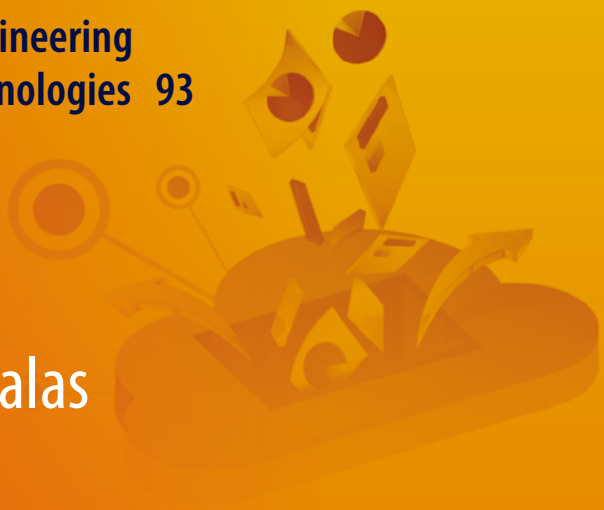
Lecture Notes on Data Engineering
and Communications Technologies 93



P. Karrupusamy
Valentina Emilia Balas
Yong Shi *Editors*

Sustainable Communication Networks and Application

Proceedings of ICSCN 2021

Springer

Lecture Notes on Data Engineering and Communications Technologies

Volume 93

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

Indexed by SCOPUS, INSPEC, EI Compendex.

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

More information about this series at https://link.springer.com/bookseries/15362

P. Karrupusamy · Valentina Emilia Balas · Yong Shi
Editors

Sustainable Communication Networks and Application

Proceedings of ICSCN 2021

Springer

Editors
P. Karrupusamy
Department of Electrical and Electronics Engineering
Shree Venkateshwara Hi-Tech Engineering
Erode, Tamil Nadu, India

Valentina Emilia Balas
Department of Automatics and Applied Software
Aurel Vlaicu University of Arad
Arad, Romania

Yong Shi
Department of Computer Science
Kennesaw State University
Kennesaw, GA, USA

ISSN 2367-4512 ISSN 2367-4520 (electronic)
Lecture Notes on Data Engineering and Communications Technologies
ISBN 978-981-16-6604-9 ISBN 978-981-16-6605-6 (eBook)
https://doi.org/10.1007/978-981-16-6605-6

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 conference proceedings of ICSCN 2021 to all the participants, technical program committee members and editors.

Foreword

It is with deep satisfaction that I write this Foreword to the Proceedings of the ICSCN 2021 held in Surya Engineering College (SEC), Erode, India, at July 29–30, 2021.

This conference was bringing together researchers, academics and professionals from all over the world, experts in sustainable networking technology, sustainable applications, sustainable computing and communication technologies.

This conference particularly encouraged the interaction of research students and developing academics with the more established academic community in an informal setting to present and to discuss new and current work. The papers contributed the most recent scientific knowledge known in the field of ultra-low-power sustainable system, sustainable vehicular ad hoc networks, Internet-enabled infrastructures for sustainability, sustainable mobility and vehicle management. Their contributions helped to make the conference as outstanding as it has been. The local organizing committee members and their helpers put much effort into ensuring the success of the day-to-day operation of the meeting.

We hope that this program will further stimulate research in sustainable big data frameworks, energy and power constrained devices, low-power communication technologies, sustainable vehicular ad hoc networks, smart transport systems and smart data analytics techniques through this exciting program.

We thank all authors and participants for their contributions.

Dr. E. Baraneetharan
Conference Chair, ICSCN 2021
Associate Professor, Dept. of EEE
Surya Engineering College
Erode, India

Preface

This conference proceedings volume contains the written versions of most of the contributions presented during ICSCN 2021. The conference provided a setting for discussing recent developments in a wide variety of topics including communications, networks and sustainable applications. The conference has been a good opportunity for participants coming from various destinations to present and discuss topics in their respective research areas.

ICSCN 2021 tends to collect the latest research results and applications on intelligent data communication technologies and networks. It includes a selection of 65 papers from 262 papers submitted to the conference from universities and industries all over the world. All the accepted papers were subjected to strict peer-reviewing by 2–4 expert referees. The papers have been selected for this volume because of quality and the relevance to the conference.

ICSCN 2021 would like to express our sincere appreciation to all authors for their contributions to this book. We would like to extend our thanks to all the referees for their constructive comments on all papers; especially, we would like to thank the organizing committee for their hard work. Finally, we would like to thank the Springer publications for producing this volume.

Erode, India — P. Karrupusamy
Arad, Romania — Valentina Emilia Balas
Kennesaw, USA — Yong Shi

Acknowledgements

ICSCN 2021 would like to acknowledge the excellent work of our conference organizing committee and keynote speakers for their presentation on July 29–30, 2021. The organizers also wish to acknowledge publicly the valuable services provided by the reviewers.

On behalf of the editors, organizers, authors and readers of this conference, we wish to thank the keynote speakers and the reviewers for their time, hard work and dedication to this conference. The organizers wish to acknowledge Thiru. Andavar. A. Ramasamy, Ln. K. Kalaiyarasan, Dr. S. Vijayan and Prof. E. Baraneetharan for the discussion, suggestion and cooperation to organize the keynote speaker Dr. R. Kanthavel, Department of Computer Engineering, King Khalid University, Kingdom of Saudi Arabia, of this conference. The organizers also wish to acknowledge the speakers and participants who attend this conference. Many thanks are given to all persons who help and support this conference. ICSCN 2021 would like to acknowledge the contribution made to the organization by its many volunteers. Members contribute their time, energy and knowledge at a local, regional and international level.

We also thank all the chair persons and conference committee members for their support.

Contents

About the Editors

Dr. P. Karrupusamy working as a Professor and Head in Department of Electrical and Electronics Engineering at Shree Venkateshwara Hi-Tech Engineering College, Erode, India. In 2017, He had completed doctorate in Anna University, Chennai and in 2007, he had completed his post graduate Power Electronics and Drives in Government College of Technology, Coimbatore, India. He has more than 12 years of teaching experience. He has published more than 60 papers in national and international journals and conferences. He has acted as conference chair in IEEE and Springer international conferences and Guest editor in reputed journals. His research area includes Modeling of PV arrays, Adaptive Neuro-Fuzzy Model for Grid Connected Photovoltaic System with Multilevel Inverter.

Dr. Valentina Emilia Balas is currently Full Professor at "Aurel Vlaicu" University of Arad, Romania. She is author of more than 300 research papers. Her research interests are in Intelligent Systems, Fuzzy Control, Soft Computing. She is Editor-in Chief to *International Journal of Advanced Intelligence Paradigms* (IJAIP) and to IJCSE. Dr. Balas is member of EUSFLAT, ACM and a SM IEEE, member in TC—EC and TC-FS (IEEE CIS), TC—SC (IEEE SMCS), Joint Secretary FIM.

Dr. Yong Shi is currently working as an Associate/Tenured Professor of Computer Science, Kennesaw State University and Director/Coordinator of the Master of Computer Science. He is the responsible for directing the Master of Computer Science program, reviewing applications for the Master of Computer Science. He has published more than 50 articles in national and international journals. He acted as an editor, reviewer, editorial board member and program committee member in many reputed journals and conferences. His research interest includes Cloud Computing, Big Data, and Cybersecurity.

Inter-networking: An Elegant Approach for Configuring Layer-2/Layer-3 Devices for Attaining Impulsive Outcomes

Arvind K. Sharma, Savita Wadhawan, Richa Datta, and Sudesh K. Mittal

Abstract A switching/routing network consists of unique but similar units with a maximum of 48 units, and each unit has an equal amount of capacity for both transmitting and receiving data. However, not all of the units are used all of the time, resulting in a loss of bandwidth as well as available units, thus designing and developing a suitable network (topology) with specific segregation is required to simplify the real networking in any organization. And, in order to successfully communicate between networked devices on the same or various IP networks while eliminating any faults, specific protocols must be followed. However, there are numerous tools and techniques available for achieving the desired goal; however, not every technique is appropriate for all switching/routing scenarios, so determining which tool, methodology, or technique to use and how to use it in light of the requirements at hand is a critical task, along with resolving any discrepancies or issues. We will address the fundamentals of layer-2/layer-3 switching in this article, as well as actual implementation for well-designed and developed inter-networking. *Keywords: Access Control List (ACL), Address Resolution Protocol (ARP), Administrative-Distance (AD), Bandwidth, Broadcast domain, Collision domain, Convergence, Distance vector (DV), Ethernet, Feasible-Distance (FD), Inter-network, Loop, Link state, Media Access Control (MAC), Packet Internet Groper (Ping), Spanning Tree Protocol (STP), Switching, Trunk, Virtual LAN (VLAN), Routing, Security.*

A. K. Sharma (✉) · S. Wadhawan
M.M. Institute of Computer Technology and Business Management (MCA), MM(DU), Mullana, Ambala, Haryana 133207, India
e-mail: arvind.sharma@mmumullana.org

S. Wadhawan
e-mail: savitawadhawan@mmumullana.org

R. Datta
M.M. Institute of Computer Technology and Business Management (HM), MM(DU), Mullana, Ambala, Haryana 133207, India

S. K. Mittal
Research and Innovation Network (CURIN), Chitkara University, Rajpura, Punjab 140401, India
e-mail: sudesh.mittal@chitkara.edu.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_1

1 Introduction

In an interconnected environment, when multiple network resources are attempting to communicate either regular or control information in the form of frames, the scope of communication at layer-2 is most likely confined to that broadcast domain only. To protect against unwanted traffic and to make the best use of available bandwidth and intra-network communication, i.e., communication between multiple BD., may be required, and for that reason, a planned methodology with specific rules and regulations is required to streamline LAN networking. So, while establish communication across various broadcast domains is permitted, it must be regulated in order to ensure network and network resource consistency, dependability, scalability, and security. The VLANs, Spanning Tree Protocol, Access Control List, Router on Stick, Etherchannel, and Port Security are few techniques which are used for attaining all mentioned features. The concept of virtual local area network is quite helpful to segregate the single broadcast domain into multiples and to utilize the wasted ports/interfaces in fruitful manner, Access Control Lists are used to filter the traffic which coming toward dedicated networks, Spanning Tree Protocols is used to avoid any network loops in order to protect against any type of congestion, Etherchannel is used for utilizing the blocked interfaces due to STP configuration and to increase the overall speed of communication, etc. Apart from this when there is need to make communication possible between multiple BDs which are not concerned to particular organization, i.e., movement is from one IP network to other, on that time, routing comes in rescue where layer-3 devices/protocol used for that purpose. In this paper, our most of the focus is on layer-2 switching, layer-3 routing, and on various techniques which can be used in systematic manner to take overall benefits from layer-2/layer-3 configuration on devices like network switch/bridges/routers, etc.

Organization of this paper is like Sect. 2 briefed about 'Switching and Routing Technologies', Section 3 devoted to 'LAN Switch's Working Methodology' with set of practical work, and Sect. 4 discussed about Efficient Routing Methodology, and Section 5 about short analysis of protocols then we'll wind the paper with suitable Conclusion and Future Work.

2 Switching and Routing Technologies

Switching is an important concept in networking used to transfer information from one network to another, achieved with a device known as switch. Switch is a small device used to connect multiple resources in a network broadly categorized into two types: managed and unmanaged. The unmanaged switch does not allow to make changes and works out of box, whereas managed switches allow you to program it and provide great flexibility to access the networks. Further, switching techniques are classified into three categories: circuit switching, message switching, and packet switching [1, 2] (see Fig. 1). In circuit switching, data transmission

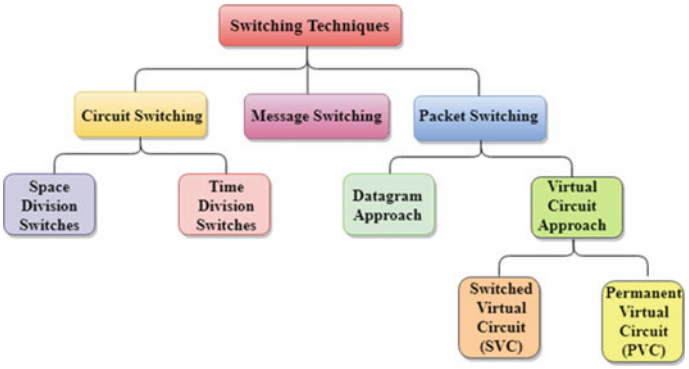


Fig. 1 Types of switching

between sender and receiver take place with the establishment of a dedicated path. The circuit switching is implemented either with Space Division /Time Division Switches, whereas, in message switching, there is no dedicated path between sender and receiver, i.e., complete message is transferred as a unit and this message is routed through intermediate nodes. At each node, the entire message is stored and then forwarded to next node, known as, store and forward network.

On the other side, in packet switching, message is divided into small packets and sent individually. These packets are given a unique number to identifying them at receiver end. If packet is not received at receiver end or received but corrupted, then that packet will be sent again. Further, packet switching follows two kinds of approaches: first is datagram approach in which packets are moved independently as datagram toward destination on different paths known as connectionless switching. Second is virtual circuit approach in which packets are moved from source to destination after establishing a virtual path between them and also known as connection oriented switching. The significant point here is when we are considering the movement of packets definitely we are working with layer-3 approaches, devices, and protocols in order to route the packet from source to destination with best suitable path without compromising with the efficiency and security of the overall network design. So for the purpose to work with layer-3, we have two types of protocols available with us which are named as Routing/ Routed protocols about which we will discuss in more detail in Sect. 4 of this paper.

3 LAN Switch's Working Methodology

The main purpose behind switching is to use the hardware address of devices on local area network for connectivity and make network segments. In these segments (individual networks), lots of devices are connected which participate in particular type of communication viz simplex, half-duplex, full-duplex, depends on the requirements,

Fig. 2 Network switch

and if we're considering a particular segment the significant point here is that it either be a part of single broadcast domain (if default VLAN used) or multiple broadcast domains (if multiple VLAN used) in layer-2 devices, especially switch [3]. Even though we've many devices available like Hub, Bridge, and Switch etc. in order to work with layer-2 but in the rest of the paper we'll focus only on uses of Switch (see Fig. 2) due to its popularity. Now before going into depth of all the techniques used in switching let's first understand the fundamental of switch's working.

The switch or network switch is a layer-2 (data-link layer) device which is helpful to make physical topology possible so that connected devices can communicate with each other efficiently as oppose to Hub/Bridge which used in order day. Now the question arises, why older devices are not that much good to use as switch is? The reason behind is in case of Hub there is only one collision domain, i.e., at a time only one device will be able to use the bandwidth and rest will be in waiting stage, and if two or more devices on the same time try to use bandwidth collision will occurs, but this is not the case in switches because in switch apart from single collision domain, collision domain is as equal as number of port are there. So every connected device have equal opportunity to use the shared bandwidth on the same time, i.e., multiple devices are able to send or receive data on the same time.

3.1 Switch Configuration

Even though the switch looks similar to order device, i.e., bridge, but it have many features which is the main reason for its popularity; yes both device prepare filter tables (MAC table, i.e., Media Access Control table) and break up collision domain, no doubt in that but bridge uses dedicated software for preparing such tables and switch uses application-specific integrated circuits. Keep in mind, term Switch/Bridge can be interchangeably used. Now let us move to three primary function of switch.

Address Learning: The actual meaning of address learning is to first find out correct address (MAC/hardware address) and then use that particular address for further communication between devices. Now, thing is when a switch is first powered on the filter table is blank, i.e., no information available about which device is connected to which port/interface of that switch. So question here is how switch updates the blank filter table with address of the device? Okay! When any device starts to transmit with logical address (IP address, already known) of its own and of the receiver,

then during this initial process of transmission an interface of switch receives a frame (including IP address of both source and destination, MAC address of source, etc.), then switch places the frame's source address in the filter table, allowing it to remember for future which interface the sending device is located on. Now filter table is updated with sender's MAC address, but no information about receiver's MAC address available still, only its IP address know. So how received frame at switch will be forwarded to correct device? Now broadcasting comes in rescue, switch will broadcast the received frame to all the connected devices excluding sending device with destination IP address, and when frame received by each device and if receiver finds out the IP address available in frame belongs to it, it will respond back else it will drop the frame. So if switch received response from any of the connected device back corresponds to its broadcast, it will again update the filter table with MAC address of receiver with correct interface/port and next time when same two devices tries to communicate there will be no broadcasting at all, and there will be one-to-one communication between these unless interfaces changed. The switch does not need to flood the frame as it did the first time, because now the frames can and will be forwarded only between the two devices.

In order to understand the address learning activity of switch, first configure a network switch and attach few host PCs with it, check the MAC table of switch which will be empty definitely, then give IP address from particular domain to all the PCs and apply ping from any PC to any other PC, and again check the MAC table of switch which will be updated. Now without wasting time, let us move to configuration part with suitable example. As per the below given activity (see Fig. 3), we have one network switch with default name switch in the configuration connected with four PCs (PC0, PC1, PC2, PC3), and after changing its name in configuration mode to Library, we just checked the MAC table with command to show MAC-Address-Table, which is empty (see Fig. 4), i.e., no information is available till date at which fast Ethernet port which device is attached.

Now after receiving empty MAC table, we just given IP address to connected PCs viz PC0: 192.0.0.2/24, PC1: 192.0.0.3/24, PC2: 192.0.0.4/24, PC3: 192.0.0.5/24, etc., and checked the connectivity between PC0 → PC3 with PC0 > ping 192.0.0.3 command (see Fig. 5) and again run the command do show MAC-Address-Table we get MAC table in switch get updated (see Fig. 6) with two entries one for PC0

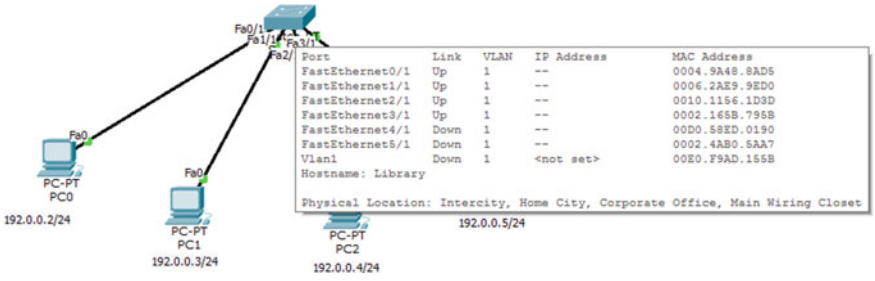


Fig. 3 Network switch in LAN

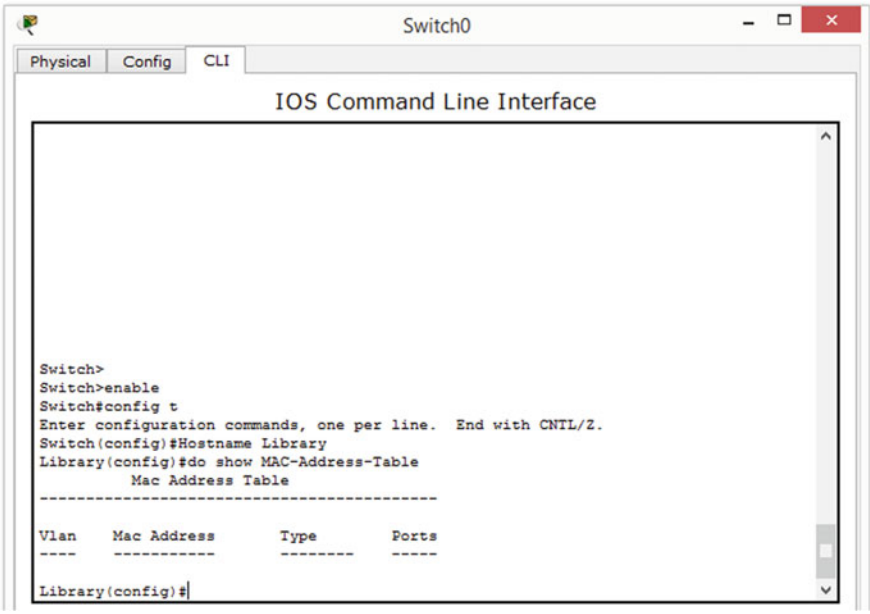


Fig. 4 Empty MAC table in network switch

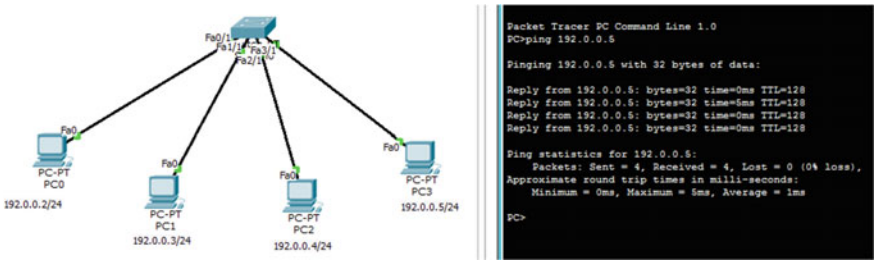


Fig. 5 Ping from PC-0 → PC-3

and other for PC3 with corresponding port numbers, actually initially when ping command started the packet/frame broadcasted by the switch to find out the destination PC and when it received the response from desired PC entry marked in the MAC table for future 1-to-1 communication between PC0 and PC3.

Filter/Forward Decision: It means first scan the filter table and then decide from which interface forwards the frame toward appropriate receiver, i.e., when a frame arrives on switch's interface the destination MAC address is compared with the data available in filter table. If the destination MAC address is known/listed in the table, the frame is forwarded with correct exit interface. This process always preserves bandwidth on network segments and is called frame filtering.

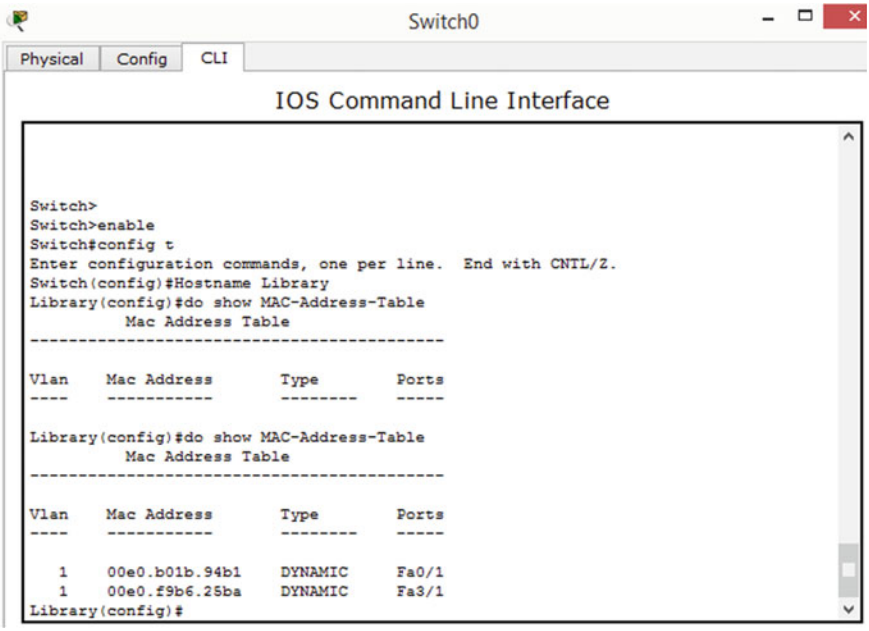


Fig. 6 Updated MAC table after ping

Loop Avoidance: Whenever any network administrator configuring a network, sometimes redundant links between switches are also provided by him, notion is good because it helps to prevent against complete network failures if one link stops working but, if redundant links are not provided by taking necessary precautions, they often cause more problems than they solve, i.e., frames going to be flooded down among all redundant links simultaneously, which creates network loops (bandwidth exhausted). So in order to deal with such evil situations, Spanning Tree Protocol (STP) comes in rescue.

3.2 Spanning Tree Protocol

The main purpose of this protocol is to identify and then stop network loops from occurring in network. Now thought comes in mind how it is actually handling such a crucial task? This protocol by default enables in all switches (We are considering here CISCO switches), and it vigilantly monitors the network and makes sure that no loop occurs, by shutting down redundant links if required. This protocol first uses Spanning Tree Algorithm (STA) for creating topology table/database and then by scanning that database search destroys redundant links [4, 5]. Let us discuss about few terms of layer-2 network before understanding how STP works.

1. **STP**: It is a bridge protocol that uses the STA to create spanning tree topology database and to find redundant links dynamically for further shutdown operation in order to stop loops. Switch/Bridge exchange Bridge Protocol Data Unit (BPDU) messages with other bridges to detect loops.
2. **BPDU**: The information exchanged with BPDUs used by switches for selection of the root switch/bridge after comparing the parameters available in, concerned to each switch.
3. **Root Bridge**: The bridge with best ID is called root bridge, the election of root bridge is the key point in STP, and this bridge is a focal point in network. All the significant decisions like, which port has to be set in blocking mode, and which port has to be set in forwarding mode, are made with the perspective of root bridge. The bridge with the lowest bridge ID becomes the root bridge in the network.
4. **Bridge ID**: This ID is concluded by combining the bridge priority and MAC address of bridge, with the help of which STP keeps tracks all the switches in the network *(32,768 is default priority on all Cisco Switches).*
5. **Non Root Bridge**: These bridges exchange BPDUs with all bridges and update the STP topology database on all bridges.
6. **Root Port**: The root port of a bridge is always that port from where either link directly connected to the root bridge or the shortest path toward the root bridge. The election of root port is made by checking the **Port Cost** if more than one link connected to root bridge. If multiple link have same port cost, the **Lowest Port Number** will be elected as root port.
7. **Port Cost**: This cost is determined by the bandwidth of a link connected to that port.
8. **Designated/Forwarding Port**: The port (s) from where the cost to reach to destination is best/lowest is called designated port, from where all the frames forwards (see Table 1; Fig. 8).
9. **Non Designated/Blocked Port**: The port (s) with a higher cost than the designated port will be put in blocking mode, but these ports will always listen to frames.

The STP first elects a root bridge which takes over network topology decisions and after mutual decision made by bridges that who will hold the root bridge responsibilities. Every bridge finds root port and designated port. While deciding who will be the root bridge, ports go through five levels called **Port States** which are discussed below.

Table 1 Cost of Ethernet networks

S. no.	Link speed	IEEE cost
1	10 Mbps	100
2	100 Mbps	19
3	1000 Mbps/1 Gbps	04
4	10 Gbps	02

1. **Blocking**: The port which is in block state will not forward frames and just listens to BPDUs. *All ports of a switch are in blocking state by default when it is powered up.*
2. **Listening**: The port listens to BPDUs for making it sure that no network loops ever occur before initiating frame forwarding (No MAC table populate yet).
3. **Learning**: The port learns all the paths in the topology after listening. The ports which are in learning states populate the MAC table, but still do not forward frames.
4. **Forwarding**: The port sends/receives all data frames on the switch's port (i.e., from root/designated ports).
5. **Disabled**: A port in the disabled state will not participate in the frame forwarding/STP (i.e., nonoperational port).

These five states are part of convergence, which occurs when ports on bridges/switches have transitioned to either the forwarding or blocking modes. Keep in mind, no data going to be forward until convergence completed. Convergence is significant for assuring that all devices have the same database at the end. *It usually takes 50 s to go from blocking to forwarding mode, and it is not recommended to change the default STP timer but we can.* So till date, there is no solution available to tackle this particular problem, and this will be the new area of research or enhancement in STP because this is a severe limitation as per our point of view.

As per the below given activity (see Fig. 7), three switches Switch0, Switch1, and Switch2 are connected with each other, and we can easily identify that there is loop existed in the topology. So when such thing happens automatically, STP start running on all the switches and identify the root bridge and with respect to this root bridge two ports will remain in forwarding state and one will be in blocking state, e.g., Port Fa0/7 of Switch0 in forward state and Fa0/6 in blocking state (See Fig. 7),

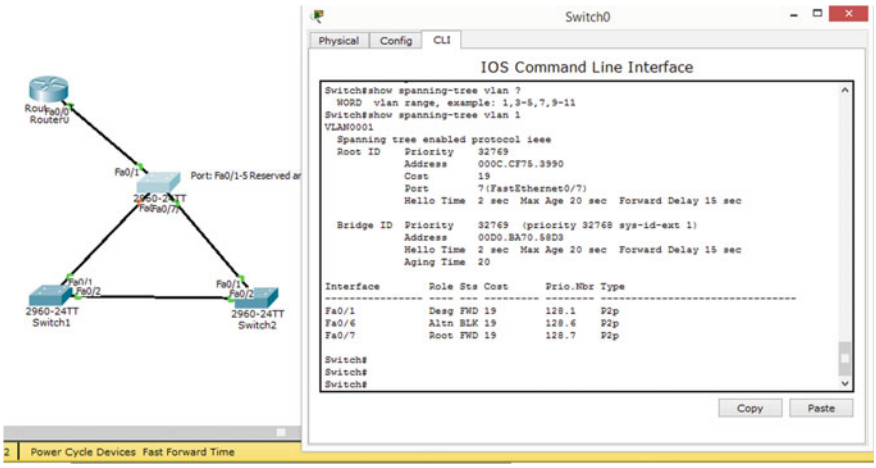


Fig. 7 STP configured switch

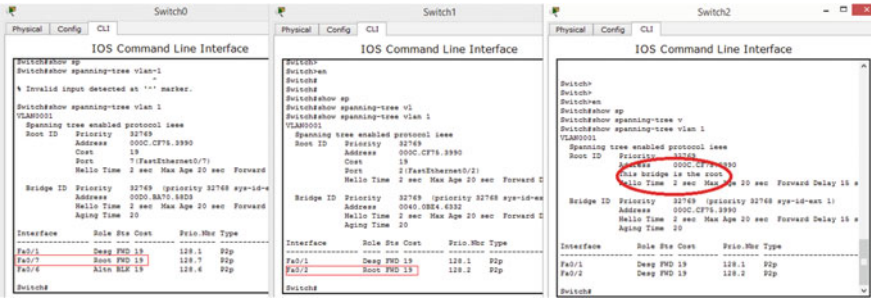


Fig. 8 Root bridge and root ports in switches

Similarly in Switch1 and Switch Port, Fa0/1 and Fa0/2 are in forwarding state and cost is 19 as we are using fast Ethernet cables for connectivity (see Table 1). The Port Fa0/7 of Switch0 and Fa0/2 of Switch1 is the root port, i.e., a port which is directly connected to root Bridge which is Switch3 (see Fig. 8) having both the port Fa0/1, Fa0/2 are designated ports.

3.3 Virtual LAN (VLAN)

The switches are very helpful to break up collision domain (CD) and whenever we are going to connect devices of a particular workgroup/network definitely the use of layer-2 switches is highly recommended, so that coming bandwidth can be consumed in a well manner. But by default, all the devices which are connected are a part of single broadcast domain (BD), i.e., default virtual LAN (VLAN) [4], so how to use switches efficiently is a catch. Let us move to few scenarios.

1. If for particular network connectivity we used a switch with 48 ports and we only have 30 devices connected that means rest of the 18 port is a waste, so how to utilize these wasted ports?
2. If we are using switch (s) having 24 ports for every department, and now if user increased due to requirements in certain departments, does we will purchase new switch in order to manage this change?
3. If switch (s) with only 24 ports is available for every department for connecting workstation of that department, but we do not have that much space to adjust more than 20 workstations physically in particular department, but space for rest of the workstation is available which is far from actual department in premises, how to manage this thing securely and efficiently, so that no one else than authorize person of that department can access the dedicated network/resources?

The single solution of all the above-mentioned questions is to break up the BD in pure switched inter-network, but how do we break up broadcast domains? The answer is by creating a Virtual Local Area Network (VLAN) apart from available

default VLAN for every network in single or multiple switches (other solution is to use routers but it is a costly device as compare to switch and only limited to ISPs so we will not move with it). So let us understand what is a VLAN. The VLAN is defined as logical grouping of specific network users/resources connected to administratively defined ports on a switch (s), i.e., the ability to create smaller BD within layer-2 switched inter-networks, by assigning distinct ports of switch to different subnetworks (VLAN is treated like BD). Now the frames broadcast onto the network are only transferred between the ports logically grouped into the same network or VLAN. But for inter-VLAN communication (different BD), layer-3 device is still required, i.e., the use of layer-3 devices we cannot neglect. So if we are using router for inter-VLAN communication, the method used is called router on a stick. The VLAN are of two type a discussed below.

1. **Static VLAN**
   It is a usual way of implementing VLAN which is secure too. The switch port association with corresponding VLAN is configured manually by an administrator and retain there until administrator make changes. It is easy to set up and monitor such VLAN configuration and definitely working well where the movement of resources within the network is purely controlled.
2. **Dynamic VLAN**
   This type of VLAN automatically creates association between logically grouped ports with devices by using its MAC address. It required intelligent management software for creating such associations, i.e., MAC addresses have been entered into central VLAN management software, and if a device is then attached to an unassigned port, the VLAN management database can scan the hardware address and then assign and configure the switch port to the appropriate VLAN.

3.3.1 VLAN Links

The frames always switched either in single or multiple networks, having same or different type. So what to do with each frame depending on its hardware address and type, handled by any switch itself by thorough monitoring. The key point here is every frames handled differently with respect to the type of link on which it flowing. There are two distinct types of links used in switched inter-network which are access and trunk link.

1. **Access Links**: This type of link is concerned to one VLAN only, and device attached to such a link do not know about any VLAN membership at all. Before sending frame to access link, switches remove VLAN information from it and devices belongs to this link cannot communicate with other devices of different VLAN, i.e., communication limited to 1 VLAN/BD only.
2. **Trunk Links**: This type of link carry data of multiple VLAN, in order to make inter-VLAN communication possible by using layer-3 devices. Trunking allows making a particular port a part of multiple VLAN on the same time.

3.3.2 VLAN Identification Methods

Actually various type of frames are traveling through links, so switches use below-mentioned methods to identify which frames belong to which VLAN.

1. **IEEE 802.1Q**: This is IEEE standard for frame tagging and actually including a field into the frame for VLAN identification, i.e., if trunking occurs between different brand of switches, we have to use 802.1Q for the trunk to work properly.
2. **Inter Switch Link (ISL)**: This is CISCO proprietary and used for Fast/Gigabit Ethernet links. ISL routing can be used on a router interface, server interface, switch port, etc. It is only limited to CISCO devices.

In order to understand the concept of VLAN, let us move with few basic operations. Initially, we are creating VLAN on switches (Sw-1, Sw-2, Sw-3), i.e., VLAN No. 77, 88, 99 and assigning port to different VLANS, but same VLAN of all three switches have same port numbers (77: Fa0/1 - Fa0/4, 88: Fa0/5 - Fa0/8, 99: Fa0/9 - Fa0/12). The purpose of VLAN creation here is to segregate the broadcast domain in switched network (see Fig. 9), so that devices of different networks will not interfere in any other network's communication and stay secure from unauthorized access ever.

Now after segregating the bigger network to smaller once, it may be the requirement to make it possible that different devices of different VLAN be able to communicate, so for that apart from default access ports in switches we need a trunk port, which have to be connected to layer-3 device (here we are using router for configuring router on stick method). Yes, the layer-3 device is must because it is only able to handle/bypass the BD for transferring data from one network to other.

Now (see Fig. 10) as per the technique behind inter-VLAN communication, first of all there is need to give IP address on all the created VLAN in all the switches, i.e., IP from subnetworks to VLANs of every switch, VLAN-77 of Sw-1, Sw-2, Sw-3 take IP from same subnetwork and it will not be the case that VLAN-88, VLAN-99 of Sw-1, Sw-2, Sw-3 are also taking IPs from same subnetwork as from VLAN-77 is taking. After allocation of IP address, there is need to enable 802.1Q encapsulation on layer-device with gateway IP address so and making the port of Sw-1 (Fa0/24) as trunk port in order to make it possible traffic from distinct BDs can transferred

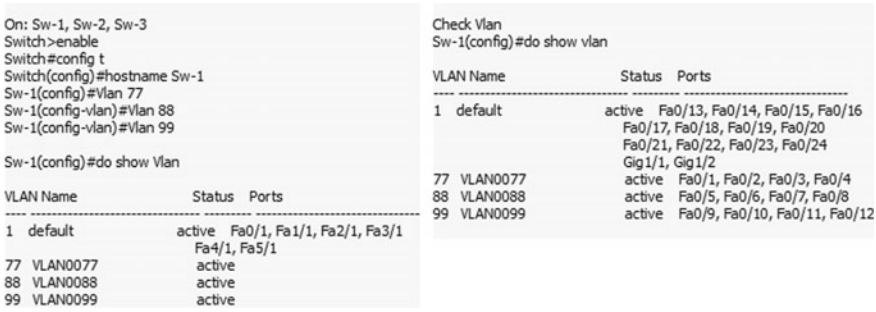


Fig. 9 VLAN creation and port allocation

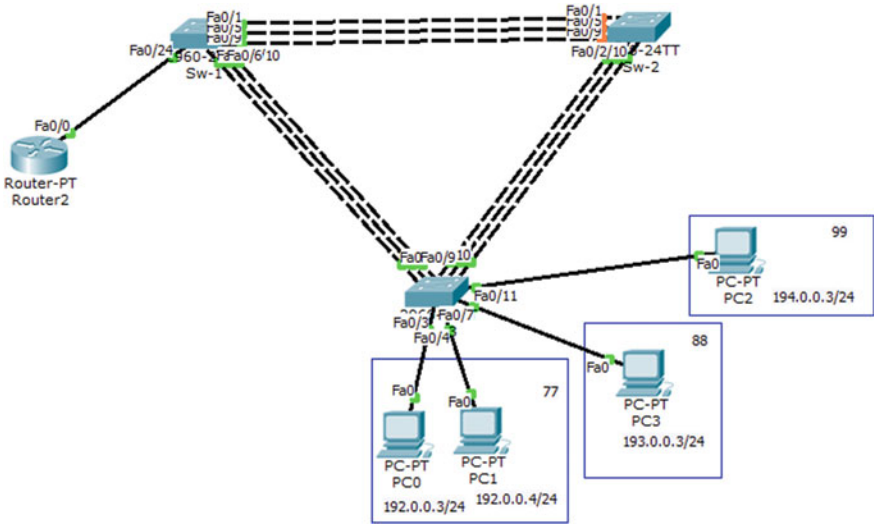


Fig. 10 Inter-VLAN communication with router on stick

without any problem (see Fig. 11). It is the choice of network administrator either to use router or layer-3 switch for inter-VLAN communication, but we are using here router for that purpose.

Step 4

On: Sw-1, Sw-2, Sw-3
Give Ip Addresses on VLan 77, 88, 99 on all switches

77: 192.0.0.2 255.255.255.0
88: 193.0.0.2 255.255.255.0
99: 194.0.0.2 255.255.255.0

```
Sw-1(config-if)#interface Vlan 77
Sw-1(config-if)#ip address 192.0.0.2 255.255.255.0
Sw-1(config-if)#no shut
Sw-1(config-if)#
Sw-1(config-if)#interface Vlan 88
Sw-1(config-if)#ip address 193.0.0.2 255.255.255.0
Sw-1(config-if)#no shut
Sw-1(config-if)#
Sw-1(config-if)#interface Vlan 99
Sw-1(config-if)#ip address 194.0.0.2 255.255.255.0
Sw-1(config-if)#no shut
```

Step 5

Now ping can work on inter VLan but not on Outer Vlan with PCs, so what to do ?

Create A Trunk Port with Router (use port which is not a part of any Vlan)

```
Router(config)#interface fa0/0
Router(config-if)#no shut
Router(config-if)#interface fa0/0.77
Router(config-subif)#encapsulation dot1Q 77
Router(config-subif)#ip address 192.0.0.1 255.255.255.0
Router(config-subif)#no shut

Router(config-subif)#interface fa0/0.88
Router(config-subif)#encapsulation dot1Q 88
Router(config-subif)#ip address 193.0.0.1 255.255.255.0
Router(config-subif)#no shut

Router(config-subif)#interface fa0/0.99
Router(config-subif)#encapsulation dot1Q 99
Router(config-subif)#ip address 194.0.0.1 255.255.255.0
Router(config-subif)#no shut
```

Step 6

Configure Trunk Port on Switch: Sw-1 for Fa0/24

```
Sw-1(config)#interface Fa0/24
Sw-1(config-if)#
Sw-1(config-if)#switchport mode trunk
```

Now all the PCs of different network are able to ping with each other which was not possible previously…

Fig. 11 Router on stick configuration with 802.1Q encapsulation

3.3.3 VLAN Trunking Protocol (VTP)

VTP is a CISCO proprietary; the conception behind VTP is to manage all configured VLAN across a switched inter-network for maintaining consistency of the network [4]. It permits an administrator to add, delete, rename, etc., and the VLAN information which then propagated to all other switches in same VTP domain. The switches can share VTP information with other switches only if these are configured into the same VTP domain. VTP information is shared between switches with trunk port. VTP worked with three different modes, i.e., if switches are configured with these modes, which are under given below.

1. **Server Mode**: The switch which is in server mode (default mode in CISCO switches) be able to create, add, delete, and rename VLAN or alter VLAN information in a VTP domain. The changes made to a switch in server mode advertised to the entire VTP domain and at least one server in VTP domain is must to propagate VLAN information.
2. **Client Mode**: In client mode, switches receive information from VTP server and also send/receive updates but cannot able to make any type of changes.
3. **Transparent Mode**: The switches in transparent mode do not participate in the VTP, and only forward VTP advertisements through any configured trunk link, i.e., just work as intermediary only.

In order to understand the working of VTP let's move with suitable example, (see Fig. 12) here we have three switches which are in different modes Client/Server/Transparent. The switch in server mode have full control like creation of VLAN, advertisements, etc., and switch with client mode is only able to send/receive the updates, but the key point is that in VLAN information, we can watch from both these mode configured switches; on the other hand, switch with transparent mode is only able to forward the updates, and no VLAN information will be shown here. Now, let's look at how to configure the VTP and see what data is available in various types of switches (see Fig. 13).

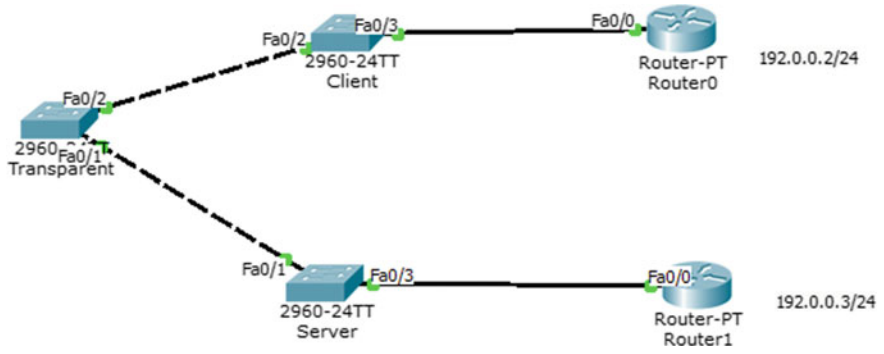


Fig. 12 VTP topology

```
Step-2: In Server, Create Domain and give Password on that, and change the VTP mode to Server

Switch(config)#VTP Domain Arvind
Changing VTP domain name from NULL to Arvind
Switch(config)#VTP Password Arvind123
Setting device VLAN database password to Arvind123
Switch(config)#VTP mode Server
Device mode already VTP SERVER.
```

```
Step-3: In Transparent, Set the same Domain Name and Password and change the mode to Transparent

Switch>en
Switch#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Switch(config)#VTP Domain Arvind
Changing VTP domain name from NULL to Arvind
Switch(config)#VTP password Arvind123
Setting device VLAN database password to Arvind123
Switch(config)#VTP mode Transparent
Setting device to VTP TRANSPARENT
```

```
Step4: In Client, Set Same Domain Name and Password, and change the mode to Client

Switch#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Switch(config)#VTP Domain Arvind
Changing VTP domain name from NULL to Arvind
Switch(config)#VTP password Arvind123
Setting device VLAN database password to Arvind123
Switch(config)#VTP mode Client
Setting device to VTP CLIENT mode.
```

Fig. 13 Configure VTP on switches

```
Step5: Make all the Connected port as Trunk Ports in all three Switches (F0/1 on Server, Fa0/1, Fa0/2 on Transparent, Fa0/2 on Client).

Server:
Switch(config)#interface Fa0/1
Switch(config-if)#Switchport mode dynamic desirable
Switch(config-if)#no shut

Transparent:
Switch(config)#
Switch(config)#interface range Fa0/1, Fa0/2
Switch(config-if-range)#Switchport mode dynamic desirable
Switch(config-if)#no shut

Client:
Switch(config)#interface Fa0/2
Switch(config-if)#Switchport mode dynamic desirable
Switch(config-if)#no shut
```

```
Step-7: Create VLAN on Server

Switch(config-vlan)#Vlan 2
Switch(config-vlan)#Vlan 3
Switch(config-vlan)#Vlan 4
Switch(config-vlan)#Vlan 5
```

Fig. 14 Create VLAN on server after making all ports as trunk ports in switches

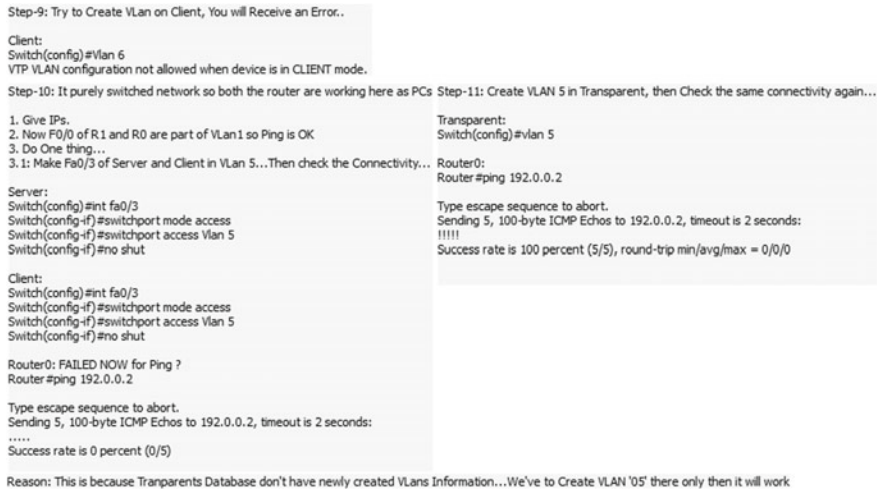


Fig. 15 Client unable to create VLAN and communication possible

The server will be able to create VLAN, but client will not see (Figs. 14 and 15), and one more point is even though VLAN information is available on client and server but still communication is not possible from Router0 →Router1 because the switch with transparent mode enabled do not have any VLAN information there, so in order to make the communication possible, we are manually adding VLAN-5 in that and now we can see that the success rate of communication is 100% when ping from Router0 →Router1 (see Fig. 14).

3.4 Etherchannel

Sometimes network administrators connect switches with multiple links (as already mentioned in Sect. 3.1 inside loop avoidance), the conception is good, there is nothing harm in that if such configuration well managed; reason for such configuration is that if one link get down due to any issue, second connected link comes in rescue and will start sending/receiving data, i.e., network still be operational. But main problem with such a configuration is that only one link will be in forwarding state and rest remains blocked (only listening) w.r.t. STP, so it means bandwidth allocated to blocked links is nothing but just a waste. The question is how to stop that waste. The right answer for that is to configure Etherchannel. Etherchannel was invented by CISCO as a LAN switch-to-switch technique for grouping several fast/gigabit Ethernet ports into one logical channel [6]. This particular technique has its own benefits, but the main purpose is just one to increase the speed between switches by reducing bandwidth wastage. Few benefits are as given below.

1. The maximum task concerned to the configuration can be performed on the Etherchannel interface apart from every port which ensures configuration consistency throughout the switch-to-switch links.
2. Load balancing can be possible on the links which are part of Etherchannel.

3.5 Security

The security is always a major concern when we are talking about inter-networking. There will be various resources which are connected for sharing information; the security for both data and devices is a mandate [7] in secure system. But here we are not talking about data security, we are focusing on device (especially layer-2) security, i.e., how to ensure layer-2 devices security, what will be the techniques used for, so below we are providing few significant ways to ensure security of switches.

- Shut down all unused ports in switch and shift these in a VLAN which is un-operational.
- Restrict access to the switches especially for altering configuration by enabling secure passwords/secret (i.e., fingerprint [8–11]) online VTY connection, so that Telnet be possible by authorize users.
- Dedicated VLAN should be used for trunk port (s) and avoid using default VLAN-1 for everything.
- Disable dynamic truncking protocol (DTP) from non-trunking ports.
- Enable port security feature on switches with required modes viz protect, restrict, and shutdown, for preventing unauthorized access [8], like if traffic is coming from certain port of another switch immediately take action, as just shut down the port which is receiving traffic. The DHCP Snooping, IP Source Guard, and ARP security are also features for port security in switches.
- In order to prevent DOS attacks, just disable unused services/protocols.
- Enable STP for Loop-Guard or avoiding loops from inter-network.
- The use of Access Control Lists (ACL) is highly recommended for filtering unwanted traffic.

4 Efficient Routing Methodology

The routing is the process of packet movement from one IP network to other IP network by utilizing layer-3 devices viz router/layer-3 switches. It seems simple that we are just exchanging information in the form of packets between multiple parties, but actually long list of complex protocols/algorithms are working behind in order to attain efficiency in overall task of routing. The protocols which are part of layer-3 process categorized as Routing Protocols and Router Protocols [4, 12, 13], let us dive into these.

1. **Routing Protocol**: These protocols are used by router to find the networks in the inter-network environment dynamically with all the best suitable paths for reaching to destined resource from the source. Finding the best path is not only the task which these protocols performed, but to creating/maintaining the routing table, topology table, and neighborhood table with consistency also the crucial task which these protocols managed. The protocol like Routing Information Protocol (RIP-v1, RIP-v2), Interior Gateway Routing Protocol (IGRP), Enhanced Interior Gateway Routing Protocol (EIGRP), Open Shortest Path First (OSPF), and Border Gateway Protocol (BGP) are few famous protocols having its own way of working with pros/cons [14]. All these protocol used with respect to the requirement in hand. The RIP is distance vector, OSPF is linked state, and EIGRP is a hybrid protocol.
2. **Routed Protocol**: These protocols work with routing protocols to send actual packet from source to destination, actually routing protocols are finding the path, and these paths are used by routed protocols by assigning these to interfaces and to determine the method of packet delivery. IP and IPx are names of routed protocols.

Now before going through the practical implementation, let us understand three type of routing, actually as per the needs, the routing process is divided into three categories known as static routing, dynamic routing, and default routing. The static routing takes places when we are manually assigning routes on each router and these routes are going to be used for sending packets between devices. This type of routing is good when either we have very less number of devices used or if we want to make a permanent secure connection between device and service provider (ISP/Server etc.). But if we have large number of devices, it is not good to manually assign route on every router which corresponds to implemented topology, as it is quite time-consuming, error-prone methodology. The default routing is used to send packets with a remote destination network having no entry in the routing table at all, i.e., for Stub Networks having only single exit interface. The last one dynamic routing which is the mostly used technique in layer-3 devices, for finding the best suitable path and managing different table just with the help of configured routing protocols, less time consuming, less error prone than others. Let us move with few basic terminologies before moving toward to actual configuration of routing protocols.

4.1 Address Resolution Protocol (ARP/RARP)

This protocol is very helpful for finding the hardware address of a device with the corresponding IP address and to maintain the ARP table on layer-3 devices. ARP is IP $\rightarrow$ MAC finder, and RARP is MAC $\rightarrow$ IP finder, but RARP is now obsolete, and BootP is used instead for reverse ARP services.

4.2 Administrative Distance (AD)

This value is used to find out the trustworthiness about the routing information which received from the neighbor router. The AD value is always in the range 0–255, i.e., information with 0 AD is the most trusted one. But sometimes routing information received from multiple routers may have same AD, so in such cases in order to find the most trusted information, then routing protocol metrics viz hop count/bandwidth going to be utilized for scrutinizing the best path, i.e., route with the lowest metric will be placed in the consistent routing table. The load-balancing method is used only if AD + Metrics representing the same value when routing information received in order to select the best path. The one point we are going to mention here concern to AD and FD values, sometimes due to the topology or other requirements, it may be the case the protocol having higher AD can work well then protocol having lower AD, so by default if two protocols are running on same device, definitely one will override other due to lower AD value, and for such case, we do not have any automatic solution unless/until manually reducing the AD value of desired protocol. More enhancements can be done in this direction in future.

4.3 Type of Routing Protocol

There are three type of routing protocols [14, 15], as given below:

1. **Distance Vector**: With the help of this protocol in order to find the best suitable path to a remote network from the source, the judgment on overall distance covered going to be take place, i.e., to count how many router (hop) are coming in between when packet reaching to destination from source. The route having least n/o hops to be considered the best route. RIP is the example of this category.
2. **Link State**: By using this protocol router, create three distinct tables, i.e., neighborhood table, topology table concerned to inter-network, the routing table, and with the help of topology table/neighborhood table, the state of links going to be found which further helps to find the best route, which at the end put in routing table. OSPF is the example of this category.
3. **Hybrid**: These protocols utilize the aspects of both distance vector and link-state protocols. EIGRP is the example of this category.

