster of a storm contains two different nodes master and worker nodes. Master and worker nodes act as the nimbus and the supervisor nodes. Nimbus works to perform functions like job tracker and supervisor works as task trackers like Hadoop's MapReduce framework. The responsibility of nimbus is to distribute code across the different layers of the storm cluster, schedule it and assign the task to main nodes and it even monitors the complete system. Whole process is implemented in different parts of the storm as in the different layers of the topology.

- **Tool 7: Splunk**

Splunk is the proprietary big data tool. It is a real-time platform that exploits machine-produced data. It allows us to combine big data and cloud technology. The facilities of Splunk allow users to search data, monitor growing data and analyse all the machine-produced data through Web interfaces. The output produced by it is shown in interesting manners such as graphs, charts and reports. The results shown are user friendly as developer can easily analyse just by visualizing the reports created by it. It is totally different from other big data processing tools as it can perform real-time search for faster results. The main objectives big data tool Splunk serves are providing metrics for many applications, diagnosing the problem for system and it supports intelligently for business-oriented operations.

6 Key Technologies

Some key technologies enable big data analytics for business:

- Predictive analytics: extracting information from existing data.
- No SQL databases: allows high-performance analysis of information.
- Knowledge discovery tools: provide automated analysis solutions.
- Stream analytics: quick processing to extract real-time insights of data.
- Distributed storage: enables real-time data analytics
- Data virtualization: creating virtual structures for big data systems
- Data integration: delivers trusted data from various sources
- Pre-processing of data: involves transforming raw data into an understandable format
- Data quality: ability of a given set of data to fulfil an intended purpose.

7 Big Data Applications

Some applications of big data analytics for business include:

- Government Sector
- Social media
- Technological issues
- Science
- Research
- Detecting frauds
- IT sector
- Call centre analytics.

8 A Comparative Analysis of Big Data Tools

Features	Apache Cassandra	Hadoop	Plotly	Mongo DB	Drill	Storm	Splunk
Mode of software	Free open source distributed	Open source and free source	Open source	Open source and free software	Open source and distributed framework	Open source and distributed	Proprietary tool and real-time platform
Data processing	SQL queries, streaming data, machine learning	Batch processing system	Interactive graphs and user-defined images	No SQL database, document-oriented database	Distributed execution environment for large-scale data	Supports stream processing	Machine data processing
Language support	Supports Java, Python, Node.js	Supports Java, C, C++, Ruby, Perl, Python	Supports JavaScript, Python, R Plotly	Supports Python, JavaScript, C, C++, PHP, Ruby, Perl	Supports ANSI SQL (industry standard query language)	Supports non-JVM Languages	supports Python, Java, JavaScript
Data flow	Data is written into a commit log as well as in memory store	MapReduce computation data flow does not have any loops. It is a chain of stages	Track the flow of individual items through Plotly Sankey	Data streams focussed on massive flow of data from multiple fire horses and then routing to systems	It makes it to use SQL query non-relational databases and file systems	Design as directed acyclic graph with streams used to process the data	Pipeline executes a series of processors which operates on the data

(continued)

(continued)

Features	Apache Cassandra	Hadoop	Plotly	Mongo DB	Drill	Storm	Splunk
Security	It provides secure communication enabling SSL encryption ensuring data is secured	Supports Kerberos authentication. It manages leverage active directory Kerberos and LDAP for authentication	It uses secure Web socket to open interactive communication between browser and server by encrypting data in transmission with SSL	Secured by default it encrypts our data both in transit and at rest and makes it easy to control access with role-based user management	It provides interactive and secure SQL analytics at the scale of petabytes	It is a fault tolerant processing system with real-time computation capabilities	Splunk ES is a premium security solution requiring a paid licence. It quickly detects and responds to internal and external attacks
Latency	Lower latency by replicating across multiple data centres	It offers higher latency	It offers reduced latency due to caching	It offers low latency analytics as it cannot analyse semi- or unstructured data	It offers low latency	Extremely low latency	It offers moderately high latency as it works with real-time architecture
Fault tolerance	It is tolerant of both network partitions and nodes dying	MapReduce is highly fault tolerant. No need to start the application from scratch in failure	It has trusted advisor fault tolerant review	Replica set deployment architecture supports fault tolerance	JDBC connection improves fault tolerance	If any node dies, storm will automatically restart for fault tolerance	It offers appreciably high fault tolerance

9 Conclusion

After studying and comparing all the big data tools, we have analysed the rate at which data is growing at high pace; so we need an appropriate technology/tools to tackle the problems such as size, velocity, value and veracity. So the shift was necessary to jump from one phase to the other. Now everywhere big data tools have spread its wings by reaching to the companies like e-commerce, data visualization, data integration, distributed storage and many more. Big data has changed the perspective of thinking from not just only data collection but till the processing to transforming, modelling [7], analysing and taking correct decisions. We analysed that no tool can suppress all have equal value in terms of ease to use to users, economic growth, higher trends of development. Big data tools like Apache Cassandra, Hadoop, Plotly, Mongo DB, Drill, Storm and Splunk are used to facilitate analysis. In this paper, we discussed all the popular and useful big data tools including both open source or proprietary and also discussed how they could help us to process big data and thus the raw data is converted into useful code of information. We have also analysed the tools on the basis of their characteristic features.

10 Limitation

Big business houses have started shifting from conventional ways to new techniques of big data technologies but still there is a lack of performance and fault tolerance. At times, security is also a major concern when the confidential information sometimes gets leaked to third party alliance. Inconsistency in data sets is also observed which frequently changes. There are large diverse challenges that are growing in front of big data solution makers.

11 Future Scope

Within the next 5 years, the cloud-based big data analytics (BDA) solution will grow faster as the size of data is growing. All adoption of technologies with pervasive demand of big data analytics in IOT (Internet of things) and AI (artificial intelligence) are integrated together. Visual data discovery tools will grow faster as for better services to layman in an enterprise serving mankind in better way. This is a growing demand of the nation for big business houses to expand further and help in boosting Indian economy.

References

1. Kharb, L.: A perspective view on commercialization of cognitive computing. In: 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, pp. 829–832 (2018)
2. Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., Khan, S. U.: "big data" on cloud computing: Review and open research issues. Inf. Syst. **47**, 98–115 (2015) https://doi.org/10.1016/j.is.2014.07.006
3. Laney, D.: 3D data management: Controlling data volume, velocity and variety. META Group Res. Note. **6**(70)
4. Goes, P. B.: Design science research in top information systems journals. MIS Q. Manag. Inf. Syst. **38**(1) Marr, Bernard (6 March 2014). Big Data: The 5 V's Everyone Must Know
5. Ram, B. K., Kumar, S. A., Prathap, S., Mahesh, B., Sarma, B. M.: Chapter 19 remote laboratories: For real time access to experiment setups with online session booking, utilizing a database and online interface with live streaming, Springer Nature (2018)
6. Acharjya, D. P., et al.: A survey on big data analytics: Challenges, open research issues and tools. Int. J. Adv. Comput. Sci. Appl. **7**(2), (2016)
7. Kambatla, K., Kollias, G., Kumar, V., Gram, A.: Trends in big data analytics. J. Parallel Distrib. Comput. **74**(7), 2561–2573 (2014). 2016/11/21

Author Index

V. Bindhu et al. (eds.), *International Conference on Communication, Computing and Electronics Systems*, Lecture Notes in Electrical Engineering 637,
https://doi.org/10.1007/978-981-15-2612-1