4.4 Loops and Loop Avoidance

When multiple devices continuously sharing similar information due to any reason with each other and making the routing table inconsistent, it is called loops. The slow convergence rate of DV protocols may result in inconsistent routing tables/loops. There are two ways to avoid the loops which are:

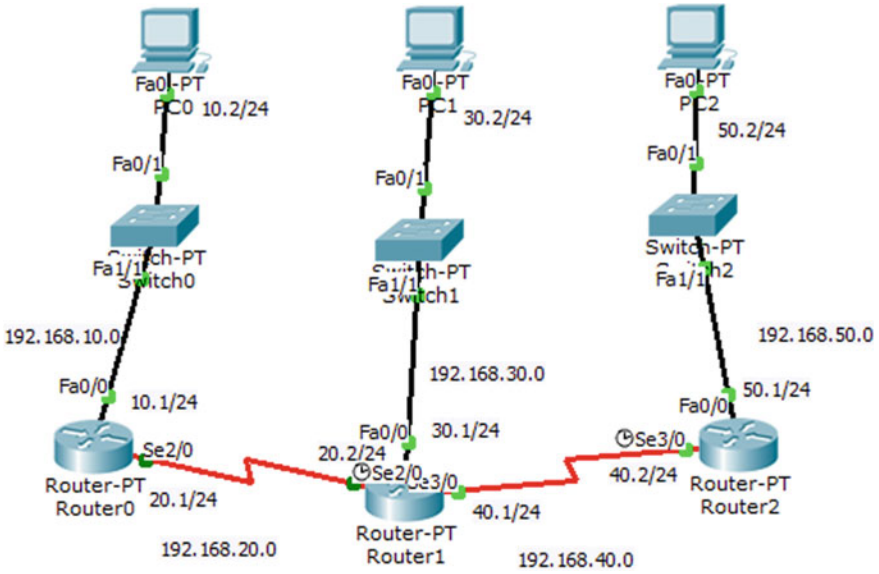


Fig. 16 Topology for static and default route configuration

1. **Maximum Hop Count**: The hop count increases indefinitely every time packet moving through a router. So by defining a maximum hop count, the loops can be avoided.
2. **Split Horizon**: This method reduces the chances of conveying incorrect routing information, by enforcing the rule that routing information will not be sent back in the direction from where it is coming.

Now we are not going to too much theory of all these concepts, and without wasting time, let us see the command used for giving routes on router with above-mentioned three routing techniques/categories of protocols (Fig. 16).

4.4.1 Static Route

Router0(config)#ip route 192.168.40.0 255.255.255.0 192.168.20.2

4.4.2 Default Route

Router0(config)#ip route 0.0.0.0 0.0.0.0 192.168.20.2

The command for configuring static route and default route is almost similar (see Figs. 17 and 18), the meaning of every single parameter is also similar (see Tables 2 and 3), it is just one line command for every network, but for every network, we have to make entry in every router, quite tedious task!.

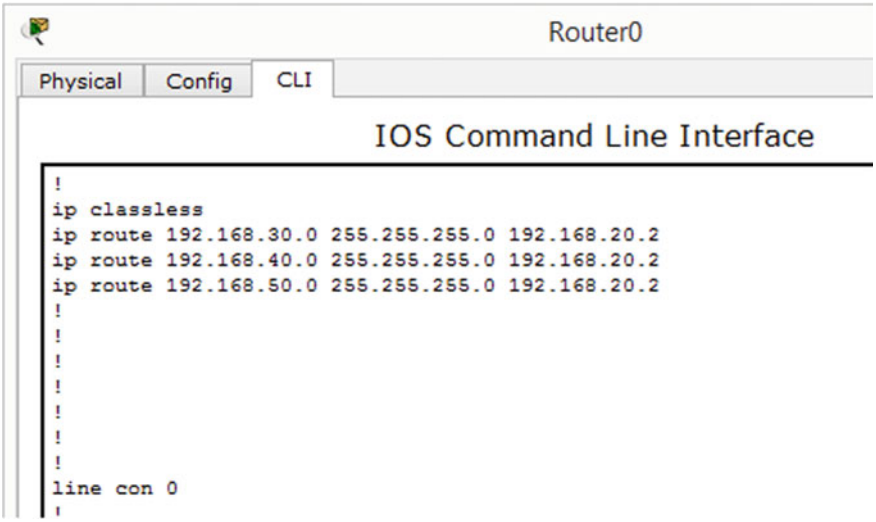


Fig. 17 Static route

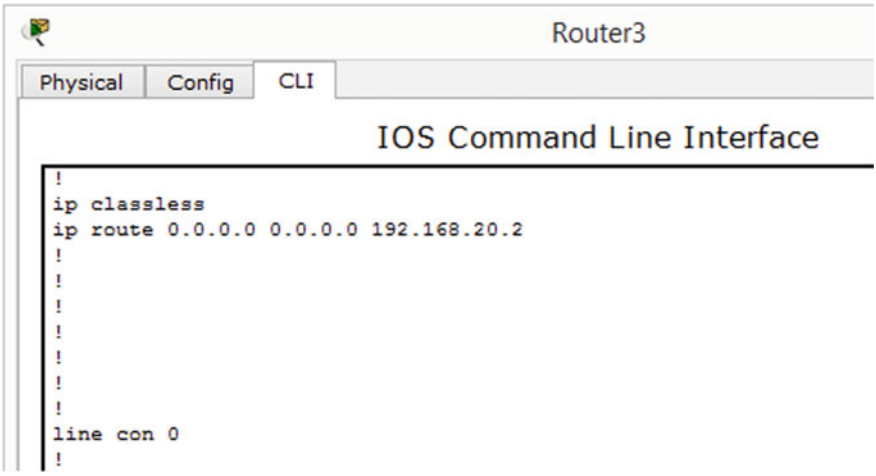


Fig. 18 Default route

Table 2 Segregated command of static route

IP route	192.168.40.0	255.255.255.0	192.168.20.2
Command	Network ID	Mask	Next hop

Table 3 Segregated command of default route

IP route	0.0.0.0	0.0.0.0	192.168.20.2
Command	Network ID	Mask	Next hop

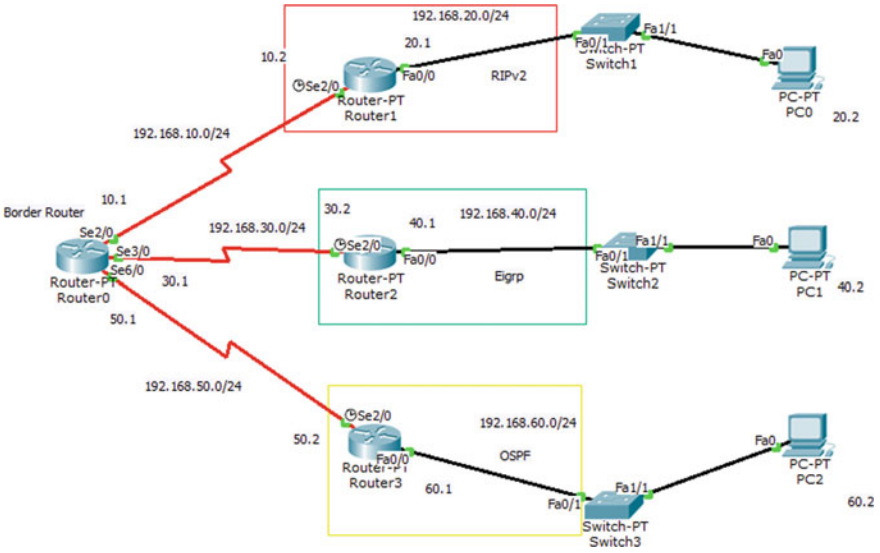


Fig. 19 Topology with RIPv2, EIGRP, and OSPF configuration

4.4.3 Dynamic Route

When protocols are going to be used for finding networks and updating routing tables, it is dynamic routing; True? Yes it's true. It is easier to use, but we have to pay cost in terms of router CPU processes/bandwidth on the network links. Let us know how to configure dynamic routing for getting dynamic routes with RIPv2, OSPF, EIGRP in single topology (see Fig. 19).

In dynamic routing, two significant tasks are when multiple routing protocols used for configuration purposes, first is to assign IP address and then just run the desired routing protocol, then second task is to redistribute the networks (see Fig. 20), so that all the routing protocols can maintain the consistent routing/topology table on routers for making connectivity possible (see Fig. 21), which protocol we are going to configure w.r.t. that command is bit different.

5 Comparative Analysis of Layer-2/Layer-3 Protocols

The protocols concerned to layer-2 and layer-3 of inter-network environment playing crucial role in various perspectives, some among these performing similar tasks but with different working strategies which make it distinct from one another. The use of all such protocols fully depends on the type of requirements in hands, and we are not going to use any merely. So let us have a look to short comparative analysis among all these with different parameters given below.

```
Step1: Give Ip Address on all Router Interfaces...

Step2:
On Router 1 Run Rip
Router(config)#Router Rip
Router(config-router)#Version 2
Router(config-router)#Network 192.168.10.0
Router(config-router)#Network 192.168.20.0

On Router 2 Run Eigrp
Router(config)#Router Eigrp 10
Router(config-router)#Network 192.168.30.0 0.0.0.255
Router(config-router)#Network 192.168.40.0 0.0.0.255

On Router 3 Run OSPF
Router(config)#Router OSPF 1
Router(config-router)#Network 192.168.50.0 0.0.0.255 Area 1
Router(config-router)#Network 192.168.60.0 0.0.0.255 Area 1

Step3: Configure Redistribution on Router 0

Router(config)#Router Rip
Router(config-router)#Version 2
Router(config-router)#Redistribute Eigrp 10 Metric 15
Router(config-router)#Redistribute OSPF 1 Metric 15
Router(config-router)#Network 192.168.10.0

Router(config-router)#
Router(config-router)#Router Eigrp 10
Router(config-router)#Redistribute Rip Metric 1 1 1 1 1
Router(config-router)#Redistribute OSPF 1 Metric 1 1 1 1 1
Router(config-router)#
Router(config-router)#Network 192.168.30.0 0.0.0.255
```

Fig. 20 Configuration of dynamic routing protocol

The detailed description about both the common layer-2/layer-3 protocols is available above (see Tables 4 and 5) with respect to the different functionality and parameter. We can easily get insight from the above description about various such protocols. In the next section, suitable conclusion is provided.

6 Conclusion and Future Work

This paper is fully devoted toward to understand the fundamental and implementation of layer-2 switching and layer-3 routing with various available methods and techniques, and what are the issues at various stages with respect to requirements in

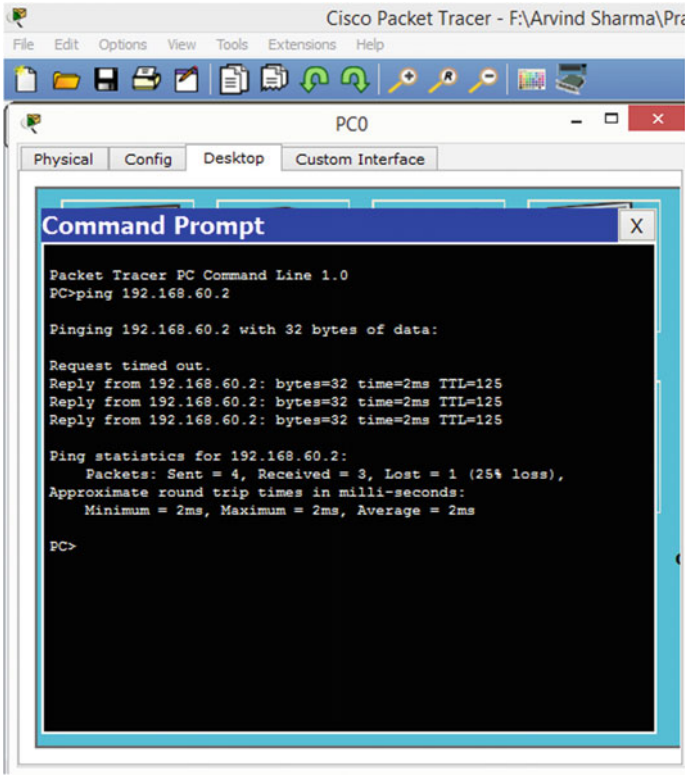


Fig. 21 Connectivity check from PC0 → PC2

hand and what will be the appropriate solutions to resolve, such issues are discussed thoroughly with suitable practical examples. All the aspects of virtual local area network, Spanning Tree Protocol, Etherchannel, redistribution, etc., discussed too. Even though the protocol discussed are working well with respect to distinct requirement, but still few limitation exists in few among these, as in STP the task of route bridge election becomes quite tedious and requires bit more computation when priority and bridge ID are same, i.e., more work with port cost and other metrics required. And also if by mistake convergence timer reduced from default 50 s. to less, definitely few more issues concerned to loop are going to be arising and we do not have any solution for that. So especially in our future work, we will look after such type of problems and make some improvements in exiting STP to tackle such problems. On the other hand in routing protocols due to computed AD/FD values, sometimes one protocol overriding other one, and it may be the case that this overriding becomes a cause of performance degradation, so in this direction, we are also going to make few betterments in existing protocols likewise to improve the functionality of distance vector protocol which can be done in near future, so that we are not only dependent on link-state category always. The main objective of

Table 4 Switching protocols

Protocol	Functionality
SDLC	**Name**: Synchronous data-link control **Task to be performed**: 1. Communication 2. Supports multipoint links 3. Supports error correction/recovery 4. Carry systems network architecture traffic 5. Give surety about data units arrival/right flow
HDLC	**Name**: High-level data-link control **Task to be performed**: 1. Predecessor of SDLC 2. Support point-to-point, multipoint links and WAN protocols 3. Part of X.25 network 4. Provides best effort reliable/unreliable services
SLIP	**Name**: Serial line interface protocol **Task to be performed**: 1. Communication 2. Used specifically for including framing byte in IP packets 3. Supporting DLC facility for ISP/home-group users over a dial-up links 4. Providing encapsulation for TCP/IP which work with serial links/ports 5. Do not support error detection/correction
PPP	**Name**: Point-to-point protocol **Task to be performed**: 1. Predecessor of SLIP 2. Support both IP/other packets for transport 3. Specifically worked with dial-up/leased-lines 4. Support error detection 5. Include the functionality of NCP/LCP for bringing links to state up/down, managing negotiation 6. Like HDLC support serial interfaces
LCP	**Name**: Link control protocol **Task to be performed**: 1. It is a part of PPP which is used for configuring/making/testing of data frames transmission
NCP	**Name**: Network control protocol **Task to be performed**: 1. Older protocol invented for ARPANET 2. Support file sharing 3. Every protocol which uses PPP, by default NCP becomes its part

this paper was just to understand in depth about how to design/develop well versed layer-2/layer-3 network in order to make communication possible between various resources in efficient manner.

Table 5 Comparison of routing protocols

Protocol	Type	Class	Metric	AD value	Classless/classful	Algorithm	Routing updates	Convergence
RIP-V1	Interior gateway	Distance vector	Hop count	120	Classless	Bellman Ford	In every 30 s	Slow
RIP-V2	Interior gateway	Distance vector	Hop count	120	Classful	Bellman Ford	In every 30 s	Slow
OSPF	Interior gateway	Link state	Link cost	110	Classful	Dijkstra	When change occur in topology	Fast
IS-IS	Interior gateway	Link state	Link cost	115	Classful	Dijkstra	When change occur in topology	Fast
EIGRP	Interior gateway	Hybrid	Lowest best composite	90 (For internal) 170 (For external)	Classful	Dual	When change occur in topology	Very fast
BGP	Exterior gateway	Path vector	Path attributes (i.e., autonomous systems paths)	200 (For internal) 20 (For external)	Classful	Best Path	When change occur in topology	Average

References

1. O'Mahony, The application of optical packet switching in future communication networks. Commun. Mag. IEEE **39**, 128–135 (2001)
2. A. Grami, *Communication Networks* (Elsevier, Introduction to Digital Communication, 2016), pp. 457–491
3. IEEE Standard for Ethernet, in IEEE Std 802.3–2015 (Revision of IEEE Std 802.3–2012), pp.1–4017, 4 March 2016
4. Lammle, *CCNA Routing and Switching Study Guide* (Sybex, United States, 2013
5. M.S. Guragain, Evaluating the Future of the Spanning Tree Protocol, Bachelor's Thesis, Turku University of Applied Sciences, 2016
6. CISCO, https://www.ciscopress.com/articles/article.asp?p=2348266&seqNum=3
7. Sungheetha, novel shared key transfer protocol for secure data transmission in distributed wireless networks. J. Trends Comput. Sci. Smart Technol. (TCSST) **02**(2020), 98–108 (2020)
8. Arvind, Cryptography and network security hash function applications, attacks and advances: a review, in *IEEE, 3rd International Conference on Inventive Systems and Control (ICISC 2019),* (2019), pp. 177–188.
9. Arvind, Cryptographic Keyed Hash Function: PARAŚU-256. J. Comput. Theor. Nanosci. **17**(11), pp. 5072–5084(13) (2020)
10. S. Sridhar, Intelligent security framework for iot devices cryptography based end-to-end security architecture, in *IEEE, 2017 International Conference on Inventive Systems and Control (ICISC),* (2017), pp. 1–5.
11. V. Suma, Security and privacy mechanism using blockchain. J. Ubiquit. Comput. Commun. Technol. (UCCT) **1**(01), 45–54 (2019)
12. D. Niu, Research on routing protocols in Ad Hoc networks, in *IEEE 2009 International Conference on Wireless Networks and Information Systems*, (2009), pp. 27–30
13. CISCO, https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/iproute_pi/configuration/15-mt/iri-15-mt-book/iri-iprouting.html
14. B. Valery, The analysis of the characteristics of routing protocols in IP networks, in *IEEE 2010 International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET)*, (2010), pp. 185–185
15. CISCO, https://www.ciscopress.com/articles/article.asp?p=2180210&seqNum=7

A Two-Level Security System Based on Multimodal Biometrics and Modified Fusion Technique

B. Elisha Raju, K. Ramesh Chandra, and Prudhvi Raj Budumuru

Abstract Biometrics is physical or behavioral characteristics to identify a human digitally. The identification of humans involves training and testing process. In general, the training process requires a large database to store the extracted features of multiple samples of a human, and the lighting conditions of acquiring samples also make it more complex. These challenges of using single biometric can be addressed by using multimodal biometrics. In this paper, multimodal biometrics of humans like face, fingerprint, and iris are considered for their easy and secure recognition among the group of people. The testing process involves the comparison of the features of real-time sample with the features of training samples, and consequently, the recognition rate of this algorithm increases. This algorithm uses discrete wavelet transform (DWT) for the filtering and compression, singular value decomposition (SVD) for the extraction of features from the compressed and filtered results of DWT. The proposed method gives better performance like 98% of accuracy and 100% recognition rate.

Keywords DWT · SVD · Fusion · Recognition rate · Biometrics · Training and testing

1 Introduction

Biometrics provides a reasonable confidence for the authentication of a person with less friction of the user. Once the user gets the access to control the devices operated by using biometrics, then the working of the device made easy. The biometrics used in the authentication of the user is mainly physical and behavioral characteristics. Face, fingerprint, iris, palm print, etc., are the physical characteristics used as traits for the authentication of the user. Similarly, behavioral characteristics like signature, walking, and oral signals are also used to authenticate the users as per the feasibility of the organizers. The physical characteristic fingerprint is preferable when there is a permissible contact between the users, but due to the effect of **COVID-19**, social

B. Elisha Raju (✉) · K. Ramesh Chandra · P. R. Budumuru
Vishnu Institute of Technology, Bhimavaram, Andhra Pradesh 534202, India
e-mail: elisharaju.b@vishnu.edu.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_2

distance needs to be maintained in order to prevent the spreading of the virus. So, the trait fface must be used as an authentication to detect and recognize that the user belongs to the society.

The face recognition process [1] consists of two phases, training and testing phase. In the first phase, the data is collected by using the image sensor like high-definition cameras, and for that data, preprocessing is carried to filter the noise and to adjust the properties of image like brightness, contrast, and illumination. The preprocessed image is observed to detect and crop the face. The detected face is considered as an input to the feature extraction algorithms to extract the required features. In normal extraction algorithms, single sample of each user, i.e., a pose, is not sufficient to detect the user in real-time applications, so 8 to 10 samples are collected to have a better efficiency.

The fingerprint technology used for authentication consists of the process as follows: the sensors are used to acquire the finger image, and then, ridges are detected to distinguish the different type of users. The ridges characteristics are used to mark the local discontinuities in the ridge flow for different type of fingerprints. Now, the preprocessing for the ridges detected image is processed to get the equalization in the enhancement of the image and then binarization followed. The binarized image is used to the extraction of the templates, and these templates are used as features to store in the database. This database is processed to detect the minutiae which are extracted for the post-processing. With the advancement of latest technologies, the biometrics has been taken a vital role in the field of data security and authentication. The major applications of the biometrics include banking, computer or network security, mobile phone security, residential and business control, attendance system, etc. The benefits provided by the biometrics system are user-friendly system, fraud reduction, best authenticity and security, more customer convenience instead of carrying cards, and finally accuracy which is more.

Multiple biometric sources are used to acquire the information of a person recognized by the biometric systems said to be multimodal biometric systems. This system is more accurate because of the multiple pieces of substantiation. In the testing process of these systems, the matching accuracy will substantially increase because of the combination and fusion methodologies adopted. It also addresses the problem of non-universality or inadequate population exposure, providing security by enrolling individuals legitimately, so that the difficulty is increased for a charlatan to burlesque the multimodal traits.

The paper organized as follows: Sect. 2 gives the related works as literature survey, Sect. 3 briefs the methodology of the proposed algorithm, results, and discussion is explained in Sect. 4 and finally concluded the proposed algorithm in the Sect. 5.

2 Literature Review

The Marcialis and Roli 2004 are considered multi-sensor system like an optical and a capacitive senor used to obtain the fingerprint information of a user [2]. Rose

et al. prepared a multi-algorithm system that extracts the features of fingerprint using algorithms based on texture and minutiae [3]. Hill et al. discussed a multi-sample system in which multiple samples of same traits are captured and further processed to get the better accuracy [4]. Multi-modal system was carried out by Brunelli and Falvigna [5] in which face and voice of the user considered as traits for authentication.

In 2011, Jameer Basha et al. [6] introduced a latest outline at rank-level fusion for iris and fingerprint. The experimental tests were carried by using the three implemented fusion methods given by: highest rank method, Borda count method, and logistic regression method. The execution time to match in highest rank method is equal to 0.45 s which is the best time achieved in their work with equal false acceptance rate (FAR) and false rejection rate (FRR), 0% and 0.25%, respectively.

In 2012, Radha and Kavitha [7] introduced multimodal biometric system using fingerprint and iris as modalities with a fusion scheme at feature extraction level. The scheme uses a concatenated feature vector from both iris and fingerprint. The feature vectors of both modalities are extracted using the log Gabor filter. The final match score is generated by using the Hamming distance (HD). The results observed for the database of 50 users with FAR of 0% and FRR of 4.3%, and the matching time is 0.14 s.

In 2013, Abdolahi et al. [8] used the fuzzy logic and weighted code for the multimodal biometric system with fingerprint and iris as the traits. The results are combined at the decision level by converting the fingerprint and iris images into a binary code. Fingerprint code is weighed as 20% and iris code as 80%. The work achieved 2% FAR and FRR and 98.3% accuracy.

3 Methodology

The methodology involved in this multimodal biometrics consists of the processing the data of the users, so that the database is maintained for the future verification. In unimodal biometric system, the process carried is consists of two major phases: training phase and testing phase (see Fig. 1). The training phase consists of obtaining the data from the sensors used for the acquaintance of the traits from the users. The obtained traits are processed in a way such that they are clearly processed to get the noise-free image [9], and then, different algorithms are applied on the images to get the required features. The features are stored as database in order to use as the features to match the real-time obtained features of the user in the testing phase. The testing phase used to verify the users based on the features extracted at the time of verification, and these features are matched with the features in the database. So that the matching score generated at the verification process gives the instructions whether to accept or reject the user into the organization which is illustrated in the algorithm.

Algorithm

- Input training images X_f [$f = 1, 2, \ldots F$], testing images X_t.

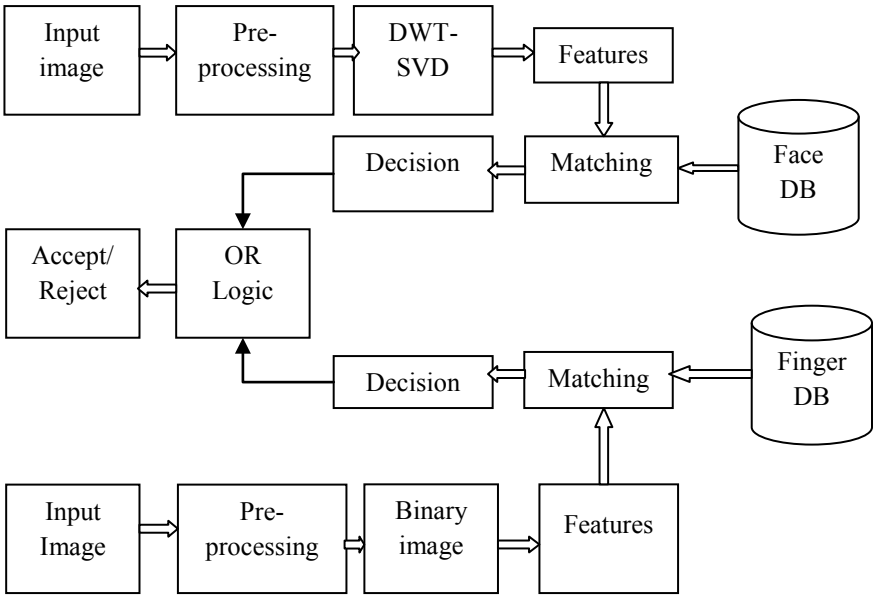


Fig. 1 Block diagram of proposed algorithm

- Normalize and noise reduction in each image by

$$X_{f}\prime = X_f * \mathrm{GF}(a, b, \vartheta),$$

where * is the convolution operation in a and b

- The two-dimensional digital Gaussian filter $GF\ (a, b,\vartheta)$ is given by

$$G_F(a, b, \vartheta) = \frac{1}{\sqrt{2\pi}\vartheta} \exp\left(\frac{-a^2 + b^2}{2\vartheta^2}\right)$$

- The feature vector of training images X_f is C_n^f, and the feature vector of testing images X_t is C_n^t $(n = 1, 2, \ldots N)$.
- For each vector of training C_n^f and testing C_n^t, the chi-square distance is measured.
- Matching result is computed for C_n^t.
- The recognition result of test image is generated and the fusion at the decision level.
- Output of recognition result.

In this paper, the work is followed as that the unimodal biometrics are used individually by using the algorithms specified to each. The face recognition algorithm is used as the main trait used to identify the user even fingerprint is also registered to authenticate the user. In face recognition algorithm, the entire phenomenon is conducted in two phases: training phase and testing phase. The training phase consists

of the following process: image acquisition, preprocessing, feature extraction, and storing in database.

3.1 Image Acquisition

The array of sensors placed in the arrangement of 2D array used in the digital cameras to acquire the response of each sensor proportional to the integral of the energy projected onto the surface of the sensor. The projected energy is converted into the electrical signal operated by the digital system to show the image on the screen of the digital system. Thus, the image obtained by this process is given as input to the preprocessing. In this acquisition process, the user is captured in different poses to store in the database, so that the greater number of samples of the user makes, the face recognition becomes more effective to process the verification based on the samples collected.

3.2 Pre-processing

The acquired image from the image sensor is considered for the preprocessing to control the contrast and illumination of the image. The preprocessing involves the re-sampling of the image in order to give assurance that the image coordinate system is correct. The noise from the sensor is also reduced with the Gaussian filter so as to decrease the false information. The relevant information is detected by the contrast enhancement [10]. The image structures are enhanced through scale-space representation by locally appropriate scales.

3.3 Feature Extraction

In this process, the noise-free mage is considered for the extraction of features. The image data consists of the features at the different levels of complexity. These features are extracted to recognize the objects present in the image, and the features of such type of images include lines, edges, and ridges. The interest points like corners, blobs, or points are localized, used as the features [11]. The texture, shape, and motion are the features to be considered as the more complex features.

The extracted features are used to detect the required area of interest points, and then, the segmentation is followed to separate the specific object of interest. The decision is made to observe which image points or regions of image are relevant for further processing.

The testing phase consists of the same process upon the feature extraction, now analysis is made to verify the data to satisfy the model-based and application-specific

assumptions. Now based on the specific parameters such as pose and size of the object, image recognition process classifies the detected object into different categories. The decision is made based on the image registration involved in the training phase and the image recognition in the testing phase.

Once the decision is made for the each individual biometric, we can fuse [12] the results based on the logic OR in which if any one of the biometric satisfies the matching score and recognizes the user we can make the decision and also a single biometric is sufficient, i.e., mostly concentrated on the face than the fingerprint in order to maintain the social distance and the contact free recognition for the organizers. The work flow of the proposed algorithm is illustrated through the block diagram.

In the proposed algorithm, the unimodal biometrics is processed individually to recognize the user. In the face recognition algorithm, user is captured with different poses and it is preprocessed to remove the noise by the Gaussian filter, and then, noise-free image is given as input to the DWT [13, 14] which is used to reduce the dimensionality of the image and it gives four bands as output like Low-Low (LL), Low-High (LH), High-Low (HL), High-High (HH) as shown in Fig. 2. The features are extracted using the SVD for the LL band, and those features are Eigen values of the diagonal matrix stored in the database. In the testing process, we acquire the Eigen values of the user by capturing the real-time image and calculate the chi-square distance between the features to find the matching score of the users.

A rectangular matrix M of size $a \times b$ will be decomposed into three matrices using SVD

$$M = \text{FSOT} \tag{1}$$

An orthonormal set can be formed from the column vectors of F, where F is an orthogonal matrix of size $a \times a$. The orthogonal matrix O of size $b \times b$ forms an orthonormal set by using the column vectors of O.

The diagonal elements of matrix S are said to be singular values, and these values are decreasing gradually. Eigen values are formed by the square values of the singular value. The matrix form of singular values is as follows:

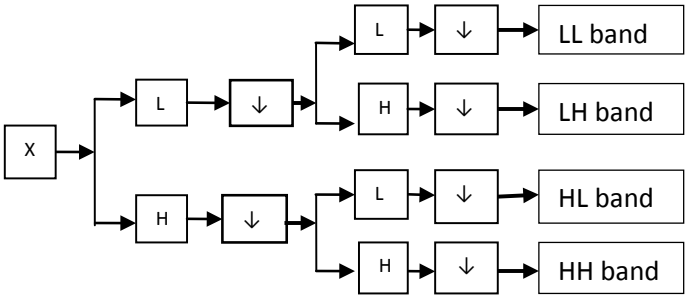


Fig. 2 Block diagram of DWT

$$S = \begin{bmatrix} \rho_1 & 0 & \cdots & 0 & 0 \\ 0 & \rho_2 & & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ 0 & 0 & \cdots & \rho_{n-1} & 0 \\ 0 & 0 & & 0 & \rho_n \end{bmatrix} \tag{2}$$

ρ_1, ρ_2, ρ_3..., ρ_n are called singular values and the relation between the singular values is given by

$$\text{Where } \rho_1 \geq \rho_2 \geq \rho_3 \geq \ldots \geq \rho_n \geq 0$$

Similarly in the fingerprint recognition algorithm, the preprocessing is carried by the Gaussian filter and minutiae extracted and stored in database and in the testing process, same technique is used to extract the features and then matching score is examined to recognize the users. In this proposed algorithm, face recognition is mainly focused to recognize the users instead of the fingerprint, so 80% of the proposed algorithm is to use the face as the main trait to process the work.

4 Results and Discussion

Experiments are conducted to evaluate the performance of the proposed algorithm. In face recognition algorithm, DWT-SVD based feature extraction methods are used to recognize the users, and in the fingerprint algorithm, minutiae extraction is carried on the ridges and valleys of the fingerprint. The dimensionality reduction algorithm employed in this proposed algorithm gave us the better results compared to the existing algorithms like LDA. The limitations in the existing face recognition algorithms like a greater number of samples to be stored and recognition of the users with low lighting conditions are addressed in this proposed algorithm, and also the fingerprint is preprocessed to get the noise free image using the Gaussian filter and it given for the minutiae extractor to extract the point s on the ridges and valleys of the fingerprint of the user.

4.1 Database

Our proposed method used two public domain face datasets: ORL and AT & T datasets and fingerprint dataset: FVC2002 dataset. These datasets are used to find the performance of the descriptor. In the ORL and AT & T datasets, each individual having ten images with different poses and FVC 2002 dataset has 10 to 20 images for each individual. The image samples of ORL database are represented for the reference (Fig. 3).

Fig. 3 Image samples of ORL database

4.2 Performance Metrics and Comparison

Set of performance measures are used to obtain the reported results like recall, precision, F-Score, and accuracy. The proposed algorithm gives the better indications compared to the existing algorithms which are illustrated in the following figures. The computation of the accuracy is given by

$$\text{Accuracy} = \frac{\text{correct recognition}}{\text{total images}} \tag{3}$$

The recall is the proportion of the results that are correctly classified, and it is for the true positive (TP) classifier which is given by

$$\text{recall}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \tag{4}$$

The overall recall is given by

$$\text{recall} = \frac{1}{n} \sum \text{recall}_i \tag{5}$$

Fig. 4 Simulation results on AT&T database

In the above equation, the TP indicates the positively detected users with face and fingerprint, and FN is false negative which are wrongly recognized, and n is the number of samples of each unimodal biometrics trait.

Precision gives the number of correctly recognized out of all the samples from the database, and it is also called as the truly analytical value which is given by

$$\text{precision} = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \tag{6}$$

Here, TP is the truly detected, and FP is the false positive, i.e., a wrong person detected instead of positive detection.

The F-Score is the combination of the precision and recall. The F-Score is commutated as follows (Fig. 4)

$$F\text{-Score} = \frac{2(\text{Precision*Recall})}{\text{Precision} + \text{Recall}} \tag{7}$$

The simulation is carried on the AT&T database, and the performance metrics are represented for both fingerprint and face database (Fig. 5).

The performance of the proposed algorithm is compared with the existing method proposed by Vaidehi et al. [15] in which different biometrics are used and different fusion techniques are employed, but the proposed algorithm is giving better results in terms of accuracy and recognition rate.

5 Conclusion

The proposed algorithm used the biometrics face and fingerprint of the users. DWT and SVD algorithms are used as the feature extractors for the face and the minutiae

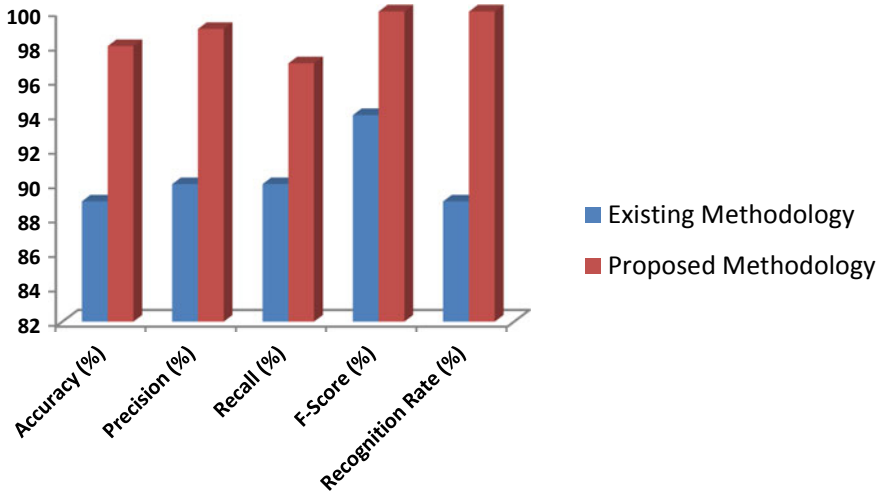


Fig. 5 Comparative results of proposed methodology

extractor used for the fingerprint. The datasets for face are ORL and AT & T and for the fingerprint FVC2002. In each dataset, 10 to 12 samples of users are taken for the process involved in the proposed algorithm. The biometrics are processed individually up to the matching score computations, and then, they are combined at the decision level to give the better accuracy of 100% for the standard datasets, and it is 98% for my own dataset. This work can be extended by the deployment of the different biometrics, and also the security will increase if the proper lighting conditions were maintained for the real-time applications.

References

1. B Elisha Raju, P. Satyanarayana Murty, Face recognition using DWT-SVD based feature extraction methods. ICRTCET-2018, GJESR, (2019), pp. 499–504. ISSN 2348–8034
2. G.L. Marcialis, F. Roll, Fingerprint verification by fusion of optical and capacitive sensors‖. Pattern Recogn. Lett. **25**(11), 1315–1322 (2004)
3. A. Ross, A.K. Jain, J. Reisman, A hybrid fingerprint matcher. Pattern Recognit. **36**(7), 1661–1673 (2003)
4. H. Hill, P.G. Schyns, S. Akamatsu, Information and viewpoint dependence in face recognition‖. Cognition **62**(2), 201–222 (1997)
5. R. Brunelli, D. Falavigna, Person identification using multiple cues. IEEE Trans. Pattern Anal. Mach. Intell. **17**(10), 955–966 (1995)
6. A. Jameer Basha, V. Palanisamy, T. Purusothaman, Efficient multimodal biometric authentication using fast fingerprint verification and enhanced iris features. J. Comput. Sci. **7**(5): 698–706 (2011)
7. N. Radha, A. Kavitha, Rank level fusion using fingerprint and iris biometrics. Indian J. Comput. Sci. Eng. **2**(6), 917–923 (2012)

8. M. Abdolahi, M. Mohamadi, M. Jafari, Multimodal biometric system fusion using fingerprint and iris with fuzzy logic. Int. J. Soft Comput. Eng. **2**(6), 504–510 (2013)
9. T. Vijayakumar, R. Vinothkanna, Retrieval of complex images using visual saliency guided cognitive classification. J. Innov. Image Process. (JIIP) **2**(02), 102–109 (2020)
10. P.R. Budumuru, G.P. Kumar, B.E. Raju, Hiding an ımage in an audio file using LSB audio technique, in *2021 International Conference on Computer Communication and Informatics (ICCCI)* (2021), pp. 1–4
11. E.E.B. Adam, Evaluation of fingerprint liveness detection by machine learning approach-a systematic view. J. ISMAC **3**(01), 16–30 (2020)
12. D. Egfin Nirmala, V. Vaidehi, Comparison of Pixel-level and feature level image fusion methods, in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)* (2015), pp. 743–748
13. K. R. Chandra, B. Prudhvi Raj, G. Prasannakumar, An efficient image encryption using chaos theory, in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)* (Madurai, India, 2019), pp. 1506–1510
14. M. Dileep, B. Prudhviraj, G. Prasannakumar, A new image encryption and data hiding technique using wavelet transform. Int. J. Res. Eng. Technol. **5**(5), 195–198 (2016)
15. V. Vaidehi, N. Babu, H. Avinash, M. Vimal, A. Sumitra, P. Balamuralidhar, M. Chandra, Face recognition using discrete cosine transform and fisher linear discriminant, in *2010 11th International Conference on Control Automation Robotics & Vision* (2010), pp. 1157–1160

EPO Based Clustering and Secure Trust-Based Enhanced LEACH Routing in WSN

L. Rajesh and H. S. Mohan

Abstract *Background*: In wireless sensor network (WSN), one of the best method is data aggregation for energy conservation. In numerous WSN applications, privacy preserving has become a very important issues with the development of WSNs. *Purpose*: Sensors are vulnerable for security threats due to the open deployment. Moreover, most of the routing algorithms does not consider security. Various methods like trust management, key management, firewalls, and intrusion detection for providing security were considered in the existing work. Between them, enhanced security is given by the trust management. Thus, in this paper, we used the secure trust-based enhanced LEACH routing (STELR) protocol to resolve security concerns. *Methods*: Here, cluster head is selected by using emperor penguin optimization (EPO) algorithm for data aggregation, and STELR is used to enable the security for routing. We can enable secure communication by checking the trustworthiness of a node. *Results*: The work is implemented in MATLAB platform, and the results are measured in terms of packet delivery ratio (PDR), throughput, end-to-end (E2E) delay, detection accuracy, energy dissipated, and trust probability (TP). *Conclusions*: Finally, the security concerns are resolved, which shows proposed approach that gives better performance when compared to existing techniques like LEACH and SDILR protocols.

Keywords Wireless sensor network · Emperor penguin optimization · LEACH protocol · Trust value · Cluster head · Key signature generation · Key signature verification data aggregation and routing

1 Introduction

Numerous applications are in wireless sensor network for monitoring various environments. It includes less cost and small sensing devices along with wireless radio transceiver [1]. WSN contain hundreds or thousands of sensor nodes [2]. The sensor

L. Rajesh (✉) · H. S. Mohan
Department of Information Science and Engineering, SJB Institute of Technology, Bengaluru 560060, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_3

senses and collects the data which is transmitted to one or more than one sink nodes. The ability of sensor is to communicate directly or each other to an external base station [3]. This network does not need any external supply for data gathering which is one of the main advantages [4]. The WSN main applications are military surveillance, building safety monitoring, forest fire detection, wild habitat monitoring, etc. [5].

In remote sensing application, limited battery power is the key challenges [6]. To reduce energy consumption, one of the main solution is cluster-based data aggregation which deals numerous copies of data and incorporate more efficient information by reducing data packets [7, 8]. One of the fundamental process is data aggregation which can be consider to diminish communication overhead and energy consumption in WSNs for save the limited resources [9].

Nowadays, Internet of Things (IoT) is one of the part of WSNs. The self and openness organization characteristics expose its weakness to the attackers that lead to loss worse damage and loss original construction [10, 11]. Feasible and efficient secure data aggregation construct a basis for WSNs applications in various filed like economic, political, military, etc. [12]. To protect WSNs from hostile attacks, a lot of efforts have been made by the researchers. However, the existing methods have lower privacy preserving capability, high communication cost, and large computational, and also few technical challenges need to solve like secure data aggregation, efficiency, and energy in WSNs [13].

The key contribution is given below,

- Placement of sensor node in the network for the secure communication without eavesdropping attack.
- Emperor penguin optimization algorithm is used for selecting the cluster head based on energy and distance.
- Secure trust-based enhanced LEACH routing (STELR) is used to enable the security for routing.

Organization: The related work is given in Sects. 2 and 3 which gives network model, thread model, emperor penguin optimized cluster head selection, and STELR. Sect. 4 involves the performance of our work. The conclusion is given in Sect. 5.

2 Related Work

Data security and data aggregation issues was addressed by Ranjani et al. [14] for WSN of IoT application. To give secure data transmission, an energy-efficient cluster-based data aggregation (ECBDA) method was used. Bayesian fusion algorithm and data aggregation was performed by cluster head to enable the security. Secure communication was enabled by checking the node trustworthiness. Based on the behavior of node, the algorithm of Bayesian fusion calculates the trust. The parameters like network residual energy, energy consumption, delay, and throughput was mainly considered.

Table 1 Analysis of existing comparison

Author	Method	Jitter	Network lifetime (ms)	Delay
Ranjani et al. [14]	Efficient sink selection scheme (ES3)	0.05 s	90	1.5 s
Kumar et al. [15]	SDARP protocol	4 s	88	8 s
Sahu et al. [16]	Token-based data security protocol	–	85	–
Fu et al. [4]	DCA-SF protocol	–	75	–

Data security and network traffic issues were concentrated by Kumar et al. [15], and in order to diminish these issues high data gathering, a security-based data-aware routing protocol was implemented for adhoc sensor network application. Two phases were conducted in the security model. Optimal cluster head was used in the first phase to monitor the cluster member and CH behavior. This model was enhanced in the second phase with the data gathering algorithm. The parameters like energy efficiency, network lifetime, and end-to-end delay was calculated.

In case of inter-cluster communication, the data security operation was performed by Sahu et al. [16] which depends on token cluster identification for WSN application. Before communication was initiated, the cluster sender checks the identifications of the receiver cluster. By cluster head node, each cluster was represented. The head node identification number was called as the token number. Parameters like initial energy, sensing range, and number of cluster and cluster head were calculated.

For detecting a selective forwarding attack, a data clustering algorithm was proposed by Fu et al. [4] for WSN application. By clustering their cumulative forwarding speeds, the malicious cluster heads were captured and isolated. The algorithm has been improved by altering the parameters of the data clustering algorithm. The parameters like high detection speed ratio, low false detection rate, and missed detection rate was calculated.

To ameliorate the traffic flow, Bestak [17] modeled an intelligent traffic control device using adhoc network. In the cloud, the smart resource management for massive data processing system was developed by Chandy [18]. The parameters like network utilization, disk utilization, memory utilization, and CPU utilization was measured and considered. The performance like jitter, network lifetime, and delay is compared with several existing techniques, which is given in Table 1.

3 Proposed Methodology

In this paper, we addressed the security issues by using STELR protocol. Initially, the sensor node is placed in the network for secure communication without attacking eavesdropping. Here, cluster head is selected by using EPO algorithm based on distance and energy for data aggregation, and STELR is used to enable the security

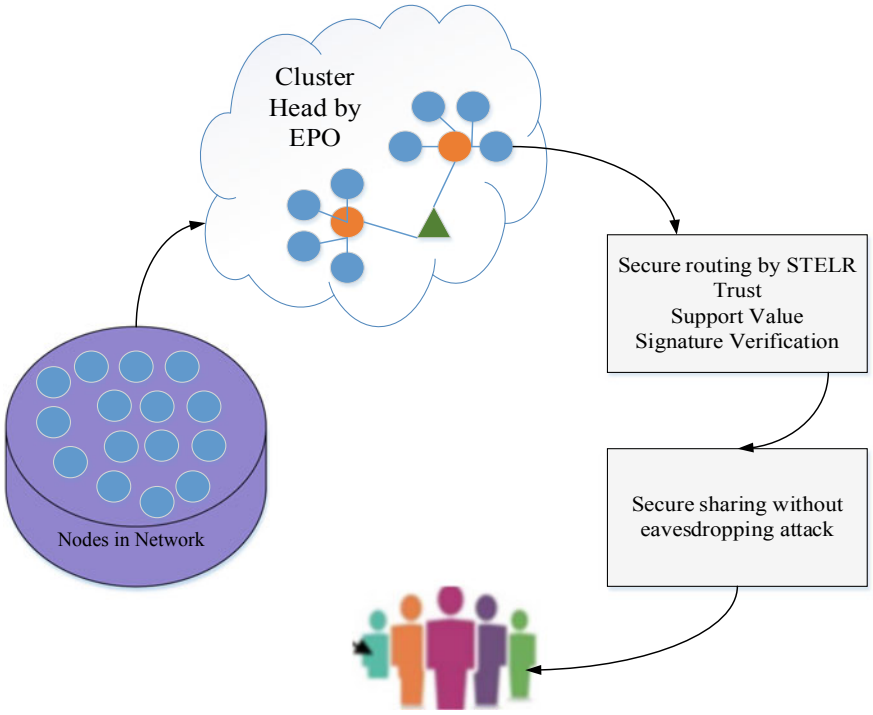


Fig. 1 Procedure for proposed work

for routing. We can enable secure communication by checking the trustworthiness of a node. To ensure the avoidance of eavesdropping, support value-based trust authentication is given to the primary users after the cluster formation in the proposed STELR protocol. The support value-based trust authentication will verify the trust value, key signature generation value, and key signature verification value. The above flow of work and working procedure is given in Fig. 1.

3.1 Network Model

In WSN, the sensor network is modeled as $G(V, E)$ which is connected graph, where set of sensor node is represented as V, number of sensor node is denoted as $|V| = N$, and set of wireless link connecting to the sensor node is represented as E. Within transmission range, the sensor nodes needs to be communicated, and for that, every sensor node is fitted out along with wireless transceiver.

Network data aggregation process, the sensor nodes can be allocated in to three groups like sink, aggregation, and ordinary node. The aggregation node and ordinary

node is same; however, both tasks are not similar. The final aggregation results is received by sink node and forwarding the query request.

The function of data aggregation is represented as $Y_1(t) = f_1(d_1(t) + d_2(t) + ... + d_n(t))$, where the individual sensor reading for node i at time instant t is represented as $d_i(t) = (i = 1, 2, 3..., N)$. Typically, the aggregation function is a sum of count, min, max, average, sum, etc. Nowadays, most of the researchers have only focused on the sum function. Some statistical function like standard deviation, average, and count are all based on sum (). To estimate by the sum function, the nonlinear function like min () and max () can also be used, and then, it will be forwarded to the sink, and in result, more energy will be saved.

3.2 Threat Model

In the field of the WSNs, the security issues have become more and more important. Different attacks can be launched by the hacker to undermine the data security. In WSN, to protect data privacy, we are mostly attentive on avoiding eavesdropping attacks.

Threats from untrusted eavesdropping listening or intercepting to packets is included by the threat model. The attacker had the capability to learn all communications in attack processing and by monitoring the wireless link to obtain user privacy information. Let us assume, the security mechanism is adopted in WSN to access the attackers by capturing a normal node. Data privacy will be threatened, if the attacker captures the privacy node data than the other nodes.

3.3 Emperor Penguin Optimized Cluster Head Selection

The cluster head for the cluster member is selected in the cluster head selection phase, having a minimum distance and the greatest residual energy. Here, by using fitness of EPO, the cluster head is selected.

3.3.1 Cluster Boundary Determination

The sensor node is chosen to establish the cluster's boundary. To define the cluster boundary, the Euclidean distance is used. The boundary requirement is less than the Euclidean distance.

$$\Psi = \Delta\Phi \tag{1}$$

where Ψ denotes the cluster boundary and Φ denotes the Euclidean distance

$$\text{Fitness Function} = \text{CH} = \sum_{\max} E + \sum_{\min} D \tag{2}$$

where E represents the energy of nodes and D represents the distance between the nodes.

3.3.2 Energy Calculation Around Boundary

The sensor node's energy between the boundary positions is determined. To model quantitatively, we assume the energy while the power is on and when the power is off. The energy is in charge of sensor nodes in various areas. The energy calculation around the boundary is as follows:

$$E^{'} = \left(T - \frac{\text{Max}_{\text{iter}}}{x - \text{Max}_{\text{iter}}} \right) \tag{3}$$

$$E = \begin{cases} 0, \text{ if } P > 1 \\ 1, \text{ if } P < 1 \end{cases} \tag{4}$$

where current iteration is defined as x, maximum number of iteration is denoted as Max_{iter}, time is denoted as T, and power is denoted as P.

3.3.3 Distance Between Sensor Nodes

Following the construction of the boundary, the distance between the sensor nodes and the best optimal solution are computed. The other search agent will update their location based on the current best optimal solution. The following is the distance between the sensor nodes:

$$\vec{D}_{\text{ep}} = \text{Abs}\left(S\left(\vec{A}\right).\vec{P}(y) - \vec{E}.\vec{P}_{\text{ep}}(y) \right) \tag{5}$$

In which the distance between SNs are denoted as $\vec{D}_{\text{ep}}$, the force of node which move toward the best optimal solution is represented S, best optimal solution is defined as $\vec{P}$, current iteration is indicated as y, $\vec{A}$ and $\vec{E}$ are utilized to avoid the collision, the position of sensor node is denoted as $\vec{P}_{\text{ep}}$.

3.3.4 Optimal Solution

The cluster head is chosen based on sensor node distance and energy. The cluster heads are updated based on the best optimal solution. This is in charge of choosing the cluster head for each cluster in a given space. The following equation is used to update the CH for each cluster in an SN.

$$\vec{P}_{\text{ep}}(y+1) = \vec{P}(y) - \vec{A} \cdot \vec{D}_{\text{ep}} \tag{6}$$

The updated CH is represented as $\vec{P}_{\text{ep}}(y+1)$, and updated equation is compared with the fitness Eq. (5), if it satisfied the fitness function, the CH is selected with optimal solution or else the condition will be continued. The proposed cluster formation reduces the energy consumption because only the cluster head nodes are used for the transmission. Here, the secondary user accesses the data over time and is referred to as an eavesdropping attack. In order to avoid this, support value-based authentication is given to the data transmission node.

3.4 Secure Trust Based Enhanced LEACH Routing (STELR)

The support value (SV)-based trust authentication (TA) (SV-TA) is provided to main users in the proposed STELR protocol to avoid eavesdropping attacks. When the information is transmitted, the security is given by the trust authentication on the primary user nodes by creating support value based signature verification. This will not help to get the malicious nodes (MN) and got the access of primary user nodes. The following steps are trust, key signature creation, and verification stages.

3.4.1 Trust Based Data and Communication

Based on communication and data trust, the trust value will be calculated. By the beta distribution factor, the communication trust is obtained. Beta probability density function can be used to determine the probability of all binary events. By using beta density function, the communication trust is defined.

$$f(y_i; \alpha_i, \beta_i) = \Gamma(\alpha_i + \beta_i) \Big/ \Gamma\alpha_i \Gamma\beta_i y_i^{\alpha_i - 1}(1 - y_i)^{\beta_i - 1} \tag{7}$$

where number of successful transmission is denoted as α_i, number of cluster member failure transmission cm_i is denoted as β_i.

Data trust is calculated by using the Gaussian distribution function.

$$f(y_i; M, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y_i - M)^2/2\sigma^2} \tag{8}$$

$$M = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{9}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - M)^2} \tag{10}$$

where variance and mean of the sensed data is represented as σ and M from the sensors group in a cluster. The sensed value of cm_i is represented as y_i.

3.4.2 Key Signature Generation

Before sending the information, the key signature is generated. The key is necessary for the primary user to access the information. If the data is not accessed by the user, then the node is attacked. After new node is searched, it depends upon the key access to transfer the data. In subsequent stages, the key signature generation process is described.

Stage 1: The random number of group public key $g_{\mathrm{pk}} = (g_1, g_2)$ and message $N \in \{0, 1\}^*$ is the input for the key signature, i.e., $\mathrm{Sign}\left(g_{\mathrm{pk}}, N\right)$.
Stage 2: The random numbers δ and ϕ are selected.
Stage 3: By using Eqs. (11) and (12), the helper values B_1 and B_2 are computed,

$$B_1 = g_1 \oplus \delta \tag{11}$$

$$B_2 = g_2 \oplus \phi \tag{12}$$

Stage 4: Support value is computed by using the calculated helper values for the encryption which is shown in Eq. (13)

$$S = (B_1 + B_2)\big/B_1 * B_2 \tag{13}$$

Stage 5: Challenge value is computed as $C_V \leftarrow H\left(g_{\mathrm{pk}}, N, S, B_1, B_2\right)$, where hash function is represented as H.
Stage 6: The signature output is as, $\sigma \leftarrow (\delta, \phi, C_V, B_1, B_2)$.

3.4.3 Verification of Key Signature

In signature verification, the SV based signature verification is used to check the nodes. The SV based signature verification will be described in the following stages.

Stage 1: By using group public key, message, and support value, the signature verification nodes' steps are initialized, i.e., Verify$\left(g_{\text{pk}}, \sigma, N\right)$.

Stage 2: Re-drive the $\hat{B_1}$, $\hat{B_2}$ (helper values) to the first step by Eqs. (14) and (15) in the signature verification.

$$\hat{B_1} = B_1 \oplus \delta \tag{14}$$

$$\hat{B_2} = B_2 \oplus \phi \tag{15}$$

Stage 3: By Eq. (16), the support value is computed

$$\hat{S} = \frac{\left(\hat{B_1} + \hat{B_2}\right)}{\left(\hat{B}_1 * \hat{B_2}\right)} \tag{16}$$

Stage 4: By using $\hat{C} \leftarrow H\left(g_{\text{pk}}, N, \hat{S}, \hat{B_1}, \hat{B_2}\right)$, challenge value can be determined. Where hash function is represented as H.

Stage 5: Observe the challenge $\left[C_V == \hat{C_V}\right]$.

The check step will be valid σ where the challenge value (CV) is the same, and the data is transmitted with the user-verified signature. If the CV is not equal, the user results in an invalid σ, so that the data are not shared with the user and are treated as an attack. By using this, information is safely shared with the verified nodes, and access to the other nodes is not possible. This operation provides the security of information sharing without a tapping attack by the primary user's altogether.

4 Results and Discussion

This work is carried out in MATLAB platform, along with the performance results of PDR, E2E delay, accuracy of detection, energy dissipation, and TP. To show the efficiency of the proposed method, work is compared to existing techniques like improved LEACH routing (SDILR) and low-energy adaptive clustering hierarchy, such as the secure distance-based method (LEACH). Compared to the secure data aggregation (SCDA), Gaussian reputation system (GRSSN), and hierarchical trust management (HTM), the performance of energy dissipation and trust probabilities is comparable.

The E2E performance shown in Fig. 2 is calculated by varying numbers of nodes. Our proposed method is compared with SDILR and LEACH techniques. At the node

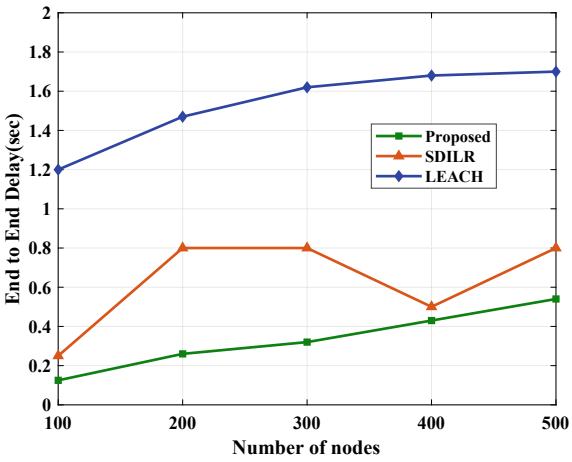


Fig. 2 Performance of E2E delay

500, the existing technique of LEACH and SDILR is 1.7 and 0.8 s, and proposed method is 0.5 s. Compared to existing techniques, our proposed work achieves less delay performance.

The throughput performance is calculated by varying number of nodes which is shown in Fig. 3. Our proposed method is compared with SDILR and LEACH techniques. At the node 500, the throughput performance of existing techniques is LEACH and SDILR which is 10,500 bits and 10,800 bits, and proposed method is 11,400 bits. Compared to existing techniques, our proposed work achieves high throughput performance.

The performance of proposed PDR is shown in Fig. 4 by comparing existing techniques like SDILR and LEACH. The PDR is calculated by different number

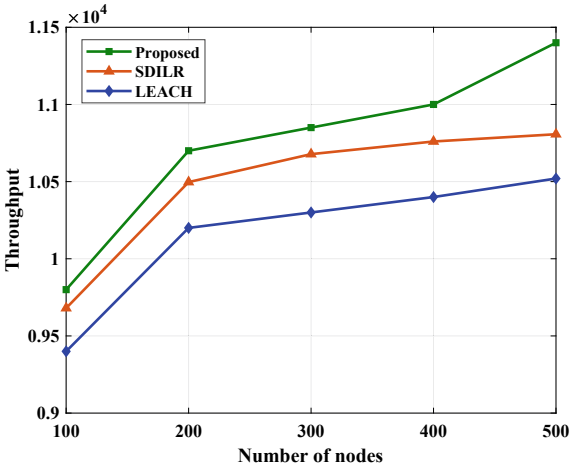


Fig. 3 Performance of throughput

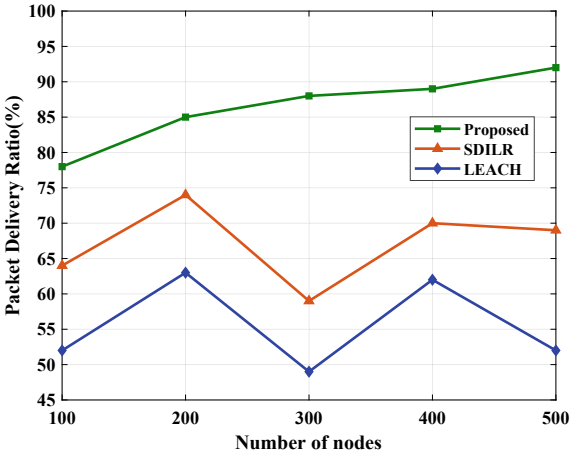


Fig. 4 PDR performance

of nodes. The existing SDILR and LEACH for PDR is 52% and 69.6%, and the proposed technique is 92%. Our proposed method achieves high PDR than existing techniques.

Detection accuracy performance is depicted in Fig. 5 by varying number of nodes. At the node 500, the detection accuracy performance of existing techniques are LEACH and SDILR which is 82 and 93.5%, and proposed method is 97%. Compared to existing techniques, our proposed work achieve high detection accuracy.

The performance of energy dissipated is shown in Fig. 6 by comparing existing techniques like SCDA, GRSSN, and HTM. Different number of MNs is determined by the energy dissipation. The existing SCDA, GRSSN, and HTM for energy dissipation is 34, 41, and 46 J, the proposed technique is 30 J.

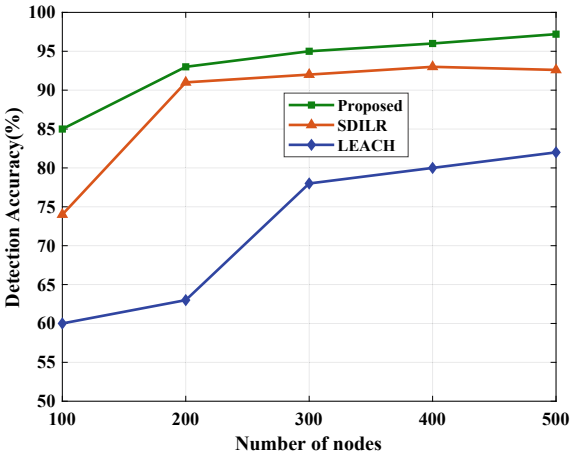


Fig. 5 Performance of detection accuracy

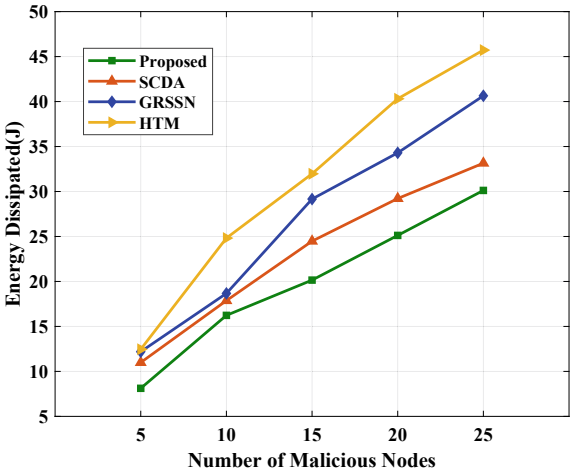


Fig. 6 Energy dissipated performance

The performance of TP is shown in Fig. 7 by comparing existing techniques like SCDA, GRSSN, and HTM. TP is calculated by varying number of MNs. Most of the nodes are dropping its energy soon in SCDA, GRSSN, and HTM. With an increasing number of MNs and fewer regular nodes, the existing method calculates the TP well. Our proposed method calculates the trust properly with proper cluster head with higher amount of regular nodes.

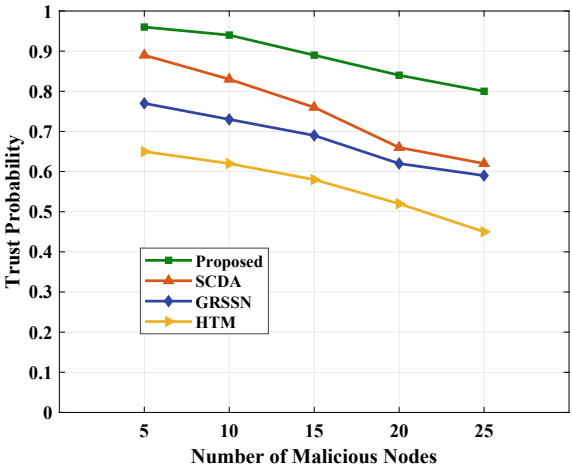


Fig. 7 Number of MNs verses TP

5 Conclusion

In this paper, we addressed the security issues by using STELR protocol. Initially, the sensor node is placed in the network for secure communication without attacking eavesdropping. Here, cluster head is selected by using EPO algorithm based on distance and energy for data aggregation, and STELR is used to enable the security for routing. The work is implemented in MATLAB platform, and the results of the PDR, E2E delay, throughput, detection accuracy, energy dissipated, and TP are achieved. At the node 500, the E2E delay is 1.7 s, throughput is 11,400 bits, PDR is 92% and detection accuracy is 97%. In future, trust model will be extended with meta-heuristic optimization algorithm for reduce the energy consumption.

References

1. W. Fang, X. Wen, J. Xu, J. Zhu, CSDA: a novel cluster-based secure data aggregation scheme for WSNs. Clust. Comput. **22**(3), 5233–5244 (2019)
2. M. Shobana, R. Sabitha, S. Karthik, Cluster-based systematic data aggregation model (CSDAM) for real-time data processing in large-scale WSN. Wireless Pers. Commun. **4**:1–19 (2020)
3. M.R. Senouci, A. Mellouk, A robust uncertainty-aware cluster-based deployment approach for WSNs: coverage, connectivity, and lifespan. J. Network Comput. Appl. **146**, 102414 (2019)
4. H. Fu, Y. Liu, Z. Dong, Y. Wu, A data clustering algorithm for detecting selective forwarding attack in cluster-based wireless sensor networks. Sensors **20**(1), 23 (2020)
5. S. Sujanthi, S.N. Kalyani, SecDL: QoS-aware secure deep learning approach for dynamic cluster-based routing in WSN assisted IoT. Wireless Pers. Commun. **114**(3), 2135–2169 (2020)
6. A.S. Reegan, V. Kabila, Highly secured cluster based WSN using novel FCM and enhanced ECC-ElGamal encryption in IoT. Wireless Pers. Commun. 1–17 (2021)
7. K. Haseeb, A. Almogren, N. Islam, I. Ud Din, Z. Jan, An energy-efficient and secure routing protocol for intrusion avoidance in IoT-based WSN. Energies **12**(21), 4174 (2019)
8. V. Gomathy, N. Padhy, D. Samanta, M. Sivaram, V. Jain, I.S. Amiri, Malicious node detection using heterogeneous cluster based secure routing protocol (HCBS) in wireless adhoc sensor networks. J. Ambient. Intell. Humaniz. Comput. **11**(11), 4995–5001 (2020)
9. P. Sharmila, A.S.M. Priyadharson, A cluster-based secured data transmission protocol for efficient data gathering in WSN. Int. J. Veh. Inf. Commun. Syst. **4**(4), 331–343 (2019)
10. T. Wang, K. Hu, X. Yang, G. Zhang, Y. Wang, A trust enhancement scheme for cluster-based wireless sensor networks. J. Supercomput. **75**(5), 2761–2788 (2019)
11. V. Vijayalakshmi, A. Senthilkumar, USCDRP: unequal secure cluster-based distributed routing protocol for wireless sensor networks. J. Supercomput. **76**(2), 989–1004 (2020)
12. T. Kalidoss, L. Rajasekaran, K. Kanagasabai, G. Sannasi, A. Kannan, QoS aware trust based routing algorithm for wireless sensor networks. Wireless Pers. Commun. **110**(4), 1637–1658 (2020)
13. M. Pavani, P.T. Rao, Adaptive PSO with optimised firefly algorithms for secure cluster-based routing in wireless sensor networks. IET Wireless Sens. Syst. **9**(5), 274–283 (2019)
14. S.S. Ranjani, S. Radhakrishnan, C. Thangaraj, Secure cluster based data aggregation in wireless sensor networks, in *2014 International Conference on Science Engineering and Management Research (ICSEMR)* (IEEE, 2014), pp. 1–6
15. K.V. Kumar, T. Jayasankar, V. Eswaramoorthy, V. Nivedhitha, SDARP: security based data aware routing protocol for ad hoc sensor networks. Int. J. Intell. Networks **1**, 36–42 (2020)

16. B. Sahu, P. Parida, A.K.. Parida, S.K. Mishra, Token based data security in inter cluster communication in wireless sensor network, in *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)* (IEEE, 2020), pp. 1–6
17. R. Bestak, Intelligent traffic control device model using Ad Hoc network. J. Inf. Technol. **1**(02), 68–76 (2019)
18. A. Chandy, Smart resource usage prediction using cloud computing for massive data processing systems. J. Inf. Technol. **1**(02), 108–118 (2019)

Design and Simulation of UWB Antenna for 4G Applications

Kishor Bhangale, Jayasree Annamaraju, Sakshi Kulkarni, Gayatri Bokey, and Triveni Dhamale

Abstract Vertical disk monopole antenna increases the structural complexities while integrating on PCB; hence, there is need of planar antenna structures. Also, there is need of compact, efficient, cheap, and robust antenna that has ability to replace multiple antenna over wide frequency range and minimum interference. In this paper, a 30 $\times$ 20 mm^2 simplified form of Ultra-Wide Band Microstrip Patch Antenna (UWB) is designed that is optimal for various 4G applications. This paper presents the design steps, the slotting and the simulation of the antenna. A single Ultra-Wide Band Antenna (UWB) has the ability to replace multiple narrow band antennas. We use this principle in our research and design a UWB Antenna over the 4G network range. We establish a range of frequencies over 3.5–5.2 GHz so that it covers LTE, WiMAX, WLAN, etc. Various performance parameters for this antenna shall be observed in this paper.

Keywords Antenna design · Compact UWB · CST · LTE · 4G applications · Simulation · WiMAX · WLAN

1 Introduction

We are in a world, bound as a single community linked by communication, also known well as a global village. In this global village, wireless communication plays a huge role. Wireless communication is the fastest developing and most growing technology in the communication sector [1, 2].

Wireless communication is the process of transmitting data in any format from one point to other, without the use of any physical medium. Wired communication can, at its maximum, reach across few meters to few kilometers. But with the use of wireless communication, we can transmit information over a vast area from nearest to the farthest like countries across. Some of the most known Wireless Communication Systems are Bluetooth, Wi-Fi, etc., [3, 4].

K. Bhangale (✉) · J. Annamaraju · S. Kulkarni · G. Bokey · T. Dhamale
Pimpri Chinchwad College of Engineering and Research, Ravet, Pune, India
e-mail: kishor.bhangale@pccoer.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_4

The main component of these wireless systems is the antenna that acts as an interface between radio waves, more specifically, the electromagnetic waves [5]. Over the decades, there has been a lot of research and development in the field of antennas. This research has led to the discovery of various types of antennas. One of the most significant discovery in this field is that of Ultra-wide Band Antennas (UWBs). While various antennas have the ability to function over particular frequency ranges, a single UWB antenna has the ability to replace multiple antennas. This gives user the advantages of compactness, cost-efficiency and wide-range of applications among others [6]. Though these antennas provide advantages over their fellow antennas, their design complexity is one of the major concerns. The design of a UWB Antenna needs complex slotting techniques, which makes it complicated for a designer to begin with, and this also impacts the PCB printing later on. The second major concern is the interference [7]. UWB uses a large frequency range which leads to interference with other standards that function over similar frequencies [8, 9].

This paper presents a solution for several disadvantages mentioned earlier by designing a UWB Antenna using a simple slotting technique over a 4G application range. It presents the design and simulation of the UWB antenna using CST microwave studio and performance evaluation of the antenna.

The remaining paper is structures as follow: Sect. 2 gives details about related work on design of UWB antenna for various applications. Section 3 provides design steps of UWB antenna. Section 4 presents the simulation of UWB antenna using CST microwave studio. Section 5 focuses on experimental results and discussions. Finally, Sect. 6 discusses conclusion and future scope.

2 Related Work

Various structures and designs have been presented in the recent year for the UWB antenna design. Ibrahim et al. [10] presented dual elemented UWB Antenna with a decoupling circuit which provided low return loss, high bandwidth and gain. The design was prone to interference in significant amount. Gayatri et al. [11] focused on the design of UWB planar printed circuit board (PCB) that can achieve large capacity as high as hundreds of Mbps or several Gbps. The large capacity of this antenna design leads to larger antenna size. Nirmal et al. [12] proposed compact wideband MIMO antenna with triangular slots and meshed metal strips. It has resulted in high mutual coupling over wideband and enhanced impedance but it has larger complexity in design. Mahmud et al. [13] offered the UWB antenna design having very simple Hibiscus petal pattern patch with tapered microstrip fed a line and partial trapezoid ground plane to cover the UWB bandwidth from 3.1 to 10.6 GHz. It has shown fairly good agreements between the simulations and measurements, better impedance match and mutual coupling. Its gain is comparatively lower than other antennas working on similar frequency ranges. Singh et al. [14] presented Particle Swarm Optimization (PSO) algorithm for easy calculation of antenna dimensions and parameters. It has given improved bandwidth, high gain and increase in efficiency

but has inability to balance amplifying antenna parameters. Lim et al. [15] explore the compact UWB operating between 4.1 and 10 GHz using time domain Gaussian pulse excitation analysis. Simulation results demonstrated reasonable agreement with the measurement results and good ultra-wideband linear transmission performance has also been achieved in time domain. It has shown comparatively slower performance than other similar domain antennas. Thus, there is need to design the robust, efficient, compact antenna which can results in minimum interference and higher gain.

3 Desıgn of UWB Antenna

Figure 1 gives the flowchart of the design steps of the UWB antenna that encompasses the frequency range consideration, design material consideration, antenna patch design and simulation on CST microwave studio.

3.1 Determination of Frequency Range

Every antenna operates over a particular range of frequencies. We need to determine the range over which the UWB Antenna that we are designing is expected to function.

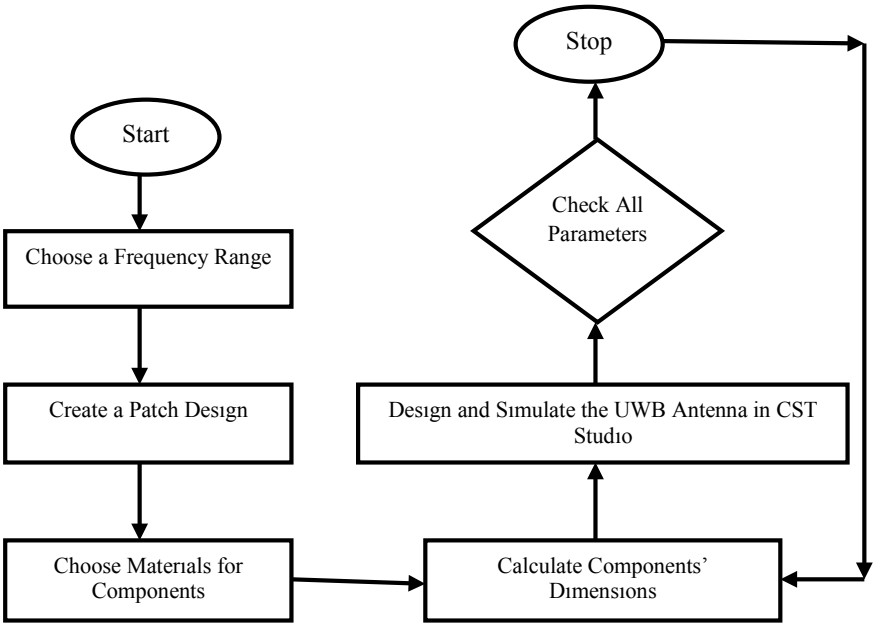


Fig. 1 Flow chart of proposed UWB antenna design steps

Since we are designing one that needs to function as a WLAN, LTE and WiMAX, we study their allocated frequency ranges and determine a common frequency range that can include all the three ranges. The IEEE 802.6 standard frequency range, over which WiMAX data transmission is allowed, is 3.4–3.6 GHz [16] while the IEEE 802.6 standard frequency range of LTE covers both 3.5 and 5 GHz [17]. For WLAN, the standard frequency range of WLAN covers frequencies ranging across 2.4 and 5 GHz [18]. Taking all these frequency ranges into consideration, we establish a range of frequencies over which all the above-mentioned purposes can be fulfilled, which ranges over a broad category of 3.5–5.2 GHz.

3.2 Determination of Materials

The antenna has three layers—Ground, Substrate and Patch. This UWB employs over a half ground [19]. Annealed Copper is used for the construction of ground of this antenna due to its high tensile strength and high conductivity [20]. The substrate is twice the size of the ground and is made of Lossy FR-4 material. Lossy FR-4 is used because it is a great electrical insulator with high dielectric strength. Also, it is light-weight and moisture resistant [21]. For the conducting patch, we use annealed copper again due to its properties. We need to determine the dielectric constant (ε), the dielectric height (h) and the operating resonant frequency (f_r).

3.3 Antenna Dimension Calculations

Width of Patch. At lower width, there is neither any variation in the broadening nor a good return loss obtained. As the width increases, the bandwidth increases slightly but there is no significant change in the resonant frequency. Width of patch is given by Eq. 1 [22–24].

$$W_p = \frac{c_0}{2f_r}\sqrt{\frac{2}{\varepsilon_r + 1}} \tag{1}$$

where

W_p—Width of the patch
C_0—Speed of light
f_r—Operating resonant frequency
ε_r—Value of the dielectric substrate's refractive index.

Effective Refractive Index. The effective refractive index can be defined as a number quantifying the phase delay per unit length in a waveguide [22–24]. Equation 2 shows the formula of effective refractive index.

$$\varepsilon_{\text{reff}} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + 12\frac{h}{W_p}\right)^{-\frac{1}{2}}, \frac{W_p}{h} > 1 \tag{2}$$

where

$\varepsilon_{\text{reff}}$—Value of the dielectric substrate's reference refractive index

W_p—Width of patch

h—Height of substrate.

The lesser the effective refractive index, less is the dispersion of electromagnetic waves incident on the antenna.

Length of Patch. The length of patch basically represents the free- space wavelength. The length of patch can be calculated using Eq. 3 and 4.

$$\frac{\Delta L_p}{h} = 0.412\frac{(\varepsilon_{\text{reff}} + 0.3)\left(\frac{W_p}{h} + 0.2643\right)}{(\varepsilon_{\text{reff}} - 0.258)\left(\frac{W_p}{h} + 0.8\right)} \tag{3}$$

where

ΔL = Increase in size of antenna due to fringing

$$L_p = \frac{C_0}{2f_r\sqrt{\varepsilon_{\text{reff}}}} - 2\Delta L_p \tag{4}$$

The length of the antenna decreases as the relative dielectric constant of the patch increases [22–25].

Length and Width of Ground Plane. The effect of length and width of ground plane is that with an increase in the ground plane size, both the gain and antenna efficiency increase [25, 26]. The length and width of the ground plane can be calculated using Eqs. 5 and 6.

$$L_g = 6h + L_p \tag{5}$$

$$W_g = 6h + W_p \tag{6}$$

where L_g and W_g refer length and width of the ground, respectively.

The design parameters and dimensions for 4.35 GHz frequency are computed using Eqs. 1–6 and shown in Table 1.

Table 1 UWB antenna parameters and dimensions

Parameter	Dimensions
W_p	16 mm
$\varepsilon_{\text{reff}}$	4.3
L_p	25 mm
L_g	14 mm
W_g	20 mm

4 Simulation of UWB Antenna Design

There are numerous electromagnetic analysis software's available. We chose CST Microwave Studio software for this antenna design. The CST Microwave Studio has facility of design, analysis, and optimization of the electromagnetic components of the communication elements. It can provide the 3D visualization of antenna design, radiation pattern, electro-mechanical and thermal effect on the components. The design steps for simulation can be given as:

- In CST Studio, we begin with choosing a template. For our antenna, we choose the Microwave and RF/Optical template.
- In the subdivisions, we choose the Antenna.
- Among the workflows, we choose the Planar workflow.
- And then, from the various solvers, we choose the Time domain solvers.
- Now, in the unit's section, make sure it shows GHz along frequency.
- In the settings section, choose the maximum frequency a bit more and minimum frequency a bit less than the original frequencies but not exact.
- Choose Far-field monitor and define it at the mean frequency.

Front view consists of a patch, feedline and a substrate as shown in Fig. 2. The patch is in the shape of two rectangles of which one has a rectangular slot.

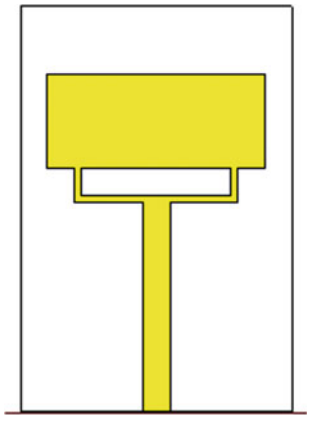


Fig. 2 Substrate and patch of the antenna view

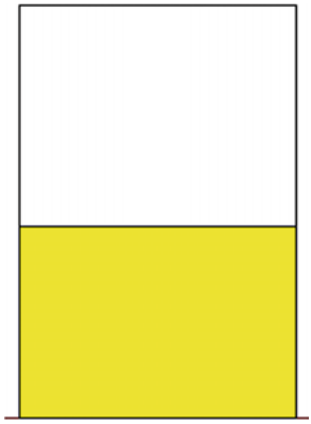


Fig. 3 Half ground of the Antenna view

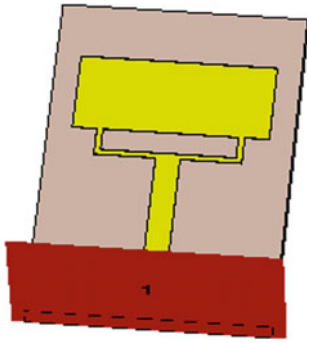


Fig. 4 Waveguide port view

The back view consists of the half ground, i.e., the ground of this antenna covers half of the substrate area as shown in Fig. 3. We choose half ground plane for this antenna so as to reduce the interference and crosstalk while simultaneously reducing electrical noise in the antenna.

To simulate the antenna, a waveguide port has to be attached to the feed line. The red rectangular section marked 1 is the waveguide port in Fig. 4.

5 Experimental Results and Discussions

The performance of the proposed UWB antenna is evaluated based on various evaluation metrics to verify the functioning of our antenna.

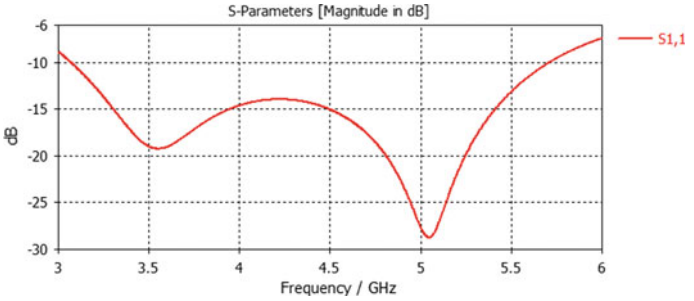


Fig. 5 S-Parameters of proposed UWB antenna

5.1 S-Parameters

The S-parameters represents how much power is reflected from the antenna. The coefficient S11 is hence, also known as reflection coefficient. The S-Parameters of the antenna are as shown in Fig. 5.

The return loss is lowest in the region of 3.5 and 5.2 GHz which implies that the loss is minimum in the working region of the antenna.

5.2 Efficiency

Ideally, an antenna's efficiency must be 100%. The radiation efficiency of this antenna is observed to be 79.5% at 3.5 GHz and 71.5% at 5.2 GHz, which is close to the ideal value. Figure 7 shows the relation between radiation efficiency and the total efficiency of the antenna.

Total efficiency of an antenna is always lower than radiation efficiency. This is clearly shown in Fig. 6 where the red line depicts radiation efficiency of the antenna, and green line depicts the total efficiency of the antenna.

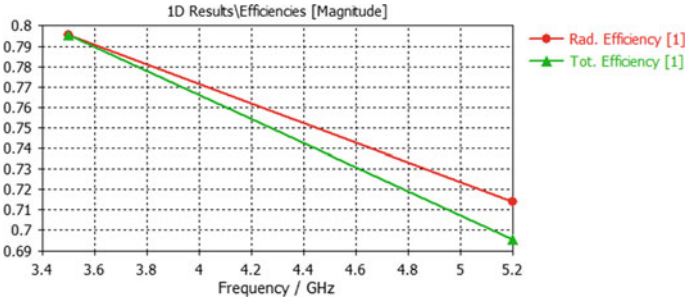


Fig. 6 Efficiency plot of UWB antenna for various frequencies

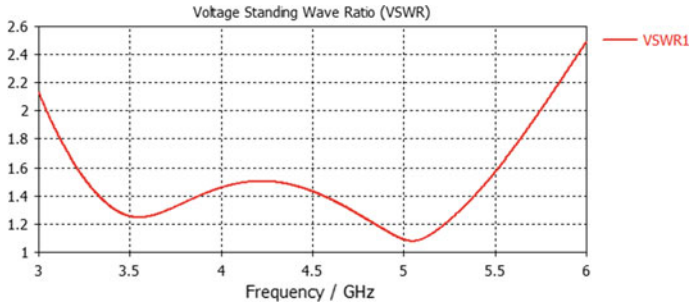


Fig. 7 VSWR of proposed UWB antenna

5.3 Voltage Standing Wave Ratio (VSWR)

Voltage Standing Wave Ratio (VSWR) is a measure of how efficiently radio-frequency power is transmitted from a power source to load. Ideally, VSWR is supposed to be in the range of 1 to 1.5. Figure 7 shows the VSWR of proposed UWB antenna. The VSWR of the antenna is observed to hold minimum values of 1.25 at 3.5 GHz and 1.1 at 5.2 GHz. The mean VSWR value is determined as 1.175.

5.4 Far-Field Monitor

In Far-field region of the antenna, the EM fields are dominated by radiating fields. The E and H fields are orthogonal to each other and to the direction of the propagation as with plane waves. Antennas usually transfer signals to large distances which are considered to be far-field region. Far-field monitor monitors the performance of antenna in this far-field region and observes the amount of power radiated. The far-field patterns at 3.5 and 5.2 GHz in Figs. 8 and 9 shows that the antenna radiates in all directions almost equally.

Far-field ($f = 3.5$). The area in red depicts the maximum power radiated while the power radiated in the blue area is almost negligible. Figure 8 shows that the antenna, at 3.5 GHz, radiates maximum power in the region in closer proximity to the antenna and is slightly lesser at the poles of the region.

Far-field ($f = 5.2$). Figure 9 shows that the maximum power transmitted comparatively lowers as the frequency increases. It is observed that the antenna radiates higher power closer to the axes.

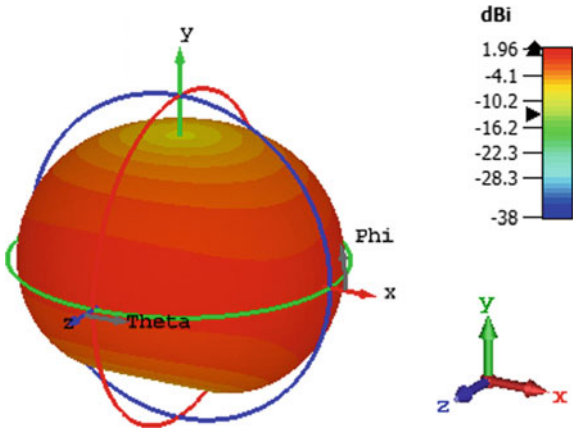


Fig. 8 Far-field radiation of proposed UWB antenna (f = 3.5)

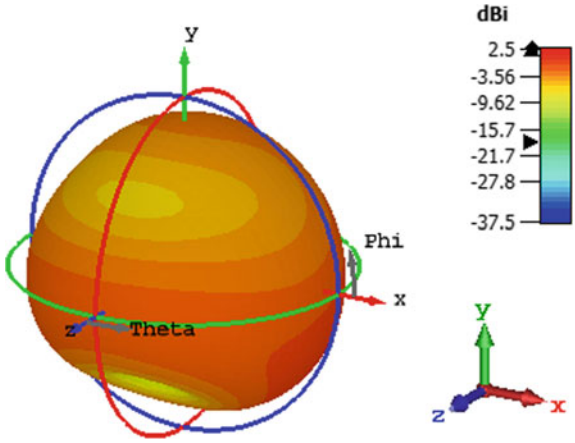


Fig. 9 Far-field radiation of proposed UWB antenna (f = 5.2)

6 Conclusıons and Future Scope

The above parameters clearly prove that the antenna's performance is up to the mark and gives the expected results. The antenna has higher efficiency than other antennas working on similar frequency range. It also has nearly ideal VSWR with low return loss. This antenna can be used for various applications, namely various broadband connections, voice-over IP, video conferencing, and advanced interactive gaming of recent times. They can also function as hotspots in restaurants, and hotels where WLAN, WiMAX and LTE internetworking are active. Research and development in UWB Antenna have huge scope in near future. Developments concerning the elimination of interference and to reduce the return loss while reducing the design complexity can be done.

References

1. P. Yadav, A. Upadhyay, V.S. Prasath, Z. Ali, B.B. Khare, Evolution of wireless communications with 3G, 4G, 5G, and next generation technologies in India, ın *Advances in Electronics, Communication and Computing* (Springer, Singapore, 2021), pp. 355–359
2. A.A. Salih, S.R. Zeebaree, A.S. Abdulraheem, R.R. Zebari, M.A. Sadeeq, O.M. Ahmed, Evolution of mobile wireless communication to 5G revolution. Tech. Rep. Kansai Univ. **62**(5), 2139–2151 (2020)
3. K.M. Luk, K.F. Lee, Antennas in wireless communications. Proc. IEEE **100**(7), 2104–2108 (2012)
4. M.A. Chung, W.H. Chang, Low-cost, low-profile and miniaturized single plane antenna design for an Internet of Thing device applications operating in 5G, 4G, V2X, DSRC, WiFi 6 band, WLAN, and WiMAX communication systems. Microw. Opt. Technol. Lett. **62**(4), 1765–1773 (2020)
5. K. Kaur, A. Kumar, N. Sharma, A review of ultra wideband antennas with band notched characteristics. Available at SSRN 3565018, (2020)
6. U. Ullah, I.B. Mabrouk, S. Koziel, M. Al-Hasan, Implementation of spatial/polarization diversity for improved-performance circularly polarized multiple-input-multiple-output ultra-wideband antenna. IEEE Access **8**, 64112–64119 (2020)
7. H. Schantz, The art and science of ultra-wideband antennas (Norwood MA: Artech House Inc, 2005)
8. T. Saeidi, I. Ismail, W.P. Wen, A.R. Alhawari, A. Mohammadi, Ultra-wideband antennas for wireless communication applications. Int. J. Antennas Propag. (2019)
9. M.N. Srifi, M. Essaaidi, Ultra-wideband printed antennas design. Ultra Wideband Commun.: Novel Trends-Antennas Propag. **195** (2011)
10. M.A. Ibrahim, W. Swelam, M.H. Abd El-Azeem, H. el Hennawy, Ultra-wide band microstrip antenna for 4G applications. IOP Conf. Series: Mater. Sci. Eng. **610**(1), 012025 (2019)
11. T. Gayatri, G. Srinivasu, M.K. Meshram, V.K. Sharma, Analysis and design of a planar UWB antenna for spectrum sensing in 3.1–10.6 GHz, ın *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (2020), pp. 1–6
12. P.C. Nirmal, A. Nandgaonkar, S. Nalbalwar, R.K. Gupta, Compact wideband MIMO antenna for 4G WI-MAX, WLAN and UWB applications. AEU-Int. J. Electron. Commun. **99**, 284–292 (2019)
13. M.Z. Mahmud, M.T. Islam, M. Samsuzzaman, A high performance UWB antenna design for microwave imaging system. Microw. Opt. Technol. Lett. **58**(8), 1824–1831 (2016)
14. J.M. Singh, M. Mishra, P. Sharma, Design & optimization of microstrip patch antenna. Int. J. Emerg. Trends Technol. Comput. Sci. **2**(5), 1–5 (2013)
15. K.S. Lim, M. Nagalingam, C.P. Tan, Design and construction of microstrip UWB antenna with time domain analysis. Prog. Electromagnet. Res. **3**, 153–164 (2008)
16. IEEE 802.11–2016.: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specification (IEEE, 2016).
17. LTE Frequency bands, https://www.everythingrf.com/community/lte-frequency-bands. Retrieved on 25 May 2021
18. Y.S. Jeong, S.H. Lee, J.H. Yoon, W.Y. Lee, W.Y. Choi, Y.J. Yoon, Internal mobile antenna for LTE/GSM850/GSM900/PCS1900/ WiMAX/WLAN, ın *2010 IEEE Radio and Wireless Symposium (RWS)* (2010), pp. 559–562
19. N. Chattoraj, Design and development of a small compact ultra-wideband antenna. IOP Conf. Ser.: Mater. Sci. Eng. **44,** 012003 (2013)
20. A.S. Mekki, M.N. Hamidon, A. Ismail, A.R. Alhawari, Gain enhancement of a microstrip patch antenna using a reflecting layer. Int. J. Antennas Propag. (2015)
21. S.S. Islam, M.A. Rahman, M.R.I. Faruque, M.T. Islam, Design and analysis with different substrate materials of a new metamaterial for satellite applications. Sci. Eng. Compos. Mater. **25**(1), 59–66 (2018)

22. A.S.S. Ramna, Design of rectangular microstrip patch antenna using particle swarm optimization. Int. J. Adv. Res. Comput. Commun. Eng. **2**(7) (2013)
23. M.M. Ahamed, K. Bhowmik, A.A. Suman, Analysis and design of rectangular microstrip patch antenna on different resonant frequencies for pervasive wireless communication. Int. J. Sci. Technol. Res. **1**(5) (2012)
24. A.K. Gautam, S. Yadav, K. Rambabu, Design of ultra-compact UWB antenna with band-notched characteristics for MIMO applications. IET Microwaves Antennas Propag. **12**(12), 1895–1900 (2018)
25. S. Kumar, H. Gupta, Design and study of compact and wideband microstrip U-slot patch antenna for wi-max application. IOSR J. Electron. Commun. Eng. (IOSR-JECE) **5**(2), 45–48 (2013). e-ISSN: 2278–2834,p- ISSN: 2278–8735
26. N. Anveshkumar, A.S. Gandhi, Design and performance analysis of a modified circular planar monopole UWB antenna, ın *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (2017), pp. 1–5

A Comparative Analysis Between Hyperledger Fabric and Ethereum in Medical Sector: A Systematic Review

K. B. Jyothilakshmi, Vandana Robins, and A. S. Mahesh

Abstract In a medical record management system, a complete database is used to store and obtain patient medical information. The electronic medical record has supplanted the conventional paper medical record as the primary source of healthcare information for clinical, legal, and administrative purposes. The risks of data being misused are high because the patient has no control over the data. As a consequence, we need a patient-centred approach that is fully decentralized, capable of detecting data theft, preventing data abuse, and enabling patients to exercise control over their access. The perfect approach to all issues and needs is Blockchain technology. Blockchain has the potential to have an effect on billing, record sharing, medical testing, identity theft, and financial data theft in the future. In addition, the implementation of smart contracts in healthcare could make things much simpler. When this command is used, the Blockchain can be used to create and validate records. This paper compares and contrasts the most widely used Blockchain systems. The Hyperledger framework and the Ethereum platform are the two major general-purpose blockchain frameworks in use. The primary difference between Hyperledger and Ethereum is the purpose for which they were developed. The primary goal of this paper is to find the contrast between these two blockchain platforms for medical record management systems in order to determine which blockchain platform is best for developing healthcare applications.

Keywords Blockchain · Hyperledger Fabric · Ethereum · Smart contract · Healthcare application · Consensus · Permissioned · Permissionless · Distributed ledger · Hash

K. B. Jyothilakshmi (✉) · V. Robins · A. S. Mahesh
Department of Computer Science and IT, Amrita School of Arts and Sciences, Kochi, India

Amrita Vishwa Vidyapeetham, Coimbatore, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_5

1 Introduction

In recent years, health information management networks have faced a number of problems; the most significant of which is the management of patient record data. With the abundance of wearable medical records outside of medical facilities, modern medical practices rely heavily on patient's historical medical records, which are becoming increasingly huge. Since medical records data is so important to a patient's diagnosis and care plan, it is critical to ensure that the data's credibility is protected. Health data can be kept secure, private, and inter-operable using blockchain technology. The Hyperledger Fabric framework and the Ethereum framework are two common trendy-purpose frameworks. This paper contrasts the medical record management system's general-purpose mechanisms. Hyperledger Fabric is a plug-and-play blockchain platform that can be used to create blockchain-based products, solutions, and applications for private businesses. Ethereum is an open-source and distributed decentralized blockchain network. With the support of smart contract functionality, it allows for the development of decentralized applications. The most important difference between Hyperledger Fabric and Ethereum is the reason for which they were produced. The most significant aspect of Ethereum is that it allows smart contracts to be executed, enabling decentralized applications to be created on top of all. It was the primary blockchain-based network to implement the idea of smart contracts, which is why it is so popular for developing decentralize applications based on smart contracts, such as healthcare apps. The Ethereum platform has a few drawbacks when it comes to creating healthcare software. It faces some of the flaws as Bitcoin: miners must be compensated in ether, a cryptocurrency that is immune to a 51 percent assault. Hyperledger Fabric, in contrast to other systems, is one of the most comprehensive solutions for designing healthcare applications. Hyperledger Fabric has security features and robust privacy. As a result, the developers have a robust toolkit at their disposal for rapidly enforcing a wide range of privacy and security policies. Unlike the other frameworks, Fabric supports smart contracts and does not rely on a computationally intensive consensus protocol, Fabric does not accept cryptocurrency as a payment method. The purpose of this work is to compare and contrast these frameworks in terms of medical data management system requirements in order to decide which framework is more effective and stable.

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines when performing and reporting this systematic literature review. Our aim is to find the best blockchain framework for a record management system, so we choose this type of literature review. Also, to summarize the benefits, drawbacks, and potential directions of each blockchain platform in the healthcare sector. Unlike a meta-analysis, this study did not provide any data synthesis. There was no quality assessment since the aim was to establish a shared interpretation of the actions and solutions instead of to evaluate the performance of various blockchain platforms.

The following activities were described as part of our review.

1. Research Questions.

2. Search Strategy.
3. Article selection.

2 Related Works

(1) Comparison of blockchain platforms: a systematic review and healthcare examples
Tsung-Ting Kuo, Hugo Zavaleta Rojas, Lucila Ohno-Machado

They identified healthcare blockchain applications as well as platforms that have been proposed or implemented by state-of-the-art healthcare blockchain studies in this study. The authors conducted a systematic evaluation of 21 parameters for ten common blockchain platforms in order to assist health informatics researchers in selecting the best appropriate platform for their unique application. They explored essential blockchain implementation elements, introduced these platforms, and compared the most significant technological elements to healthcare in their study. Their findings demonstrate that while most platforms have comparable functionalities, they differ in technological qualities like as transaction speed, block mining centralization mitigation, user privacy or anonymity enhancement, and support for permissioned or private blockchain networks. Based on their findings, which show that most healthcare or biomedical lead projects are still in the conceptual stages, they predict that implementing blockchain technology in the healthcare or medicine domains will necessitate a societal revolution that will be far more difficult than the technological difficulty. Vision, entrepreneurship, and a little initial investment that will likely pay off in a few years will be required to think differently and change the way our sector functions for decades. The intrinsic benefits of blockchain (decentralization and immutability of the ledger) need scenarios in which a minimum number of users choose to utilize it, with some agreeing to actively contribute "blocks" and perform other activities (for example, predictive modelling) for healthcare or biological applications. If the correct incentive system can be put in place, the transition will undoubtedly occur.

(2) Comparison of Smart Contract Blockchains for Healthcare Applications
Hongru Yu, Haiyang Sun, Danyi Wu, Tsung-Ting Kuo

They addressed practical considerations while developing a hospital blockchain and smart contract system in their study. They examined the technical aspects of smart contract blockchain platforms, chose platforms for their research, built a blockchain network for each platform, tested the blockchains, and summarized their experience, and time spent implementing them. MultiChain is the simplest blockchain platform to set up, according to its implementations. Ethereum took longer to start up as a permissioned network than MultiChain. Hyperledger Fabric's network has more layers to boost manageability and security, but it can take longer to set up than other platforms. In comparison to Hyperledger Fabric, installing the software required for MultiChain and Ethereum is also simple. On the other side, Ethereum and Hyperledger offer full

smart contract functionality, whereas MultiChain's smart contract support is limited. The smart contracts are simply programmable and extremely readable.

The following are the key differences between the three platforms when it comes to biomedical/healthcare applications. A big community of developers from all over the world supports and maintains Ethereum. It is a smart choice when contemplating the long-term viability of the platform on which the applications are built because of its open nature. The blockchain platform's versatility, security, and manageability are enhanced by Hyperledger Fabric's well-designed, multi-layered access control system and Hyperledger Fabric's own certifications. MultiChain is intended to be a permissioned blockchain that is both simple and powerful to use, and it forks from the well-known Bitcoin Blockchain to inherit established features.

3 Blockchain

A shared (or distributed) database is what a blockchain is. It is a complex framework for packaging data in a way that you can trust, and it can only be changed by specific users. Once data is generated, it is broadcast and validated on a peer-to-peer network before being compiled into a block of information. The block is mixed with several information (a hash key) from the preceding block using a cryptographic method called as hashing before being added. Since it combines the previous block's Hash key, each new block is connected to all of its predecessors in the form of a chain. The block's transactions are organized into a Merkle tree, with each transaction verifiable to the centre.

Another important feature of blockchain is persistence. It is almost difficult to erase data when it has been admitted into the blockchain based on distributed ledger, which is held across several branches. It can only be accessed and changed by those who have the correct key. Each time the data is modified, a fresh information block is validated and posted to the ledger. Permissioned participants and validation tools work together to validate or deny new data if anyone attempts to tamper with a transaction or block in the chain. This procedure ensures that blockchain is a safe, stable, and reliable source (Fig. 1).

3.1 Types of Blockchain

Blockchains can be divided into three categories.

1. public (permissionless),
2. consortium (public permissioned),
3. private.

They vary in terms of who has access to, writes to, and reads the details on blockchain.

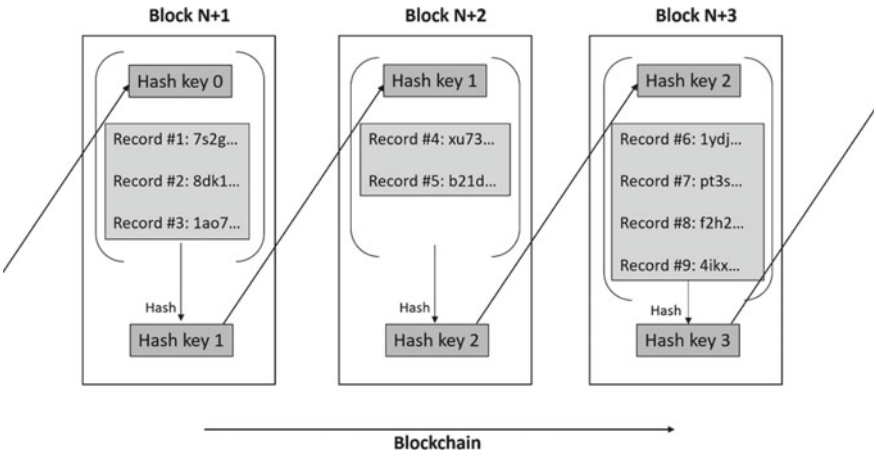


Fig. 1 Data blocks in a blockchain are connected together

Public

Everyone can access the data in a shared chain, and anyone can enter, contribute, and modify the core software. The public blockchain is commonly used in cryptocurrencies, such as Bitcoin and Ethereum, which are both public permissionless blockchains. Anyone can view, perform transactions, and participate in the consensus process on a public blockchain. It is totally distributed, decentralized, and transparent, which means that whatever transactions take place in the public blockchain network can be seen by anybody with basic internet connection. There is no middleman on the public blockchain, which means that there is no control over the network's operations.

Consortium

A consortium blockchain is partially centralized since only a limited number of relevant categories of organisations have rights to display and interact in the consensus protocol. This sort of blockchain is a variation of private blockchain. It tries to take away the private blockchain's sole autonomy which means that the network is managed by more than one individual or corporation. Blockchain consortium is a form of hybrid between public and private blockchains since, although being a permissioned network, it has a decentralized structure. Each organization in this network receives the same treatment, ensuring that all parties engaged are treated fairly. The network is not ruled by a single body. It is a platform that allows numerous firms to collaborate and share data while also keeping track of rules and records.

Private/permissioned

The network in a private blockchain is distributed but also centralized. Only a few nodes are allowed to join the network, and they are all regulated by a single central authority. The transaction can only be carried out by trusted members of the network.

It is only partially decentralized and opaque. There is a centralized authority in charge of everything. A centralized and trustworthy authority is in charge of giving approvals and providing authorized parties access to the ledger in a private blockchain. The most significant benefit of a private network is that it is transparent only to those who are actually inside the business, not to those who are on the outside.

4 Challenges in the Healthcare Sector

Regardless of age, gender, or community, providing high-quality healthcare is considered a basic need. The healthcare system faces significant obstacles in delivering high-quality services.

1. Healthcare regulatory changes.
2. Data management.
3. Data security and integrity.
4. Inefficient administration.
5. Information and integrated health services.
6. Cybersecurity.
7. Rising healthcare costs.

5 Benefits of Using Blockchain Technology in Healthcare

5.1 Single, Longitudinal Patient Records

Longitudinal patient records, including inpatient, ambulatory, and wearable data, can be created using blockchain, allowing clinicians to come up with better ways to provide care by compiling episodes, illness registries, lab reports, and treatments.

5.2 Master Patient Indices

When working with medical information, it is common for records to be mismatched or duplicated. Furthermore, each EHR has a unique schema for each region, resulting in a variety of methods for gaining access and exploiting even the most basic data sets. All data collection, not just the primary key, is hashed to a ledger with blockchain. The user will search for the address; there could be several addresses and keys, but they would all refer to the same patient identification.

5.3 Claims Adjudication

Since blockchain is a validation-based exchange, the claims can be instantly verified, and the network can determine how a contract is implemented. There will also be less mistakes or frauds since there would be no central authority.

5.4 Supply Chain Management

Healthcare companies may use blockchain-based contracts to monitor resource cycles throughout their full lifespan, including when the operation is progressing, whether the contract is operating, and whether any delays have occurred.

5.5 Interoperability

The concept of blockchain, interoperability, can be realized by using sophisticated APIs to render EHR compatibility and storage systems. If the blockchain network was shared with licensed providers in a secure and coordinated manner, the cost and burden of data reconciliation will be removed.

5.6 Tracing and Securing Medical Supplies

With complete transparency, blockchain can assist protect and identify the trail of pharmaceutical supply. It can even track the labour expenses and carbon emissions associated with the production of these items.

5.7 Data Security

The blockchain's secure characteristics can significantly improve the security of health data. Each person has a public identification or key and a private key that can only be opened when and for the length of time required.

5.8 Faster, Cheaper, Better Patient Care

Blockchain has the potential to build a single system for storing and retrieving health records in a secure and timely manner by authorized users. Countless mistakes can be avoided, faster diagnoses, and interventions are possible, and treatment may be individualized to each patient by eliminating miscommunication between different healthcare personnel involved in caring for the same patient.

5.9 Tracking Diseases and Outbreaks

The special attributes of blockchain can aid real-time disease monitoring and disease pattern investigation, which can aid in determining the disease's source and transmission patterns (Fig. 2).

Aside from these, blockchain has the potential to revolutionize sales cycle management, drug supply management, and fraud prevention.

6 Blockchain Platforms Used in Healthcare Applications

A blockchain is a series of data-storage blocks. From medical records to pharmaceutical supply chains to smart contracts for payment processing, the blockchain revolution has a huge opportunity to drive digital transformation. Every modern healthcare system is built on the foundation of electronic medical records. With each

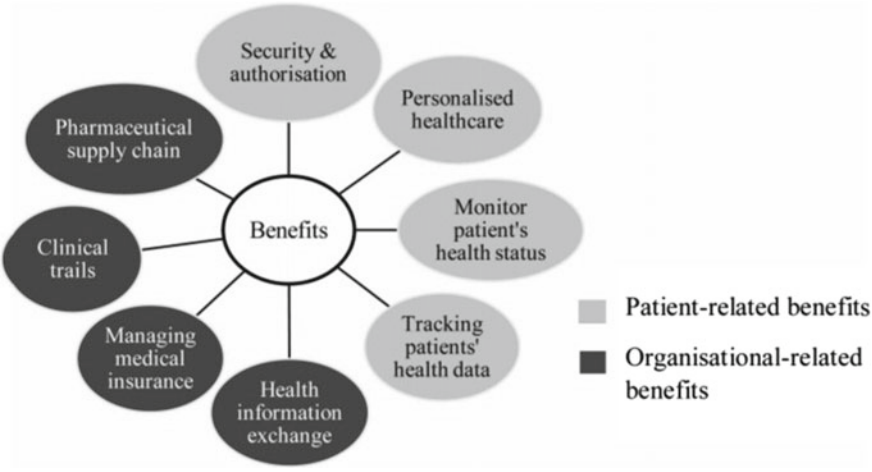


Fig. 2 Benefits of blockchain in medical sector

visit to the hospital, medical records, on the other hand, get longer and more complicated. Each hospital and physician's office appears to have its own record-keeping system. It is difficult for healthcare providers to obtain them. Blockchain is the only solution to these problems. With blockchain, data is more stable. Quorum, Hyperledger, Ethereum, EOSIO, Corda, and IBM are some of the most common blockchain platforms. One of the most crucial steps in beginning to plan and develop a proper blockchain health initiative is to choose the right underlying blockchain framework. The following are the two main platforms.

8:1 Ethereum

Ethereum is a network that is open to the public. It is an open-source, decentralized blockchain with smart contract features, which means it is a background-running computer program. The first blockchain platform to implement the concept was Ethereum, which is why it is so famous for creating smart contract-based decentralized applications like healthcare apps. All of the Ethereum system is transparent, which adds a layer of protection for each and every user. Even if it is open to the public, no third parties will be able to view it, so it is clear that no one can make adjustments to the ledger, ensuring that no corrupting actions will take place without anyone knowing. Patient information and medical records. In the creation of healthcare apps, privacy is a critical concern. It is not necessary for all transactions to be fully open to everyone. Ethereum has some drawbacks when it comes to designing healthcare applications. It has a poor transaction throughput and is thus more vulnerable to cyberattacks. The Ethereum system has a high transaction cost and a slow efficiency. For patient-managed medical data networking systems, MedRec and Patientory recommend using an Ethereum-based platform. According to Nebula Genomics, genomic data is exchanged and analysed using an Ethereum-based blockchain network. Clinical systems, such as medical data exchange and electronic remote client tracking, have also been proposed for Ethereum.

8:2 Hyperledger

Hyperledger Fabric is a blockchain technology platform that has been used to create blockchain-based medical record management systems. Hyperledger Fabric is a flexible, safe, and adaptable infrastructure. As opposed to other architectures, Hyperledger Fabric provides the most comprehensive approach for designing healthcare applications. It is well-equipped for sensitive and confidential data operations. This technology is well suited to dealing with traffic congestion. As a result, benefits such as open-source, high-quality code, modular design, and high performance make it more profitable. Hyperledger Fabric is permissioned, which is a key feature. The private blockchain, unlike the public blockchain, is not available to everyone. The person who wishes to gain access must first obtain permission. Multiple channel support, which restricts data access to some consortia members, and private collections, where data can be shared peer-to-peer and stored in private databases, but only accessible from chain code on registered peers and hashed to verify authenticity, are key components in ensuring data fabric privacy. Since a permissioned blockchain is a closed system, Hyperledger Fabric safeguards patients' privacy and confidentiality

while also keeping medical records safe and up to date. Participants have individual personalities and are familiar with one another. Fabric's smart contract programming supports a number of programming languages, including Java, Node.js, and JavaScript.

The Hyperledger network is suggested for use in an oncology clinical data exchange system for patient care. Hyperledger was also used to create a system for enforcing Institutional Review Board rules. Hyperledger is now being considered for a mobile healthcare application as well as medical data management and access applications. Hyperledger has developed a working group to promote technological and business partnerships for healthcare blockchain applications.

7 Merits and Demerits of Blockchain Platforms in Healthcare

7.1 Merits

7.1.1 Hyperledger Fabric

For a wide variety of industries, Hyperledger Fabric facilitates distributed ledger solutions on permissioned networks. Its modular architecture maximizes blockchain solutions' security, resilience, and flexibility.

Permissioned Membership

Hyperledger Fabric is a permissioned network architecture with known identities for all members.

Performance, Scalability and Level of Trust

The Hyperledger Fabric platform is based on a modular architecture that divides transaction processing into three stages: distributed logic processing and agreement ("chain code"), transaction arranging, and transaction authentication and commitment. There are several benefits of this separation: it requires less levels of confidence and verification between node types, and it improves network scalability and efficiency.

Data on a need—to—Know Basis

Businesses need privacy of some data elements due to competition, data security rules, and regulations on personal data confidentiality, which can be accomplished by data partitioning on the blockchain. Channels, which are powered by Hyperledger Fabric, allow data to be shared with only those who need to know.

Rich Queries Over an Immutable Distributed Ledger

For the blockchain application, the ledger is a sequenced archive of state transitions. Each transaction generates a collection of asset key–value pairs that are generated, modified, or deleted in the ledger. The peer's file system, which also has LevelDB embedded, is appended with the timeless source of truth for v1.0.

Modular Architecture Supporting Plug-In Components

The Hyperledger Fabric architecture's modularity allows network designers to plug in their desired component implementations, which is a benefit. "Bring your own name" is one of the most requested fields for modularity. Some multi-company networks also have identity management in place and choose to reuse rather than recreate it. Consensus or encryption, where certain countries have their own encryption protocols, are other elements of the architecture that can be conveniently plugged in.

Protection of Digital Keys and Sensitive Data

Support for HSM (Hardware Security Module) is important for securing and managing digital keys in order to provide effective authentication. For key generation, Hyperledger Fabric supports both updated and unmodified PKCS11, which helps with cases like identity management that need more protection. HSM improves the security of keys and confidential data in situations involving identity management.

7.1.2 Ethereum

Ethereum provides interoperability with technologies that have the potential to transform the healthcare industry. Among other blockchain applications, it allows data security, approval control, medical tool, and product tracking.

Decentralization

After Bitcoin, Ethereum is the world's second most decentralized cryptocurrency. There is no centralized body in charge of anything. This distinguishes it from other smart contract systems such as NEO and Tron. The ultimate defence against intrusion is to use a decentralized network like Ethereum.

A Robust Developer Community

Ethereum has the world's biggest developer group, far bigger than Bitcoin. As a result, Ethereum has a significant advantage over other protocols. Blockchain-based cryptocurrencies are still a relatively young technology, and much work remains to be done in order to make crypto useful to the average citizen. Ethereum will be the first project to achieve product-market fit, paving the way for widespread adoption.

Interoperability

When we create an app on Ethereum, we can link it to hundreds of other protocols right away. This is referred to as money legos in the Ethereum culture.

7.2 Demerits

7.2.1 Hyperledger Fabric

1. It lacks highly qualified programmers.
2. It has a dearth of use cases.
3. It has a complex structure.
4. It comes with the bare minimum of APIs and SDKs.
5. It is not a fault-tolerant network at all.

7.2.2 Ethereum

Ethereum is a fantastic forum, but it is far from flawless. Here are some major issues that Ethereum is currently dealing with.

Slow Speeds

Decentralized protocols, such as Bitcoin and Ethereum, are notoriously sluggish. Bitcoin has an average transaction speed of 7 TPS, while Ethereum has a speed of 15 TPS. That is twice the speed of Bitcoin, but it is not nearly fast enough.

The Programming Language

Solidity is a programming language used by Ethereum developers to build apps and tokens. This is a brand-new programming language with a number of well-known flaws. In practice, this means that developers may have to write smart contracts in a new language that they may not be familiar with. As a consequence, it is relatively normal for smart contracts to be written with security flaws.

Unstable

The Ethereum blockchain is still undergoing a number of changes, including switching from a PoW to a PoS consensus process.

Dependability

ERC20 tokens, like every other token, are based on Ethereum and built on top of it. This poses a danger, given that Ethereum is constantly changing.

Hard Forks

Since you are reliant on another blockchain, there is a chance you'll run into problems due to hard forks.

Sovereignty

Since the tokens are built on the Ethereum blockchain, there is no way to influence their future growth. The program and use case should determine whether or not ERC20 should be used.

8 Hyperledger Verses Ethereum: Key Difference

8.1 Purpose

The main difference among Hyperledger and Ethereum is how they are supposed to be used. Since smart contracts are decentralized and open to the public, Ethereum can run them on the EVM for applications.

Hyperledger, on the other hand, uses Blockchain for industry. It can also promote the use of pluggable components to improve security, robustness, and flexibility. Hyperledger is also built on a standardized system giving users a lot of flexibility about how they utilize it.

8.2 Mode of Accessibility

You do not need permission to use Ethereum because it is a public Blockchain network. Anyone can install the Ethereum platform in order to participate in Ether extraction as well as explore and monitor the network's transactions.

Hyperledger maintains strong access control. The platform and its resources are only accessible to registered members. Furthermore, before joining the Hyperledger network, each participant must obtain permission.

8.3 Confidentiality

Ethereum, as previously said, is a decentralized system which does not need any permissions in order to run. All activities are recorded on the Blockchain network and are available and open to all peers, making it clear and transparent.

Hyperledger, on the other hand, is a permissioned blockchain network that is extremely stable. They can only be seen by people who have been granted access to a network transaction. You must first obtain permission to access a specific service on the Hyperledger system.

8.4 Programming Language

The programming language is another significant distinction between Hyperledger and Ethereum. Smart contracts, which are composed in steadiness, a strong contract-oriented language, are used by Ethereum.

Smart contracts generated by Hyperledger are referred to as "chain code." The chain code can also be used to execute business rules that the network participants

have settled on. After that, it is known as a smart contract. These chain codes are developed in Golang, a Google-developed computer language.

8.5 Consensus Mechanism

Users can choose from a number of consensus protocol in the Hyperledger Fabric venture, including the Kafka consensus algorithm. It also provides protocols for Solo and Raft. Raft is crash-tolerant, whereas the Solo parameters are suitable for programmers.

In the Ethereum consensus scheme, an improved version of the proof of work method is used. Even though the Ethereum consensus method is extremely powerful and reliable, it consumes a significant number of resources and can cause delays. On the other hand, the Ethereum consensus method PoW is considered to be better to the Raft protocol

8.6 Cryptocurrency

Ether is the native or built-in token of Ethereum. Each participant can mine for ether by paying gas.

Hyperledger does not have its own cryptocurrency, and mining for it is not needed. This improves the network's usability and enables it to handle large interaction rates allowing it to automate business processes all over the network.

Figure 3 Contrast of two Blockchain technologies (Ethereum and Hyperledger Fabric).

9 Technical Requirements

When it comes to selecting a technology stack for designing healthcare applications, there are several considerations to consider. It is vital to evaluate the functional specifications for health information technology systems when contrasting the various blockchain systems used in building new applications. The conditions for various healthcare use cases vary, but there are certain issues that are universal. These problems are briefly discussed in this section, and they are used to defining and compare the most widely used blockchain systems for medical record management in the following section. The top priority in developing a healthcare application is the patient safety and their data on health conditions. According to the existing blockchain model, not all transactions in health care should be entirely available to all. There are constitutional and regulatory standards that must be met

	Ethereum	Hyperledger Fabric
Description of Platform	Modular blockchain platform	Generic blockchain platform
Governance	Ethereum developers	Linux Foundation
Privacy features	Private transactions, experimental zero-knowledge proofs	Private channels, private transactions, zero-knowledge proofs
Security	Public, but supports permissioned networks	Permissioned, granular access control, less prone to attacks
Consensus	Compute-intensive PoW, PoS	Multiple approaches
Speed	15 transactions per second	3000 transactions per second
Scalability	Low transaction throughput	Higher transaction throughput
Transaction cost	High	Low
Incentive	Cryptocurrency required	No cryptocurrency required
Smart contracts	Solidity	Java, Node.js, Go, Higher-level abstraction (composer, JavaScript)
Currency	-Ether -Tokens via smart contact	-None -Currency and tokens via chain code

Fig. 3 The contrast of two blockchain technologies

when processing healthcare data. The European General Data Protection Regulation (GDPR) sets stringent guidelines for the processing of patient data in order to protect patient's privacy. As a result, any blockchain architecture that is used to build systems and applications should have a comprehensive collection of confidentiality features. To avoid data theft in any form, HIT systems must be designed and

installed. Any entity in the application should be identifiable, as well as its activity. The legislation mandates that security requirements for healthcare apps, including anonymity, be met. Know your consumer (KYC) and anti-money laundering (AML) strategies, for example, necessitate the knowledge of real-world identities through business systems. To monitor how users communicate with the service and the activities connected with it, strong authorization capabilities and detailed access control mechanisms are required. Transaction efficiency is another factor to consider when selecting a user interface for building health systems. Healthcare technologies must be scalable in terms of consistency and transaction performance in some cases, such as remote patient monitoring (RPM) systems. The total number of transmissions involved in the agreement process determines the transaction throughput or scalability of blockchain frameworks. How transactions on the blockchain Web are authenticated until they can be recognized as legal contracts is consensus is concerned. As a consequence, the various Blockchain frameworks' methods for achieving consensus are essential considerations. Another critical aspect to consider when selecting technology platforms for designing healthcare applications is operational price. Medical solutions must be set up on networks with consistent and transparent transaction prices, if not no costs in any respect. Finally, the blockchain platform that would be used to build healthcare applications must be able to facilitate smart contract execution. Smart contracts, as previously mentioned, are not a core feature of blockchain, but their use will aid in the coding and reconstruction of the set of rules that control relations between the various roles in the healthcare sector.

10 Limitation

The study's main drawback is that we only compared two platforms, so the scope may be restricted. However, based on our findings, an informatics researcher, IT specialist, or a technical head in healthcare or other institutions may access various functional aspects of the platforms, such as setup/learning time and technical features. The correct platform is chosen based on the application's specifications. For example, Ethereum maintenance and Hyperledger Fabric fine-grained access control.

11 Conclusion and Future Work

This work emphasizes the advantages of blockchain in healthcare and which platform is better in managing healthcare information, whilst additionally acknowledging that these advantages have been slow to arrive due to Blockchain's early stage of growth and dependence on cryptocurrency-based solutions.

According to a review of the literature on blockchain-based healthcare applications, two major blockchain systems are widely used in the development of these applications. The identified frameworks were compared to the needs of healthcare

applications with the intention to aid health informatics researchers, clinicians in deciding on the best network for building medical applications.

Both Hyperledger and Ethereum, according to our research, are extremely versatile, but their strengths lie in different areas. Ethereum is a generic framework for any application thanks to its strong smart contracts system. However, it is clear and does not need permission which has an impact on the performance, scalability, and safety.

Hyperledger, on the other hand, uses a permissioned way of operation and authentication to address efficiency usability and security concerns. We can also tailor Hyperledger to a variety of applications, almost like a toolbox.

Hyperledger Fabric contains more layers in its network to increases manageability and security, which can result in long setup time compared to Ethereum. The installation of the software Ethereum are also easier comparing to Hyperledger Fabric.

Considering the healthcare applications, the characteristics of the two platforms are follows

Ethereum is supported and maintained by a huge group of developers from all over the world. Because of this openness, Ethereum is a good option when it comes to the long-term viability of the network on which the applications are built. While Ethereum does not support private networks, it can still send and receive personal and private data. Private data may be stored off-chain in these situations, with the best guidance to the facts being registered on a public network. As a result, only the data's pointers are accessible to network members, while the real private data is hidden from all but approved users.

Hyperledger Fabric is well-designed, multi-layered access control framework along with Hyperledger Fabric's own certificates leads to high versatility, security, and manageability of the Blockchain platform. It is mostly used for everything on the blockchain network, including the exchange of monetary value. For order in minimal specifications, a chain code is executed on proportion-based transactions. It improves the trustworthiness of the connection. It also enhances network scalability and consistency. It is a public ledger that is encoded. It has SQL-like querying capabilities. It is simpler to appeal and check the decision. At the enterprise stage, it is extremely adaptable and scalable. It is yet another reputable blockchain platform. On any blockchain network, it could share a variety of properties. It offers a high level of confidentiality. It engages in a variety of multilateral transactions.

Hyperledger Fabric is a permissioned blockchain technology framework that has been actively employed in the implementations of blockchain-based systems for healthcare data management. To ensure privacy of data subjects, Fabric mainly relies (i) on multiple channels support, which make it possible to limit the access to the data to certain participants of the consortia, and (ii) on private collections where sensitive data can be exchanged peer-to-peer and stored in the private databases, yet accessible from chaincode on authorized peers and hashed to verify authenticity. Storing only hash on-chain is also used to provide verifiability of vast amounts of anonymized data for data-driven research and applications.

Therefore, **Hyperledger Fabric—this Blockchain architecture has the most potential for medical application growth and healthcare record management**. Because in healthcare sector data privacy and security is most important. With the help of Hyperledger we can easily maintained our data also.

In the future, the Hyper Fabric will be used to construct an application for a specific healthcare use case, with the output being tested to better understand and illustrate the Hyper Fabric's scalability, latency, and throughput. In the future, data sharing through various blockchain networks will be a fascinating research subject.

References

1. H. Yu, H. Sun, D. Wu, T.T. Kuo, Comparison of smart contract blockchains for healthcare applications, in *AMIA Annual Symposium Processing 2020* (2019), pp. 1266–1275. Published 2020 Mar 4
2. T.-T. Kuo, H.Z. Rojas, L. Ohno-Machado, Comparison of blockchain platforms: a systematic review and healthcare examples. J. Am. Med. Inf. Asso. **26**(5), 462–478 (2019)
3. V. Patel, A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. Health Inf. J. 1398–1411 (2019)
4. T.K. Mackey, T.T. Kuo, B. Gummadi et al., 'Fit-for-purpose?'—challenges and opportunities for applications of blockchain technology in the future of healthcare. BMC Med **17**, 68 (2019)
5. C.C. Agbo, Q. Mahmoud, J. Eklund, Blockchain technology in healthcare: a systematic review. Healthcare (2019)
6. S. Khezr, M. Moniruzzaman, A. Yassine, R. Benlamri, Blockchain technology in healthcare: a comprehensive review and directions for future research. Appl. Sci.
7. T.-T. Kuo, R.A. Gabriel, L. Ohno-Machado, Fair compute loads enabled by blockchain: sharing models by alternating client and server roles. J. Am. Med. Inf. Assoc. **26**(5), 392–403 (2019). https://doi.org/10.1093/jamia/ocy180. PubMed PMID: 30892656
8. R. Ribitzky, S.J. Clair, D.I. Houlding, C.T. McFarlane, B. Ahier, M. Gould, H.L. Flannery, E. Pupo, K.A. Clauson, Pragmatic, interdisciplinary perspectives on blockchain and distributed ledger technology: paving the future for healthcare. Blockchain Healthc. Today
9. P. Mamoshina, L. Ojomoko, Y. Yanovich et al., Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare. Oncotarget **9**(5), 5665–5690 (2018)
10. P. Zhang, J. White, D.C. Schmidt, G. Lenz, S.T. Rosenbloom, Fhirchain: applying blockchain to securely and scalably share clinical data. Comput. Struct. Biotechnol. J. **16**, 267–278 (2018)
11. A. Kovach, G. Ronai, MyMEDIS: a new medical data storage and access system (2018)
12. O. Choudhury, H. Sarker, N. Rudolph, M. Foreman, N. Fay, M. Dhuliawala, I. Sylla, N. Fairoza, A.K. Das, Enforcing human subject regulations using blockchain and smart contracts. Blockchain Healthc. Today (2018)
13. X. Liang, J. Zhao, S. Shetty, J. Liu, D. Li, Personal, indoor, and mobile radio communications (PIMRC), in *IEEE 28th Annual International Symposium on: IEEE; 2017. Integrating blockchain for data sharing and collaboration in mobile healthcare applications,* vol. 2017, pp. p. 1–5
14. X. Yue, H. Wang, D. Jin, M. Li, W. Jiang, Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. J. Med. Syst. **40**(10), 218 (2016)
15. K.A. Clauson, E.A. Breeden, C. Davidson, T.K. Mackey, Leveraging blockchain technology to enhance supply chain management in healthcare. Blockchain Healthc. Today (2018)
16. P. Mamoshina, L. Ojomoko, Y. Yanovich et al., Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare. Oncotarget **9**(5), 5665–5690 (2018)

17. R. Ribitzky, J.S. Clair, D.I. Houlding, et al. Pragmatic, interdisciplinary perspectives on blockchain and distributed ledger technology: paving the future for healthcare. Blockchain Healthc. Today (2018)
18. T.-T. Kuo, H.-E. Kim, L. Ohno-Machado, Blockchain distributed ledger technologies for biomedical and health care applications. J. Am. Med. Inform. Assoc. **24**(6), 1211–20 (2017)
19. T.-T. Kuo, H.-E. Kim, L. Ohno-Machado, Blockchain distributed ledger technologies for biomedical and health care applications. J. Am. Med. Inform. Assoc. **24**(6), 1211–1220 (2017)
20. J.M. Roman-Belmonte, H. De la Corte-Rodriguez, E.C. Rodriguez-Merchan, How blockchain technology can change medicine. Postgrad Med. **130**(4), 420–427 (2018). https://doi.org/10.1080/00325481.2018.1472996 (Epub 2018 May 10 PMID: 29727247)
21. M. Hölbl, M. Kompara, A. Kamišalić, L. Nemec Zlatolas, A systematic review of the use of blockchain in healthcare. Symmetry **10**, 470 (2018)

Internet of Things-Based Devices/Robots in Agriculture 4.0

Gulbir Singh and Kuldeep Kumar Yogi

Abstract Agriculture 4.0 focuses majorly on precision agriculture. Precision agriculture can be achieved in several ways such as refinement of cultivation practices, choices of crops, reduction of risk and volatility, water management, optimized use of pesticides, land/crop monitoring with minimal environmental impact. The best way to achieve precision agriculture through the Internet of Things-based devices in agriculture. The rapid developments on the Internet of Things-based devices have impacted every industry including "Agriculture." This revolutionary change in agriculture is changing the present agricultural methods, and creating new opportunities, and challenges. The Internet of Things-based devices and communication techniques along with wireless sensors are analyzed in this chapter in detail. The specific sensors available for precision agricultural applications like the preparation of soil, checking the status of the crop, pest, and insect identification, and detection, irrigation, spraying of fertilizers are explained. The use of Internet of Things-based devices helps the farmers through the crop stages i.e., sowing to harvesting is explained. At last, this chapter concludes and provides the challenges faced while implementing Internet of Things-based devices in agriculture.

Keywords Internet of Things · Internet of Things-based devices · Sensors · Communication technology · Agriculture 4.0 · UAVs

1 Introduction

The economy of the world majorly depends upon the agriculture. The world's population depends upon agriculture for its survival, and agriculture is the major source to fulfill human needs. Agriculture has a major role in the economic development of countries. The major roles are listed below:

G. Singh (✉) · K. K. Yogi
Department of Computer Science, Banasthali Vidyapith, Vanasthali, Rajasthan, India

K. K. Yogi
e-mail: ykuldeep@banasthali.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_6

- Agriculture contributes to the national income of a nation.
- It is the major source of the food supply chain to the humans and the industry.
- It is the major source of raw materials for different industries.
- It is very helpful to reduce inequality.
- It is a source of foreign exchange for the nations.
- It contributes to the capital formation.

Precision agriculture is an approach to achieve smart farming. Smart farming refers to a farm management technique that makes use of advanced technologies intending to increase the quantity and quality of agricultural products. This technique includes the usage of Internet of Things-based devices, data management, soil and crop preparation, crop sowing, irrigation, pest/yield monitoring, and last cultivation.

Over the past years, the smart farming has become more useful to farmers as it grants access to farmers to modern technologies and devices that help in the increase in the agricultural product's quantity and quality, which this technique helps in reduction of farming cost.

The benefits of smart farming in agriculture include the high rate of crop production, the decrease in the use of fertilizers, pesticides, and water, reduce environmental pollution, and the increase in the safety of farms and farmers.

Recent researches suggested a vast range of smart devices to fetch and monitor the information related to crop and field status. Many manufactures now providing a vast range of the Internet of Things-based sensors, communicating devices, agricultural robots, unmanned aerial vehicles (UAVs), and other heavy agricultural machinery.

The Internet of Things-based technologies are used in agriculture to retrieve information about the crop field through a variety of sensors for monitoring, trespassing, and controlling the field area [1]. The Internet of Things-based technologies with the use of remote sensing devices provide a way to sense an object in a controlled environment. It helps to reduce the gap between the computer-oriented system and the real world. The main advantage of introducing Internet of Things-based devices in agriculture is to maximize accuracy and efficiency with a low human intervention.

According to a report published by Business Insider Intelligence, the usage of Internet of Things-based devices which are used in smart agriculture is continuously increasing day by day. The following Fig. 1 shows the rapid growth of the Internet of

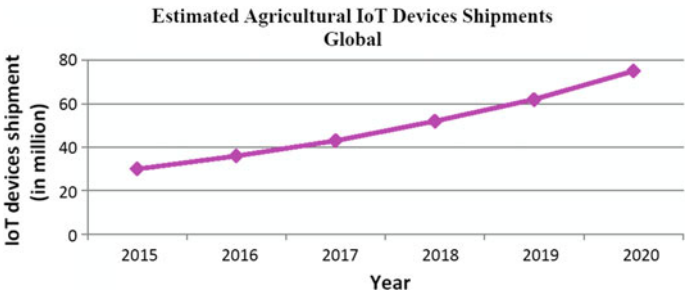


Fig. 1 Growth in the usage of Internet of Things-based devices from 2015–2020

Things-based devices from 2015–2020. The graph shows that the growth is enhanced up to 150% [2].

This article provides knowledge that can help different engineers and researchers for the implementation of Internet of Things-based technologies to obtain desired smart agriculture. This article also provides an overview of the major applications of the Internet of Things in the agriculture field. The use of different Internet of Things-based devices/sensors is also discussed followed by a conclusion and challenges arise in implementing the Internet of Things in the agriculture.

2 Literature Review

Many research centers and software as well as software industries focus on the impact and usage of Internet of Things-based technologies and suggested major concerns in this field. The section focuses on the past work done by authors/researchers that can use the Internet of Things-based devices in agriculture.

In Fig. 2, the usage and application of Internet of Things-based devices in agriculture are shown, and Table 1 shows the research work done by different researchers in agriculture using Internet of Things-based technologies.

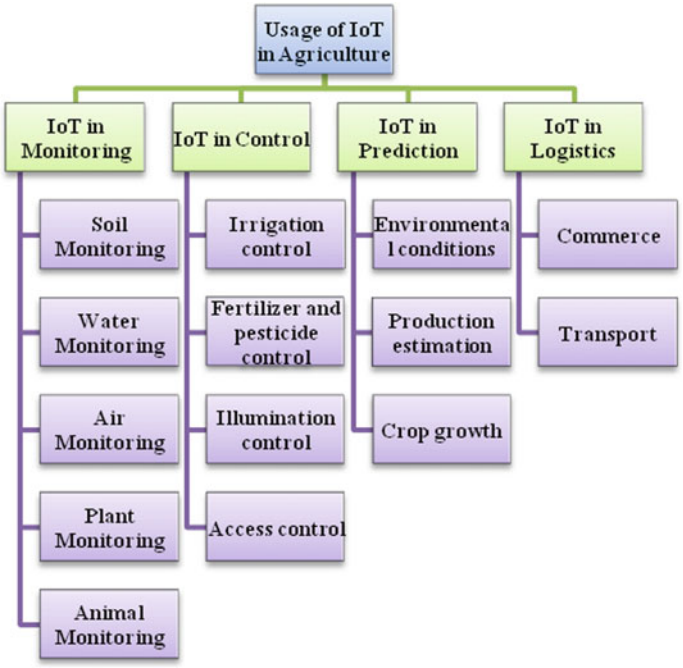


Fig. 2 Usage of Internet of Things-based devices in agriculture

Table 1 Past research which was done by the researcher using Internet of Things-based technologies in agriculture

Authors and Year	Subdomain	Observations
Cheng-Jun [3]	Monitoring of field	The author proposed a system for crop monitoring, and using an online tool, it captures multiple variables like soil moisture, pH levels of water, humidity, temperature, etc. The proposed system consists of three levels of architecture first, the perception layer using WSN support; second, the network layer for data communication, and the last layer as an application layer Web server to analyze the data
Mafuta et al. [4]	Monitoring of soil	The authors have suggested an Internet of Things-based system with WSN to monitor the soil temperature, and moisture. The proposed system uses communication technology like ZigBee, GPRS, and the Internet
Singh et al. [5]	Air pollution monitoring	The authors have proposed a system with the help of Adriano and a variety of gas sensors. The authors have developed a Web application to store and process the processed data
Singh and Yogi [6]	Crop disease monitoring	The authors have a critical review on crop/plant disease detection using Internet of Things-based devices. They also focus on the use of machine learning and Artificial Intelligence for plant disease detection
Jain et al. [7]	Animal monitoring	The authors have proposed a system for animal monitoring, where the Internet of Things-based system is responsible for monitoring the Swamp Deers. The proposed system collects the data of animals' behavior and climate at the same time
Shuwen and Changli [8]	Irrigation control and monitoring	The authors have suggested a solar power-based irrigation monitoring system that uses the ZigBee protocol for data communication

(continued)

Table 1 (continued)

Authors and Year	Subdomain	Observations
Pahuja et al. [9]	Pesticides and fertilizers monitoring	The authors have proposed a microclimate monitoring system for the greenhouse. They have used the WSN system to collect and process the data
Cozzolino et al. [10]	An infrared sensor in the field	The authors have done a critical review of the usage of infrared sensors in the agriculture field. They have used infrared sensors to monitor the harvesting of cereal crops
Bhatnagar et al. [11]	Internet of Things devices in agriculture	The authors have analyzed the usage of Internet of Things-based devices in the agriculture field. They have suggested the challenges faced in the implementation of these devices in agriculture
Unold et al. [12]	Internet of Things-based health monitoring system for COW	The authors have proposed a framework composed of hardware, cloud system, and user interface. The proposed system was tested in real-time, and has proved more effective for monitoring animals
Gadre and Deoskar [13]	Challenges, transformation, and benefits of Industry 4.0	The paper covers the strategic role that can help manufacturers as a guide for industry 4.0
Nejkovic et al. [14]	A semantic approach for RIoT autonomous robots	The authors have proposed a semantic approach for an autonomous robot of RIoT
Singh and Gupta [15]	Secured and optimized ad-hoc on-demand distance vector protocol for secured and optimized communication for disaster-response applications	The authors proposed a new secured and optimized ad-hoc on-demand distance vector protocol for secured and optimized communication for disaster response applications. With the help of IoT-based devices, the proposed protocol is reflected the better outcome in disaster response and prevention applications like climate and weather observation for agriculture

(continued)

Table 1 (continued)

Authors and Year	Subdomain	Observations
Manocha and Gupta [16]	Satellite image enhancement technique	The authors suggested a new satellite image enhancement framework (SIE-EVD) to reduce the blur or noise of an image of agricultural land captured by IoT-based devices without losing high-frequency details of the image

3 Application of Internet of Things-Based Devices in Agriculture

The traditional framing can be changed or transformed to modern/smart/precise farming by the implementation of Internet of Things-based devices, latest sensors. The solution to the traditional farming issues can be solved by following the practice of smart farming. Figure 3 lists all the major applications of Internet of Things-based devices in smart agriculture.

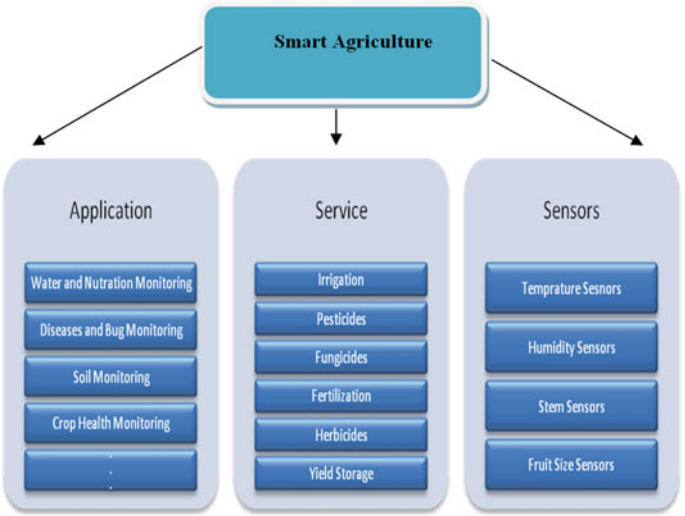


Fig. 3 Major application of Internet of Things-based technologies in smart agriculture

3.1 Soil Monitoring

Soil is the major building block for plants as soil provides all the nutrients and water supply to the plant that helps in plant growth. The first step of soil monitoring will be done through soil sampling, as the sampling process examines the field-specific details for farming. The major objective of soil analysis is to provide information on the nutrients status, so that the corresponding measures or actions can be taken when the crop faces issues related to nutrients deficiency.

The main factors that need to analyze in the soil are soil type, fertilizers/pesticides application, cropping history, topography and irrigation level, etc. This factor gives detailed information related to the physical, chemical, and biological status of the soil. Soil mapping helps in solving many issues related to sowing different varieties of the crop in specific farmland.

The Internet of Things-based technologies like sensors and vision-based devices can be used to identify the depth and distance for sowing the seeds. To perform the sowing activity in farmland, the robots can be taken into consideration, as they can be implemented at every location of the farm and can use global and local map generated from the vision-based and GPS-based system is connected to a computer.

3.2 Farm Irrigation

Irrigation can be done by various methods like sprinkler irrigation, drip irrigation and can be recommended to solve the issues of water wastage issues in farmlands. While in traditional farming, the irrigation methods such as furrow irrigation and flood irrigation were used that can badly affect the quality and quantity of the crop. It would be a tough job to determine the water requirement of field crops where other factors are involved such as soil moisture, crop irrigation methods, and crop type.

The current condition of irrigation can be enhanced by embracing the Internet of Things-based technology. Rapid growth is expected in crop production and efficiency with the help of the Internet of Things-based technologies like CWIS-based irrigation management systems. The CWIS is a wireless sensor-based system for monitoring irrigation where other field sensors are used to collect the field data from the farm.

3.3 Fertilizers Spraying

Fertilizers provide nutrients to plants for their fertility and growth. The plants need three major micronutrients nitrogen for the growth of the leaf, phosphorus for the flowers, fruits, and root development, and potassium for the water movement and stem development [17]. The shortage of the nutrients and applying them wrongly

may affect the harmfully of the plants. The use of fertilizers in excessive amounts may harm financial losses but also harm the crop, soil, and the environment.

The fertilization is done with the help of Internet of Things-based technologies in smart agriculture to help inaccurate estimation of the required amount of nutrients, and this also minimizes the negative effects on the crop and the environment. The Internet of Things-based fertilization technique helps in identifying the nutrients requirements by crop with high accuracy and minimum labor requirement. In smart farming, Internet of Things-based technologies like geo-mapping, GPS accuracy, VRT technology, and autonomous vehicles are used for the fertilization process.

3.4 Pest and Crop Disease Management

According to a report generated from FAO (Food and Agriculture Organization) estimated that 20 to 40% of global crops are lost annually due to disease and pests. To control a huge loss, the use of agrochemicals and pesticides was taken into consideration to reduce the loss in past years.

The use of Internet of Things-based devices like wireless sensors, drones, and robots helps the farmers to reduce the use of pesticides in fields or only used when they are required with the precise amount. These pesticides are very harmful to humans, environment and even sprayed in huge amounts also harmful to crops. The advanced Internet of Things-based technologies help in pest/disease management in real-time, provide live monitoring, forecasting of pests and diseases, and hence are proved more efficient [18]. The advanced pest and disease detection techniques depend upon three levels such as sensing, evaluation, and treatment. The image processing technique is a method where the real-time pictures have been captured from the farm and using advanced tools and techniques the pest and disease forecasting can be done.

Techniques like VRT chemigation [19] and vehicle precise spray are used in smart farming and also can be utilized for disease and pest treatment.

3.5 Harvesting, Forecasting, and Yield Monitoring

To monitor agriculture yield, the yield monitoring mechanism can be used to analyze the different aspects related to agriculture like harvested grain quality, moisture content, and grain mass flow.

Crop forecasting is the technique used for the prediction of yield and production before the harvesting of the crop. This forecasting helps the farmers to plan the near future and decisions making. The yield monitoring contains several development stages, and it uses the fruit conditions such as color, size, etc.

4 Equipment and Technologies Used in Smart Agriculture

The major tasks in modern agriculture can be done by the use of large-scale heavy and urban tools like harvesters, robots, tractors, etc., which fully or partially supports remote sensing, and other related communication technologies. In smart agriculture, the task is majorly performed by vehicles equipped with GIS and GPS technologies, so that they can perform precisely, independently, and accurately. Figure 4 shows the major applications of Internet of Things-based devices in smart agriculture.

4.1 Wireless Sensors

Wireless sensors play a major role in collecting information regarding crop conditions and their related data. The wireless sensors can work standalone when required and also integrated into every advanced tool of agriculture. In the following, the major sensors type have been discussed according to their working purpose and procedure.

Fig. 4 Major technologies used for smart agriculture

4.1.1 Optoelectric Sensors

Optoelectric sensors are generally used to differentiate plants, and they can be used for the detection of herbicides, weeds, and other unwanted plants in crops flew [20]. The optoelectric sensors are combined with location information, and it helps in mapping weeds and resolution. Optoelectric sensors use reflection spectra to differentiate the vegetation and soil.

4.1.2 Airflow Sensors

The airflow sensors can be used for measuring the moisture percentage and soil air permeability. They are also used to detect the structure of the soil to identify the different types of soil. The measurements are done at singular or dynamic locations while in motion, fixed location, or in the mobile node. It pushes the desired quantity of air required by the ground at a predefined depth in the soil. It can identify the different soil properties like soil compaction, moisture level, soil structure, etc.

4.1.3 Electromagnetic Sensors

These sensors can be used to find the electrical conductivity, electrical response, and electromagnetic responses in the actual situation. Electromagnetic sensors use electric circuits to evaluate the capacity of soil particles to accumulate or conduct electric charge that is carried by the following techniques: contact or non-contact. The electromagnetic sensors are also capable to measure the nitrates and organic matter in the soil [21].

4.1.4 Optical Sensors

The light reflection phenomena are used by optical sensors to determine the soil organic substances, soil color, soil moisture, and minerals in the soil. The optical sensors reflect light on different portions of soil for testing the soil. These sensors are also used for crop assessment, especially to monitor fruit maturation. The optical sensors are combined with microwave scattering, then they can help distinguish grove canopies such as olives and similar crops [22].

4.1.5 Acoustic Sensors

Acoustic sensors are used for various in the agriculture applications like soil cultivation, weeding, fruit harvesting, etc. The main benefit of these sensors is that they are low cost and have a fast response when considering mobile equipment. These

sensors compute the changes in noise as they connected with other materials, e.g., soil particles [23].

4.1.6 Electrochemical Sensors

These sensors have a major role in agriculture as they help in measuring soil features and identify the nutrient levels like pH levels. Normal soil chemical examination is a very costly and time-taking process and can be easily exchanged with electrochemical sensors.

4.1.7 Mechanical Sensors

The mechanical sensors are used to measure the soil compaction (resistance). The mechanical sensor can be put into the soil to get the information of the force assessed using a strain gauge. The pressure unit is used to identify the soil's mechanical compaction level.

4.1.8 Mass Flow Sensors

The mass flow sensors are used to monitor yield as they provide the yield information to compute the quantity of grain flow. The yield monitoring system contains server modules such as grain moisture sensors, hardware for data storage, and dedicated software to analyze data collected from the agriculture field.

4.2 Internet of Things-Based Tractors

Due to the growth in the agriculture industry, the rural labor resources have come under heavy stress and pressure as a result of which the tractors and other heavy machinery start to enter agriculture to provide efficient and progressive efforts. To fulfill the rising demand of farmers, the major agriculture-based industries have started to provide better solutions to farmers' requirements. The self-driving tractors are now available in the market and provide the ability to reduce the revisiting of the same row/area of field by reducing the overlap. These Internet of Things-based tractors provide a better precision along with the reduced error, mainly when spraying pesticides and other related tasks which are unavoidable when a human operator the machine.

This advanced machinery, most farmers are unable to afford due to their heavy cost. To overcome this issue, HELLO TRACTORS has designed a cost-efficient monitoring device that can be mounted to any ordinary tractor and that device

Fig. 5 Internet of Things-based tractor in agriculture field

provides software and analytical tools. Figure 5 shows an Internet of Things-based tractor in the agriculture field.

4.3 Harvesting Robots

The harvesting robots provide a major role in the harvesting of crops. The harvesting of some crops can be done a single time, or in some crops, it is performed several times. Doing harvesting early or late may affect the production of the crop, so it is very critical to perform harvesting of the crop at the right time. To automate to harvesting process, the role of robots raised for achieving precision harvesting. Several scientists and researchers have done their research to increase the capability of fruit detection by its size, shape, color, and localization [24, 25]. The automated harvesting of crops/fruits requires the use of dedicated sensors that are efficient for unambiguous, and precise details of that particular fruit. This process requires very sophisticated and specialized tools to identify the fruit's conditions. Considering this issue, many robots were used for specific crops. Few of the major robots which are used for crop harvestings such as Octinion, SW 6010, and FFRobot. Figure 6 shows different harvesting robots used in agriculture.

Fig. 6 Harvesting robots

4.4 Unmanned Aerial Vehicles Used in Agriculture

The Internet of Things-based devices has developed to higher levels and is used in several industries. In the agricultural field, the communication between the end devices has their own restrictions due to limited range, power, and bandwidth, and also the modern communication systems are taught to implement in the rural areas. Considering these issues, the remote-controlled aerial vehicles (UAVs) area unit the alternative solution to these issues. The UAVs can communicate across the whole field through the wireless sensors spreaded in the field. The UAVs will simply collect the information from the fields for further processing. The UAVs are referred to as drones, and they are equipped with high-resolution cameras and sensors and might fly through thousands of hectares of fields. Figure 7 shows UAVs.

5 Challenges

- Precision agriculture requires the implementation of new technologies for crop production, for a normal farmer the setting up of Internet of Things architecture and sensor networks for his field could be a tough task. So, the lack of knowledge in implementing these technologies can be dangerous.
- The wireless network connectivity in many remote or rural areas around the world can affect smart farming.
- Lack of configuration and scalability problem.
- Energy depletion risk.

Fig. 7 UAVs

- Technical failure and corollary damages.
- Loss of human employment.
- Security of Internet of Things-based devices from natural disasters like heavy rain, storms, fire, etc.
- Mobility and analysis of cost is a critical challenge for smart agriculture.

6 Conclusion

In this research, we have discussed the work of researchers/engineers and the major application areas of Internet of Things-based devices in smart agriculture. This article also discussed the various equipment and technologies used in smart agriculture. Finally, this research highlights various challenges that arise while implementing Internet of Things-based devices in agriculture. This study is very useful and helpful for researchers, agriculturists, and, professionals working in the field of modern agriculture. The usage of the Internet of Things is useful in the advancement of farming and agriculture by introducing new concepts like live tracking, pest control, irrigation control, soil investigation, etc. This paper a critical review of Internet of Things-based devices deployed in the agricultural fields. In the introduction section, we have discussed the growth of the Internet of Things-based devices in agriculture as the graph in Fig. 1 shows it is increased up to 150%. In the next section, we have provided the applications of Internet of Things-based devices in agriculture like IoT in monitoring, IoT in control, IoT in prediction, etc. In the next section, we have discussed the equipments and technologies used in agriculture like various wireless sensors used, IoT-based tractors, harvesting robots, and unmanned aerial vehicles. Then, we have discussed the challenges faced while implementing IoT-based devices in agriculture.

Acknowledgements The authors are very thankful to the Department of Computer Science and Engineering, Banasthali Vidyapith, Rajasthan, India, for their motivation, and support to complete this research.

References

1. N. Sales, O. Remédios, A. Arsenio, Wireless sensor and actuator system for smart irrigation on the cloud, in *2015 IEEE 2nd World Forum on Internet of Things (WF-Internet of Things)* (IEEE, Milan, Italy, 2015), pp. 693–698
2. N. Putjaika, S. Phusae, A. Chen-Im, P. Phunchongharn, K, Akkarajitsakul, A control system in an intelligent farming by using arduino technology, in *2016 Fifth ICT International Student Project Conference (ICT-ISPC)* (IEEE, Nakhonpathom, Thailand, 2016), pp. 53–56
3. Z. Cheng-Jun, Research and implementation of agricultural environment monitoring based on internet of things, in *2014 Fifth International Conference on Intelligent Systems Design and Engineering Applications* (IEEE, Hunan, China, 2014), pp. 748–752
4. M. Mafuta, M. Zennaro, A. Bagula, G. Ault, H. Gombachika, T. Chadza, Successful deployment of a Wireless Sensor Network for precision agriculture in Malawi, in *2012 IEEE 3rd International Conference on Networked Embedded Systems for Every Application (NESEA)* (IEEE, Liverpool, UK, 2012), pp. 1–7
5. G. Singh, G. Kumar, V. Bhatnagar, A. Srivastava, K. Jyoti, Pollution management through internet of things: a substantial solution for society. Humanities Soc. Sci. Rev. **7**(5), 1231–1237 (2019)
6. G. Singh, K.K. Yogi, a review on recognition of plant disease using intelligent image retrieval techniques. Asian J. Biol. Life Sci. **9**(3), 274–285 (2020)
7. V.R. Jain, R. Bagree, A. Kumar, P. Ranjan, wildCENSE: GPS based animal tracking system, in *2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing* (IEEE, Sydney, NSW, Australia, 2008), pp. 617–622
8. W. Shuwen, Z. Changli, Study on farmland irrigation remote monitoring system based on ZigBee, in *2015 International Conference on Computer and Computational Sciences (ICCCS).* (IEEE, Greater Noida, India, 2015), pp. 193–197
9. R. Pahuja, H.K. Verma, M. Uddin, A wireless sensor network for greenhouse climate control. IEEE Pervasive Comput. **12**(2), 49–58 (2013)
10. D. Cozzolino, K. Porker, M. Laws, An overview on the use of infrared sensors for in field, proximal and at harvest monitoring of cereal crops. Agriculture-Basel **5**, 713–722 (2015)
11. V. Bhatnagar, G. Singh, G. Kumar, R. Gupta, Internet of things in smart agriculture: applications and open challenges. Int. J. Stud. Res. Technol. Manag. **8**(1), 11–17 (2020)
12. O. Unold, et al., IoT-based cow health monitoring system, in V. Krzhizhanovskaya, et al. (Eds.), *Computational Science—ICCS 2020. Lecture Notes in Computer Science*, vol. 12141. (Springer, Cham, 2020)
13. M. Gadre, A. Deoskar, A.: Industry 4.0—digital transformation, challenges and benefits. Int. J. Future Gen. Commun. Networking **13**(2), 139–149 (2020)
14. V. Nejkovic, N. Petrovic, M. Tosic, N. Milosevic, Semantic approach to RIoT autonomous robots mission coordination. Robot. Auton. Syst. **126**, 103438 (2020)
15. K. Singh, R. Gupta, SO-AODV: a secure and optimized Ad-Hoc on-demand distance vector routing protocol over AODV with quality metrics for disaster response applications. J. Inf. Technol. Res. **14**(3), 87–103 (2021)
16. N. Manocha, R. Gupta, SIE-EVD: a novel satellite image enhancement technique with quality metrices for effective visual display using CBIR. Int. J. Adv. Trends Comput. Sci. Eng. **9**(1.3), 470–474 (2020)
17. H. Kiiski, H. Dittmar, M. Drach, R. Vosskamp, M.E. Trenkel, R. Gutser, G. Steffens, Fertilizers, 2. types. Ullmann's Encyclopedia of Industrial Chemistry (2009)
18. S. Kim, M. Lee, C. Shin, Internet of things-based strawberry disease prediction system for smart farming. Sensors **18**, 4051 (2018)
19. R. Oberti, M. Marchi, P. Tirelli, A. Calcante, M. Iriti, E. Tona, M. Ho evar, J. Baur, J. Pfaff, C. Schtz, Selective spraying of grapevines for disease control using a modular agricultural robot. Biosyst. Eng. **146**, 203–221 (2016)
20. D. Andújar, Á. Ribeiro, C. Fernández-Quintanilla, J. Dorado, Accuracy and feasibility of optoelectronic sensors for weed mapping in wide row crops. Sensors **11**, 2304–2318 (2011)

21. M.A.M. Yunus, S.C. Mukhopadhyay, Novel planar electromagnetic sensors for detection of nitrates and contamination in natural water sources. IEEE Sens. J. **11**(6), 1440–1447 (2011)
22. I. Molina, C. Morillo, E. García-Meléndez, R. Guadalupe, M.I. Roman, Characterizing olive grove canopies by means of ground-based hemispherical photography and spaceborne RADAR data. Sensors (Basel). **11**(8), 7476–7501 (2011)
23. Q. Kong, H. Chen, Y.L. Mo, G. Song, Real-time monitoring of water content in sandy soil using shear mode piezoceramic transducers and active sensing—A feasibility study. Sensors **17**, 2395 (2017)
24. A. Zujevs, V. Osadcuks, P. Ahrendt, Trends in robotic sensor technologies for fruit harvesting: 2010–2015. Procedia Comput. Sci. **77**, 227–233 (2015)
25. S. Bargoti, J. Underwood, Deep fruit detection in orchards, in *2017 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, Singapore, 2017), pp. 3626–3633

A Study on Transliteration Techniques and Conventional Transliteration Schemes for Indian Languages

Jayashree Nair, Riyaz Ahammed, and Anakha Shaji

Abstract Transliteration is the process of transcribing text scripted in one system of writing into another. It tries to script the same word in a language into another, thereby enabling reader to rebuild the original spelling of the unknown transliterated word. The primary objective of transliteration is to produce a different way of representing texts using a different language or writing system. India, being a multilingual nation, transliterators have played a major role in various modes of communication. Apart from the readability of unknown languages, transliterated texts or words are also used in Computational Linguistics Tools such as Stemmers, Morphological Generators, Dictionaries, etc., for easier processing and handling. Forward transliteration means transliterating the source language to its prescribed targeted language and the conversion of the transliterated word to its original form is known as back-transliteration. The motivation of the proposed work is to provide a forward and backward transliteration on different Indian language and on Roman script. This paper presents a study on the existing transliteration schemes and the various attempts made in building transliterations and back-transliterations for Indian languages from a computational linguistics perspective.

Keywords Forward transliteration · Back transliteration · Phoneme · Grapheme · Devanagari · Harvard-Kyoto · Transliteration · Conventional schemes

J. Nair (✉) · R. Ahammed · A. Shaji
Department of Computer Science and Applications, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, India
e-mail: jayashree@am.amrita.edu

R. Ahammed
e-mail: riyazahammed@am.students.amrita.edu

A. Shaji
e-mail: anakhashaji@am.students.amrita.edu

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_7

1 Introduction

Transliteration means to the process of converting the words of one language with the letters and symbols of another. From a linguistic view, it is an aligning of a writing system into another, syllable by syllables or by words. It conserves the sound or phonetic syllables irrespective of the language. Thus, readers who don't know the source language can read it in their desired language. Transliteration also define rules for dealing with syllables in a source language which do not compare with letters in a destination script or vice versa. The primary objective of transliteration is to provide a different way for understanding or representing text using another script retaining the syllables sounds in the text.

Various techniques that have been invented for transliteration is established on the basis of the languages. Transliteration between Indian languages aid to help people learn one language through another. The common phonetic base makes this easy. But, transliteration between the languages will have to be handled with proper care, for there are can be source syllables which may not be present in the target languages.

Transliteration is used extensively in natural language applications such as information retrieval, machine translation, and computational linguistics tools such as stemmers and morphological analyzers. This study attempts to examine the many transliteration systems and rule-based procedures used in the Indian languages of Sanskrit and Malayalam. The majority of current transliteration systems employ a productive model for transliteration modification and analyses the job of generating a suitable transliteration for a given word. From transcription, you should be familiar with transliteration. The syllables of the source language are combined with the syllables of the target language.

1.1 Indian Languages

Apart from their "mother tongues," Indian are multi-lingual. India houses approximately 1652 "mother tongues" including 103 languages that are foreign according to Census 1961 & Nigam 1972 [19]. Each Indian state has its own official language. India spaces roughly 70,000 new articles published yearly with 600 million readers. [19] India is home to Indo-European, Dravidian, Austroasiatic, and Sino-Tibetan languages. "Assamese", "Bangla", "Dogri", "Gujarati", "Hindi", "Kashmiri", "Konkani", "Maithili", "Marathi", "Nepali", "Oriya", "Punjabi", "Sanskrit", "Sindhi", and "Urdu" are all Indo-Aryan languages; "Kannada", "Malayalam", "Tamil", and "Telugu" are Dravidian languages; and "Manipuri (Meitei)", spoken in Manipur, and "Bodo", spoken in north eastern India, are "Santali" is a Munda language that belongs to the Sino-Tibetan language family's Tibeto-Burman branch[19]. The Austronesian languages are spoken in Southeast Asia, with the exception of the Khasian languages spoken in Meghalaya, north eastern India, and the Nicobarese languages spoken in the Nicobar Islands in the Andaman Sea, just to the north-

west of the Indonesian island of Sumatra-both of which are classified as part of the Mon-Khmer subfamily of Austroasiatic[19].

1.2 Malayalam Language

Malayalam is a language spoken in southern India, particularly in Kerala and the union territories of Lakshadweep and Pondicherry. Malayalam is the official language of Kerala[10]. It is a Dravidian language. It is one of 22 Indian languages, spoken by 2.88% of Indians and 34 million people worldwide [10]. Vattezhuthu was the script used to write Malayalam in the early days. The present Malayalam script is based on the Vattezhuthu script, which has been expanded with Grantha script letters to accommodate Indo-European characters. Loanwords from the Aryans[10]. Malayalam alphabets contains 15 vowels which is known as swarah aksharas, 36 consonants known as vyenjana aksharas and other symbols like chillu aksharas. Three major regional dialects exist in Malayalam, as well as a variety of minor dialects. There are some language variations based on social status, particularly caste. The Malayalam language has developed diglossia, a distinction between formal, literary language and casual speech as a result of these influences. Malayalam is a descendant of either a western variety of Tamil or a Proto-Dravidian branch from which modern Tamil is also descended. The first written evidence of the language is from approximately 830 CE. An early and broad flow of Sanskrit language influenced the Malayalam script.

1.3 Sanskrit Language

The Vedas, written in what is known as Vedic Sanskrit, are the most ancient literature in the Sanskrit language, which is an Old Indo-Aryan language. The earliest manuscripts, including the Rigveda, which academics usually date to approximately 1500 BCE, came from the north-western region of the subcontinent, in the area of the ancient seven rivers [11]. Sanskrit is a classical language of Southern Asia that belongs to the Indo-Aryan branch of the Indo-European languages. In the state of Uttarakhand, Sanskrit is the official language. It is the holy language of ancient India and the language of Hindu mythology and also it is a text with some historical touch for Buddha's and Jain's [11]. It belongs to the family of Devanagari language. Vedic Sanskrit is the pre-traditional way of Sanskrit. The initial verified Sanskrit text is "Rigveda", holy writ of Hindus, from the second half millennium BCE [11]. The writing system used for Sanskrit is Brahmi Script. Sanskrit alphabets contains 16 vowels known as swaras which comprise of 5 short ones, 8 long ones and 2 support vowels, 34 consonants. In Sanskrit, the vowels form a separate group from the consonants.

2 Machine Transliteration Techniques

Machine Transliteration can be generally classified into 4 type namely, Grapheme model, Phoneme model, Correspondence and Hybrid Model.

2.1 Grapheme Based Transliteration Model

Grapheme model is the simplest method because the mapping of grapheme of source to targeted grapheme is direct and no need of any additional understanding [4]. So it is also called direct method. The units used in this model are grapheme which is word segments. But in this case, the transliterated word may not be phonetically equivalent [4]. Decision trees that translate each source grapheme into target graphemes are learnt and then used directly to machine transliteration in the approach based on a decision tree.This approach has the advantage of considering a wide variety of contextual data, including the left three and right three contexts. However, no consideration is given to the phonetic features of transliterationciteb4. 121 Oh, Choi, and Isahara Several transliteration methods based on this model have been proposed, including those based on a source-channel model (Lee and Choi 1998; Lee 1999; Jeong et al. 1999; Kim et al. 1999), a decision tree (Kang and Choi 2000; Kang, 2001), a transliteration network (Kang and Kim 2000; Goto et al. Li et al. 2004).

2.2 Phoneme Based Transliteration Model

Phoneme model uses phonemes as the unit which is the pronunciation units. In this model, there are 2 steps. First, the source language grapheme to phonemes of source language then, these phonemes are converted to graphemes of target language [4]. Using a pronunciation dictionary, the approach converts an English word into English pronunciation. After then, the English phonemes are split into parts, each of which matches to a Korean grapheme specified by a set of rules [4]. Finally, it uses an expanded Markov window to automatically convert each piece of English phonemes into Korean graphemes[4].

Below shows some example of converting grapheme to phoneme: **LEMON EH M AH N**[4].

2.3 Hybrid and Correspondence Model

Using of grapheme and phoneme model in transliteration system leads to the implementation of hybrid and correspondence model transliteration model [9]. The cre-

ator practices the correspondence by linking of an original grapheme and original phoneme during it fabricates the target language graphemes, after that it easily merge grapheme and phoneme through "linear interpolation " [9]. But these 2 models are very difficult to execute. So we can say that transliteration is a phonetic process rather than one to one [4].

3 Background Study

This sections list outs the jargon used in the domain of transliteration (Fig. 1).

3.1 Transliteration

Transliteration the processes of mapping of one system of writing using another. It helps the user to learn one script with the help of other. From a grammatical view, transliteration is an aligning of a writing system into another, syllable by syllables or by words. Transliterated text can be used in computational linguistic tools like Morphological generator, dictionary, etc. When the materials being processed and listed are in multiple languages, transliteration is necessary in the documentation. In India, where papers are written in 15 languages and dialects, transliteration is a severe challenge [15]. There is always the risk of misplacing and obscuring entries unless a consistent pattern of transliteration is followed.

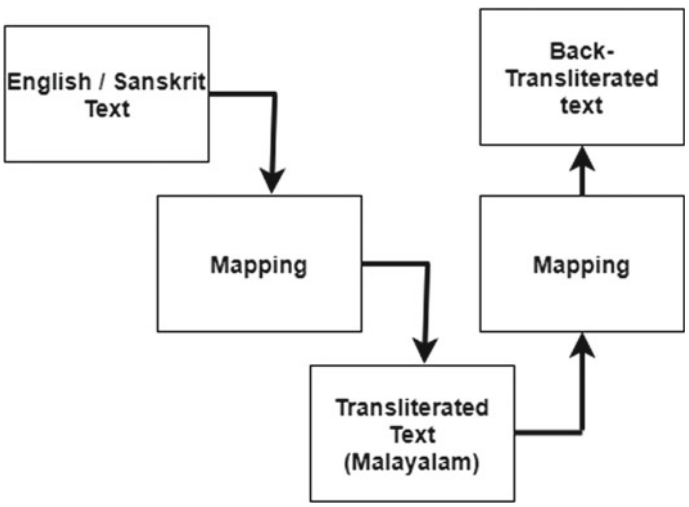


Fig. 1 Architecture diagram of forward and backward transliteration

3.2 Forward Transliteration

Forward transliteration is a method in which the source language (English) is converted to the targeted language (Malayalam). **For example, "Transliteration" is an English word which is represented using Roman alphabet and "ട്രാൻസ്ലിറ്ററേഷൻ" is the transliterated word for the term Transliteration which is represented using Malayalam alphabets**. This process is known as Forward Transliteration.

3.3 Backward Transliteration

Backward Transliteration is a method which converts the transliterated word to its original form. **For example, when the back transliteration is made the word "ട്രാൻസ്ലിറ്ററേഷൻ" is transliterated to its original form as "Transliteration" using the Roman Alphabets**. This process is known as Backward Transliteration.

3.4 Devanagari

It is also called as Nagari, which is from left to the right syllabaries depends on the Brahmi script in ancient, used in the subcontinents of India [14]. The Sanskrit, Prakrit, Hindi, Marathi, and Nepali languages are written using this script, which originated from the North Indian monumental script known as Gupta and subsequently from the Brahmi alphabet, from which all modern Indian writing systems are descended. Long, horizontal strokes at the tops of the letters differentiate Devanagari, which are often linked in current usage to make a continuous horizontal line through the script when written[14]. It has been in use since the 7th century CE, and its mature form first occurs in the 11th century. There are around 47 characters with 14 vowels and 33 consonants in Devanagari [15]. It is 4̂th mostly widely used writing system in the world. As it differs from Latin alphabets, the script has no idea of syllables case. Devanagari is written from left to right, with a strong preference for regular circles inside perpendicular boundaries. It is distinguished by a straight line, known as a "shirorekha", that runs parallel to the top of all letters [14].

3.5 Natural Language Processing (NLP)

NLP is the technology which helps computers read, understand and process human's natural language[15]. Natural language processing belts all the thing which a machine needs to take in natural language and also create natural language [16]. Natural language came into existence because when the user wants to communicate with

the computer they need to learn the machine language some time it is difficult to learn this machine language. To fill that gap, this NLP is introduced. Some of the Natural Language Application machine translation, speech recognition, information retrieval, etc.

3.6 Transliteration Schemes

A transliteration scheme is a system of rule which defines the mapping of a source language letters with that a target language letters. Following are the most popularly used scheme applied in Malayalam and Sanskrit.

3.6.1 Harvard-Kyoto

It is for Sanskrit and the languages that uses the Devanagari script transliteration into American Standard Code for Information Interchange [13]. It is mainly used casually in electronic mails. Harvard-Kyoto was introduced for depositing a good amount of Sanskrit work into machine understandable format with using no diacritics like International Alphabets of Sanskrit Transliteration. Rather than diacritics it uses capital letters [13]. After all it uses both capital and small letters, proper noun's first letter capitalization format cannot be go with.

3.6.2 ITRANS

The "Indian languages Transliteration" is an American Standard Code for Information Interchange transliteration scheme for "Indic scripts", especially for Devanagari writings [12]. It was created to meet the requirement for encryption of Indian texts, and it has a wider variety of applications than the 'Harvard-Kyoto' system for Devanagari transliteration, with which it coexists, although not entirely. With the huge execution of 'UNICODE', the traditional International Alphabet of Sanskrit Transliteration is used increasingly also for e-texts [12]. Transliteration of Indian languages Romanization is straightforward to examine and gather since it simply employs diacritics generated on a standard English language keyboard. The 'ITRANS' method is with no use of diacritics, as compared to other transliteration schemes [12]. Whilst using ITRANS, for actual names, the syllables at the beginning capital letter are not feasible as long as, ITRANS uses both upper and lower case syllables.

3.6.3 ISO 15919

It is a transliteration system that was created in the 2001 "ISO 15919" standard [14]. It combines a significant number of "Brahmic graphemes" with the "Latin script"

using special symbols[14]. The Devanagari component is quite similar to the Sanskrit instructional standard and the International Sanskrit Transliteration Alphabet [14].

3.6.4 ALA-LC Romanisation

It is a transliteration system that is recognized by the "Library of Congress" and the "American Library Association", and is mostly utilized by "North American libraries" [14]. The table for transliteration in this system is language-dependent, with a list for "Hindi", "Sanskrit", and "Prakrit".

3.6.5 WX

It is a Roman-based Indian language transliteration method that is primarily used by the NLP community in India] [14]. It was developed at IIT Kanpur for the computational processing of Indian languages. The technique's primary premise is outlined below.

- There is only a single mapping for all the consonants and vowels into Roman. From computation point of view there is an advantage for it because it is a prefix code [14].
- For non-extracted short vowels along with consonant's small letters are used, whereas extracted consonants capital letters are used [14].

3.6.6 Mozhi

For Malayalam script, it is the most popular method for Romanization. For insert method, editor for Malayalam Mozhi is used. And it is slightly depending on ITRANS technique for Devanagari. In this system, it won't use the diacritics, instead it applies the syllable case difference to identify dissimilarity in vowel strength and non-identical group of consonants.

4 Related Works

This section inclusive of all the related work in transliteration and some important paper literature on this field. A consolidated report on these are shown in the Table 1.

4.1 A Phoneme Based Model for English to Malayalam Transliteration [4]

This paper transliteration of English—Malayalam based on Phoneme. The Phoneme model get some advantage of pronunciation and Out-of-vocabulary (OOV) problem [4]. This Phoneme model split text is into words and with the help of a pronunciation dictionary it will be converted to English phonemes. English phonemes are converted to graphemes of Malayalam and combine the graphemes of Malayalam into words in Malayalam [4]. The graphemes of Malayalam are stored in the form of Unicode's [4] (Fig. 2).

4.2 Tamil to Malayalam Transliteration [3]

This paper propose a machine transliteration technique suggested to allocate most expected destination language letter to every syllables in the original language

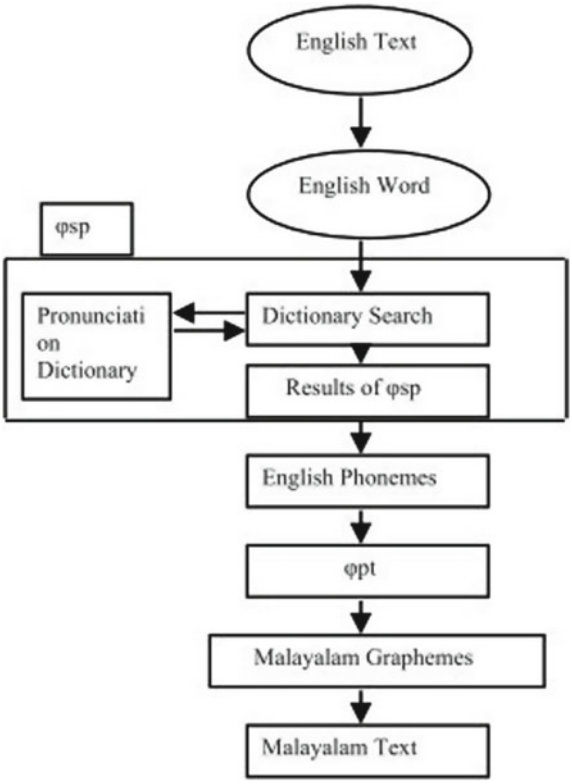


Fig. 2 A phoneme based English to Malayalam transliteration design [4]

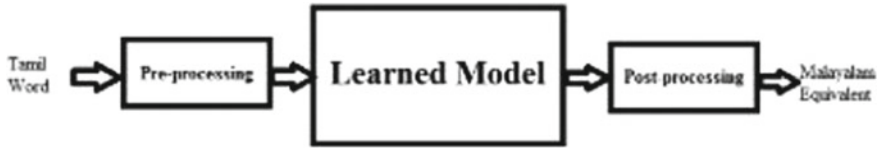


Fig. 3 The proposed architecture of Tamil to Malayalam transliteration [3]

transliteration from Tamil phonemes and graphemes [3]. It is a GUI system which uses Support Vector Machine(SVM) or Trigrams'n'Tags(TnT). Using SVM "binary classifiers" property 'binarization' of the content is executed and differentiate between samples of this class [3]. TnT is a "statistical classifier" that train the different languages corpus and optimize the speed. In this system, the transliteration is made with 3 steps that is, pre-processing, learning, post-processing [3]. The pre-processing takes the input and post processing provides the output [3] (Fig. 3).

4.3 A Transliteration Engine for Asian Languages[5]

This paper explains the transliteration engine's development and its design which suits syllabary and alpha syllabary scripts and is a web based engine [5]. An alphabet set that represents consonants, vowels, and syllables made up of consonants and vowels is known as an alpha syllabary [5]. The transliteration engine able to produce text phrases and Hyper Text Mark-up Language Unicode strings which allows to sink with the HTML document directly. In script mappings words in Roman alphabet and the target is a set of words in Unicode so the engine variety of scripts and use the Hepburn Romanization system [5]. By looking for the longest match of original syllable sequence the engine process the input text within the script mapping [5]. The engine can't do back transliteration. The information will be lost when the more number of sound symbol of the original is mapped with single sound symbol of the target [5].

4.4 A Roman to Devanagari Back Transliteration Algorithm Based on Harvard-Kyoto Convention[7]

The paper is about the Roman script back transliteration to Devanagari. Following the "Harvard-Kyoto" (HK) protocol, a Sanskrit script method was used to convert "American Standard Code for Information Interchange" (ASCII) encoded English to Devanagari [7]. Back transliteration is the process of taking back the transliterated words to its original format. Rule-based grapheme model using scripts of alphabet

and symbols are used in this paper. In Sanskrit each letters are connected with a 16 bit unique Unicode's "UTF-16" encoding method, in the creation of the Devanagari Language [7].

4.5 A Rule-Based Grapheme to Phoneme Converter for Malayalam [6]

The paper a rule based approach to the convert or map Input grapheme to correspond phoneme in Malayalam by the help of predefined rules. The consonants combine to form conjunct consonants [6]. First sentence tokenize sentence to into graphemes and remove all punctuation. Using Dictionary approach to overcome the out-of-the-vocabulary problems [6]. In transliteration, each alphabet of Malayalam has predefined symbol (may be combination of alphabets) as a grapheme. Combination of alphabets. For example, "the symbol for "ന്ധ" (pronounced as "yha") is "yh". When a word is given, the system has to identify the graphemes and need to separate them" [6] (Table 1; Fig. 4).

4.6 Rule Based Transliteration Scheme for English to Punjabi [2]

Rule based approach is used in this paper [2]. For syllable production of the string in English, first identifies the Vowels and Consonants, then identifies vowel and consonant mix and view it as single syllable, then identifies consonants accompanied by vowels and view them as separate syllables, identifies the vowels accompanied by 2 unbroken consonants as different syllable. Examine vowel neighbouring by 2 consonants as different syllable. The after the processing will transliterate each to its corresponding Punjabi script [2]. N-gram probabilities are calculated based on the relative frequency "p(wn/ wn-1)= c(wn-1, wn)/c(wn-1)" where the probabilities of a word "wn" given a word "wn-1" is count "(wn-1, wn)/count (wn-1)" [2].

4.7 English to Malayalam Transliteration Using Sequence Labelling Approach [1]

In this paper, transliteration viewed as a sequence labelling problem using SVM for classification. Syllables in the original script might mismatch with the destination language. The original script is segmented into units of transliteration and convert it with the destination language units. The transliteration problem is view as "Sequence Labelling Approach" because it contains Segmentation and Alignment. For example,

Table 1 A consolidated report charting all the important transliteration papers

S. No	Paper	Method	Algorithm	Remark
1.	English to Malayalam transliteration using sequence labelling approach [1]	Support vector machine	Supervised learning	This system had an accuracy of 90% used around 20000 names for training and 1000 names for testing [1]
2.	Tamil to Malayalam transliteration [3]	SVM or Trigrams'n'Tags, Tkinter for GUI	Supervised learning	For transliterating known letters TnT is more while, for unknown letters SVM is more efficient [3]
3.	Rule based transliteration scheme for English to Punjabi[2]	Statistical machine translation Tools(Moses & GIZA++)	Rule Based Model	Provides an accuracy of 88.19%. They used around 6328 proper names, locations, organization for testing from that 6109 were correctly transliterated
4.	A Phoneme based model for English to Malayalam transliteration [4]	Grapheme and Phoneme	Rule based model	English was the input and Malayalam was the output. Input can the typed or read from file [4]
5.	A Transliteration engine for Asian languages [5]	Web based transliteration engine	Modified hepburn system	The source for transliteration is Roman letters and target is UNICODE. No back transliteration is done
6.	Rule-based grapheme to phoneme converter for Malayalam[6]	Grapheme and Phoneme	Rule Based Model	The out of vocabulary issues are solved using dictionary based approach and manual editing is while rule based works better [6]
7.	A roman to devanagari back-transliteration algorithm based on Harvard-Kyoto convention [7].	Grapheme model	Rule based model	The text "Atngahridaya" was personally reviewed with a 100% accuracy in back transliteration using a rule-based method [7]

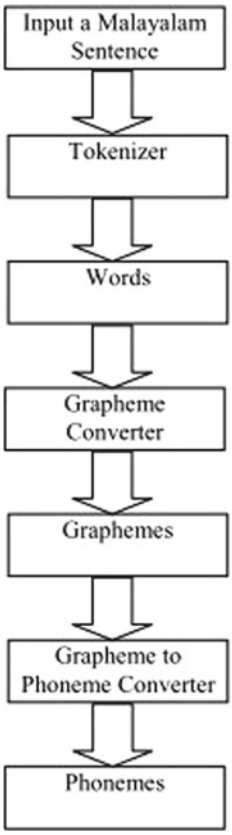


Fig. 4 Block diagram of rule based grapheme to phoneme converter for Malayalam [6]

a source string 'N' is segmented as "N_1. N_2, N_3, . . . , N_m" and the equivalent target string 'M' is segmented as "M_1, M_2, M_3, . . . , M_n" where each segments are treated as a label. Each of the segments in source script is aligned with its corresponding lingual segment in target. Pre-processing. Training and transliteration are the three phases of transliteration [1].

5 Conclusion

This paper presented a survey on the different Malayalam and Sanskrit Transliteration Schemes and techniques. A literature survey on some of the existing works in this area is also projected here. Machine transliteration is the process of converting a writing system into another. It is useful for cross-lingual information retrieval, multilingual text and speech processing. Transliterated text can be used in Computational Linguistic Tools like Morphological Generators, Dictionaries, etc., for easier pro-

cessing and handling. The existing work had many gaps like there are some words in Malayalam which cannot be transliterated with the Roman and Devanagari scripts so, such problems are corrected in our system.

References

1. S. Sumaja, R. Loganathan, K.P. Soman, English to Malayalam transliteration using sequence labeling approach. Int. J. Recent Trends Eng. **1**(2), 170 (2009)
2. D. Bhalla, N. Joshi, I. Mathur, Rule based transliteration scheme for English to Punjabi (2013). arXiv preprint arXiv:1307.4300
3. K. Raju, et al., Tamil to Malayalam transliteration, in *2015 Fifth International Conference on Advances in Computing and Communications (ICACC)* (IEEE, 2015)
4. C. Sunitha, A. Jaya, A Phoneme based model for English to Malayalam transliteration, in *International Conference on Innovation Information in Computing Technologies* (IEEE, 2015)
5. S. Manoharan, A transliteration engine for Asian languages. WEBIST 1:376–379 (2008)
6. S.S. Nair, C.R. Rechitha, C. Santhosh Kumar, Rule-based grapheme to phoneme converter for malayalam. Int. J. Comput. Ling. Nat. Lang. Process. **2**(7), 417–420 (2013)
7. J. Nair, A. Sadasivan, A Roman to devanagari back-transliteration algorithm based on Harvard-Kyoto convention, in *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)* (IEEE, 2019)
8. P.J. Antony, K.P. Soman, Machine transliteration for Indian languages: a literature survey. Int. J. Sci. Eng. Res. IJSER **2**, 1–8 (2011)
9. J. Oh, K. Choi, H. Isahara, A comparison of different machine transliteration models. J. Artif. Intell. Res. **27**, 119–151 (2006)
10. G. B. Nair, Malayalam, in *Concise Encyclopedia of Languages of the World* ed. by K. Brown, S. Ogilvie, pp. 680–683 (2009)
11. H.W. Bailey, Buddhist Sanskrit. J. Roy. Asiatic Soc. Great Britain Ireland **87**(1/2), 13–24 (1955). (Cambridge University Press). https://doi.org/10.1017/S0035869X00106975.JSTOR 25581326
12. Wikipedia contributors. ITRANS. In Wikipedia, The Free Encyclopedia (2020). Retrieved 10:27, November 17, 2020, from https://en.wikipedia.org/w/index.php?title=ITRANS&oldid=987803481
13. Harvard-Kyoto system source-Dominik Wujastyk (1996). Transliteration of Devan$\bar{a}$gar$\bar{i}$". Retrieved 24 Nov 2014
14. Wikipedia contributors, Devanagari, in Wikipedia, The Free Encyclopedia (2020). Retrieved 12:09, November 17, 2020, from https://en.wikipedia.org/w/index.php?title=Devanagari&oldid=987957208
15. M.J. Garbade, A simple introduction to natural language processing. Medium. Tillgängling online: https://becominghuman.ai/a-simple-introduction-to-naturallanguage-processing-ea66a1747b32 (2019-11-01) (2018)
16. A. Chopra, A. Prashar, C. Sain, Natural language processing. Int. J. Technol. Enhancements Emerg. Eng. Res. **1**(4), 131–134 (2013)
17. A. Govindankutty, From proto-Tamil-Malayalam to west coast dialects. Indo-Iranian J. vol. XIV, Nr. **1**(2), 52–60 (1972)
18. W. Bright, The Devanagari script, in *The World's Writing Systems*, pp. 384–390 (1996)
19. https://www.education.gov.in/hi/sites/uploa_files/mhrd/files/upload_document/languagebr.pdf
20. P.P. Joby, Expedient information retrieval system for web pages using the natural language modeling. J. Artif. Intell. **2**(02), 100–110 (2020)
21. I.J. Jacob, Performance evaluation of caps-net based multitask learning architecture for text classification. J. Artif. Intell. **2**(01), 1–10 (2020)

22. S. Bhaskaran, et al., Indian language identification for short text, in *Advances in Computational Intelligence and Communication Technology* (Springer, Singapore, 2021), pp. 47–58
23. J. Nair, R. Nithya, M.K. Vinod Jincy, Design of a morphological generator for an English to Indian languages in a declension rule-based machine translation system, in *Advances in Electrical and Computer Technologies* (Springer, Singapore, 2020), pp. 247–258

Two-Stage Folded Resistive String 12-Bit Digital to Analog Converter Using 22-nm FINFET

G. Vasudeva and B. V. Uma

Abstract Data converters play an important role in interfacing digital section with the analog section. During the conversion, there should be minimum loss of information and less propagation delay. FINFETs are devices that operate faster with high current density compared with CMOS circuits. In this work, operational transconductance amplifier (OTA)-based 12-bit digital to analog converter is designed and implemented demonstrating advantageous in terms of DAC specifications. The essential building block in most communication and control system is data converters, including analog to digital converter (ADC) and digital to analog converter (DAC). In this work, a new architecture for digital to analog (DAC) is proposed, designed, and evaluated for its performance. The 12-bit DAC is designed using two stages of 6-bit DAC. Each of the 6-bit DAC comprises of two-step voltage divider-type DAC and folded resistive string network. Device mismatches and area optimization are achieved by using folded resistive string approach, and two-stage DAC improves resolution with coarse and fine voltage generation logic from the two-step voltage divider method. Schematic capture is carried out using Cadence tool. From the simulation results, it is observed that the proposed DAC has a maximum operating bandwidth of 100 MHz, and the gain at 3 dB is 41.86 dB. The power dissipation of proposed DAC is 4.33 mW and, hence, suitable for high speed ADC application. The INL and DNL of the DAC design have been calculated as + 0.034 to −0.001 V and + 0.06 to −0.05 V.

Keywords OTA · High speed · Resistive string folded structure · High resolution · FINFET · Wide bandwidth · Low-power circuits

G. Vasudeva (✉) · B. V. Uma
Research Scholar, Electronics and Communication Engineering Department, RV College of Engineering, Affiliated to Visvesvaraya Technological University, Belagavi, Bangalore 560059, Karnataka, India

B. V. Uma
e-mail: umabv@rvce.edu.in

Professor and Dean Student Affairs, Electronics and Communication Engineering Department, RV College of Engineering, Affiliated to Visvesvaraya Technological University, Belagavi, Bangalore 560059, Karnataka, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_8

1 Introduction

Digital to analog converters (DAC) are one of the primary building blocks of analog to digital converters (ADC). There are several types of DAC architectures such as R-2R, binary weighted, current steering, and resistor string of which resistor string DAC is one of the simple DAC architectures that is used for high-speed data conversion. An N-bit DAC converts N-bit data into analog voltage and requires 2^N taps of resistive ladder. The 2^N number of resistors is connected in series between VDD and VSS, and at every node of the resistor, a switch is used to tap the voltage across the resistor. As the N-bit increases the number of resistors required are 2^N and hence occupies large area. To address the area requirement several other DAC architectures are designed. Resistor string DACs are used as reference voltage in ADCs and in circuits that require rail-to-rail output voltage generations. Kim et al. [1] have presented a 12-bit D/A based on current steering logic, designed using 14-nm FINFET CMOS technology for 4G applications. The spurious tones that arise in the current source mismatch in FINFET are addressed using switching-order shuffling and dynamic element matching. A 6-bit thermometric coding and 6-bit binary coding along with bit segmentations methods are adopted in DAC design. The power consumption of the D/A is limited to 6 mW, and SFDR is 80 dBc. Current steering DAC has limitations of occupying more layout area and is sensitive to mismatches in device properties. With technology scaling, these limitations have not been addressed and have further escalated in nanometer technology.

Murmann et al. [2] have presented FINFET technology-based high-speed DAC for mm-wave applications. Charge redistribution logic is used for design of DAC structure with parallel array logic. The limitations of current steering structure such as pulse timing, level signal mismatches, matching impedance, and power generation are addressed using time-interleaved switched capacitors with operating speed of 120 Gsps. Ting et al. [3] have presented design of segmented resistive string DAC for generation of reference signal or stimulus for ADC. Capacitive loading is reduced using folded string ladder network, and sub-resistor strings are used to reduce area of DAC structure. Aspokeh et al. [4] have presented a 5-bit resistive string DAC structure using capacitive logic which reduces area and power dissipation in 13-bit ADC. Kommangunta et al. [5] have presented a low-power area optimized resistive string DAC operating at 500 ksps, and the design results in reduction in resistor area by 28%, and OPAMP-based buffer is used to minimize offset errors. INL and DNL of 0.024 LSB and 0.004 LSB are achieved, respectively, with power dissipation limited to 65.23 μW. Mahadavi et al. [6] have proposed the 12-bit DAC design with resistive string logic and merged capacitor technique. It is suitable for high speed and high-resolution applications with 800 Msps and power dissipation limited to 1.37 mW. Yenuchenko et al. [7] have developed a 10-bit DAC with identical weighting elements and switches which uses a unified implementation based on transmission gate cells. It is demonstrated to operate at 1.8 V supply voltage with reduced number of switches. Debashis [8] in his thesis has presented an 8-bit DAC which is realized using segmented 3-bit DAC and 5-bit DAC, and the output of the

DACs is processed by the capacitive charge sharing DAC to generate differential input to SAR ADC. Variations in resistor mismatches would be detected for the various floorplans in the arrangement. The INL and DNL curves are used to calculate the fraction of systematic and random mismatches. The causes of variances in these incompatibilities are determined depending on the varied floorplans employed during the layout. Yang et al. [9] have presented a 12-bit CMOS dual-ladder resistor string which saves a considerable chip area. The D/A converter with the power supply of $\pm$ 5 V which adopts a self-adjusted reference circuit to provide differential reference voltages for obtaining better accuracy and symmetry. The results show the INL and DNL which are less than 2 LSB and 0.25 LSB. AB-Aziz et al. [10] present a paper on 12-bit pseudo-differential current source resistor strings hybrid DAC. They discuss a hybrid DAC architecture that combines the concepts of binary weighted resistor DAC and the thermometer coding DAC. This hybrid architecture gives a better INL of 0.375 LSB and DNL of 0.25 LSB, respectively. Huang et al. [11] present a compact 8-bit two-stage DAC with two voltage selector and an optimized area efficient 10-bit DAC with one voltage selector for AMOLED column driver ICs with CMOS technology. The measured DNL is 0.44LSB and INL is 0.68LSB for the 8-bit DAC, and measured DNL is 0.126 LSB and INL is 0.256 LSB for the 10-bit DAC. The proposed two-stage DAC architecture in this paper keeps the size of the 10-bit data driver even smaller than that of the conventional 8-bit driver. Lu et al. [12] introduce a 10-bit RFR-DAC for high color-depth LCD driver ICs and suggest a unique RFR-DAC architecture with a 10-bit resolution for liquid crystal display applications in their work. The proposed RFR-DAC combines a 6-bit RDAC with a 4-bit FR-DAC to provide a novel two-voltage selection and one-voltage selection methods that eliminate the requirement for unity gain buffers to isolate parallel-coupled resistor strings. The technology utilized is CMOS, with the worst DNL being 0.11 LSB and INL being 0.92 LSB using a two-voltage selection method and DNL being 1.37 LSB and INL being 1.45 LSB via a one-voltage selection strategy.

Use of resistors in DAC circuit leads to errors or performance degradation due to mismatches, and due to voltage drop, additional buffers are required to drive the output nodes. Thermometric coding in DAC requires large number of switches that leads to glitches and reduces DNL. Capacitive DAC offers device matching and is power-efficient; however, the number of capacitors exponentially increases with increase in N-bits. C-2C DAC is another structure to achieve high speed and good resolution. The disadvantage in the case of binary weighted resistor DAC as it requires large range of resistors with necessary high precision for low resistors. Also requires low switch resistances in transistors and can be expensive. Hence, resolution is limited to 8-bit size. Power dissipation of binary weighted circuit is extremely high. Current steering DACs are a more common integrated DAC compared to resistor DACs. The drawback of R-2R DAC is it requires two sets of resistors with precision resistance value (R and 2R). For a R-2R-based 12bit DAC, we require 12 switches minimum 6(R) and 6 (2R) resistors and 2 (2R) resistors. At any given point of time, more than one switch will be ON equivalent to the corresponding binary number to generate the equivalent analog voltage. This process increases delay and nonlinearity issues affecting INL and DNL, and suitable matching is not carried out. Based on

various architectures reported in literature, the most successful DAC architecture is the hybrid DAC that combines two different DAC structures or uses segmented principles. In this paper, a novel method for DAC architecture is designed using 22-nm FINFET technology. For high-speed applications and resolution more than 8 bit, resistive string DAC is recommended, because of its reduced delay (only one switch will be on at any point of time) and single value of R. There is uniformity in the circuit element minimizing reliability issues. The folded resistive string is more optimum in terms of area during layout implementation and avoids all device mismatching issues. The two-step voltage divider circuit generates Vout if coarse and fine resolutions including conversion accuracy. The 12-bit DAC is designed using two stages of 6-bit DAC. Each of the 6-bit DAC comprises of two-step voltage divider-type DAC and folded resistive string network.

The paper is organized as follows. Section 1 presents introduction to DAC. Section 2 highlights fundamentals of FINFETs and its device parameters, small-signal model. Section 3 discusses OTA concepts and FINFET-based OTA schematic for DAC. Section 4 discusses voltage divider folded resistive string DAC. Section 5 discusses about two-stage voltage 12-bit folded resistive string DAC. Results and discussions is presented in Sect. 6, and conclusion is presented in Sect. 7.

2 FINFET

FINFET-based analog and digital circuit design is advantages as the FINFET devices support high drain current, operates at low switching voltages generating low leakage current leading to low-power applications. Hu et al. [13] introduced FINFET devices and have demonstrated their advantages over MOSFETs. One of the FINFET device is the double gate-FET (DG-FET) that uses two controlling gates for current flow in the device [14, 15]. Figure 1 presents the generic structure of FINFET device with top gate and bottom gate that control the flow of current in the channel between the source and drain regions.

The small-signal circuit diagram for FINFET is shown in Fig. 2. The intrinsic circuit comprises of parasitic capacitances C_{gd}, C_{gs}, and C_{ds} along with the parasitic resistance R_{gd}, R_{gs}, R_{ds}, and R_{sub}. The capacitances C_{pg} and C_{pd} are considered at

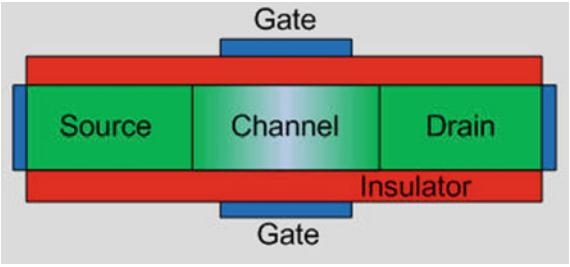


Fig. 1 Structure of FINFET (double gate) [14]

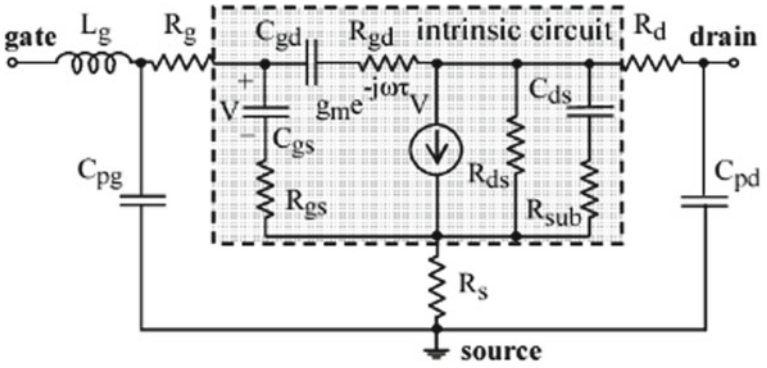


Fig. 2 Small-signal equivalent circuit for FINFET [16]

low frequencies with pinch-off condition. The parameters L_g, R_g, R_s, and R_d are computed considering V_{gs} above pinch-off.

The small-signal model and model file of FINFET considered from predictive technology model (PTM) are considered for design of ADC. The device parameters are presented in Table 1. The source doping and drain doping concentration based on Gaussian doping are considered at 1e + 19/cm^{-3}, the dielectric constant of channel is 11.7, and the dielectric constant of insulator is 3.9. The bandgap and affinity of channel material are considered at 1.12 eV and 4.05 eV with gate contact work function of 4.6 eV. Mobility of electrons and saturation velocity are considered at 1400 cm^2/Vs and 1.07e + 07 cm/s. Considering these structural and electrical properties for FINFET, the device model is simulated for its input and output characteristics.

Figures 3a and b present the input and output characteristics for the FINFET considered with the structural parameters as in Table 1. The parameters chosen in this work are compared with the parameters that have been considered in [17]. The technology selected in this work is 22 nm. The input is obtained by setting the drain

Table 1 FINFET device parameters

Parameter	Value (proposed work)	Value [17]
Channel length	22 nm	32 nm
Oxide thickness 1	2.5 nm	1.6 nm
Oxide thickness 2	2.5 nm	1.6 nm
Gate length	22 nm	–
Source/drain extension length	50 nm	32 nm
Gate to source/drain overlap	2 nm	–
Work function	4.6 eV	4.5 eV
Source/drain doping	1×10^{19} cm^{-3}	2×10^{20} cm^{-3}

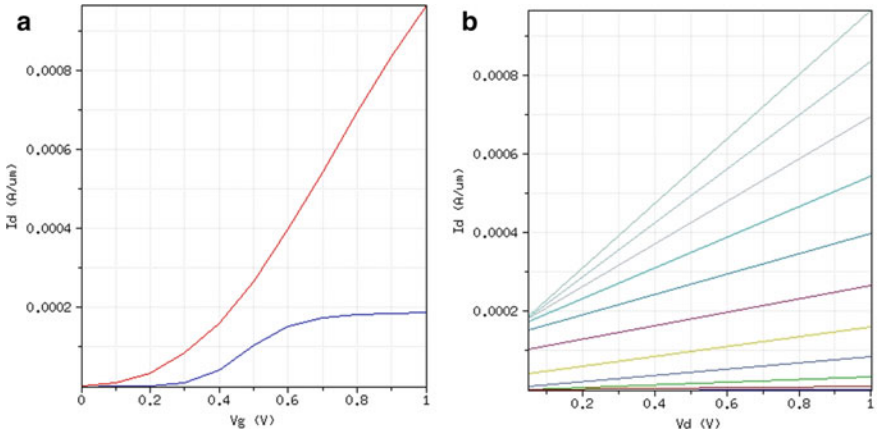


Fig. 3 **a** Input characteristics of FINFET. **b** Output characteristics of FINFET

voltage at 0.5 and 1 V. The output characteristic is obtained by setting the gate voltage between 0 to 1 V with incremental step of 0.1 V.

Figures 3a and b represent the V-I characteristics of FINFET considered at 22-nm technology with high-K dielectric. From the I_{ds}–V_d characteristics in the figure, it is observed that a small change in gate voltage doubles the I_{ds} current. Higher current-driving capability of FINFET leads to high-frequency operation. Power dissipation and transfer characteristics of inverter device are evaluated. The maximum power dissipation is observed to be less than 800nW, and transition width is less than 0.12 V. The leakage current during positive switching and negative switching current is observed to be less than 9 μA. Considering the advantages of FINFET over MOSFET, analog and digital sub-systems are designed, and these blocks are integrated into DAC logic.

3 OTA-Based Voltage Follower

OTA is transconductance device in which the input voltage controls the output current flow. The transconductance of the device gm makes the OTA a voltage controlled current source. The advantage in OTA is the transconductance parameter is controlled by the amplifier bias current. The output current is a function of the difference between the applied voltage, and the current at the output is converted to voltage by connecting a resistive load. The OTA-based circuits do not require negative feedback; the transconductance is one of the design parameters of OTA-based circuit designs. In this work, DAC circuit is designed considering advantages of OTA. OTA-based voltage follower is one of the main circuit used in DAC. High immunity toward short channel effects such as drain-induced barrier lowering (DIBL) and enhanced subthreshold swing FINFET device is preferred for DAC circuit designs. Low power

dissipation and high performance are achieved by use of FINFET in place of CMOS circuits. Design of FINFET-based OTA has advantages such as high gain, low power consumption, wide unity gain bandwidth, better slew rate and CMRR compared with CMOS-based OTA. The primary building blocks of OTA are presented in Fig. 4. The input stage performs level shifting, second stage is the folded cascode stage with feedback, and the last stage is the output stage.

The simple OTA circuit is presented in Fig. 5 that is realized using eleven transistors. The transistors $F1$ and $F2$ are the differential pair and form the transconductance cell that converts the input voltage V^{+}in and V^{-}in (differential input voltages) to current [18]. The differential current output (I_{out}) of the differential pair is converted to single-ended current at the output by using the current mirrors $F3$ to $F8$, $F10$, and $F11$. $F9$ transistor is used to bias the differential pair and is used as current sink circuit. The cut-off frequency of the OTA is decided by setting the appropriate bias current and the load capacitance of the OTA. The transconductance gain g_{m} of the OTA is controlled by setting the current that enters the transistor $F9$, and the gate voltage V_{b} is appropriately set. Design methodology based on $g_{\text{m}}/I_{\text{D}}$ method is the most popular approach [19] that identifies the transistor geometries based on data sheets

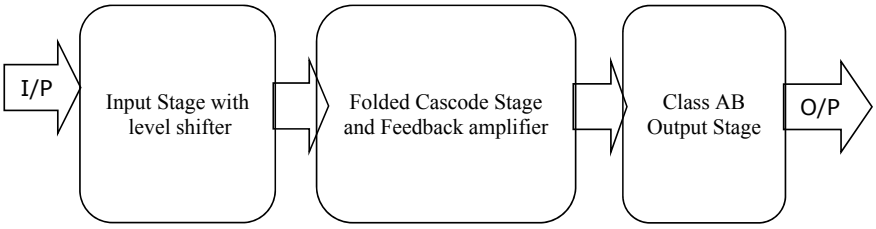


Fig. 4 OTA block diagram

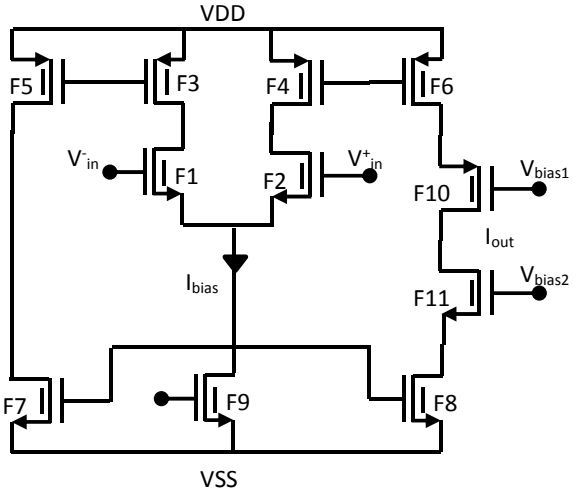


Fig. 5 FINFET-based OTA circuit [20]

and simulation results. In this method, the design of OTA circuits based on datasheet and several design variables that were required for the design were assumed without clear rules. The design specifications meeting input range, common mode rejection, and noise parameters were not considered in this approach. Even, the channel length variations regarding g_m/I_D were not considered in the design process.

Figure 5 presents the circuit schematic of OTA with input stage, differential pair, and output stage including the bias circuits. Operational transconductance amplifier captured in Cadence environment consists of three sub-blocks such as differential amplifier, gain boosting block, and common mode feedback as shown in Fig. 6. The inputs areV_1, V_2, V_3, and the outputs are V_{OM} and V_{OP}.

The N-channel FINFET and P-channel FINFET transistor sizing involved in designing OTA are shown in Table 2.

The simulation results of DC analysis and differential input of OTA are shown in Fig. 7. The DC gain is of 68 dB, but the phase margin is observed to be less than the required (13^0). The FINFET geometries are varied to achieve the desired phase margin.

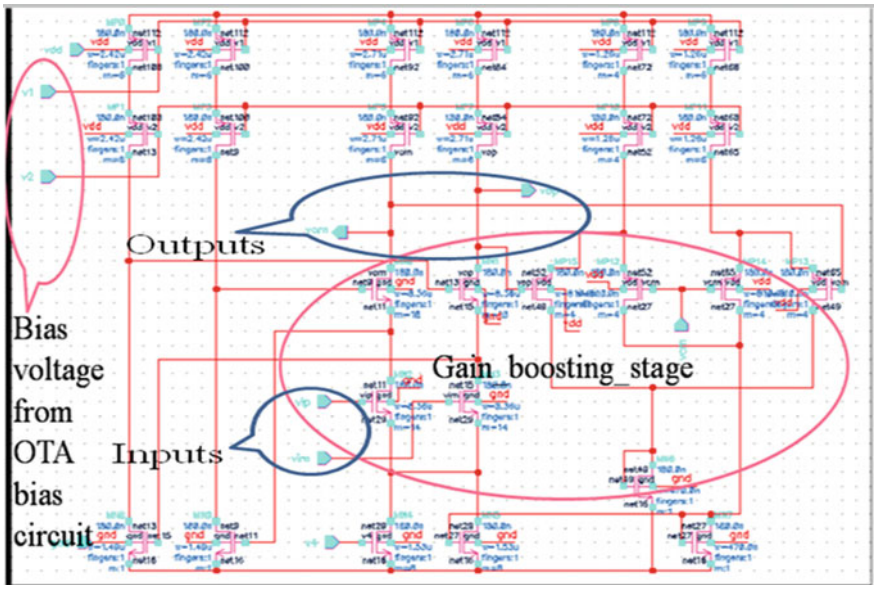


Fig. 6 Schematic diagram of FINFET-based OTA for DAC

Table 2 Transistor sizing for OTA design

Component	(W_n/L_n)nm
N-channel FINFET	$(80.3/22)$
P-channel FINFET	$(180/22)$

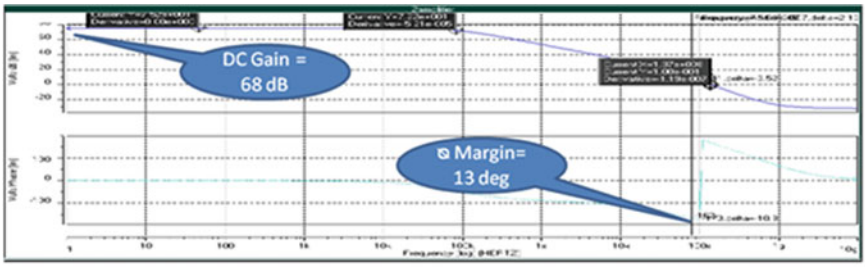


Fig. 7 DC analysis and differential input simulation of OTA

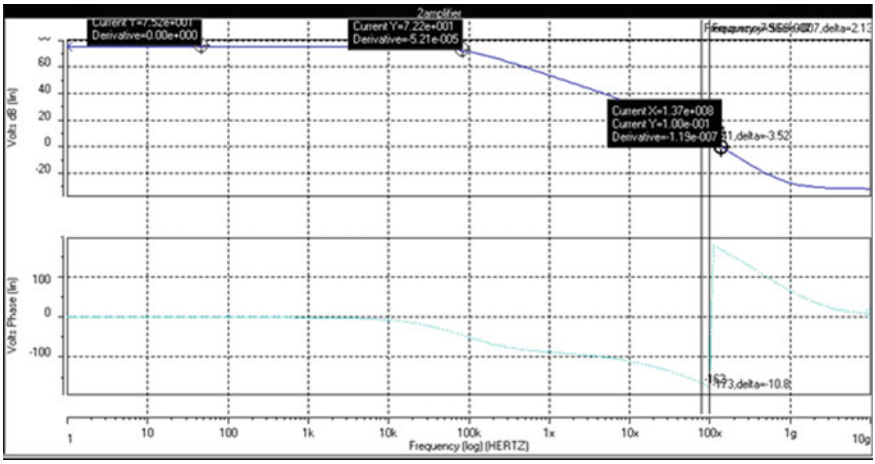


Fig. 8 Gain and phase margin analysis of improved OTA for DAC

The improved OTA design has a unity gain bandwidth (UGB) of 100 MHz, and the gain at 3 dB is 41.86 dB and phase margin of 52^0. The FINFET transistor geometries are increased by a factor of 1.2 times the designed geometries to achieve the required phase margin (Fig. 8).

4 Voltage Divider Folded Resistive String DAC

DAC input–output relation is mathematically expressed as in Eq. (1) [21]. The N-bit input data is $\{b_1, b_2, b_3, b_N\}$, b_1 is MSB and b_N is LSB, N is the data input length, and V_{ref} is the reference voltage.

$$V_{\text{OUT}} = V_{\text{ref}}\left(b_1 2^{-1} + b_2 2^{-2} + \ldots + b_N 2^{-N}\right) \tag{1}$$

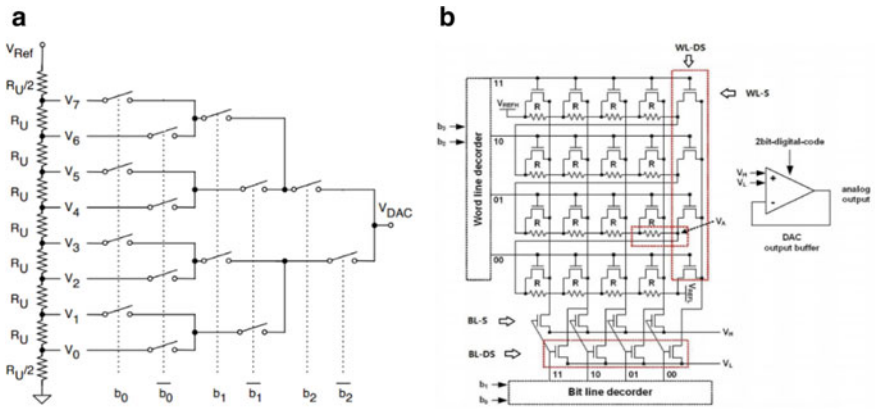


Fig. 9 **a** Resistor string DAC ladder network [22]. **b** Resistor string DAC folded network [22]

The resistor string-based digital to analog converter is shown in Fig. 9a [22]. The resistor string has 9 resistors and 14 transistor switches. The input bits b_0, b_1, and b_2 are inputs are used as control inputs to the 14 switches that allow the corresponding ladder voltage to the output. The voltage follower at the output amplifies the ladder voltages to generate the equivalent analog voltage. The number of switches led to noise due to switching which is addressed by use of folded resistor string structure. Figure 9b [22] presents the resistor string DAC designed using folded string logic. The 4-bit input is split into two groups of b_1, b_2 and b_3, b_4 that are inputs to two 2:4 decoders along the x and y directions, respectively. The output of decoders is used to control the 2D switch matrix to enable one of the node voltages to be connected to the output voltage.

The input bits b_0 and b_1 through the 2:4 decoder enable one of the vertical lines, and the corresponding bits b_2 and b_3 are used to enable one of the horizontal lines through the 2:4 decoder. The ladder voltage along the folded resistor string is directed into the voltage follower as equivalent analog output. The number of switches for a 4-bit DAC is limited to 16 in the folded resistor string DAC. The folded resistor structure is implemented using matching circuits during layout design minimizing mismatches in the device parameters. The folded resistor string DAC is area-efficient but introduces propagation delay and limited resolution. To overcome the limitations of folded resistive string DAC, in this work two-stage resistor string voltage divider type D/A Converter is designed as shown in Fig. 10.

In the first stage, the resistive ladder network and two voltage follower circuits are used. The decoder output controls the switches ON and OFF. In the first stage, two switches are switched on simultaneously between every resistor in the ladder. The second stage is used to generate fine voltage levels between the two voltages references generated from the first stage. In this work, the advantages of voltage divider type and folded resistive string DAC are combined, and a new DAC architecture is designed. Detailed discussion on proposed DAC structure is presented in next section.

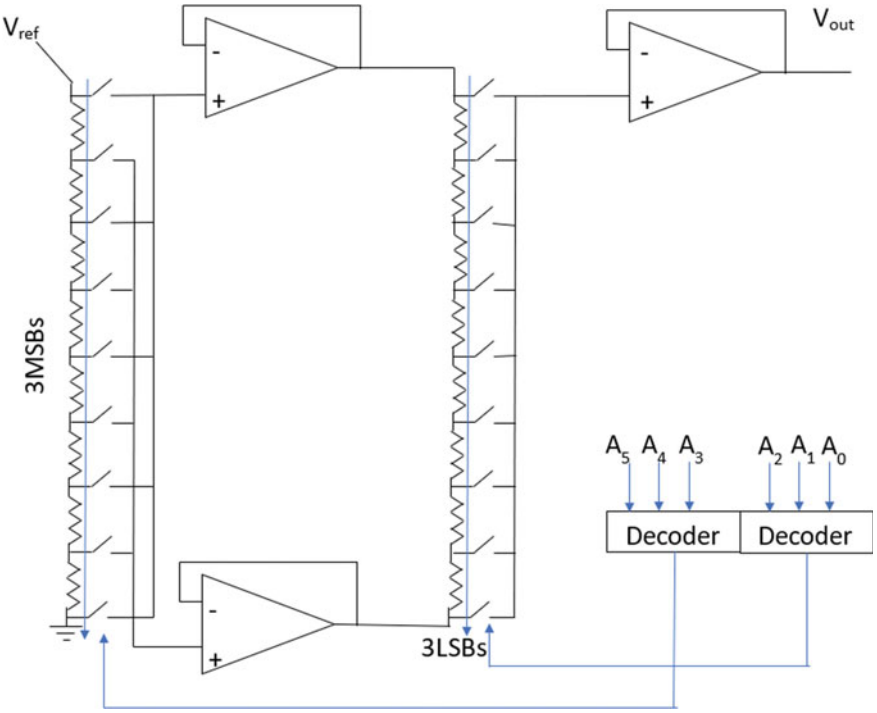


Fig. 10 Two-stage resistor string voltage divider type D/A converter

5 Two-Stage Voltage 12-Bit Folded Resistive String DAC

The proposed 12-bit DAC block diagram is shown in Fig. 11. The DAC structure is split into two groups of 6 bits. The first stage generates V_{out1} corresponding to 6

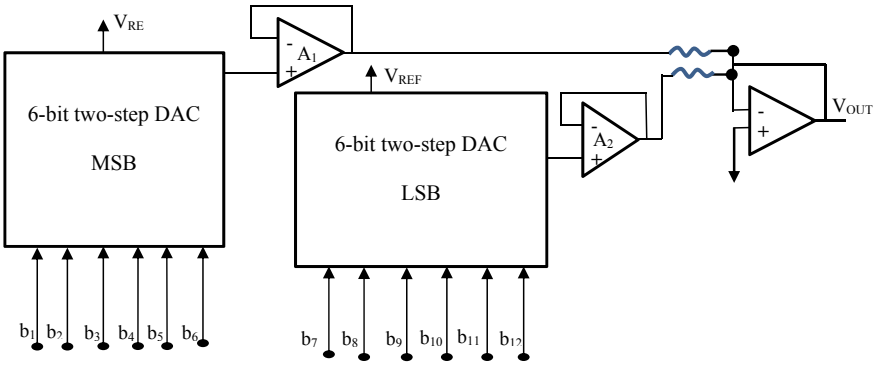


Fig. 11 Proposed 12-bit DAC structure

MSB, and the second stage generates V_{out2} for 6 LSB. The output of two-stage DAC V_{out1} and V_{out2} is accumulated in the adder circuit to generate the final analog output V_{OUT}.

Considering Eq. (1), for $N = 12$ bit, the DAC output is expressed as in Eq. (2),

$$V_{OUT} = V_{ref}\left(b_1 2^{-1} + b_2 2^{-2} + \ldots + b_{12} 2^{-12}\right) \tag{2}$$

Equation (2) is split into two sections of 6 bit each and is presented as in Eq. (3),

$$V_{OUT} = V_{ref}\left(b_1 2^{-1} + b_2 2^{-2} + \ldots . b_6 2^{-6}\right) + V_{ref}\left(b_7 2^{-7} + b_8 2^{-8} + \ldots . b_{12} 2^{-12}\right) \tag{3}$$

Rewriting Eq. (3) by combining the data bits into two groups of 6 bit, Eq. (4) is obtained. The second term in Eq. (3) is reduced by identifying 2^{-6} as common term. The second term is reduced to binary power of 2^{-1} to 2^{-6} as the first term. The V_{ref} in second term is scaled by 2^{-6}.

$$V_{OUT} = V_{ref}\left(b_1 2^{-1} + b_2 2^{-2} + \ldots . b_6 2^{-6}\right) + V_{ref}\, 2^{-6}\left(b_7 2^{-1} + b_8 2^{-2} + \ldots . b_{12} 2^{-6}\right) \tag{4}$$

The first term in Eq. (4) is realized using first-stage DAC with input bits b_1 to b_6, and the second term is realized using second-stage DAC with inputs b_7 to b_{12} and $V_{ref} = V_{ref}/2^6$. The expression in Eq. (4) is simplified and is presented in Eq. (5).

$$V_{OUT} = V_{out1} + V_{out2} \tag{5}$$

where $V_{out1} = V_{ref}\ (b_1\ 2^{-1} + b_2\ 2^{-2} + \ldots . b_6 2^{-6})$, $V_{out2} = V_{ref2}\ (b_7\ 2^{-1} + b_8 2^{-2} + \ldots . b_{12} 2^{-6})$, and $V_{ref2} = V_{ref}/2^6$. The variation between the first stage and second-stage reference voltages is 83%, which will impact the second-stage DAC performance as the V_{ref} is at a lower value. To address these limitations, the second-stage DAC is set with the same reference voltage as in first stage. The output of first stage V_{out1} is amplified by 2^6 and is accumulated with the second-stage DAC output to generate the final output as represented in Eq. (6).

$$V_{OUT} = \mathrm{A}V_{out1} + V_{out2} \tag{6}$$

where A is the gain introduced into the first-stage DAC output and is set to 64. The output of first-stage DAC is amplified by inverting amplifier (not shown in Fig. 11) with gain of 64. The advantages of two-stage DAC proposed are that the number of resistors and the transistors switches is 64/68 of the generic resistive string DAC. The limitations of two-stage DAC are that it requires three voltage followers or buffers. The two buffers between second stage and first stage introduce delay, and appropriate circuit design methodology is required to be considered to overcome these challenges. In this work, to improve the processing speed and further reduce

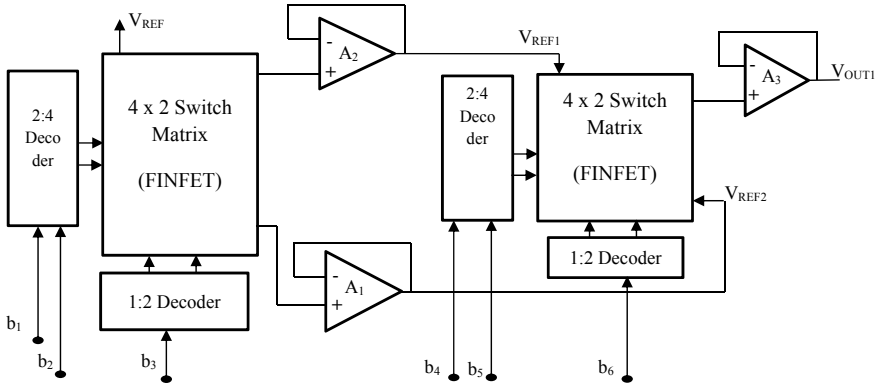


Fig. 12 Proposed 6-bit voltage divider resistive string DAC

the performance degradation in DAC circuit, a new architecture is proposed. The proposed DAC circuit topology is shown in Fig. 12. The 6-bit DAC is realized using voltage divider-type DAC shown in Fig. 10. The 6-bit DAC is split into two 3-bit DACs as shown in Fig. 10. The first stage of 6-bit DAC comprises of 2:4 decoder, 1:2 decoder, 4 × 2 switch matrix, folded resistive string, and two amplifiers A1 and A2. The second stage of 6-bit DAC comprises of all the building blocks of first stage, but there is only one amplifier. The two-stage DAC provides high resolution and low power consumption due to two-stage operation and isolation between switches and output node, respectively. The presence of two amplifiers and resistors in two stages introduces voltage offset and mismatches leading to performance degradation.

The novelty in this proposed work is the realization of the resistive ladder network. The resistive ladder or resistive string is realized using folded string technique to minimize device mismatches and area. The proposed DAC circuit is realized using folded string resistive structure shown in Fig. 13. The voltage offset and delay in the amplifier circuit are addressed by the design of voltage follower using OTA circuit.

The proposed folded string two-stage voltage divider-type DAC comprises of two folded string resistive network with eight resistors in each string and nine transistors connected across. The line 2:4 decoder generates the control signal for the switches based on two input bits, and the word 1:2 decoder generates the word line from one bit input. The two amplifiers $A1$ and $A2$ are used to accumulate the node voltages and generate V_{H} and V_{L} as the two reference voltages for Stage 2. The second-stage folded string two-stage voltage-type structures comprise of 2 × 4 resistor string and eight switches connected across the resistor. The 2:4 decoder along the horizontal axis and the 1:2 decoder along the vertical axis enable the corresponding resistive node voltage that exits between two references V_{H} and V_{L}. The node voltage is amplified by the voltage follower circuit $A3$ to generate the DAC voltage. The proposed circuit requires two 2:4 decoders and 1:2 decoders, two folded string resistive network, three amplifiers or voltage followers, and 21 switches.

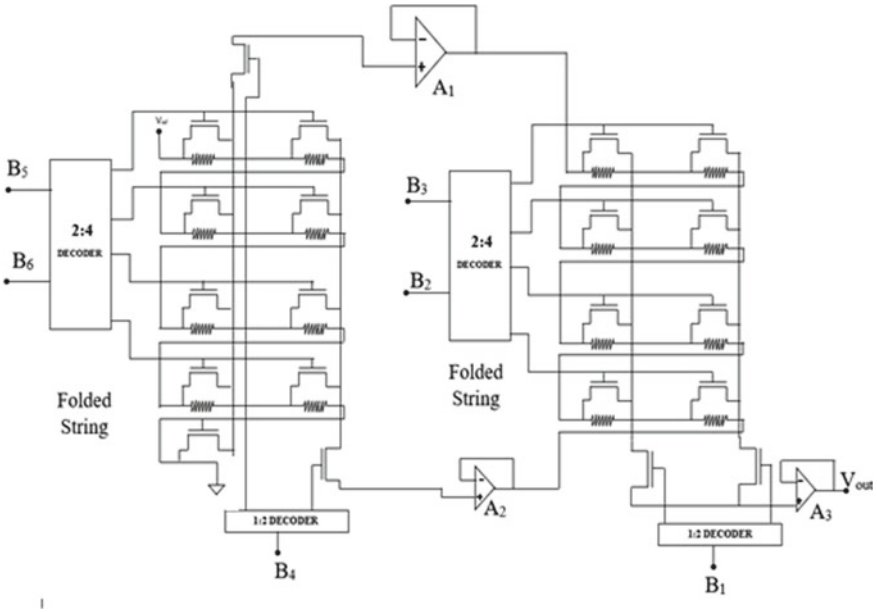


Fig. 13 Internal structure of proposed two-stage 6-bit DAC

The circuit schematic for 12-bit DAC is modeled hierarchically by first designing the 6-bit DAC and combining two 6-bit DAC into top-level module. The 22-nm FINFET model is set and modeled in Cadence environment. Prior to modeling in Cadence environment, the proposed schematic for 12-bit DAC is modeled in SPICE. The performance of the modeled DAC is evaluated considering different metrics and discussed in detail in next section.

6 Results and Discussion

Figure 14 shows the schematic diagram of transmission gated resistor string DAC for 6 bit, where it consists of 64 resistors for 6 bits. Based on the switches *B*6, *B*5, *B*4, *B*3, *B*2, *B*1, the divided voltage has been compared with the analog sampled input by comparator. The circuit presents the digital to analog converter for 6 bit with folded resistive ladder where the reference voltage 1.8 V has been divided as equal to 1 LSB value.

Figure 15 presents the simulation results of 6-bit DAC. Sixty-four combinations of 6-bit input are considered in verifying DAC functionality. Figure 15 shows the simulation result of 6-bit transmission gated digital to analog converter with its input digital bits such as *B*6—MSB, *B*5, *B*4, *B*3, *B*2, *B*1—LSB and its output shown as DAC output. Since, it is 6-bit DAC, the step output of the DAC is equal to 64 steps.

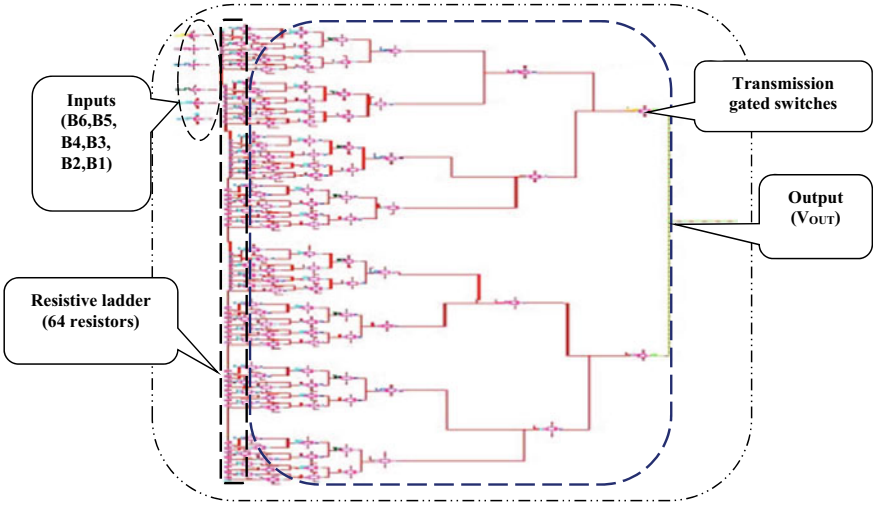


Fig. 14 Proposed two-stage folded resistor string 6-bit DAC schematic

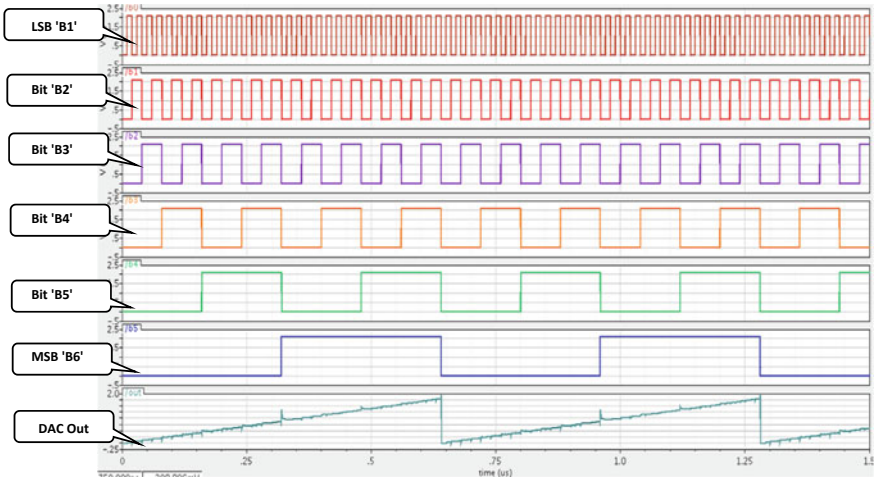


Fig. 15 Simulation result of 6-bit DAC

Figure 16 shows the output of integral nonlinearity (INL).The integral nonlinearity value has been calculated by developing graph manually. This top-level block consists of 12-bit DAC which gives analog output with respect to the 12-bit digital input. Figure 17 shows the simulation result of 12 bit along with 12-bit digital input generating the corresponding analog output.

Table 3 shows the comparison proposed DAC performance with reference design. It is observed from the simulation results that the result of the developed has been

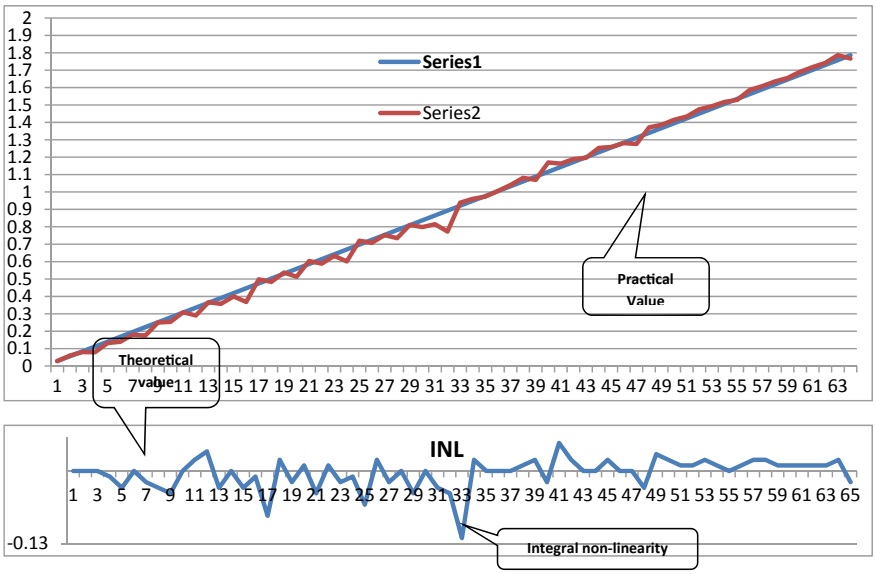


Fig. 16 Integral nonlinearity of the 6-bit DAC

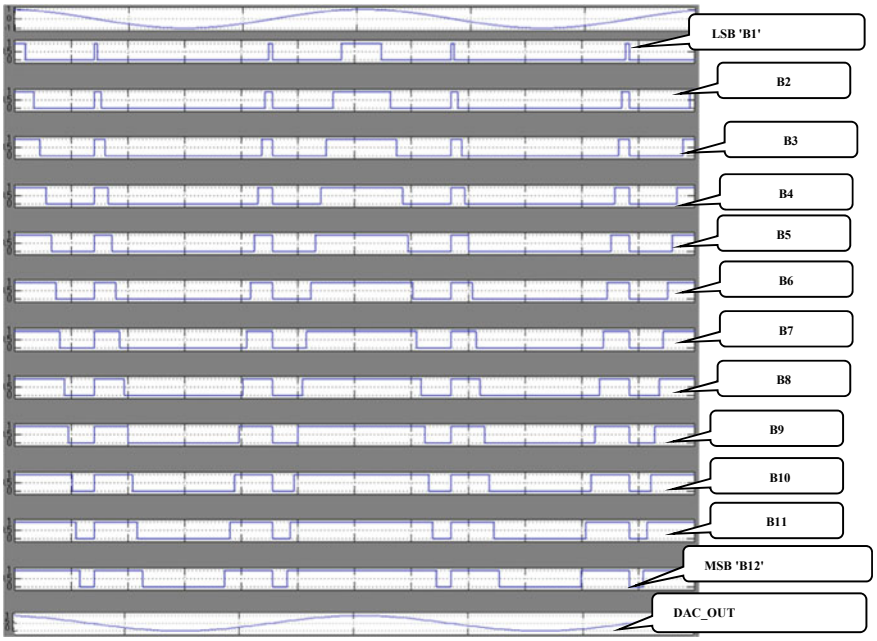


Fig. 17 Simulation results of 12-bit DAC

Table 3 Comparison of existing work and proposed DAC

Parameters	Desired specifications [23]	Obtained specifications
Resolution	12 bits	12 bits
Power supply	1.8 V	1.8 V
Technology	22 nm	22 nm
Input bandwidth	50 MHz	100 MHz
Sampling frequency	–	1 GHz
Total power dissipation	8.98 mW	4.33 mW
INL	$\leq \pm$ 0.3 V	+ 0.034 V / − 0.001 V
DNL	$\leq$ + 0.5 / −0.1 V	+ 0.06 V / − 0.05 V
Area	890 μm^2	450 μm^2

improved much better than the existing work proposed by Beaulieu [23]. The parameter in which improvement achieved is power dissipation, area, INL, and DNL. The power dissipation of designed DAC has been calculated as 4.33 mW, whereas existing work dissipates 8.98 mW. The area of the existing design has been calculated as 890 μm^2, whereas designed proposed DAC occupies 450 μm^2. Reducing the total number of resistors and design of folded resistive string circuits has contributed to low power dissipation and performance improvement in terms of circuit mismatches. Considering the power-delay product, an improvement of 75% is achieved in the proposed design as compared with reference design demonstrating the computing efficiency. Power dissipation per unit area in the proposed design is 9.63 W/m^2 which is an improvement of 4% compared with reference design.

The proposed OTA-based DAC circuit using FINFET device is designed to operate at sampling rate of up to 1 GHz. For a 12-bit DAC with input bandwidth of 100 MHz and maximum sampling frequency of 1 GHz, the power dissipation is limited to less than 4.33 mW which is an advantage for use of DAC in ADC. The INL is computed considering all possible levels of input, and it is limited to + 0.034/−0.001 V which is 100% improvement compared with desired specifications. The DNL error is also limited to less than + 0.06/−0.05 V, and an improvement of 76% is achieved compared with desired specifications. The maximum deviation in INL error is at the 33rd input data that occurs due to switching between the two folded string registers. The INL error is limited at higher input values due to both coarse and fine switching of folded string DAC circuits.

7 Conclusion

This paper presents a novel high-resolution two-stage voltage divider folded resistive string 12-bit DAC circuit designed using OTA, decoders, and switch matrix using 22-nm FINFET transistors. The advantages of two-stage DAC proposed are that the number of resistors and the transistors switches is 64/68 of the generic resistive string DAC. The two-stage DAC provides high resolution and low power consumption due to two-stage operation and isolation between switches and output node, respectively. The 6-bit MSBs are used in generating the coarse voltage, and 6-bit LSBs generate fine voltage in the second-stage DAC. The combination of first and second stage together generates high-resolution DAC output. The first-stage 6-bit DAC operates at V_{ref} to generate V_{out} voltage defining coarse conversion. The second-stage DAC divides this coarse voltage into six voltage references for data conversion defining fine resolution. The INL and DNL errors are within the set limits, and the power dissipation is less than 4.33 mW. The maximum operating bandwidth is 100 MHz with OTA gain of 41.86 dB. The limitations of two-stage DAC are that it requires three voltage followers or buffers. The presence of two amplifiers and resistors in two stages introduces voltage offset and mismatches leading to performance degradation. Further studies are carried out to overcome these limitations by design of OTA circuits that can offer minimum voltage offset and better matching circuits. The proposed DAC design is suitable for high-speed ADC (500 Mbps). The proposed DAC circuit can be used as programmable DAC and can be programmed to either work as 6-bit DAC or work as 12-bit DAC supporting high speed and high-resolution requirements, respectively.

Acknowledgements The authors thank RV College of Engineering for providing access to Cadence design suite license V6.7.11 using which the schematic capture and simulation results were carried out. We acknowledge PTM.edu for giving us permission to use FINFET model files and access to device parameters for our work. The authors acknowledge the valuable suggestions given by Prof. Cyril Prasanna raj. P.

References

1. J. Kim, W. Jang, Y. Lee, W. Kim, S. Oh, J. Lee, J. Choi, J.-H. Chun, T.B. Cho, Design and analysis of a 12-b current-steering DAC in a 14-nm FinFET technology for 2G/3G/4G cellular applications. Cir. Syst. I: Regular Papers IEEE Trans. **66**(10), 3723–3732 (2019)
2. B. Murmann, P. Caragiulo, R. Smith, High-speed DACs for millimeter-wave digital arrays in FinFET CMOS. Online: https://apps.dtic.mil/sti/pdfs/AD1075724.pdf
3. H. Ting, Z. Wu, J. Yan, H. Wu, A segmented resistor-string DAC based stimulus generator for ADC linearity testing, in *7th International Symposium on Next Generation Electronics (ISNE)*, (Taipei, Taiwan, 2018), pp. 1–4. https://doi.org/10.1109/ISNE.2018.8394734
4. T. Aspokeh, A. Amini, A. Baradaranrezaeii, M. Yazdani, Low-power 13-Bit DAC with a novel architecture in SA-ADC, in *25th International Conference "Mixed Design of Integrated Circuits and System (MIXDES),* (Gdynia, Poland, 2018), pp. 165–168. https://doi.org/10.23919/MIXDES.2018.8436764

5. V. Kommangunta, K. Shehzad, D. Verma, P. Kumar, K.-Y. Lee, Low-power area-efficient 8-bit coarse-fine resistor-string DAC, in *IEEE International Conference on Consumer Electronics—Asia (ICCE-Asia)*, (Seoul, Korea (South), 2020), pp. 1–3. https://doi.org/10.1109/ICCE-Asia49877.2020.9276800ss
6. S. Mahdavi, R. Ebrahimi, A. Daneshdoust, A. Ebrahimi, A 12bit 800MS/s and 1.37mW Digital to Analog Converter (DAC) based on novel R-C technique, in *IEEE International Conference on Power, Control, Signals, and Instrumentation Engineering (ICPCSI),* (Chennai, India, 2017), pp. 163–166. https://doi.org/10.1109/ICPCSI.2017.8392025
7. M.S. Yenuchenko, M.M. Pilipko, D.V. Morozov, A 10-bit segmented M-string DAC, in *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)* (Moscow and St. Petersburg, Russia, 2018), pp. 265–268. https://doi.org/10.1109/EIConRus.2018.8317081
8. B. Debashis, A segmented 8-bit resistor string DAC for SAR ADC, MTech thesis (2016)
9. F. Yang, C. An, L. Xie, X. Jin, A 12-bit CMOS Dual-ladder resistor string D/A converter integrated with self-adjusted reference circuit. WSEAS Trans. Circuits Syst. **13**, 129–134 (2014). E-ISSN:2224–266X
10. M.T.S.AB-Aziz, A. Marzuki, Z.A.A. Aziz, 12-bit pseudo-differential current source resistor string hybrid DAC, J. Circuits, Syst. Comput. **20**(4) 709–725 (2011)
11. Z. Huang, L. Tian, Q. Zhang, H. Wang, S. Feng, A compact-sized 10-bit two-stage DAC for AMOLED column driver ICs, IEICE Electron. Express 20150897 (2015)
12. C.W. Lu, py Yin, C.M. Hsiao, M.C.F. Chang, Y.S. Lin, A 10-bit resistor-floating-resistor-string DAC (RFR-DAC) for high color-depth LCD driver ICs. IEEE J. Solid-State Circuits **47**, 2454–2466 (2012)
13. C. Hu, T.-J. King-Liu, J. Bokor, FinFET-A self-aligned double-gate MOSFET scalable to 20nm. IEEE Trans. Electron. Devices **47**(12), 2320–2325, (2000)
14. A.O. Rozeau, Editors, *Planar double-gate transistor: From technology to circuit* (Springer, 2009), pp.1–20
15. J.P. Colinge and Editors, *FinFET and other multi-gate transistors* (springer, 2008), pp.1–13
16. P.E. Allen, D.R. Holberg, *CMOS analog circuit design* (2nd Edn, Oxford university press, New York, 2002)
17. V.B. Rahin, A.B. Rahin, A low-voltage and low-power two-stage operational amplifiers using FinFET transistors. Int. Acad. J. Sci. Eng. **3**(4), 80–95 (2016). ISSN:2454-3896
18. F. You, S.H.K. Embabi, E. Sanchez-sinencio, Low voltage class AB buffers with quiescent current control. IEEE J. Solid State Circuits **33**(6), 915–920 (1998)
19. N.M. Sabry, H. Omran, M. Dessouky, Systematic design and optimization of operational transconductance amplifier using g_m/I_D design methodology. Microelectron. J. (2018)
20. M. Nizamuddin, D. Sharma, FinFET based operational transconductance amplifier for low power applications. Int. J. Comput. Sci. Eng. **7**(5) (2019). E-ISSN:2347-2693
21. D.A. Johns, K.W. Martin, *Analog Integrated Circuit Design* (2nd Edn, 2012)
22. D. Chhetri, V.N. Manyam, J. Jacob Wikner, An event-driven 8-bit ADC with a segmented resistor-string DAC, in *2011 20th European Conference on Circuit Theory and Design (ECCTD)* (2011), pp. 528–531
23. J. Beaulieu, 12-bit Digital to Analog Converter. https://ece.umaine.edu/wp-content/uploads/sites/203/2012/05/report.pdf

Knowledge Base Creation for Prediction and Detection of Forest Fire

Bhagyashri Zodge and Jyoti Joglekar

Abstract Forests are necessary as well as important for combatting climate change. Forest fires destroy the forest's wealth and affect the environment. They pose a threat, to fauna and flora, seriously. Forest fires are difficult to detect and thus have a long response time which leads to mass destruction. Hence, it has been a prominent research area with the aim of conserving ecology. In this paper, we have created a knowledge base for forest fire risk prediction and detection purposes by extracting different forest-fire-causing factors from Landsat 8 satellite imageries. With increasing forest fires incident everywhere, the main goal of constructing a domain-based knowledge base of forest fires is to have records of forest fires and relevant geographical, environmental, and demographical information systematically, to understand the behavior of forest fires so that they are more predictable eventually leading to quick response time and thus mitigating the loss due to forest fires. In our study, we have extracted texture features such as contrast, energy, entropy, correlation, color feature, normalized difference vegetation index (NDVI), and land surface temperature (LST) of Landsat 8 satellite images. All these feature extractions will help develop information systems on forest areas which include topographic, weather, vegetation data to predict the risk of forest fire.

Keywords Knowledge base · Multispectral image analysis · Landsat 8 · Feature extraction · Texture feature · Normalized difference vegetation indices · Land surface temperature

B. Zodge (✉) · J. Joglekar
K. J. Somaiya College of Engineering, Mumbai, India
e-mail: bhagyashri.z@somaiya.edu

J. Joglekar
e-mail: jyoti.joglekar@somaiya.edu

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_9

1 Introduction

Forests are significant in maintaining the ecological balance necessary for the sustenance of human life on this earth. Forest fires if not controlled in time not only destroy the environment and the ecological balance but also put life and property in danger [1]. Forests are generally huge acres of land that are remote and hence go unnoticed and are unmanaged filled with dry leaves, woods, and trees that act as a fuel source. Hence, it becomes difficult to know when fire outbreaks will happen and at which location.

Currently, several detection and monitoring systems are been used by forest departments such as aerial and satellite monitoring and systems based on different types of detection sensors and their combination. Machine learning and deep learning are other such techniques which are growing fast and are having significant impact with its result [2]. Such type of upcoming techniques provides ways to combine multiple modes of information such as spatial statistical ground data, weather data, and satellite imagery into a unified model for classification or prediction unlike the traditional ones [2].

But for all these systems to work properly the most significant step is; the process of building a rich quality knowledge base. Without having a good amount of information regarding the forest area such as forest type, climatic conditions, vegetation type, temperature, demographics of the area, and many more such features, the database is not wholesome which in turn will not give proper results, however, advanced the techniques might be. Many countries are developing forest fire databases and information systems to manage and control forest fires effectively.

Traditionally, cross-correlations are found out which is normalized cross-correlation coefficient that tells how good is the area with the pattern with which we are going to match. Other than that color feature, intensity feature, and perimeter feature (which is pre-analysis) are some of the features used in the previous work. In this paper, texture features such as contrast, entropy, energy, and correlation are extracted from band 8 (panchromatic) of Landsat 8 which has a high resolution as compared to other bands along with normalized difference vegetation index (NDVI) using infrared bands and color feature are extracted. Land surface temperature (LST) is also derived using thermal bands, and a knowledge base is created for prediction and detection of forest fires using Landsat 8 satellite imageries.

This paper focuses majorly on developing a quality knowledge base that eventually will be used for the prediction and detection of forest fires. The prediction model will be based on multivariate time series analysis which is in evolving stage and will be used to predict forest fire for different times of the day by evaluating time series data.

2 Related Work

Tao Cheng and Jiaqiu Wang explored the spatio-temporal data for data mining for forest fire prevention in their research work that uses a dynamic recurrent neural network for spatial forecasting. To understand and make use of all the data accumulated, it is required to study patterns and recognize the unexpected ones [1].

Sriram Ganapathi Subramanian and Mark Crowley presented a paper that discussed new deep reinforcement learning (RL) approaches that were suitable for time series analysis for change detection over time by finding patterns in the form of spatially spreading wildfires [2]. Paper on feature fxtraction from satellite imagery uses enhanced thematic mapper plus (ETM +) Landsat 7 optical imageries. Here, an automatic feature extraction algorithm is developed to generate and conserve spatio-temporal information on features. This algorithm processes data and broadcasts the feature information through a Web enabled information system [3]. A paper by Jerry Gao, Kshama Shalini, Navit Gaur, Xuan Guan, Sean Chen, Jesse Hong, Medhat Mahmoud analyzes big data for the detection of forest fire. The aim was to analyze forest fires using big data models. They used three different approaches that are video-based approach, satellite-based approach, and sensor-based approach [4].

Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein published a paper on textural features for image classification where they developed a set of features for classifying pictorial data. Here, they have presented textural features that are computable with a co-occurrence matrix and can be applied to different kinds of image data [5].

Another feature extraction paper is on color feature extraction by Divya Srivastava, Rajesh Wadhvani, and Manasi Gyanchandani where they have reviewed various color feature extraction methods. This paper presents the various color feature extraction techniques and in addition, gives a comparative study [6]. Another paper that helps understand extraction of yet another feature is a calculation of land surface temperature using Landsat 8 data for such as temperature vegetation index using ArcGIS software [7]. Authors Smys. S. and Jennifer Raj discussed using long-range technology for the assessment of fire risk where they have developed a network that includes long-range nodes and sensors that measure atmospheric changes and the CO2 level in the surrounding. This network was evaluated using network simulator 2 to compare against other existing wireless technologies [8].

In this work, we are developing knowledge base through satellite images, unlike other systems which collect data through sensors or camera hence this model is cost-effective. Also for getting ground level information, the satellite images are further segmented into 3 × 3 pixels tiles and then will be processed programmatically to extract features and analyze them.

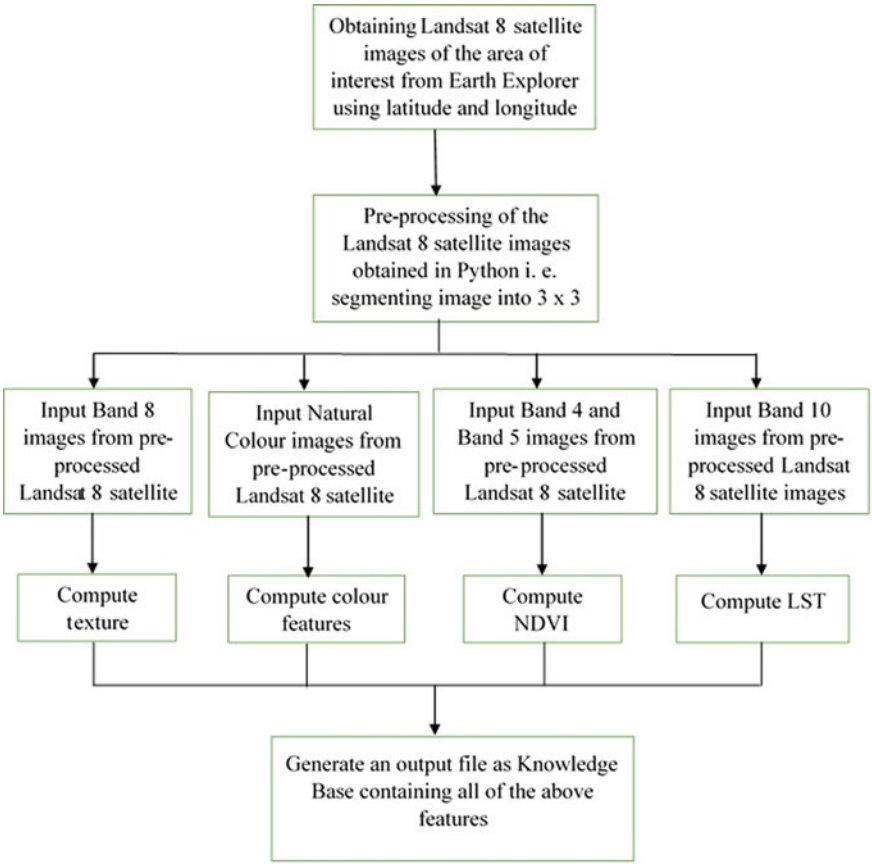
3 Methodology

3.1 Data Collection

The general difference between dataset and domain-based knowledge base is that the dataset is the complete raw data directly downloaded from the source, whereas domain-based knowledge base includes the complete processed data on which the model/algorithm is developed, and the results are analyzed.

For our study, we have selected the forest area of the Pauri region in the state of Uttarakhand, India. As there have been many major forest fire incidents reported in the Pauri region among many others, to start with we selected this region as an area of interest initially for the design and creation of a knowledge base for forest fire prediction and detection. Pauri forest area is situated at 30°08'49" N latitude and 78°46'28" E longitude. Landsat 8 satellite images dataset for Pauri, Uttarakhand is downloaded from https://earthexplorer.usgs.gov/ Website.

3.2 Flow Diagram: Design and Development of Knowledge Base



The above block diagram depicts the flow of creation of knowledge base, where once the Landsat 8 images are downloaded then it is passed to the program for further processing. The satellite images first are clipped into 90 × 90 pixels image according to the area of interest then it is further segmented into 3 × 3 pixels tiles for more closer and detailed study. Once 3 × 3 pixels images are ready then distinguishable features such as texture, color, NDVI, and LST are extracted and saved in CSV file format which is our domain-based knowledge base for prediction and detection of forest fires.

3.3 Landsat 8

Landsat 8 is a Sun-synchronous satellite and it completes one Earth orbit every 99 min. The satellite has a 16-day repeat cycle. A Landsat 8 scene size is 185 × 180 km [9].

Landsat 8 has 11 spectral bands including a pan band, different bands, their wavelength, and resolution are utilized for developing the knowledge base. In this paper, we have used band 4 (RED) and band 5 (near-infrared) having a resolution of 30 m for calculating normalized difference vegetation indices, band 8 (panchromatic) having a resolution of 15 m (higher than other bands) for extracting texture feature, and band 10 (thermal infrared) having resolution 100 m for calculating land surface temperature.

3.4 Image Analysis for Feature Extraction

Feature extraction techniques characterize interest points in an image. If the image size is huge then its important features, as a reduced version, are required to accomplish the desired task [10]. Features are said to be distinctive if they are clearly distinguished from the background, independent from radiometric and geometric distortion, interpretable, and unique [11, 12]. Common feature extraction techniques include:

- Texture feature extraction
- Color feature extraction
- Shape feature extraction
- Speeded-up robust features (SURF)
- Local binary patterns (LBP).

3.5 Texture Feature Extraction

In texture analysis, regions in an image are classified by various texture features. The intensities of pixels are used to quantify roughness, smoothness, or bumpy areas in an image [11, 12].

Texture analysis is used in various applications such as remote sensing, automated inspection, and medical image processing. A statistical method of examining texture is by developing a gray-level co-occurrence matrix (GLCM). It considers the spatial relationship of pixels [13].

Statistical texture features used in this work are energy, contrast, correlation, and entropy [5, 13, 14].

3.6 Color Feature Extraction

An image has three primary visual features, namely color, texture, and shape. There are two broad categories for color feature extraction, namely global methods and local methods [6]. In the global method, the complete image data are considered for feature extraction; while in the local method, a part of the image is considered for feature extraction.

3.7 Vegetation Indices Used for Image Analysis

By combining surface reflectance at two or more wavelengths, a specific character of vegetation is highlighted in the form of a vegetation index by using reflectance property [15].

There are various vegetation indices (VIs), with many being functionally equivalent. Some of them are as follows:

- Normalized difference vegetation index (NDVI)
- Infrared index
- Moisture stress index
- Leaf water content index (LWCI)
- Soil-adjusted vegetation index (SAVI)

The normalized difference vegetation index (NDVI) is the most commonly used vegetation index and derived from NIR and RED band intensity values [16].

$$\text{NDVI} = (\text{NIR} - \text{RED})/(\text{NIR} + \text{RED})$$

where NIR is the near-infrared band value, and RED is the red band value.

NDVI values are in the range from -1.0 to $+$ 1.0. A value less than zero does not interpret any ecology meaning. Higher values are associated with highly photosynthetically active vegetation [17].

3.8 Land Surface Temperature

LST, the earth's surface temperature, is measured in Kelvin [7]. Top of atmosphere (TOA), brightness temperature and normalized differential vegetation index (NDVI), land surface emissivity (LSE), and land surface temperature (LST) are measured using various bands of Landsat images [7].

4 Implementation Details

As we are using Landsat 8 satellite, images below are the snapshots of the scene of forest area in the Pauri region, Uttarakhand as it was on July 3, 2020 (Fig. 1).

The implementation details for knowledge base creation are given in the form of the following pseudocode as below:

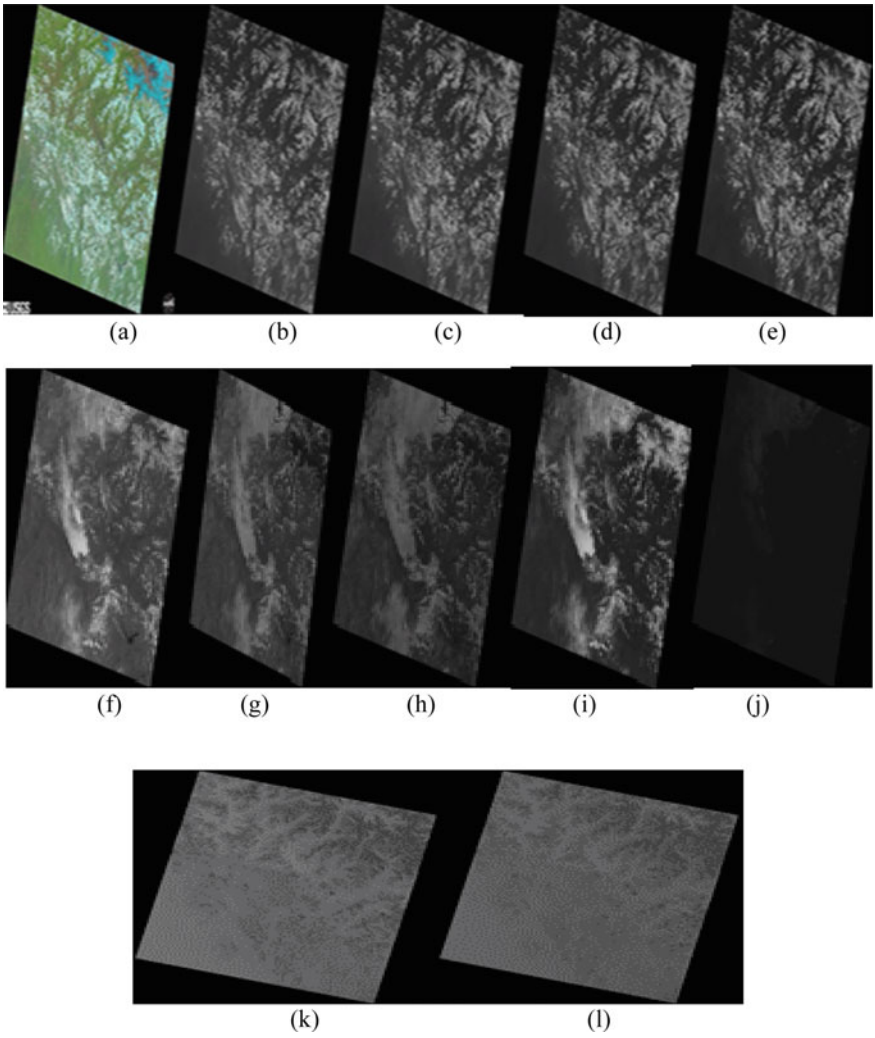


Fig. 1 Bands under. **a** Natural-color image. **b** Coastal aerosol wavelength. **c** Blue wavelength. **d** Green wavelength. **e** Red wavelength. **f** Near-IR wavelength. **g** Short-wave IR 1 wavelength. **h** Short-wave IR 2 wavelength. **i** Panchromatic. **j** Cirrus wavelength. **k** Thermal IR 1 wavelength. **l** Thermal IR 2 wavelength

```
//import libraries
//open images

#Segmentation of images

Create directories to save clipped images and segmented images band wise respectively
Get all images of the required bands
Clip those images into 90 x 90 image size and save them in the folder created for clipped images by giving path
Segment clipped images into 3 x 3 pixels tiles and save in the folder created for segmented images by giving path


#Features Extraction

extract bands for that spectral images
{
for texture features, extract band 8
using grey level co-occurrence matrix, greycoprops packages texture features are extracted
save the output in the respective CSV file
}

extract bands for that spectral images
{
for color feature, extract natural color band
set min and max value for color pixels for which count is to be determined (in this case red, green, blue, and yellow)
if the pixel value in the images come under that particular range of set min and max
{
calculate the number of pixels coming under that color pixels range
}
compare the count of the number of pixels for every color
the one with the highest count is the dominant color
save the output in the respective CSV file
}
```

```
extract bands for that spectral images
{
for NDVI, extract band 4 and band 5
using the standard formula mentioned in the methodology values are calculated
save the output in the respective CSV file
}
extract bands for that spectral images
{
for LST, extract band 10
using the steps described in the methodology for LST values are calculated
save the output in the respective CSV file
}

#Generating combined complete output file
Merge all of the above CSV files as a final output file which is the knowledge base
```

Computational complexity is high when high-resolution images are processed for various feature extractions. But with high-performance computing resources, it can be handled efficiently and the speed of the program will not be affected. Hardware specifications for the system we have used are as follows:

- AMD Ryzen.
- 48 GB RAM.
- 8 TB hard disk.
- Nvidia RTX 3080 Ti GPU.

Since we have worked with Python 3.6, some of the libraries we have used are as follows:

- Numpy.
- Python imaging library (PIL).
- Rasterio.
- Geospatial data abstraction library (GDAL).

With the above hardware and software specifications, it took 1 to 2 min for a single 3 by 3 image for all bands except band 8. For band 8 it took up to 15 to 30 min for a single image as its resolution is more as compared to other bands.

5 Results and Discussion

For our study purpose initially, 4 tiles with different types of landscapes of the forest area were clipped. The landscapes selected from forest areas are as follow (1) area with more green vegetation, (2) area with water body, (3) barren land, and (4) area of forest where the fire took place. Values from each forest area scenario (as described above) are taken as a single-entry sample and presented in table format for a particular 3×3 tile (here, for tile number 12_12.tif for 90×90 images of bands 4, 5, 6, 8, and 10 and natural color is taken) where each entry in the table corresponds to the forest area numbered in the manner mentioned above (Tables 1, 2, 3, and 4).

It is observed that the contrast of forest area with more green vegetation is higher than the contrast of forest area where the fire took place. Waterbody has the lowest contrast of all whereas barren land has the highest contrast of all. Values of energy range from 0.267261242 to 0.319438282 and values of entropy range from 2.419381946 to 3.169925001. The majority of the values for energy are on the higher side for all landscapes, whereas for entropy, majority of the values are on the lower side of the range. In addition, vegetation indices denote low to moderate types of vegetation around this forest area.

As per the features extracted by this knowledge base model whenever a landscape's contrast value is lower than green vegetation area but higher than water body area, entropy value is highest and vegetation is low, only then fire can be predicted. With visual inspection accuracy of the knowledge base created for prediction and detection of forest fire is beyond 80% taking into consideration texture features and vegetation indices. With more features extracted accuracy will surely go beyond 90%.

Thus, depending on the values obtained through feature extraction for fire and non-fire scenarios, parameters will be set for these features, and accordingly, predictions

Table 1 Texture feature

Contrast	Energy	Correlation	Entropy
2535.857	0.267261	−0.22396	3.169925
757.7143	0.267261	−0.08912	2.947703
145,725.1	0.267261	−0.08307	3.169925
1712.714	0.267261	−0.40743	3.169925

Table 2 Color feature

Red pixel count	Green pixel count	Blue pixel count	Yellow pixel count	Dominant color
0	9	0	0	G
0	0	9	0	B
6	0	3	0	R
0	9	0	0	G

Table 3 Normalized difference vegetation ındex

Min. NDVI	Mean NDVI	Median NDVI	Max. NDVI	Vegetation
−0.24125	−0.19235	−0.18772	−0.16341	Low
−0.19236	−0.14824	−0.14526	−0.12838	Low
−0.00324	0.009812	0.012497	0.023481	Low
−0.13841	−0.13337	−0.13473	−0.12767	Low

Table 4 Land surface temperature

Min LST	Max LST
34.3086	34.4332
37.21302	38.20803
26.70073	30.85189
20.37411	20.4577

can be made. That is (in this case), whenever the values of contrast, entropy, and vegetation approach, the minimum limit alerts can be sent to the concerned authorities to take immediate actions to combat fire within a short time.

6 Conclusion

This paper presents a methodology to develop an auto-enabled algorithm for a domain-based knowledge base for the prediction and detection of forest fires for a particular latitude and longitude. Unlike other systems where sensors and optical systems are used for data generation and monitoring, our model is more cost-effective. Since there will be a minimum cost of installation and maintenance of the devices. Further, this knowledge base can be used to develop models/systems for forest fire risk prediction and detection in case of forest fires. Also, more features can be added into the knowledge base which can help to predict and detect forest fires other than the ones mentioned in this paper such as soil moisture index, wind speed, relative humidity, and many more.

This methodology can be used on a larger scale to develop a knowledge base at different locations and further by studying every scenario's feature values at that location various models can be developed using different techniques for forest fire risk prediction and detection in the case of forest fires.

References

1. T. Cheng, J. Wang, Integrated spatio-temporal data mining for forest fire prediction. Trans. GIS **12**(5), 591–611 (2008)

2. S. Ganapathi Subramanian, M. Crowley, Using spatial reinforcement learning to build forest wildfire dynamics models from satellite ımages. Front. ICT **5**(6) (2018)
3. C.G. Aishwarya, T.M. Apoorva, S.R. Shetty, B.R. Vandana, V. Sangeetha, Feature extraction from satellite ımagery. Int. J. Adv. Res. Comput. Commun. Eng. **6**(5) (2017)
4. J. Gao, K. Shalini, N. Gaur, X. Guan, S. Chen, J. Hong, Medhat mahmoud: data-driven forest fire analysis (2017). IEEE 978-1-5386-0435-9/17
5. R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for ımage classification. IEEE Trans. Syst. Man Cyber. **3**(6), 610–621 (1973)
6. D. Srivastava, R. Wadhvani, M. Gyanchandani, Review: color feature extraction methods for content based retrieval ımage. IJCEM Int. J. Comput. Eng. Manag. **18**(3), 9–13 (2015)
7. D. Anandababu, B.M. Purushothaman, S. Suresh Babu, Estimation of land surface temperature using LANDSAT 8 data. Int. J. Adv. Res. Ideas Innov. Technol. **4**(2), 177–186 (2018)
8. S. Smys, J.S. Raj, Assessment of fire risk and forest fires in rural areas using long range technology. J. Electron. **2**(01) 38–48 (2020)
9. Landsat 8, https://www.usgs.gov/core-science-systems/nli/landsat/landsat-8
10. Feature Extraction, https://in.mathworks.com/discovery/featureextraction.html
11. J. Joglekar, S.S. Gedam, Area based ımage matching methods—a survey. Int. J. Emerg. Trends Adv. Eng. **2**(1) (2012)
12. J. Joglekar. Class lecture: texture feature. Department of Computer Engineering, K. J Somaiya College of Engineering, Vidyavihar, Mumbai, February. 12, 2020
13. Texture Analysis, https://in.mathworks.com/help/images/texture-analysis-1.html
14. GLCM, https://in.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html
15. Entropy, https://in.mathworks.com/help/images/ref/entropy.html
16. Vegetation Indices, Internet: http://www.harrisgeospatial.com/docs/VegetationIndices
17. J. Joglekar, Class lecture: IR ımage processing. Department of Computer Engineering, K. J Somaiya College of Engineering, Vidyavihar, Mumbai, December. 19, 2019

IoT-Based Smart Water Quality Monitoring System to Expand Sensors Life and Battery Power

Vikas Malhotra

Abstract Development in wireless communication technology increase the applications of sensors more efficiently and flexibly. In the proposed method, energy management and movable sensors have been used to increase the battery life as well as sensors life. Real-time clock-based power control methodology has been used in the proposed method for energy management. The average data of 60 samples have been collected and published to the cloud server to overcome the variation of the data from sensors. The user can vary the time and samples in the programming according to need. The mean square error with value 0.9409, 0.22, and 2326.33 of turbidity, pH and TDS has been calculated, respectively.

Index Terms IoT · Water quality monitoring · Power management · Actuator · Dissolved oxygen · Sensors

1 Introduction

In the water-scant countries, wastewater reclamation and its reuse have become the desegregated part of the water resources conversation. Due to the shortage of freshwater resources for many applications, the other option is to reuse the wastewater after water treatment. It is necessary to monitor the different contamination before, after and during water treatment so that the quality of processed water meets the standard regulation [1]. Besides, the measurement of quality of water for drinking, the quality of water measurement is also important for agriculture, human life, ecosystem and industrial purpose [2].

For the water quality measurement process, during sensor selection besides, determine the different contaminants there are more parameters such as sensor effectiveness, low cost and easy to use are also important. Industrial, commercial, agriculture and domestic pursuits are the main reasons to generate harmful materials in water [3].

Existing wastewater treatment plants are not capable to use in real-time control systems because of the non-availability of controllable actuators that work with flexibility [4]. However, the need for new technologies to monitor the wastewater on

V. Malhotra (✉)
Thapar Institute of Engineering and Technology, Patiala, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_10

a real-time basis is increasing with the increasing of variable pollutants composition in water with time [5].

The traditional technology used to determine water quality mostly requires data sampling and sampling with pretreatment makes this technology slow and cost-effective. Therefore, biosensors have been developed to determine the pollutants so that more sensitive, portable and on-site measurements can be performed. Different kind of biosensor such as optical, electrochemical, piezoelectric and thermal biosensors can be used to determine three types of water contaminants like organic materials (various aromatic substances, hydrocarbons, diphenyl methane, surfactants and chloride components), heavy metals (chromium, lead, mercury, antimony, cadmium arsenic) and microorganism (viruses, bacteria, protozoa, fungi, yeasts and algae) within the wastewater. The main source of organic material in water is the effluents of agriculture, industrial and household. The main source of heavy metals in water is mining, agriculture, sullage of chemical manufacturing, cosmetic products, painting, metal processing and nuclear waste, etc. Similarly, the main source of microorganisms in water is sewage water in the household and food industries.

In wastewater treatment, the monitoring of organic matter is important for human and environmental health. Biochemical organic demand (BOD) and chemical organic demand (COD) parameters are generally used to detect the concentration of organic matters in wastewater with electrochemical and thermal sensors, respectively. However, there is no sensor developed to determine many organics such as pesticides and herbicides from agriculture sewerage. The nonessential heavy metals are toxic and cause various chronic diseases in human beings. Different biosensors such as optical, electrochemical, piezoelectric sensors and microfluidic devices can be used to measure heavy metals but these biosensors can detect only free heavy metal ions. Single-phase and transparent samples (i.e., turbidity less than one nephelometric turbidity unit (NTU)) can be directly analyzed with biosensors otherwise pretreatment of sample is needed.

Remarkable morbidity and deaths in the world are due to a large range of diseases created from waterborne microorganisms. These microorganisms can be detected by using optical and electrochemical biosensors in real-time [6].

The parameters used to determine these contaminations in the water to analyze the quality of water can be divided into three types, i.e., physical (temperature, total dissolved solids (TDS), turbidity, color, taste, flow rate, pressure), microbiological (E Coli, total coliform) and chemical (pH, dissolved oxygen (DO), NO2, NO3, heavy metals) [7].

The main drawback of raspberry pi is that it has I2C channels with limited numbers [8]. The water quality sensors of American-based company, i.e., Atlas Scientific is widely used and these sensors have high performance [9]. For the usage of many sensors with huge data, IoT technology based on the cloud is reliable and cost-effective [10].

IoT is an ingenious technological case and is mostly used to accumulate, monitor and analysis of data from remote locations [11, 12] IoT system employ two types of systems, i.e., device-to-cloud and device-to-device communication. By using cloud storage, it is easy to manage and analyze data with online monitoring [13]. This

technology can be used for water monitoring and controlling system. In the proposed method, IoT-based device–to–cloud communication is used. The proposed technique can be incorporated with other nodes for smart cities applications as in various applications such as smart health; smart farming is generally using integrated network clients with data collection IoT devices to send the information to focal station [14].

For low maintenance cost and long device life, it is necessary to use some energy harvesting or management technique for the development of a node or device [15]. Some researchers has also worked at nodes of IoT networks for energy management using different machine learning techniques such as smart ant optimization algorithm [16], Q-learning [17], Naive Bayes, Multilayer Perceptrons (MLP), and Support Vector Machine (SVM) classifiers [18]. However, these techniques are efficient for large number of nodes because they mostly concentrate on shortest routing path within the network for data propagation. In proposed method, management of energy has been focused on single node wireless sensor network.

2 Proposed Methodology

In the proposed system, four low-cost sensors have been used (DS18B20 temperature sensor, SKU-SEN0189 turbidity sensor, SKU-SEN0244 TDS sensor and SKU-SEN0161 pH sensor). The power supply to the sensors and Wi-Fi module has been controlled by the relay as shown in Fig. 1 as the dotted line represents the power supply. The Thinkspeak cloud server is used for data publication.

In the proposed method, energy management with the use of a real-time clock (RTC) has been designed to increase battery life using power 'ON' the all sensors and Wi-Fi module with the help of relay switches, at predefine time schedules. Sensors movement has been also controlled with the help of a lead screw, a stepper motor to dip and raise the sensors from the water so that the life of sensors can be increased. Five parameters were collected from the anaerobic tank at the wastewater treatment plant. The value of pH gets alter, when the DO sensor is also immersed in the sample with the pH sensor during data collection. This is due to the interference of the DO sensor and pH sensor through the sample. Therefore, to collect the correct value of pH and DO, both sensors were powered in different time slots in the proposed method.

For every hour in 24 h the firstly, power is given to DO and temperature sensor using relays for two minutes and data from both sensors are collected and stored. Then, power is given to turbidity, TDS, pH sensors and Wi-Fi module for 18 min so that data can be collected from these sensors for 20 min in the form of 60 samples The average of these samples are further calculated in the Arduino board and then these average data of each sensor published to the Thingspeak cloud server such that one average data of each sensor has been published in every hour as shown in Fig. 2. The output variation of sensor data is reduced by taking an average of the 60 data samples.

The number of samples and timing can be varying according to requirement by changing the program in Arduino. Figure 3 represents the snapshot of the developed system.

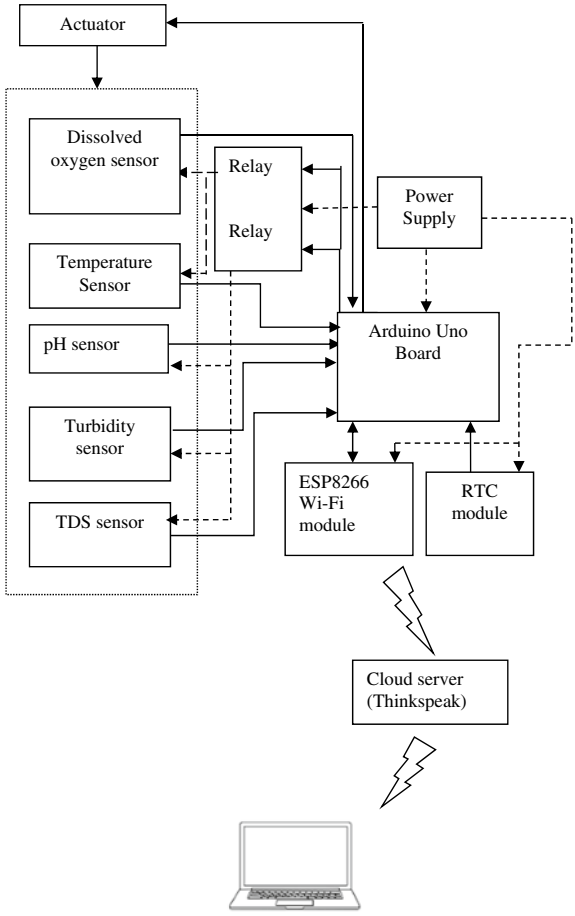


Fig. 1 Proposed system

2.1 Parameters Measured

In the proposed method, four following parameters are measured.

(1) **Temperature**: Temperature is an important parameter to measure because a large amount of variation affects the values of other parameters (pH and conductivity) [19].

(2) **Potential Hydrogen (pH)**: The Potential Hydrogen (pH) measurement is used to measure the concentration of hydrogen ions in the water and can be defined as pH $= -$ log [H+]. The water is acidic if the pH level of water is less than 7 but if the pH level of the water is greater than 7 then water is alkaline and if the pH level is 7 then water is neutral [20]. The pH sensor from DF Robot is used in the proposed method. It has an accuracy of ± 0.5 at a temperature of 25 °C.

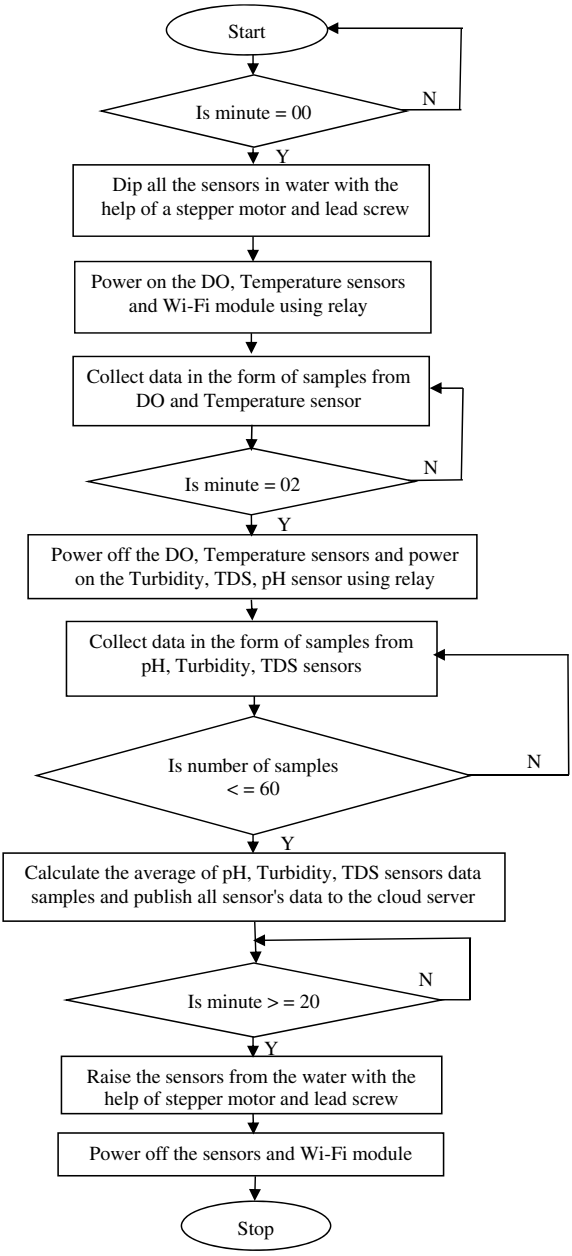


Fig. 2 Flow chart of the methodology used in the proposed system

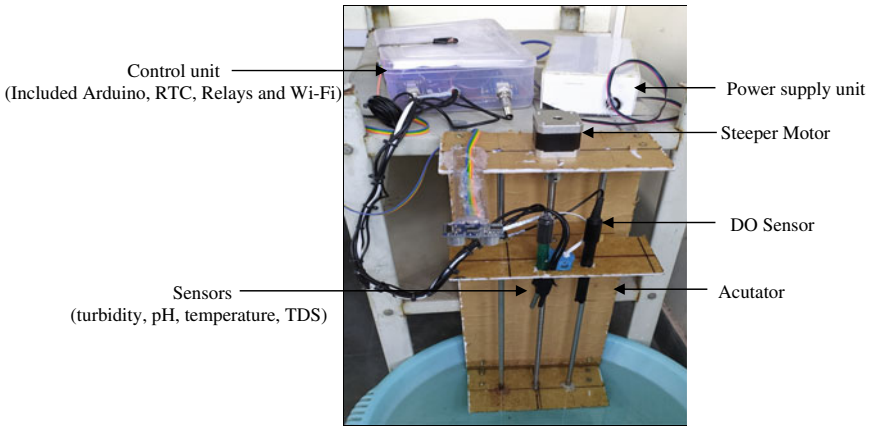


Fig. 3 Developed proposed system

(3) **Turbidity:** Colloidal materials, dissolved and suspended solids in water are the reason for the reduction in transparency of water. The measurement of this transparency is known as Turbidity. Due to the inclusion of suspended materials, turbid water has less transmission of light within the water as compared to clear water [21].

(4) **Total Dissolved Solids (TDS)**: Constituent ions of inorganic salts generate the TDS in the water. The small amount of organic matter that may be dissolved in water can be determined by TDS [22].

(5) **Dissolved oxygen (DO)**: The measure of oxygen gas condensed into the water named as Dissolved Oxygen of the water. It is additionally important parameter to investigate water quality and addressed as oxygen in milligram in per liter water (mg/L) [23].

3 Result

The 24 h data of four parameters collected using the proposed system is shown in Table1. The turbidity value of 1.39, ph value of 7.774 and highest TDS value of 478.84 have been given by the sensors, whereas reference values of pH were 7.3 measured from Hanna instruments (HI5221), turbidity value of 5.36 NTU measured from Hach 2100Q Turbidity meter and TDS value of 430 ppm measured from Handhold TDS Tester. The mean square error of 0.9409, 0.22, and 2326.33 of turbidity, pH, and TDS has been calculated, respectively, (Figs. 4, 5, 6, 7, 8 and 9).

Table 1 Average data of different parameters

Time	Temperature (*C)	Turbidity (NTU)	pH	TDS (ppm)	DO (mg/L)
2:18 PM	29.8	4.39	7.774	476.39	2.4
3:18 PM	30.5	4.39	7.774	477.08	2.0
4:18 PM	31.9	4.39	7.774	477.36	1.8
5:18 PM	32.3	4.39	7.774	477.79	1.4
6:18 PM	32.1	4.39	7.774	477.97	1.4
7:18 PM	31.9	4.39	7.774	478.12	1.4
8:18 PM	31.9	4.39	7.774	478.21	1.4
8:18 PM	31.9	4.39	7.774	478.21	1.4
9:18 PM	32.0	4.39	7.774	478.28	1.3
10:18 PM	31.9	4.39	7.774	478.33	1.3
11:18 PM	31.9	4.39	7.774	478.37	1.3
12:18 AM	31.9	4.39	7.774	478.39	1.3
1:18 AM	31.9	4.39	7.774	478.43	1.3
2:18 AM	31.9	4.39	7.774	478.46	1.1
3:18 AM	31.9	4.39	7.774	478.48	1.1
4:18 AM	31.7	4.39	7.774	478.49	1.1
5:18 AM	31.7	4.39	7.774	478.51	1.1
6:18 AM	31.7	4.39	7.774	478.54	1.1
7:18 AM	32.1	4.39	7.774	478.62	1.1
8:18 AM	32.2	4.39	7.774	478.69	1.0
9:18 AM	32.4	4.39	7.774	478.74	1.0
10:18 AM	31.2	4.39	7.774	478.83	1.0
11:18 AM	31.5	4.39	7.774	478.84	0.9
Hach turbidity instrument (Model: 2100Q)		5.36			
Hanna pH instrument (Model: HI5221)			7.3		
Handheld TDS tester				430	
MSE		0.940	0.22	2326.33	

4 Discussion

In proposed design, power is supplied the sensors and Wi-Fi module, when there is requirement of water parameters readings otherwise all the sensors and Wi-Fi module remains in Power 'OFF' situation using real-time clock (RTC) module. It helps to save unnecessary power wastage even there is no need to analyze water

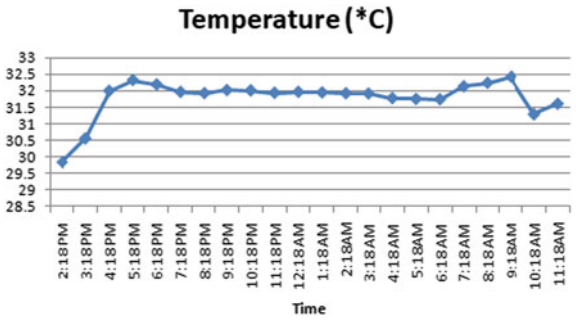


Fig. 4 Graph of temperature data

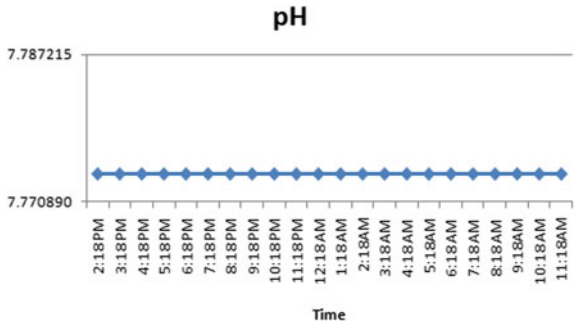


Fig. 5 Graph of pH data

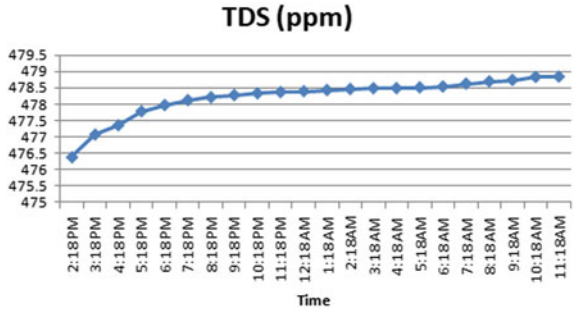


Fig. 6 Graph of TDS data

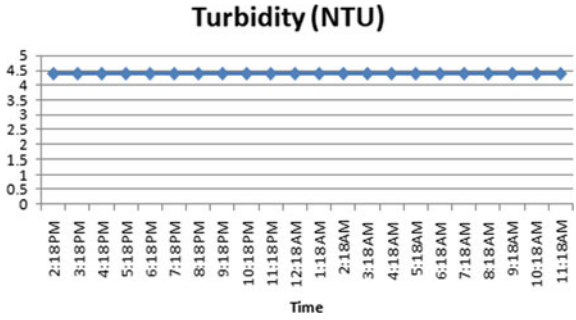


Fig. 7 Graph of turbidity data

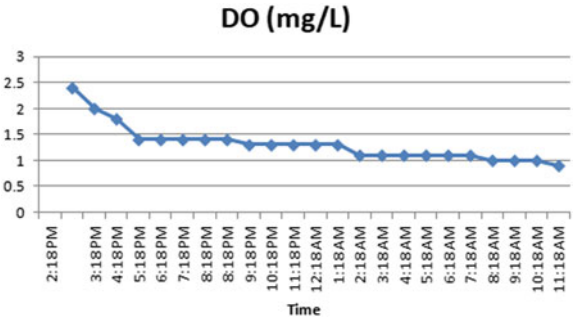


Fig. 8 Graph of different parameters data

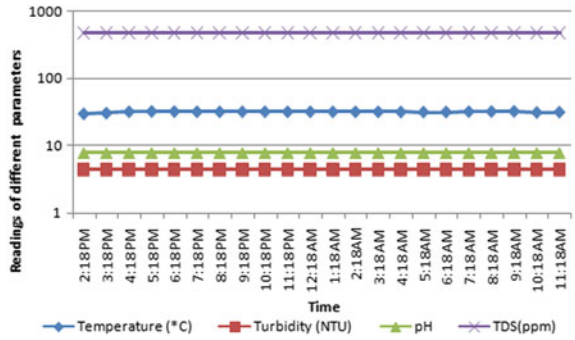


Fig. 9 Graph of different parameters data

parameters as contradiction to traditional water quality monitoring system there is no such mechanism. Furthermore, flexible actuator dips the sensors into the water, when water parameters have to be collect otherwise it rises out the sensors out of the water so that sensors life can be increased.

5 Conclusion

If the distance between dipped pH and DO sensors is less than 15 feet during data collection, then pH and DO sensors should be isolated from each other for correct value extraction of pH and DO parameters. In the proposed method, two-time slots have been used to isolate the pH and DO sensors. The battery life increased to the 2 times using proposed method as compared to the using the system without power switching. In the future work, integrated circuits (ICs) can be used in the system circuit for isolation. The proposed system can be used for water quality monitoring systems with less power requirement and increased sensor life.

References

1. N.A. Abdullah, S. Ramli, H. Mama, S. Khan, C. Gomes, Chemical and Biosensor technologies for wastewater quality management. Int. J. Adv. Res. Publ. **1** (2017)
2. T.I. Salim, H.S. Alam, R.P. Pratama, I.A.F. Anto, A. Munandar, Portable and online water monitoring system using a wireless sensor network, in *International conference on automation, cognitive science, optics, micro-electro-mechanical system and information technology* (Indonesia, 2017)
3. S. Geetha, S. Goutthami, Internet of things enabled real-time water quality monitoring system. vol. 2, Springer Open (2017)
4. P.A. Vanrolleghem, D.S. Lee, On-line monitoring equipment for wastewater treatment process: state of art. Water Sci. Technol. **47**, 1–34 (2003)
5. V.V. Daigavane, M.A. Gaikwad, Water quality monitoring system based on IoT. Adv. Wireless Mobile Commun. **10**, 1107–1116 (2017)
6. F. Ejeian, P. Etedali, H.M. Tehrani, A. Soozanipour, Z.L.M. Asadnia, A.T. Kafrani, A. Razmjou, Biosensors for wastewater monitoring: a review. Biosens. Bioelec. **118**, 66–79 (2018)
7. S.O. Olantinwo, A.T. Joubert, Enabling communication networks for water quality monitoring applications: a survey. IEEE Access **7** (2019)
8. P. Khatri, K.K. Gupta, R.K. Gupta, Raspberry Pi-based smart sensing platform for drinking water quality monitoring system: a python framework approach. Drinking Water Eng. Sci. **12**, 31–37 (2019)
9. A.T. Demetillo, M.V. Japitana, E.B. Taboada, A system for monitoring water quality in a large aquatic area using wireless sensor network technology. Sustain. Environ. Res. **29** (2019)
10. K. Saravanan, E. Anusuya, R. Kumar, L.H. Son, Real-time water quality monitoring using internet of things in SCADA. Environ. Monit. Assess. **190** (2018)
11. M.S.U. Chowdury, T.B. Emran, S. Ghosh, A. Pathak, M.M. Alam, N. Absar, K. Andersson, M.S. Hossain, IoT based real-time river water quality monitoring system. Procedia Comput. Sci. **155**, 161–168 (2019)
12. H. Kumar, Involvement of cloud computing and iot in the field of health care. Int. J. Res. Eng. Sci. Manag. **4** (2021)
13. S.C. Abraham, Internet of Things (IoT) with Cloud Computing and Machine-to-Machine (M2M) Communication. Int. J. Emerging Trends Sci. Technol. **3**, 4654–4661 (2016)
14. N. Balaji, Reliable data transmission with heighted confidentiality and integrity in IoT empowered mobile networks. J. ISMAC **2**, 106–117 (2020)
15. O. Khutsoane, B. Isong, N. Gasela, A.M. Abu-Mahfouz, water grid-sense: a LoRa based sensor node for Industrial IoT applications. IEEE Sensors J. **20** (2020)
16. J.L.Z. Chen, K. Lai, Machine learning based energy management at internet of things network nodes. J. Trends Comput. Sci. Smart Technol. **2**, 127–133 (2020)
17. Prabuchandran, S.K. Meena, S. Bhatnagar, Q-Learning based energy management policies for a single sensor node with finite buffer. IEEE, Wireless Commun. Lett. **2**, 82–85 (2013)
18. A.Y. Barnawi, I.M. Keshta, Energy management in wireless sensor networks based on naive bayes, MLP, and SVM Classifications: a comparative study. J. Sens. (2016)
19. A.N. Parsad, K.A. Mamun, F.R. Islam, H. Haqva, Smart water quality monitoring system. IEEE Asia Pacific World Congress Comput. Sci. Eng. Conf. (2015)
20. K. Spandana, V.R.S. Rao, Internet of things (IoT) based smart water Monitoring system. Int. J. Eng. Technol. **7**, 259–262 (2018)
21. K.F. Poe, W. Payne, R. Emanuel, *Water Quality and Monitoring* (University of Arizona, College of Agriculture and Life Sciences, 2005)
22. S.M.S. Nadeem, R. Saeed, Determination of water quality parameters of water supply in different areas of Karachi city. Euro. Acad. Res. **1** (2014)
23. G. Kibria, Dissolved oxygen the facts. Tech. Rep. Researchgate, August (2017)

A Study on Machine Learning-Based Approaches for $PM_{2.5}$ Prediction

V. Santhana Lakshmi and M. S. Vijaya

Abstract Clean air and water is the fundamental need of humans. But people are exposed to polluted air produced due to several reasons such as combustion of fossil fuels, industrial discharge, dust and smoke which generates aerosols. Aerosols are tiny droplets or solid particles such as dust and smoke that floats in the atmosphere. The size of the aerosol also called as particulate matter ranges from 0.001–10 μm which when inhaled by human affects the respiratory organs. Air pollution affects the health of 9% of the people every year. It is observed as the most important risk factor that affects human health. There is a need for an efficient mechanism to forecast the quality of air to save the life of the people. Statistical methods and numerical model methods are largely employed for the predicting the value of $PM_{2.5}$. Machine learning is an application of artificial intelligence that gives a system capability to learn automatically from the data, and hence, it can be applied for the successful prediction of air quality. In this paper, various machine learning methods available to predict the particulate matter 2.5 from time series data are discussed.

Keywords Aerosols · Particulate matter · Machine learning (ML) · Forecast · Predicting · Artificial ıntelligence (AI) · Time series data

1 Introduction

People when exposed to polluted air for a long time are affected with lung disease. This is due to the presence of harmful substances in the air. This section explains the role of polluted air in affecting human morbidity and mortality. Dust also called as Aeolian dust present in the dry region is injected into the atmosphere by high velocity wind. This forms aerosol in troposphere. It produces radiative forcing and

V. Santhana Lakshmi (✉) · M. S. Vijaya
Department of Computer Science (PG), PSGR Krishnammal College for Women, Coimbatore, India

M. S. Vijaya
e-mail: msvijaya@psgrkcw.ac.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_11

is responsible for the change in temperature, ocean cooling and alters the rainfall amount. Aerosols are small particles that float in the air. The open ocean is a significant source of natural aerosols. It produces 1015–1016 g of sea-salt aerosols annually. Sea-salt aerosols, along with wind-blown mineral dust, forms natural tropospheric aerosols. The radiative processes in the clean atmosphere, namely reflection, transmission and absorption, are affected by the physical and chemical properties of background aerosols. The primary environmental risk factor that poses the greatest threat to human life is air pollution. A person when inhales the polluted air for a long time is affected by the disease such as asthma, bronchitis, ventricular hypertrophy, Parkinson's diseases and Alzheimer [1]. Forecasting the air quality is essential so that people receive the information earlier about the increase in pollution level, thereby saving the life of the people. Time series data that are used for identifying $PM_{2.5}$ value includes meteorological data, geographical data, traffic data and satellite data in the form of images. A good choice of data and methodology plays a main role in the accurate prediction of air quality.

2 Air Pollution and Air Quality Prediction

Air gets contaminated when some solid particles and gases suspends in the air. This section introduces various factors that cause pollution and harmful substances in air. Natural sources of air pollution include dust from earth surface, sea salt, volcano eruptions and forest fire. Man-made sources of air pollution include industry emission, transportation emission and agriculture [2].

2.1 Natural Causes of Air Pollution

Volcano during eruption will emit sulphur dioxide. It produces respirable acid after reacting with water vapour and sunlight. Finally, it results in vog which is visible haze. The lava enters into the ocean to make sea water boil. As a result, it creates thick clouds of "laze". These clouds are filled with hydrochloric acid. The level of air pollution depends on the speed and direction of wind, vog and ash. This affects the respiratory organs of humans [3].

Forest fire will affect the meteorology of the surrounding area. It has severe impact on the quality of the air. Toxic gases are emitted while burning of biomass in the forest. These gases move several miles from their point of origin. The air quality in the surrounding areas also suffers as a result of this. During the fire period in Uttarakhand, the concentrations of contaminants such as nitrogen oxide, nitrogen dioxide, carbon monoxide, PM_{10} and $PM_{2.5}$ increased tremendously [4].

2.2 Man Made Pollution

The water bodies are polluted due to unauthorized release of industrial wastewater. The waste water includes toxic chemicals which is mainly responsible for health hazards in living beings. Untreated waste water contributes to the existence of toxins in water sources such as lakes, rivers and groundwater detrimental to the health of plants, livestock and human beings. Such toxins are mostly responsible for polluting sacred rivers such as the Ganga, Yamuna and others. The water was contaminated, and people could not use it for drinking, bathing, and other purposes [5].

Growth of motor vehicles and rapid urbanization results in the deterioration of air quality. As per [6] in 2020, over 21.5 million vehicles were sold domestically. Exhaust from vehicles such as hydrocarbons, sulphur dioxide, lead/benzene, nitrogen dioxide, carbon monoxide and particulate matter are the major sources of outdoor air pollution worldwide. Public were affected by respiratory diseases and cardio-vascular diseases in India. As per the report of India Times, the vehicle tailpipe emission were linked to about 361,000 premature deaths from $PM_{2.5}$ and ozone in 2010 and about 38,500 in 2015.

2.3 Air Pollutants

Air pollution occurs when harmful substances are suspended in the atmosphere. Harmful substances include lead, carbon monoxide, ozone, nitrogen dioxide and sulphur dioxide.

Carbon monoxide is an odourless and colourless gas that is present in the air. It is toxic for humans to ingest more carbon monoxide. The normal carbon monoxide concentration that does not affect humans is about 0.2 parts per million (ppm).When the level of carbon monoxide consumed by a person increases, it lowers the amount of oxygen in red blood cells that is transported by haemoglobin throughout the body. As a result, essential organs such as the heart, brain and nervous tissues are deprived of sufficient oxygen to function. Bush fires and volcanoes are the natural sources of carbon monoxide.

Lead is a heavy metal present in the air. Humans are exposed to lead through inhalation process. Once enters within the body, it flows through the bloodstream until it reaches the bones. Lead has an impact on the nervous system, kidneys, digestive system and reproductive system, depending on the level of exposure. Lead poisoning reduces the blood's ability to contain oxygen. Motor vehicle waste and certain manufacturing processes are the primary sources of lead and carbon monoxide.

Nitrogen dioxide has an unpleasant odour. Plants contain a small amount of nitrogen dioxide. Sometimes, it is formed naturally in the atmosphere by lightning. Nitrogen dioxide has a larger role in the production of photo-chemical smog, which has serious health consequences. Breathing nitrogen dioxide will have significant

impact on people who are already affected with respiratory diseases. Older people and children with asthma and heart diseases are most at risk.

Ozone gas contains three atoms of oxygen (O_3). There are two types of ozone. One that is created in the upper atmosphere shields us from ultraviolet rays, and another type is formed in the ground level. Ground-level ozone is one of the harmful air pollutants. Ozone is created by chemical reactions between nitrogen oxides (NOx) and volatile organic compounds (VOC). The pollutants emitted by the cars, industrial boilers, power plants, refineries, chemical plants and other sources react in the presence of sunlight [7]. Breathing ozone can cause coughing, throat irritation, chest pain and airway inflammation among other things. Patients with bronchitis, emphysema and asthma will get severe infection when exposed to it for a long time.

Particulate matter is abbreviated as PM. Particle contamination is another name for it. It represents a combination of solid particles contained in the air, such as ashes, ash and soot, as well as liquid droplets. Some particles are visible to the naked eye and some can be seen only through electron microscope. Particle pollution includes $\mathbf{PM_{10}}$ and $\mathbf{PM_{2.5}}$. Both are inhalable particles with diameters of 10 and 2.5 micrometres, respectively. Exposure to such particles affects the lungs, heart and leads to many respiratory problems.

Sulphur dioxide is a noxious and translucent substance with a pungent odour. Sulphuric acid, sulphurous acid and sulphate particles are formed when it interacts with other chemicals. Major sources of sulphuric acid are humans and industrial waste. Burning of fossil fuels also creates sulphuric acid. It creates irritation in nose, throat and causes coughing. It gives tight feeling around the chest. Asthma patients are at greater risk of being severely affected after being exposed to sulphuric acid.

When carbon monoxide, lead, nitrogen dioxide, ammonia, particulate matter and sulphur dioxide are mixed with air, it becomes polluted. Identifying the amount of these pollutants in air helps to solve many research problems. Correlation between amount of these pollutants in air and human health can be identified. Lot of researches are going on to identify the contribution of these factors in affecting the environment. It is possible to predict the quality of vegetation by identifying the amount of these pollutants in air. Many countries are trying to identify the correlation between vehicle emission and deterioration of air quality. Lot of computational methods are available for identifying these pollutants.

3 Computational Methods for Air Quality Prediction

Artificial intelligence, machine learning, deep learning, big data analytics, cloud and Internet of things (IoT) are the technologies causing revolutions in various industries. Applications developed by combining these technologies perform efficiently well and execute without the human intervention. These technologies help to analyse large datasets of complex data types with less computational difficulty. Apart from traditional statistical methods, machine learning and deep learning algorithms are

effective in predicting the airborne pollutants. This section describes various methods that are used for predicting air pollutant concentration.

3.1 Statistical Methods

The statistical methods use historical data to forecast the future value. For predicting air quality, statistical methods such as simple moving average, exponential smoothing and auto regressive moving average can be used. Simple moving average method is used to identify the long-term trend whether upward or downward in the time series data. It calculates the average of finite number of values from the historical data to forecast the new value. Air quality can be predicted using historical meteorological and geographical data.

Exponential smoothing predicts the future value by calculating weighed average of past values. Exponential smoothing assigns decreasing weights for older observations and more weights for recent observation. It smoothens the time series data. There are three types of exponential smoothing methods. It includes simple exponential smoothing, double exponential smoothing and triple exponential smoothing. Simple exponential smoothing was best suited for univariate data without trend. It uses smoothing parameter alpha. The speed at which the effect of previous time steps' observations decays exponentially is controlled by this parameter. The value of this parameter lies between 0 and 1. As the name implies, double exponential smoothing uses two smoothing parameters such as trend component and level component at each period. Triple exponential smoothing method also called as Holt Winters method uses third parameter when the data has seasonality in it.

Autoregressive moving average model also uses historical data to forecast the future value, but it expects the data to be stationary. If the time series data is not stationary, differencing is performed to make it stationary. In order to fit the ARIMA model three important values p, d, q has to be identified. p and q represent order of auto regression and order of moving average and d is the order of differencing. ARIMA and exponential smoothing are the most popular statistical methods used for forecasting.

3.2 Machine Learning

Machine learning is a form of artificial intelligence that builds an intelligent model based on historical data to predict the target. Machine learning algorithms are used to train models which can perform decision making without being explicitly programmed. Regression, auto regression and support vector machine are machine learning techniques that can be used to forecast pollution concentration.

Forecasting the quality of air is very much essential since it has more impact on the human health. Among all the pollutants, particulate matter plays main role in

increasing the mortality rate. Lot of methods are available to predict the $PM_{2.5}$ level in air. The simplest method is to build regression model.

Regression is a statistical technique used for identifying and analysing the relationship between dependent variables and independent variables. It is a predictive modelling technique used to predict the real value. Linear regression and logistic regression are the two types of regression algorithms. In simple linear regression, the relationship between single dependent variable and independent variable is analysed. Multiple linear regression is a variation of simple linear regression in which the relationship between single dependent variable and multiple independent variables are analysed. Regression algorithms are used to forecast the $PM_{2.5}$ value based on the multiple independent variables such as wind speed, temperature, precipitation, humidity and rainfall.

A time series model called an auto regression (AR) uses observations from previous time steps as input to the value at the next step. It is used to make predictions because there is a connection between the values in a time series and the values that come before and after them. The behaviour is modelled using historical evidence, thus the term autoregressive.

The support vector machine (SVM) is a new kind of classification algorithm that uses statistical learning theory to operate. Support vector regression (SVR) is used to investigate the concentration variation of air pollutants. The basic principle behind support vector regression is to use a non-linear mapping unknown function to map the original data x into a feature space F with high dimensionality, and then use linear regression in this space. SVR permits the user to specify the minimum amount of error that is accepted while prediction. This is one of the advantages of SVR, and the computational complexity does not depend upon the dimensionality of the dataset. So, when the time series data of previous years are used for training the model, the algorithm considers these factors as independent variables and tries to fit the line to predict the dependent variable which is air quality index.

3.3 Deep Learning

Deep learning is a subset of machine learning that simulates human problem-solving and decision-making. Deep learning algorithms do not need human supervision for extracting features. They automatically extract essential features from the dataset. Deep learning architectures such as multi-layer perceptron (MLP), convolutional neural network (CNN), long short-term memory (LSTM) a type of recurrent neural network (RNN) is used for time series forecasting.

Multi-layer perceptron is a feed-forward neural network with multiple perceptrons and many layers. Perceptron is a linear classifier. The algorithm specially deals with binary classification. It is widely used for speech recognition, image recognition etc. Multilayer perceptron neural network works efficiently in classifying whether the air quality index is dangerous to human health or safe.

One of the deep learning classification algorithms is the convolutional neural network which is specially designed to analyse images. The image is fed into convolutional layer where the feature extraction is performed. The processes such as filtering, padding and pooling are performed, and the output is fed into fully connected layer after flattening it. Face recognition is the most popular application of CNN. CNN helps to predict $PM_{2.5}$ while using remote sensing images fetched from the satellite.

Long short-term memory (LSTM) is a form of artificial recurrent neural network that is commonly used to classify long-term data dependencies. This algorithm is well suited for handling time series data. LSTM architecture has more loops in it which makes the information to persist. It can automatically identify temporal dependencies and structures such as trend, seasonal and cyclic.

A cell, an input gate, an output gate and a forget gate make up a typical LSTM unit. The three gates control the flow of information into and out of the cell, and the cell remembers values across arbitrary time periods.

Because there might be delays of uncertain duration between key events, LSTM networks are well-suited to categorising, analysing and generating predictions based on time series data.

4 Literature Survey

Yves Rybarczyk and Rasa Zalakeviciute [8] created multiple regression model for predicting the concentration of $PM_{2.5}$. Along with meteorological data, traffic data was used to identify its role in polluting the air. They identified that the traffic was positively correlated with PM2.5 value. The higher the traffic, the higher the $PM_{2.5}$ value. Traffic data collected from the Google Map was used for analysis. Two images collected from same place were used, one with traffic and one without. The difference was calculated in terms of pixels for red, green and orange. Those values were given as input to predict the $PM_{2.5}$ concentration. Rather than creating one model to identify the $PM_{2.5}$ concentration, different models were created for a day because the parameters (traffic) were not same at all the time in a day. The accuracy of the model was evaluated using correlation coefficient and root mean square value.

Aditya C R, Chandana and et al. predicted particulate matter value using auto regression. The value was predicted using previous readings of PM2.5. The dataset had two attributes date and previous $PM_{2.5}$ concentration. Once the model was developed, it obtained the knowledge to predict $PM_{2.5}$ value for the given date. Mean squared error was used to evaluate the model [9].

Wei-Zhen Lu, A.Y.T Leung et al. used support vector machine and radial basis function for predicting the pollution concentration. Six pollutant values collected in hourly basis for 1 complete year were obtained from monitoring station. Radial basis function (RBF) method and SVM method was used to predict the concentration for a day, week and month. The performance of the model was evaluated using the statistical metric, mean absolute error (MAE) [10].

Hwee San Lim, Mohd Zubir, Mat Jafri et al. used remote sensing images and regression algorithm to identify the pollution concentration. Remote sensing is a method of gathering data about the earth using instruments without coming into close contact with it. Sensors measure the energy reflected from earth. Sensors are mounted on the satellites. Geometric correction was performed to avoid the distortion in raw digital satellite images. Optical depth and reflectance value were calculated. Reflectance value was the sum of the surface reflectance and atmospheric reflectance. AOD is a quantitative represents the amount of depletion that a beam of solar radiation undergoes as it travels through the atmosphere. Regression algorithm was used for determining the pollution concentration [11].

Mehdi Zamani Joharestani, Chunxiang Cao and et al. included more attributes to improve the accuracy of the model. In addition to meteorological data, satellite data, ground-measured $PM_{2.5}$, geographical data was used for the purpose of the modelling. Satellite remote sensing, in terms of aerosol optical depth (AOD) provides information with the required, spatial coverage for the areas under examination. The Moderate Resolution Imaging Spectro-radiometer (MODIS) is the AOD product which is installed in both Aqua and Terra satellites. In each and every city, air pollution monitoring sites record the concentration of PM_{10}, $PM_{2.5}$, CO, O_3, NO_2 and SO_2. Geographic data used includes altitude, latitude and longitude details of the sites. Meteorological time series data such as maximum and minimum air temperature (T-max, T-min), relative humidity (RH), daily precipitation, visibility, wind speed (Windsp), sustained wind speed (ST-windsp), air pressure and dew point were combined with satellite data. The time series data was applied on algorithms such as random forest, XGBoost, deep neural network LSTM and CNN to predict the $PM_{2.5}$ concentration [12].

Chitrini Mozumder and K. Venkata Reddy and Deva Pratap used vegetation index to identify air pollutants. It was found that vegetation was negatively correlated with the air pollution index. Air Pollution Index (API) can be identified using vegetation indices as well as some other image extracted parameters. Remotely sensed data was available as image in IRS P6 (Resourcesat 1) Linear Imaging Self-scanning Sensor (LISS) IV. The images were captured at a resolution of 5.8m using three spectral bands in the visible and NIR (near-infrared) regions: green, red and blue. The air pollution parameter considered was API or AQI (Air Quality Index) The image parameters considered were Normalized Difference Vegetation Index (NDVI), Vegetation Index (VI), Transformed Vegetation Index (TVI) and Urbanization Index (UI). Using all the parameters, multivariate regression model was developed with API as dependent variable and the features extracted from IRS and Landsat images as independent variables. Root mean squared error (RMSE) was used to assess the performance of the model [13]. The research works carried out to predict the air pollutant are summarized in the Table 1 and the results are summarized correspondingly in Table 2.

Table 1 Summary of techniques and algorithms used for prediction

S No	Author name	Title of the paper	Dataset	Objective	Technique	Algorithm used for model construction
1	Yves Rybarczyk and Rasa Zalakeviciute	Regression models to predict air pollution from affordable data collections	Traffic data meteorological data	To predict the PM2.5 concentration	Regression analysis	Linear regression model and multiple regression model
2	Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu,	Detection and prediction of air pollution using machine learning models	Date and PM 2.5 level	To predict the PM2.5 concentration	Time series analysis	Auto regressive model
3	Wei-Zhen Lu, A.Y.T Leung	Air pollutant parameter forecasting using support vector machines	Date and six pollutants such as sulphur dioxide nitrogen oxides nitric oxide, nitrogen dioxide carbon monoxide and respirable suspend particles (RSP)	To predict the RSP concentration	Regression analysis	Support vector regression
4	Hwee San Lim, Mohd Zubir Mat Jafri Khiruddin Abdullah Wong Chow Jeng,	Air pollution determination using remote sensing technique	Digital satellite images	To find aerosol optical depth (AOD) and PM_{10} concentration	Regression analysis	New algorithm proposed based on regression
5	Mehdi Zamani Joharestani, Chunxiang Cao, Xiliang Ni, Barjeece Bashir and Somayeh Talebiesfandarani, atmosphere, July	PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data	Meteorological data, geographical data, satellite images, ground measured values	To predict $PM_{2.5}$ concentration	Ensemble learning and deep learning	Random forest, XG Boost, LSTM, CNN

(continued)

Table 1 (continued)

S No	Author name	Title of the paper	Dataset	Objective	Technique	Algorithm used for model construction
6	Chitrini Mozumder, K. Venkata Reddy, Deva Pratap	Air pollution modelling from remotely sensed data using regression techniques	Normalized difference vegetation ındex (NDVI), vegetation ındex (VI), transformed vegetation ındex (TVI) and urbanization ındex (UI) from satellite images	To predict air quality index of all pollutants and design as image	Regression	Multivariate regression model

Table 2 Summary of results and findings

S No	Algorithm used for model construction	Evaluation measure	Result	Findings
1	Linear regression model and multiple regression model	Correlation coefficient Root mean square value	R = 0.8, RMSE = 5.3	Traffic data plays a main role in predicting the $PM_{2.5}$ concentration
2	Logistic regression auto regressive model	Mean accuracy mean squared error	Mean accuracy = 99% SE = 27	Auto regression suits best for predicting $PM_{2.5}$ value
3	Support vector regression	Mean absolute error	MAE = 0.0178(day) MAE = 0.0204(week)	Compared to feed forward neural network SVR suits better for prediction since the parameters to be set by the user is only 2
4	New algorithm proposed based on regression	correlation coefficient Root mean square value	R = 0.8 RMSE = 16 $\mu g/m^3$	Landsat TM satellite data provides very useful information for estimating and mapping air pollution
5	Random forest, XG Boost, LSTM, CNN	Mean absolute error Root mean square value	MAE = 10.20 $\mu g/m^3$, RMSE = 14 $\mu g/m^3$	Satellite-derived AODs have no significant effect on forecasts. Images with high spatial resolution have to be used
6	Multivariate regression model	Correlation coefficient Root mean squared error	R2 = 0.79 RMSE = 7.77	Landsat is better than IRS in providing air pollution information of the study area

5 Observations and Discussion

Forecasting the air quality is considered as important because of the serious impact the airborne pollutants creates on human health. It is essential to build a predictive model to take some precautionary measures. Air pollutants are identified using techniques such as machine learning and deep learning. Machine learning algorithms such as regression, auto regression and SVM are broadly used. The observations made from the study are summarized below.

Even though the traditional methods such as simple moving average, ARIMA and exponential smoothing were available to identify air quality, machine learning methods and deep learning algorithm do the prediction with ease and accuracy. When considering air quality prediction, the algorithm used and the features selected decide

the effectiveness of the model. Among all the pollutants, $PM_{2.5}$ plays an important role in affecting the health of the human.

The concentration of $PM_{2.5}$ can be identified using machine learning algorithms such as regression, support vector machine, neural network, time series algorithm and deep learning. It is also found that developing one model for a day will not be accurate since the concentration of pollutants varies at different times in a day. Different models should be created to forecast the concentration in the peak time and non-peak time. To improve the accuracy of prediction, traffic information and vegetation index can also be included.

The features such as meteorological information and geographical information are used to identify the $PM_{2.5}$ value. In addition to these parameters, remote sensing images can be included in which aerosol optical depth helps to identify the pollution level. It is also identified that the concentration of pollutants such as sulphur dioxide, carbon monoxide, ozone and nitrogen dioxide contributes more for the prediction. From the observations, it is concluded that using only meteorological factors and geographical information is not sufficient to identify the concentration of $PM_{2.5}$. It is necessary to include the features such as traffic, vegetation index and remote sensing images.

6 Conclusion

In this paper, need for air quality prediction and the research directions in building efficient prediction models are reviewed. The different airborne pollutants and sources of their emission is presented, and their importance is reviewed. The significance of machine learning methods in air quality prediction is studied. Various computational methods involved in predicting the quality of the air are discussed and their results are analysed. Research on air quality prediction carried out by implementing machine learning and deep learning algorithms on time series meteorological data, geographical data, remote sensing images, traffic data and vegetation data is deliberated. Various observations made from the analysis are highlighted. It is necessary to design and develop an efficient air quality prediction model by including more features that contribute in identifying the air contamination. Further study can be carried out by identifying the impact of air pollution in vegetation and health.

References

1. A. Ghorani-Azam, B. Riahi-Zanjani, M. Balali-Mood, Effects of air pollution on human health and practical measures for prevention in Iran. J. Res. Med. Sci. (2016)
2. C. Pénard-Morand, I. Annesi-Maesano, Air pollution: from sources of emissions to health effects. Breathe **1**(2), 108–119 (2004)
3. http://magde.info/main/texts/PhD/Chap2.PDF
4. https://www.thoracic.org/patients/patient-resources/resources/volcanic-eruptions.pdf

5. S. Singh, Implications of forest fires on air quality—a perspective. Bull. Environ. Sci. Res. **5**(3–4), 1–4, (2016)
6. R.K. Lohchab, J. Kumar Saini, Industrial Pollution Management (IAHRW Publications, 2018)
7. A.A. Desai, A review on assessment of air pollution due to vehicular emission in traffic area. Int. J. Curr. Eng. Technol. **8**(2) (March/April 2018)
8. https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics
9. Y. Rybarczyk, R. Zalakeviciute, Regression Models to Predict Air Pollution, (IntechOpen, 2018)
10. C.R. Aditya, C.R. Deshmukh, D.K. Nayana, P.G. Vidyavastu, Detection and prediction of air pollution using machine learning models **59**(4) (2018)
11. W.-Z. Lu, A.Y.T Leung, et al, Air pollutant parameter forecasting using support vector machines. Int. Joint Conf. Neural Network (2002)
12. H.S. Lim, M. Zubir, M. Jafri, K. Abdullah, W.C. Jeng, Air pollution determination using remote sensing technique. Adv. Geosci. Remote Sens.
13. M.Z. Joharestani, C. Cao, X. Ni, B. Bashir, S. Talebiesfandarani, PM2.5 prediction based on random forest, XGBoost,and deep learning using multisource remote sensing data, Atmosphere (2019)

A Unique Interlinking Converter Control for Hybrid AC/DC Islanded Microgrids

M. Jayachandran, Gundala Srinivasa Rao, and Ch. Rami Reddy

Abstract This paper aims to address the power-sharing requirement and operational reliability of hybrid AC/DC microgrids. Besides, the innovative power management scheme is proposed for interlinking converter by utilizing model predictive control (MPC) strategy to integrate renewable energy sources, energy storage devices, and the utility grid. The proposed control model not only maintains AC and DC bus voltage stable but also ensures the proper power balance between AC and DC sub-microgrids irrespective of the intermittent nature of renewable energy and load fluctuations.

Keywords AC/DC microgrid · Interlink converter · Power sharing · Model predictive control

1 Introduction

Integration of renewable energy sources and storage elements into the power system will enhance the overall system performance. The proliferation of AC/DC microgrids offers higher quality, stability, and reliability with reduced operational cost and power loss to the power system [1]. Regarding AC microgrids, the distributed generators like PV and fuel cells are interconnected to the system that requires DC to AC conversion. To increase the reliability and robustness of the system, the AC

The original version of this chapter was revised: The corresponding chapter author "M. Jayachandran" affiliation has been updated. The correction to this chapter is available at https://doi.org/10.1007/978-981-16-6605-6_66

M. Jayachandran (✉)
Sri Manakula Vinayagar Engineering College, Puducherry, India
e-mail: jayachandran.escet@pec.edu

G. S. Rao
CMR College of Engineering and Technology, Hyderabad, India

Ch. R. Reddy
Malla Reddy Engineering College, Secunderabad, India

corrected publication 2022
P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_12

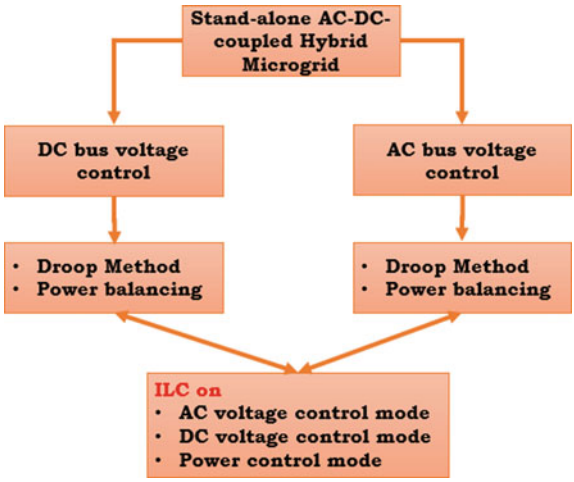


Fig. 1 Power management strategies of Hybrid AC-DC-coupled isolated microgrid system

microgrid is connected to the utility grid for sharing excess/deficient power. In case of grid failure, all the distributed generators and loads exchange power on the AC platform [2]. Besides, modern power electronic loads operate in DC platforms which require once again AC to DC conversion. In AC microgrid systems, DC-AC-AC-DC conversion is quite common [3]. Conversely, the DC microgrid can facilitate the direct connection of renewables such as PV, fuel cells, and dc loads. However, wind power generators cannot directly be connected to DC microgrids. It requires AC-DC conversion. Besides industrial loads such as induction machines connected to DC microgrids that require once again DC-AC conversion. In DC microgrid systems, AC-DC-DC-AC conversion is inevitable [4, 5]. Therefore, the researchers in recent years have mainly concentrated on the combination of both AC and DC microgrids connected with a common converter called bidirectional interlink converter (ILC) that helps to transfer excess/deficit power from one microgrid to another [6]. This hybrid system can connect along with the utility grid during normal operation. It also gets disconnected from the grid under emergencies and can operate in autonomous mode. This AC/DC hybrid microgrid system avoids multiple conversions which are present in the independent AC or DC microgrid system. However, the control challenges are very severe in hybrid microgrid systems. The primary control objective of AC/DC microgrids is to achieve a power balance between two subgrids [7, 8].

As far as AC/DC microgrid operation is concerned, the control and power management strategies are essential features to determine output powers of distributed generators and control voltage as well as frequency simultaneously [9, 10]. The following challenges are identified when integrating both AC and DC microgrid: In the case of an isolated mode of operation, conventional p-f and q-v droop control methods are not suitable for power-sharing between AC and DC microgrids [11]. Therefore, power management strategy requires a specified droop control method for sharing power demands. In this regard, the interlinking converter plays a vital role

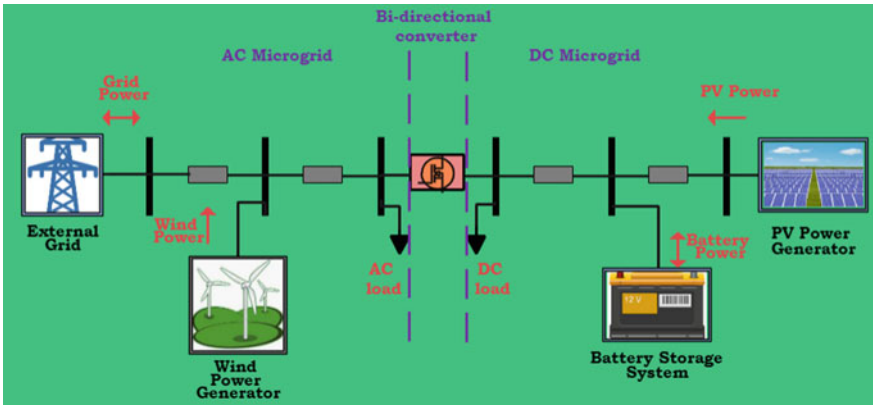


Fig. 2 Schematic diagram of Hybrid AC/DC microgrid system

in power management and control. Depending on control schemes used in AC and DC microgrids, this ILC can be leveraged on AC bus control mode, DC bus control mode, and power control mode. To equalize the power generation and demand, the bi-directional ILC converter supervises the power flow between DC and AC microgrids under islanding operation conditions [12]. Besides, harmonic power-sharing between the grids is another important challenge in the hybrid system. The trade-off between voltage regulation and reactive power sharing has gained prominence [13, 14]. The classification of microgrid power control includes centralized control, decentralized control, distributed control, and multi-agent systems. However, these methods of analysis has its own drawbacks. The centralized and decentralized control are combined for the reliable operation of microgrids. This integrated microgrid control strategy is referred to as a hierarchical control structure that has been established for power-sharing among distributed generators [15]. Most of the research so far focuses focus on the droop based hierarchical control with PWM technique. This paper investigates a novel MPC approach for bidirectional ILC control of AC/DC microgrids.

The overview of power management strategies are illustrated in Fig. 1. The main contribution of this work is to develop a power management strategy based on a predictive control approach for interlinking converters in hierarchical controlled hybrid AC/DC microgrid systems. The proposed control scheme can accomplish the optimized power flow between AC as well as DC subgrids. This power management strategy can be performed on AC bus voltage control mode, or dc bus voltage control mode, or bi-directional power control mode. The proposed control design utilizes the model of the microgrid and differs from the existing approaches reported in the literature.

2 Structure of AC/DC Microgrid System

All DC generators and DC loads are connected to the DC bus in the DC microgrid, whereas the AC bus can accommodate all AC generators, AC loads, and utility grids in the AC microgrid. These two subgrids are connected through an interlink converter as shown in Fig. 2. The key benefit of this structure is to minimize the power conversion requirements and improves overall efficiency with reduced system cost.

2.1 DC Bus Voltage Control

Primary control is implemented via a MPC control. Droop control is commonly employed on the primary level of hierarchical control to improve system stability and power-sharing accuracy.

$$V_{dc} = V_{dc-max} - m_p.P_{dc} \tag{1}$$

However, voltage deviation takes place. To eliminate the voltage deviation caused by droop control, a voltage secondary control loop is incorporated using MPC approach.

$$V_{dc} = V_{dc-max} - m_p.P_{dc} + \delta V_{dc} \tag{2}$$

where, δV_{dc} is voltage compensation at secondary control. Integration of distributed energy storage systems in DC microgrids can effectively mitigate the RES and load fluctuations. The charging and discharging operation can be controlled by a bidirectional DC-DC converter under the power mismatch at the DC microgrid.

$$\begin{aligned} V_{dc} &= V_{dc-ref} - m_o.SOC.P_o(charging) \\ V_{dc} &= V_{dc-ref} - \frac{m_o}{SOC}.P_o(discharging) \end{aligned} \tag{3}$$

2.2 AC Bus Voltage Control

To achieve load power sharing between distributed generators for parallel inverter-based ac microgrid system, droop method, as a decentralized control, is generally adopted as,

$$\begin{aligned} f_{ac} &= f_{ac-max} - m_p.P_{ac} \\ V_{ac} &= V_{ac-max} - n_q.Q_{ac} \end{aligned} \tag{4}$$

Using the predicted capacitor voltage, the AC bus voltage can be tracked by controlling the capacitor voltage. Then the cost function generates the least switching sequence applied to the inverter as,

$$J = \sum_{j=1}^{p}([V_{c\alpha-ref} - V_{c\alpha}(k+1)]^2 + [V_{c\beta-ref} - V_{c\beta}(k+1)]^2) \tag{5}$$

In order to eliminate the voltage deviation caused by droop control, a voltage secondary control loop is incorporated.

$$\begin{aligned} f_{ac} &= f_{ac-max} - m_p.P_{ac} + \delta f_{ac} \\ V_{ac} &= V_{ac-max} - n_q.Q_{ac} + \delta V_{dc} \end{aligned} \tag{6}$$

Integration of grid into AC microgrids can effectively mitigate the RES and load fluctuations. This can balance the power mismatch at the AC microgrid.

3 Bidirectional Power Flow Control

1. When both the total PV power generation and wind power generation are less than DC as well as AC power demand respectively, all PV sources operate in maximum power point and BILC operates in inverter mode. Then, all BESS supplies the excess total power demand.
2. If the total PV power generation is more than DC power demand, wind power generation is less than AC power demand, and batteries can absorb the surplus power, BILC operates in inverter mode. Then, batteries are charged with excess PV power, and DC microgrid supplies the excess AC power demand.
3. If the total PV power generation is less than DC power demand, wind power generation is more than AC power demand, and batteries can absorb the surplus power, BILC operates in rectifier mode. Then, the AC microgrid not only supplies the excess DC power demand but also charges the batteries.
4. When both the total PV power generation and wind power generation are more than DC as well as AC load power consumption respectively, BILC operates in rectifier mode. Then, all batteries are charged with renewables. Conversely, all batteries are completely charged then a PV power curtailment is performed and switch on the dump loads.

4 Proposed Power Management Strategy

The predictive control scheme for the bi-directional interlink converter is designed coordinately with droop control for the distributed generators in two sub-grids. With

the knowledge of power demand with both ac and dc loads ($P_{oad-demand}$), the solar PV power (P_{pv}), and the wind power (P_{wind}), the total net power is determined as,

$$P_{total-net} = P_{pv} + P_{wind} - P_{load-demand} \tag{7}$$

This total net power can be balanced through battery/diesel generator using BIL converter. The net power available in the DC bus calculated as,

$$P_{dc-bus}(k+1) = P_{pv} - P_{dc-load} \tag{8}$$

This net DC power must be balanced through battery energy storage system. Similarly, the net power available in the AC bus calculated as,

$$P_{ac-bus}(k+1) = P_{wind} - P_{ac-load} \tag{9}$$

This net AC power must be balanced through diesel generator.

$$J = [P_{total-net} - P_{dc-bus}(k+1)]^2 + [P_{total-net} - P_{ac-bus}(k+1)]^2 \tag{10}$$

By minimizing the objective function yields the optimized control sequence applied to the converter.

5 Simulation Results

Parameters for simulation work:
DC bus voltage: 750 V
PV system: 100 kW, 415 V
WTG: DFIG, 100 kW, 415 V
Battery: lead-Acid 100 kW, 415 V, 2 kAh for 3-Autonomy days
DC variable load: 0–100 kW, 220 V
AC variable load: 0–100 kW, 415 V, 50 Hz
Utility grid: 100 kW, 415 V.

In mode-1: When the generated power from the renewable energy source is more than the load demand, two possible cases are observed from Figs. 3, 4, and 5.
(a) At t = 3 h: If dc load demand is greater than PV power generator and ac load power demand is lesser than wind power generator, the surplus dc power demand is shared with a wind power generator. The excess wind power is charged with batteries. Therefore, BILC act as a rectifier.
(b) At t = 9 h: If dc load demand is lesser than PV power generator and ac load power demand is greater than wind power generator, PV supplies power to both dc as well

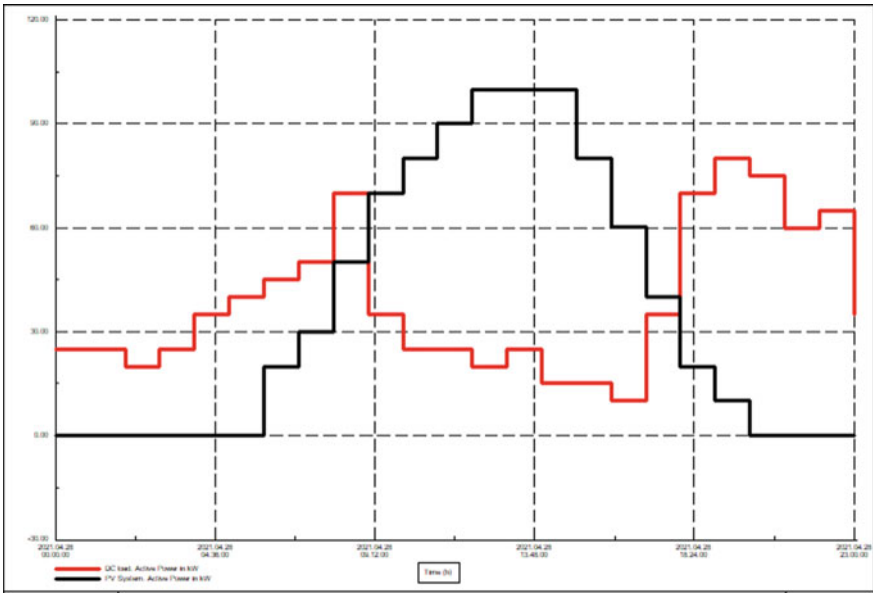


Fig. 3 PV power generation with DC load demand

Table 1 Operational modes of BILC

Mode	Case	Power flow	BILC mode
$P_{RE} > P_{Load}$	$P_{PV} < P_{DC-Load} P_{Wind} > P_{AC-Load}$	AC to DC microgrid	Rectifier
	$P_{PV} > P_{DC-Load} P_{Wind} < P_{AC-Load}$	DC to AC microgrid	Inverter
$P_{RE} < P_{Load}$	$P_{PV} < P_{Load}$	DC to AC microgrid	Inverter
	$P_{PV} < P_{DC-Load} P_{Wind} = P_{AC-Load}$	Power flow within MG	Ideal
	$P_{PV} > P_{DC-Load} P_{Wind} < P_{AC-Load}$	DC to AC microgrid	Inverter
	$P_{PV} < P_{DC-Load} P_{Wind} > P_{AC-Load}$	AC to DC microgrid	Rectifier

as ac load. Therefore, BILC works in inverter mode. Excess power generated from solar PV can charge the batteries.

In mode-2: when the generated power from the renewable energy source is less than the load demand, the four possible cases are observed from Figs. 3, 4,

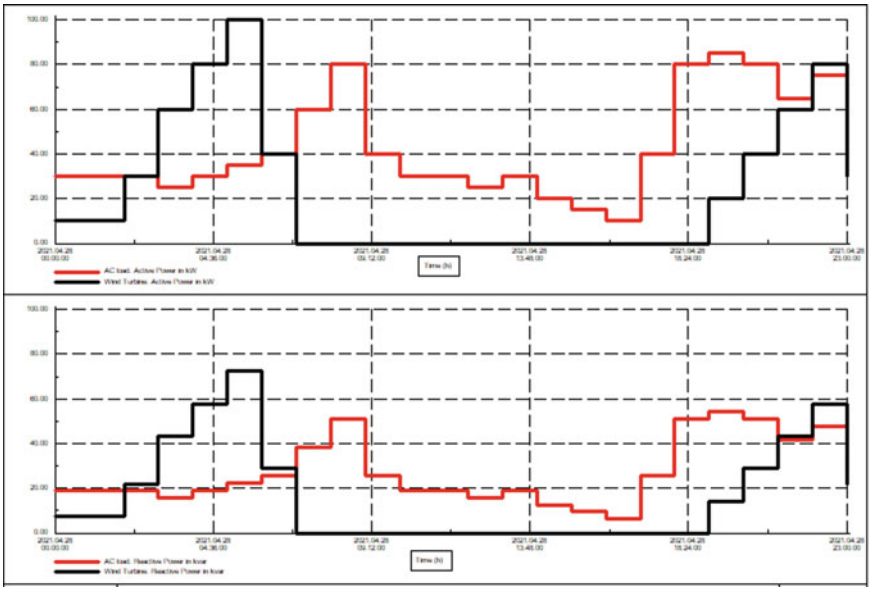


Fig. 4 Wind power generator with AC load demand

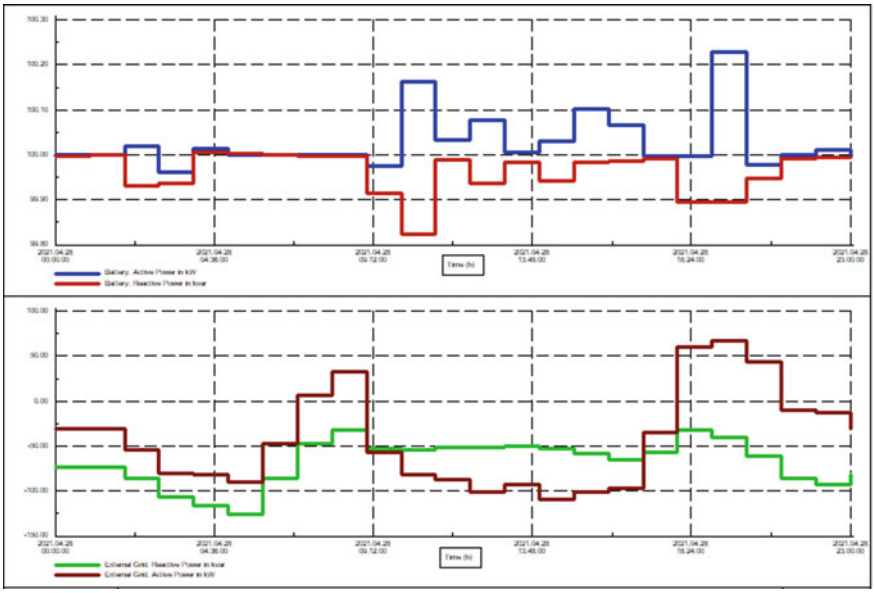


Fig. 5 Power balance using Battery as well as external grid

and 5.
(a) At t = 1 hrs: If both dc, as well as ac load demand, are more than PV power generator and wind power generator respectively, PV generator operates MPPT mode to supply power to dc load and wind power generator supply power to ac load. However, the excess load power demand (both dc and ac) can be shared by batteries. Therefore, BILC operates in inverter mode.
(b) At t = 2 hrs: If dc load demand is more than PV power generator and ac load power demand is equal to a wind power generator, ac power demand is satisfied with a wind power generator. Both PV and battery are contributed to compensate dc power demand. Therefore, BILC neither operates inverter nor in rectification mode.
(c) At t = 9 hrs: If dc load demand is lesser than PV power generator and ac load power demand is greater than wind power generator, PV supplies power to dc load and the excess ac load demand is shared with both PV and battery system. Therefore, BILC works in inverter mode. Excess power generated from solar PV can be export to the grid.
(d) At t = 22 hrs: If dc load demand is greater than PV power generator and ac load power demand is lesser than wind power generator, the surplus dc power demand is shared with wind power generator as well as a battery bank. Therefore, BILC act as a rectifier. The summary of operational modes of BILC is tabulated in Table 1.

6 Conclusion

An autonomous power flow control approach based on MPC is proposed for a bi-directional interlink converter in this paper. Employing this unique method into hybrid ac/dc subgrids not only manages the power flow but also achieves proper load sharing among distributed generators. Besides, the proposed optimized control strategy prevents circulating current among the ac and the dc microgrids. Moreover, dc and ac bus voltages are precisely regulated and maintain the system stable.

References

1. D. Kumar, F. Zare, A. Ghosh, Dc microgrid technology: system architectures, ac grid interfaces, grounding schemes, power quality, communication networks, applications, and standardizations aspects. IEEE Access **5**, 12230–12256 (2017). https://doi.org/10.1109/ACCESS.2017.2705914
2. S. Jena, N.P. Padhy, Distributed cooperative control for autonomous hybrid ac/dc microgrid clusters interconnected via back-to-back converter control, in *2020 IEEE Power Energy Society General Meeting (PESGM)* (2020), pp. 1–5. https://doi.org/10.1109/PESGM41954.2020.9281505
3. H. Abu-Rub, M. Malinowski, K. Al-Haddad, *AC-DC-AC Converters for Distributed Power Generation Systems* (2014), pp. 319–364. https://doi.org/10.1002/9781118755525.ch11

4. F. Nejabatkhah, Y.W. Li, Overview of power management strategies of hybrid ac/dc microgrid. IEEE Trans. Power Electronics **30**(12), 7072–7089 (2015). https://doi.org/10.1109/TPEL.2014.2384999
5. J. Wang, C. Jin, P. Wang, A uniform control strategy for the interlinking converter in hierarchical controlled hybrid ac/dc microgrids. IEEE Trans. Industrial Electronics **65**(8), 6188–6197 (2018). https://doi.org/10.1109/TIE.2017.2784349
6. X. Shen, D. Tan, Z. Shuai, A. Luo, Control techniques for bidirectional interlinking converters in hybrid microgrids: leveraging the advantages of both ac and dc. IEEE Power Electronics Mag. **6**(3), 39–47 (2019). https://doi.org/10.1109/MPEL.2019.2925298
7. S. Peyghami, H. Mokhtari, F. Blaabjerg, Autonomous operation of a hybrid ac/dc microgrid with multiple interlinking converters. IEEE Trans. Smart Grid **9**(6), 6480–6488 (2018). https://doi.org/10.1109/TSG.2017.2713941
8. P. Lin, P. Wang, C. Jin, J. Xiao, X. Li, F. Guo, C. Zhang, A distributed power management strategy for multi-paralleled bidirectional interlinking converters in hybrid ac/dc microgrids. IEEE Trans. Smart Grid **10**(5), 5696–5711 (2019). https://doi.org/10.1109/TSG.2018.2890420
9. G.D. Agundis Tinajero, M. Nasir, J.C. Vasquez, J.M. Guerrero, Comprehensive power flow modelling of hierarchically controlled ac/dc hybrid islanded microgrids. Int. J. Electrical Power Energy Syst. **127**, 106629 (2021)
10. V. Bindhu, G. Ranganathan, Effective automatic fault detection in transmission lines by hybrid model of authorization and distance calculation through impedance variation. J. Electronics Informatics **3**(1), 36–48 (2021). https://doi.org/10.36548/jei.2021.1.004
11. J. Hofer, B. Svetozarevic, A. Schlueter, Hybrid ac/dc building microgrid for solar pv and battery storage integration, in *2017 IEEE Second International Conference on DC Microgrids (ICDCM)* (2017), pp. 188–191. https://doi.org/10.1109/ICDCM.2017.8001042
12. T. Vijayakumar, R. Vinothkanna, Efficient energy load distribution model using modified particle swarm optimization algorithm. J. Artif. Intelligence Capsule Netw. **2**(4), 226–231 (2020). https://doi.org/10.36548/jaicn.2020.4.005
13. F. Nejabatkhah, Y.W. Li, H. Tian, Power quality control of smart hybrid ac/dc microgrids: an overview. IEEE Access **7**, 52295–52318 (2019). https://doi.org/10.1109/ACCESS.2019.2912376
14. N. Bhalaji, C. Rimi, Remaining useful life (rul) estimation of lead acid battery using Bayesian approach. J. Electrical Eng. Autom. **2**(1), 25–34 (2020). https://doi.org/10.36548/jeea.2020.1.003
15. M. Najafzadeh, R. Ahmadiahangar, O. Husev, I. Roasto, T. Jalakas, A. Blinov, Recent contributions, future prospects and limitations of interlinking converter control in hybrid ac/dc microgrids. IEEE Access **9**, 7960–7984 (2021). https://doi.org/10.1109/ACCESS.2020.3049023

A Novel Autonomous Flotation Vehicle for Continuous Water Quality Monitoring

Siddaraju

Abstract In water sources, water quality control and prediction are essential. The survival of our marine environments depends on good quality of water. Continuous water quality assessment is a valuable method for water resources agencies, as it provides real-time data for protecting the environment and pollutant emission tracking. A low-cost mobile aquatic surveillance system would allow for budget data gathering on water quality, aiding catchment administrators in maintaining aquatic ecological functions. Water sampling by itself does not capture any of the biochemical processes in contaminated water. When enough than one toxin is mixed with water and interim contaminants are formed. The aim of this device is to predict and detect gases released by such reacts. Water is pumped into the detector canisters by on board pump motors, and specimens are sent to various sensor areas, where they are captured and checked. The results indicate that by using the monitoring system, multiparametric, lengthy, and online monitoring for water pollution details can be reliably obtained and estimated.

Keywords Autonomous water vehicle · Online water monitoring · Sensors · Solar powered systems · IoT · Geo fencing · Water sensors · Gas sensors · 4G communication · Embedded technology

1 Introduction

According to the World Health Organization (WHO), nearly one-sixth of the global population or 1.1 billion individuals, do not have clean drinking water, and 2.4 billion do not have access to proper sanitation [1–3]. Water pollution can cause problems at various levels, and nearly, all forms of water pollution are detrimental to human health: Some have a long-term impact, while others have obvious and immediate consequences. Water contamination is an ever-present danger to both

Siddaraju (✉)
Department of Computer Science & Engineering, Dr.Ambedkar Institute of Technology, Bengaluru, Karnataka 560056, India
e-mail: siddaraju.cs@drait.edu.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_13

developing and developed regions. Groundwater, which is usually less impacted by industrial and domestic effluents, is overused in densely populated areas, causing residents to become more reliant on fresh water. Many rivers in Asia are among the world's most contaminated [4–6]. Unlike fresh water, which can travel for miles or even hundreds of miles across a variety of topographic environments, groundwater is usually contained within underground aquifers [7–9]. Groundwater contamination poses a unique set of difficulties compared to surface water contamination which spreads rapidly over a wide geographic region, differs potentially significantly due to the effects of other sources of pollution, rainfall and river self-purification mechanisms, and toxic chemicals groundwater via the hydrological cycle [10–13]. River water, on the other side, often follows a linear path to a single segment at the geomorphologic bottom. The preceding experimental study is motivated by the unique features of fresh water (in general rivers and lakes) [14–16].

By offering an efficient data collection model, we created a groundbreaking comprehensive control and analysis system to identify water contamination and reduce the harm to human life and economic [17–19]. In the meantime, we expect a major rise in demand for advanced environmental technologies to help India combat the increasing threat of water contamination in the immediate future [20].

The rest of the paper is organized as follows. Literature review is discussed in section two, section three presents problem definition, design objectives are described in section four, implementation of the system is discussed in section five, results are provided in the sixth section, and finally, we conclude the paper.

2 Literature Review

Unmanned surface vehicles (USVs), also known as autonomous surface crafts (ASCs), are a type of unmanned surface vehicle [21–23]. USVs can be employed on the surface of rivers, lakes, seas, and other bodies of water to accomplish a range of tasks that would be better and affordable for machines to execute than people such as marine environment monitoring, hydrologic survey, target item seeking, and scientific research. A number of major research and development activities on enhanced UVSs have been done since the end of the 1990s, particularly over the last few years [24, 25]. An ASV is designed to conduct environmental monitoring and hydrologic surveys in shallow water areas, where a survey vessel cannot reach and it is risky for personnel to do so. The ASV is powered by electricity, which is stored in a battery. Due to the battery's capacity constraint, it is essential to improve propulsive efficiency in order to accelerate the ASV or increase endurance performance. The focus of the remotely controlled system is on enhancing dependability and expanding the coverage area. Because the ASV is utilized for maritime environment monitoring or hydrologic surveys, it is preferable to carry a variety of equipment with varying data interfaces. These issues are addressed in this work.

Caccia's [26] evaluation on ASV technology describes the development of ASV experimental activities for both research and military uses. The key study topics

found in the literature have been vehicle design and implementation [27], navigation and control [28], and route planning and collision avoidance. Multiple ASVs have been suggested as a method of providing communication support for autonomous underwater vehicles and configuration control for marine security systems. A new study field is mobile adaptive sampling, in which the ASV may change its trajectory to increase measurement resolution in time and space. Zhang [29] describes their initial work in this area. Higinbotham et al. [30] are noteworthy for developing solar-powered ASVs for ocean and meteorological observation. A solar and wind powered boat and an electric catamaran are two such examples.

3 Problem Definition

At present, water samples are collected from the river/lakes and are send to laboratory for testing. This takes few days and manual work to carry out the tasks. The liquid is only considered for the test, and the gases emitted are not collected nor tested. The quality of water also differs from different places to places in the same lake but with some variation. Water quality testing on day-to-day basis is not feasible and would be costly. An automated system will be required for collecting water from different locations, test them, and send the data to remote server using 3G/4G teleNetwork.

4 Design Objectives

A remotely controlled floating platform with computation hardware and sensors is required to collect water samples and relay sensed values to remote server. The architecture also makes it simple to integrate sensors and applications into a current water tracking network by using widely accessible interface devices and communication standards.

5 Implementation

The major function of the robotic boat is to autonomously cruise and continually gather water quality data, which are then relayed back to shore in real-time. This offers several operating and scholarly advantages over existing labor laborious methods. In 24-h sample tactics, the capacity to examine and capture particular incidents and places in a timely way and a method for assessing natural variability to drive and validate modeling efforts are all examples of sampling techniques.

The water testing system consists of two chambers. First, chamber is for filtering the water from debris and organic wastes. This requires to be done as these may choke the pipes in long run. Once the water is cleared of debris, it is suitable to

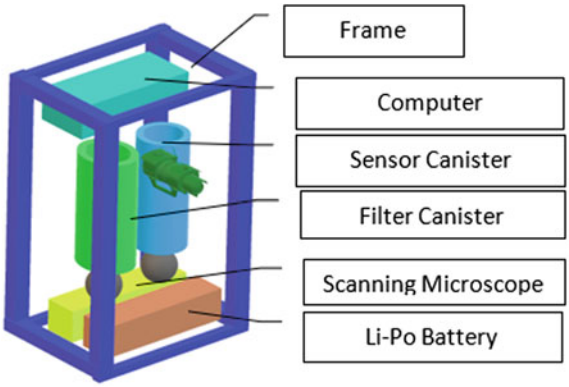


Fig. 1 Physical layout of the autonomous water monitoring system

be observed under a microscope. This is done by channeling the water through a water tight compartment with microscope and LED lights to illuminate for closer observation (Fig. 1).

To transmit all of the frames and detector parameters to the main device, the modules in this suggested vision system include a DC power supply as well as a high transmission bus. For video communications, the IP camera is installed on a pan-tilt setup with a Wi-Fi and Ethernet port positioned within the casing.

The overall bandwidth used to operate all camera at the same time must be below the bus's maximum bandwidth. The USB host device, not the numbers of USB ports, defines the total capacity.

According to the suggested computational device specification, the motherboard selected should have a reduced resource consumption (Fig. 2). The ATmega2560 microcontroller was used to connect the energy electronics only with system and control the electromechanical systems and pumps. The ATmega2560 has a designated USB bridge to the device that utilizes virtual ports. It is simple to scale and works with all big operating systems. To monitor the electronic and pump, one ATmega2560 system is used. The ATmega® AVR® design is more code productive than typical CISC microcontrollers, with speeds close to 10 times faster. The laptop's DC voltage is difficult since it needs many controlled voltage supplies (Fig. 3). The battery pack on a standard computer motherboard is 24 pin ATX. A DC-to-DC ATX adapter is

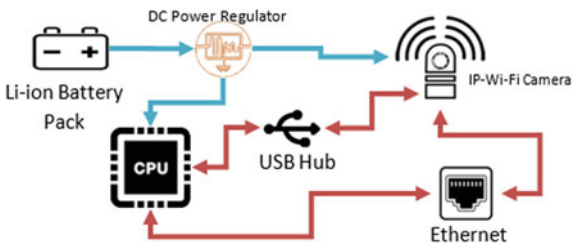


Fig. 2 Vision system of the autonomous water monitoring system

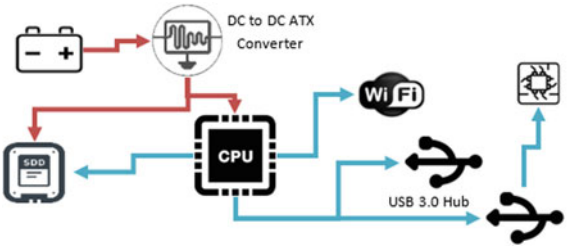


Fig. 3 Computation system of the autonomous water monitoring system

used to feed this electricity supply, which drives the device and solid-state drive. The main machine is powered by an HDPLEX 400 W DC-to-DC ATX power supply.

The adapter unit must have a special ripple noise suppression circuit, large voltage inductance, ideally WIMA stereo quality capacitors, good quality MOSFET, trusted platform chip, and copper header PCB to ensure a decent charge controller without bursts or sound. The controller device should really be sufficient to shield the device from voltage spikes, overload, overheating, and short circuits, as well as work with a wide range of input voltages. Since lithium-ion battery packs have a nonlinear deep discharge loop, large voltage responsiveness is needed.

A PL2303 USB to RS232 adapter is used to link the wheel hub motor to the device in order to design a portable connection between microprocessors and the CPU (Fig. 4).Wheel hub rotor should be fitted with its own PL2303 device. The PL-2303 connects to any USB host with an RS232-like full duplex asynchronous serial link

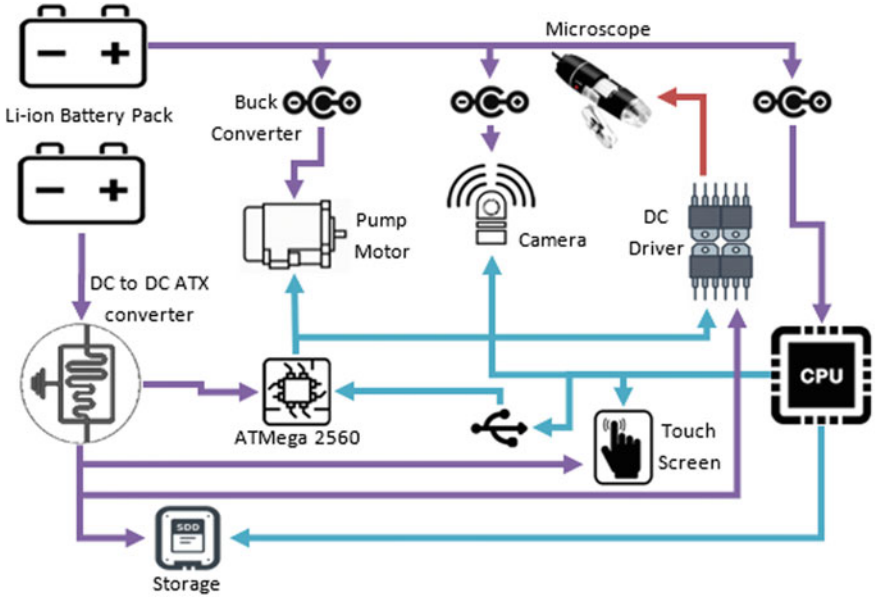


Fig. 4 Power and I/O system of the autonomous water monitoring system

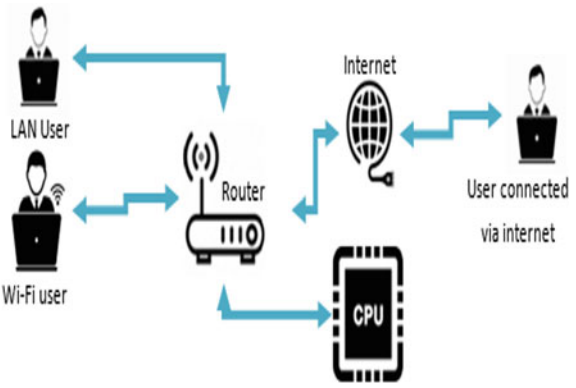


Fig. 5 Communication system of the autonomous water monitoring system

and can emulate the standard COM port on most operating systems, enabling current COM port-based programs to be conveniently migrated and rendered accessible on USB. A core motor with only an RS232/RS422/RS485 adapter will use this adapter. This interface enables end users or other computers to link to an Ethernet router for TCP port connectivity. Many IP-based devices can connect with the main CPU via Ethernet. For access to a wide range of currently available hardware, the router must accept Wi-Fi protocols such as 802.11b/g/n and 802.3/3u. This router can achieve speeds of up to 300Mbps utilizing four antennas. A consumer can bind to the device in four different ways (Fig. 5). The user will mirror the view on a different monitor by directly connecting a screen to the GPU chip.

Second, a user can link to the router with an RJ45 wired network and interface explicitly into the device's software. Finally, a user can link to the device via Wi-Fi from a length of approximately to 300 m. This mode can be used to test the device, when it is moving. Fourth, by linking the network to the net using WLAN and then attaching to the computer using remote access software tools. The router's average energy usage is 5 watts.

5.1 Power Generation and Management

The planned device's devices and operating systems are configured to accommodate voltage instability while still providing numerous controlled power supplies. Because of volatility caused by a sudden need for electricity by the hub engine or mechanical arm, multiple battery packs are needed. Different rechargeable batteries would also allow for partial system shutdown during diagnostic activities. This device has one big Li-ion battery pack that is self-contained (Fig. 6). The configuration of the pack is 3S13P. This packs use 18,650 Li-ion cells that are rated at 3.6 V and 2.6 amps. It is healthy to drive these batteries to 2 amps. $3.6 \times 2.6 \times 40 = 374$ watts can be delivered by the 4S10P specification cell. The other two Li-ion batteries have a 4S5P design and will produce a total of 187 watts. The power pack's total combined

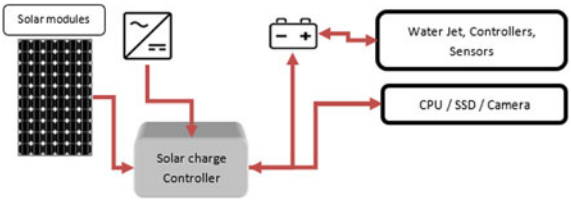


Fig. 6 Solar power system of the autonomous water monitoring system

energy output is about 748 watts. Since the voltage variation between fully loaded and discharged batteries grows as the cell size expands, the right cell configuration must be selected. When a minimum voltage is reached, some components will shut down. As a consequence, battery use can be inconsistent. As a result, a device must use an individual calls kit. Normalizing voltages for various instruments may also be done with voltage regulators and DC-to-DC converters.

The solar panels (removable) have a nominal voltage of 12 V. The voltage will be inconsistent, and the Li-ion battery pack will need 17 V DC to operate. To meet the optimum voltage for charging the battery, a step-up DC-to-DC boosting has been used. This device's application system architecture is adaptable to new features. Any new functionality can be introduced as a separate program that runs without interfering with other programs. This is accomplished by the use of bus architecture. Related to the management of computational energy, computers, command storage, series execution, content collection/encoding, telematics, goal setting, power/control swapping, celestial body location prediction (Earth, Sun, Mars, etc.) and solar array control and connectivity will all be supported by this system framework.

The following types of software can also be developed in this device platform device (but not limited to the following):

1. Architectures based on intelligence and neural networks that run on Windows and Linux with GPU support can be seen here.
2. Incorporation of a microprocessor, sensors, and 3D devices.
3. Route planning grid matrix, landscape rendering, and 3D point cloud grid
4. Flexibility and vectorization drive and control system
5. Tracking and logging
6. Coordination between processes.

Different facets of an automated device's functionality can be developed into individual modules that communicate with one another through inter-process messaging platforms. These pieces of technology can be designed in the same way as a bus and based on publish/subscribe message-passing template.

6 Results

A number of tests were carried out in order to discover and define critical hydrodynamic and propulsion parameters for ASV modeling and control. The highest vehicle

speed with normal propellers is two kts; but with a constant solar input of 50 W, the prolonged speed will be less than two kts to avoid drawing from battery packs. The maneuverability of this ASV at low speeds to retain a reference path is a particular issue, as the control efficiency of rudders grows with the square of speed.

It would take some days for the entire lake to be polluted as polluted water gets into the larger water body over a period of time. Hence, this device can be used as early-warning system. The system is capable of floating for several day, even weeks in water and relay water quality information to the server. The canisters were capable of holding water and test gases released by the water. Remote controlling the device makes it easier to navigate in water. Software capable water navigation is also capable.

7 Conclusion

The presentation of a revolutionary autonomous surface vehicle (ASV) capable of navigating across intricate inland water storages. The custom-built solar-powered catamaran can gather various water quality data throughout the water column while moving, as well as measure the geographic release of numerous greenhouse emissions. The combination of sensing devices such as a gas sensors, ph., microscope, and camera enables the ASV to benefit of all involved in formerly unmapped shallow water settings while avoiding static and mobile impediments. The ASV is directly connected with a storage-scale stationary and mobile sensor nodes, allowing for high-level activities such as quasi communications, task uploads, data-upload, throughout inspection, and sensor validation. Extensive field experiments have proved the ASV's capacity to continually sample the liquid water and function in a range of weather situations as well as at night. The ASV operation and data collection complement current manual monitoring operations by providing greater geographical and chronological monitoring of the water storage, with tens of kilometers of surveys now performed.

References

1. R.P.N. Budiarti, A. Tjahjono, M. Hariadi, M.H. Purnomo, Development of IoT for Automated Water Quality Monitoring System, in 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), Jember, Indonesia, 2019, pp. 211–216
2. Z. Lin, W. Wang, H. Yin, S. Jiang, G. Jiao, J. Yu, Design of Monitoring System for Rural Drinking Water Source Based on WSN, in 2017 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an (2017), pp 289–293
3. Heyi Wang, Yi Gao, Zhaoan Xu, Weidong Xu, An recurrent neural network application to forecasting the quality of water diversion in the water source of Lake Taihu, in 2011 International Conference on Remote Sensing, Environment and Transportation Engineering, Nanjing (2011), pp 984–988

4. H. Sui, G. Zheng, J. Zhou, H. Li, Z. Gu, Application of NB-IoT Technology in City Open Water Monitoring, in 2020 6th International Symposium on System and Software Reliability (ISSSR), Chengdu, China (2020), pp. 95–98
5. X. Wang et al., The lake water bloom intelligent prediction method and water quality remote monitoring system, in 2010 Sixth International Conference on Natural Computation, Yantai (2010), pp. 3443–3446
6. Q. Wu, Y. Liang, Y. Sun, C. Zhang, P. Liu, Application of GPRS technology in water quality monitoring system, 2010 World Automation Congress, Kobe (2010), pp. 7-11
7. A.H. Abdulwahid, IoT Based Water Quality Monitoring System for Rural Areas, in 2020 9th International Conference on Renewable Energy Research and Application (ICRERA), Glasgow, United Kingdom (2020), pp. 279–282
8. C. Sowmya, C.D. Naidu, R.P. Somineni, D.R. Reddy, Implementation of Wireless Sensor Network for Real Time Overhead Tank Water Quality Monitoring, in 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, (2017), pp. 546-551
9. W. Chen, Research on water environment automatic monitoring evaluation system for ecological compensation, in 2011 International Conference on Remote Sensing, Environment and Transportation Engineering, Nanjing (2011), pp. 6576–6579
10. C. Zhang, J. Wu, J. Liu, Water quality monitoring system based on Internet of Things, in 2020 3rd International Conference on Electron Device and Mechanical Engineering (ICEDME), Suzhou, China (2020), pp. 727–730
11. C. Saab, I. Shahrour, F.H. Chehade, Smart technology for water quality control: feedback about use of water quality sensors, 2017 Sensors Networks Smart and Emerging Technologies (SENSET), Beirut (2017), pp. 1-4
12. Y. Xu, F. Liu, Application of wireless sensor network in water quality monitoring, in 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Guangzhou (2017), pp. 368–371
13. N. Vijayakumar, R. Ramya, The real time monitoring of water quality in IoT environment, in 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], Nagercoil (2015), pp. 1–4
14. M. Qun, G. Ying, L. Zhiqiang, T. Xiaohui, Application of comprehensive water quality identification index in water quality assessment of river, in 2009 WRI Global Congress on Intelligent Systems, Xiamen (2009), pp. 333-337
15. S. Pawara, S. Nalam, S. Mirajkar, S. Gujar, V. Nagmoti, Remote monitoring of waters quality from reservoirs, in 2017 2nd International Conference for Convergence in Technology (I2CT), Mumbai (2017), pp. 503-506
16. S. Sukaridhoto, D. Pramadihanto, Taufiqurrahman, M. Alif, A. Yuwono, N. Funabiki, A design of radio-controlled submarine modification for river water quality monitoring, in 2015 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, (2015), pp. 75-80
17. S. Gong, Y. Wang, Design and realization of lake water quality detection system based on wireless sensor networks, in 2011 Second International Conference on Mechanic Automation and Control Engineering, Hohhot (2011), pp. 4858–4861
18. F. Yuan, Y. Huang, X. Chen, E. Cheng, A biological sensor system using computer vision for water quality monitoring. IEEE Access **6**, 61535–61546 (2018)
19. C. Jian, Q. Suxiang, H. Hongsheng, Y. Gongbiao, Wireless monitoring and assessment system of water quality based on GPRS, in 2007 8th International Conference on Electronic Measurement and Instruments, Xi'an (2007), pp. 2–124–2–127
20. B. Das, P.C. Jain, Real-time water quality monitoring system using Internet of Things, in 2017 International Conference on Computer, Communications and Electronics (Comptelix), Jaipur (2017), pp. 78–82
21. J.E. Manley, Unmanned surface vehicles, 15 years of development. Oceans **2008**, 1–4 (2008)
22. J. Blank, B. E. Bishop, In-Situ modeling of a high-speed autonomous surface vessel, in 2008 40th Southeastern Symposium on System Theory (SSST) (2008), pp. 347–351

23. J.R. Higinbotham, P.G. Hitchener, J.R. Moisan, Development of a new long duration solar powered autonomous surface vehicle. Oceans **2006**, 1–6 (2006)
24. J. Curcio, J. Leonard, A. Patrikalakis, SCOUT—a low cost autonomous surface platform for research in cooperative autonomy, in Proceedings of OCEANS 2005 MTS/IEEE (2005), pp. 725–729
25. J. Larson, M. Bruch, J. Ebken, Autonomous navigation and obstacle avoidance for unmanned surface vehicles, in Proceedings of the SPIE 6230, Unmanned Systems Technology VIII, 623007 (2006)
26. M. Caccia, Autonomous surface craft: prototypes and basic research issues, in 2006 14th Mediterranean Conference on Control and Automation (2006), pp. 1–6
27. A. Leonessa, J. Mandello, Y. Morel, M. Vidal, Design of a small, multi-purpose, autonomous surface vessel, Oceans 2003. Celebrating the Past ... Teaming Toward the Future (IEEE Cat. No.03CH37492), Vol 1 (2003), pp. 544–550
28. J. Alves et al., Vehicle and Mission Control of the DELFIM Autonomous Surface Craft, in 2006 14th Mediterranean Conference on Control and Automation (2006), pp. 1–6
29. B. Zhang, G.S. Sukhatme, Adaptive Sampling for estimating a scalar field using a robotic boat and a sensor network, Proceedings. IEEE International Conference on Robotics and Automation **2007**, 3673–3680 (2007)
30. J.R. Higinbotham, J.R. Moisan, C. Schirtzinger, M. Linkswiler, J. Yungel, P. Orton, Update on the development and testing of a new long duration solar powered autonomous surface vehicle. Oceans **2008**, 1–10 (2008)

Knowledge-Based Medicine Recommendation Using Domain Specific Ontology

S. Subbulakshmi, Ramar, Devajith Jyothi, and S. Sri Hari

Abstract Our research motivation is to make ease and support decision-making process in medical field by providing recommendations to medical practitioners, while carrying out any clinical diagnosis. There are many ongoing related research works which provide recommendation in medical field, our work stands out over others by providing the necessary information's as recommendations, while carrying out clinical diagnosis. Thus, proposed system was developed in such a way that it acts as a second opinion to rectify their doubts during the entire process of clinical diagnosis. Existing works do not provide necessary information when combined to form meaningful knowledge representations. We try to sort out these problems by creating a conceptual structure which combines disease symptom and medicine domain. It also tries to create and represent knowledge on therapeutic domain by combining Allopathy medicines and Homeopathy medicines. To represent with end users, we developed an interface model namely knowledge-based clinical assistance (KBCA) to provide a working prototype to derive the recommendation.

Keywords Semantic web · Knowledge representation · Ontology · Inference rules · Clinical diagnosis · Assistance · Allopathy · Homeopathy · Therapeutic

1 Introduction

Information technology plays a very crucial role in the development of every sector in our society. Thus, the importance of providing the right information to the right person from the influence of selections available from various sources of raw data is very important. Our research intention is to create a new knowledge base on

S. Subbulakshmi (✉) · D. Jyothi · S. S. Hari
Department of Computer Science and Applications, Amrita Vishwa Vidyapeetham, Amritapuri, India
e-mail: subbulakshmis@am.amrita.edu

Ramar
Muthayammal Engineering College, Namakkal, Tamil Nadu, India
e-mail: dean.cse@mec.edu.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_14

the field of medicine recommendation and clinical diagnosis. For this purpose, the medicinal and clinical knowledge were developed from various evidences, extracted from different resources, widely used and available today. These recommendations were developed to assist the users during the decision-making process.

The most significant and critical part in the health care system are the clinical diagnosis, i.e., to examine a patient carrying a disease. So this proposed work as its title indicates is especially devoted to providing support to the doctors or medical practitioners which we consider as end users of our system, to rectify their doubt while diagnosing patients. Our work provides a baseline property of clinical diagnosis by initially providing a relationship between disease module symptoms module and medicine modules. Recommendations were made in the field of disease detection, symptom detection and medicine recommendation based on the inference rules applied over the knowledge representation.

Today, ontology plays a very crucial role in knowledge representation, reasoning assertions and inference rules, to create a huge set of knowledge information. Knowledge information of web services [1] with ontologies is used to select best service compositions. Ontologies used in mobile applications [2] provide a personalized recommendation of tourist spots to users. Similarly, social media analysis [3] for digital forensic investigation is realized with a rich set ontology mapped with media information and social behavior. Intelligent zoo management system is implemented with a semantic knowledge base [4] with a set of ontologies. Ontology bundled with quality details [5] of web services and user categories are used to find the optimal web service for personalized web service recommendation. High-level description on semantic web technology is used to overcome the problems faced by biomedical librarians [6] to structure library data and to describe data in more structured format.

We use ontologies for clinical based knowledge representation which describe resources pertaining to the field of health and medicines. Such ontologies provide valuable clinical suggestions based on a patient's disease history. The Protégé tool is used for creation and representation of ontologies. Appropriate logic's, axioms and inference rules are defined to enhance automatic knowledge retrieval based on the query thrown. This research concentrates on cardiovascular organ, its diseases and appropriate medicine recommendations. Major reason, for opting this topic is, heart disease is the leading cause of death, and studies reveal that malfunction of heart leads to major causes of disability all over the world. One such example is Novel corona virus or COVID-19 which directly affects the heart and lungs, patients with COVID-19 show cardiac muscle inflammation, which eventually leads to cardiac muscle damage. Another motivation in developing this work is to promote the use of appropriate and relevant medicine for specific diseases and symptoms.

We have created domain ontologies explaining both allopathy medicines and homeopathy medicines in par with disease and symptoms. The system is implemented by creating a knowledge base and a framework for both types of medicines, which depicts a conceptual structure explaining relevant medicines for cardio vascular disease. This work thrives to create collaborative ontologies of different domains to define a rich set of complex knowledge information in the form of profuse class relationships, property characteristics and description logic's. Comprehensive ontology

bundled with semantic data is used by the knowledge-based clinical assistance model to provide valuable suggestions and recommendations to health practitioners.

This model is useful to both medical practitioners and ordinary people who want to get the details of disease and available treatments for it. System could be useful in a scenario, where a patient is not feeling well. He could specify his symptoms and get details of diseases related to those symptoms. Once disease is identified, the system provides details of medicines that could be used as treatment for those symptoms. One of the listed medicines is selected and approved by medical practitioners, which could be administered to those patients so that they could get cured with ease. The system could also be used to get knowledge about available medicines for treatment of certain diseases or symptoms. This is essential because different drugs are made available in the market, whenever new drugs are discovered by the pharmaceutical company. It helps to provide knowledge information to doctors with ease.

Following Sect. 2 explains background study of related works, Sect. 3 elaborates methodology involved in creation of model, and Sect. 4 describes design and architecture of the system. Section 5 discusses query implementation with results, and finally, Sect. 6 comprises conclusions and future enhancements.

2 Related Works

Most research work in the field of health and medicine widely uses Protégé for designing and creating knowledge-based retrievals. The research [7] is a conceptual work designed to promote the use of traditional medicines. This work tries to elaborate on herbal medicines proposed by the World Health Organization (WHO), and they created a medicine ontology which is linked with proposed symptom ontology and disease ontology. The system has restrictions in design of inference rules and logic, used for knowledge retrieval.

Work done by Waheed et al. [8] is based on Unani medicine ontology which includes domain knowledge of Unani medicine. It includes core principle, symptoms, diseases, patient treatment and supports Unani medical practitioners and researchers to improve their knowledge and to have better understanding of available drugs. Unfortunately, Unani medicine is not a widely accepted form of medicine, so system usage is restricted to a closed set of limited users.

Onto diabetic [9] creates a knowledge base on diabetics which is used to recommend trials for Diabetic patients. It generates treatment procedures according to the user's request which works on two ontologies, diabetic patient's clinical analysis ontology and semantic profile ontology for clinical trials. Ontology analyses patient information according to patient symptoms and generates a risk score. System fails to store the knowledge base of medicines, and it recommends only the details of clinical processes to be carried out to cure disease.

Medicine expert application [10] takes results of tests carried out on cholesterol, blood pressure, ECG, etc., and generates risk value as recommendations to inform the patient about the condition of his/her current health. From the risk score patient,

treatment procedures get suggested. Above system could be used only by the medical experts and it is not usable to the common users.

GalenOWL [11] is to present a drug recommendation system based on Semantic Web technology. It is a semantic-enabled online service, capable of offering real time drug-drug and drug-diseases interaction discovery. Here, medical information and terminology had been translated to ontological terms. The system is limited to creation of ontology with knowledge representation and usage of ontology retrieval is not implemented. [12, 13] are recommendations based on Clinical Practice Guidelines (CPG) to generate evidence based Clinical Decision Support System (CDSS). It describes only general procedure to be followed, but in reality treatment procedure differs for different diseases.

The aim of the work [14] is to create a knowledge acquisition tool using unified medical language and generate medical logical modules. It elaborates semantic based health care engineering which includes natural language processing, data sharing, data integration, and ontology engineering. Armel et al. [15] propose semantic representation and deep ontology creation of African traditional medicines. It elaborates representation resources pertaining to the domain of study which includes a huge set of characteristics and predicates. Systems defined in [14] and [15] mainly focus on creation of knowledge base and fail to provide retrieval service.

Our research work to create knowledge-based clinical assistance elaborates ontology creation and mapping process of symptoms to disease, disease to symptoms and medicines to symptoms. Different sets of assertions are defined and inferenced with ontology descriptions which provide valuable clinical assistance without need for complex querying processes. KBCA acts as a machine independent agent system which retrieves relevant data from a rich set of knowledge information available in the form of ontologies, once the user provides details of his medical issues.

3 Methodology

Since our project is developed as a semantic web application, ontology structure acts as the foundation of entire work. Ontologies are widely used to define and represent knowledge [16], and it also helps us to easily conceptualize and describe classes in the area of interest and domain. It adapts a different set of properties to define conceptual structure. Data-properties are authorized to provide unique identification for individuals of classes. Object properties are utilized to initiate relationships among classes, members and individuals. This work defines ontologies on the heart disease domain along with its symptom domain and medicine domain. Ontology modeling process covers knowledge representation, which provides a semantic structure and knowledge base which helps in developing various web services or web-based applications.

Fig. 1 Major domains ontology used in KBCA

3.1 Domain and Scope of Ontologies

Domains of our ontology include medicines, symptoms and diseases. Medicine domain includes homeopathic and allopathic medicines, specific for heart diseases. These domains were introduced in order to support the medicine practitioners, while carrying out medical diagnosis of patients. Symptom ontology includes all the symptoms specific to heart diseases, which were developed to create an upper layer relationship between the medicines and symptoms.

Similarly, we developed disease centric ontology named Cardio_Vascular_Disease, created to recommend diseases. Figure 1 shows the major domains specified for the work.

3.2 Define Classes and Class Hierachcy

In this phase, classes and their super and subclass hierarchy were defined. Our research follows a top down approach to design class hierarchy. The design and implementation of classes were developed according to the requirements of the ontologies. The subclasses are the classes defining each property of the super class. All the classes in protégé were defined as subclass of things according to protége conventions. This work consists of five major classes Cardio_vascular_disease, Diseases, Symptoms, Homeo Medicines, Allopathic medicines. Figure 2 describes the classes and subclass distribution of homeo medicines. Similarly, all major domains have a list of class hierarchy.

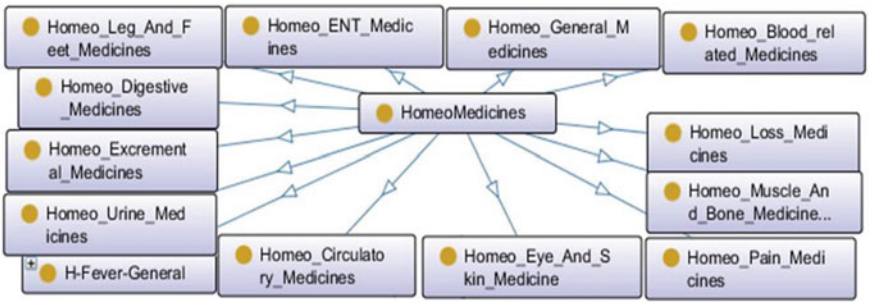


Fig. 2 Symptom class hierarchy of homeo medicine ontology

3.3 Define Object Property Data Property and Individuals

Ontologies become meaningful only when the upper-domain classes, and its hierarchy shares knowledge with other ontologies and classes. Classes and its hierarchy simply don't contain all the required knowledge to represent the proposed work, in order to make it meaningful and create new dimensions of knowledge we need class properties to identify and share information. In Protégé, we were given functionalities for creating object properties and data properties. If a property needs to relate between two individuals or between classes, then it needs to be an object property. Object properties may exist independently or may be related to one or more classes. Every object property is associated with domain and range. Domains are considered as superclasses and subclasses to which the object property belongs.

If we need to relate individuals to a literal or if we need to uniquely identify with a literal, then we make use of data property. It may be simple or complex, simple includes primitive data consisting of string, numbers, Boolean, etc., whereas complex property consists of enumerated data types. Sample set of object properties which act as a bridge between domain and range are specified in Table 1. Properties are also defined as functional, transitive, disjoint, equivalent properties based on the nature and logical requirements of domain knowledge.

3.3.1 Object Property

HasSymptoms relates with Cardio_Vascular_Disease as (Domain) and Symptoms as(Range), using this object property any subclasses which comes under Cardio_Vascular_Disease can be easily related with any specific symptoms which comes under the range. This object property is to simply relate the relationship that a particular disease can have a specific symptom. *IsaSymptomOf* property defines the inverse of HasSymptoms, it relates with symptoms as (Domain) and Cardio_Vascular_Disease as (Range), using this object property any subclasses which comes under symptoms can be easily be related with any specific diseases

Table 1 Sample set of domain and range relationship for object property

Domain	Object-property	Range
Cardio_Vascular_Disease	HasSymptoms	Symptoms
Symptoms	IsaSymptomOf	Cardio_Vascular_Disease
Diseases	DiseaseHasSymptoms	Symptoms
Symptoms	HasDisease	Diseases
Symptoms	HasMedicines	AlopathyMedicines
AlopathyMedicines	MedHasSymptoms	Symptoms
HomeoMedicines	HomeoMedHasSymptoms	Symptoms
Symptoms	SymptomHasHomeoMed	HomeoMedicines

Table 2 Sample set of data property assertions

Data properties	Data-property assertions
HomeoMedID	"H-M-Asarum6"^^xsd:string
Cardio-HeartDiseaselD	"HD-7ValvularDisease" "^^xsd: string
HD-Name	"ValvularDisease" "^^xsd:string
Medicine_Name	"Betahistamine" "^^xsd:string
MedicinelD	MD-CIR-TSB-4Beta-adrenegic-receptor"^^xsd:string
SymptomID	"FVR-stomach pain"^^xsd:string

which comes under the range. This is to simply specify that a symptoms (Domain) can lead to a particular disease (Range).

HasMedicines is another object property used to recommend that a particular Symptom (Domain) can have specific allopathic medicines (Range). Similarly, *SymptomHasHomeoMed* is another object property used to recommend that a particular Symptom (Domain) can have specific homeo medicines (Range). *MedHasSymptom* property defines the inverse property of *hasMedicines*, i.e., provide recommendation that a particular Allopathic medicines (Domain) can be recommended, for particular symptoms (Range). Similarly, *HomeoMedHas-Symptoms* property defines the inverse property of SymptomHasHomeoMed, i.e., provide recommendation that a particular homeo medicines (Domain) can be recommended for particular symptoms (Range).

3.3.2 Data Property

It describes relationships between instances or individuals or data values. It is the unique data value given to a property of an instance or individual created for a class, like unique identification, name, address, etc. Table 2 shows a sample set of data property assertion used in the system.

HomeoMedID provides a unique identity for homeo medicines. *Cardio-HeartDiseaseID* is to identify Heart Disease location. *CardioVascularSystemID* helps to identify and locate disease affecting parts of vascular disease. *HDName* holds names for heart diseases.

Medicine_Name provides names for allopath medicines which is of type string. *MedicineID* is to uniquely identify and locating allopath medicines. *SymptomID* is to identify and locating assigned symptoms. Most of the data property used in this domain are of type string.

3.3.3 Define Instance

Individuals are created after defining classes and its property assertions. It is created for domain specific classes. Each individual depicts a set of property assertions in

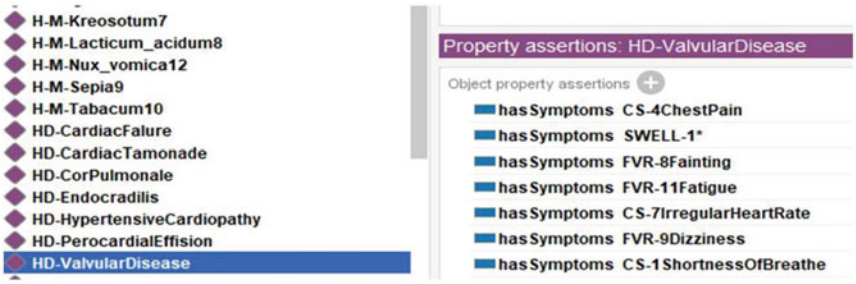


Fig. 3 Sample set of individuals and its property assertions

the form of data property values, object instance for object property and other logic defined such as functional dependency, transitive dependency, etc. Figure 3 shows a sample set of individuals created and it depicts a set of property assertions of the individual *HD-ValvularDisease* for the predicate *hasSymptoms.*

4 System Architecture of Knowledge-Based Clinical Assistance

Users of our system will be medical practitioners and common people suffering due to different causes who want to find knowledge information representing interaction between disease, symptom and medicines. On creating a knowledge base in the form of different ontologies, the actual process of clinical assistance starts with user interaction as depicted in Fig. 4 which consists of different paths.

First, the user specifies symptoms and type of medicines he prefers, then the system identifies disease with that symptoms. Here, user query flows from symptom domain to disease domain through the property h*asSymptoms*, to get the list of diseases. Similarly, if the user specifies the disease to identify symptom, flow of the system flows from disease domain to symptoms domain using the property *hasSymptomOf* and list the symptom details related to that disease. After that appropriate medicines are identified by mapping user given symptoms with symptoms and disease details stored as knowledge base for those medicines using *hasMedicines* property. It also considers other logic and inference rules defined for those medicines. So, finally it suggests a set of medicines to user, preferably to doctors in order to do confirmation about medicines before giving prescription to users. Doctors on going through medicines and patient's disease history may approve medicines.

The system is also able to work in reverse order, to identify list of diseases and symptoms for which a particular medicine. Such queries may be thrown by doctors in order to get details of available medicines and their usage. Above-mentioned functionalities are implemented for both Allopathy and Homeopathy medicines.

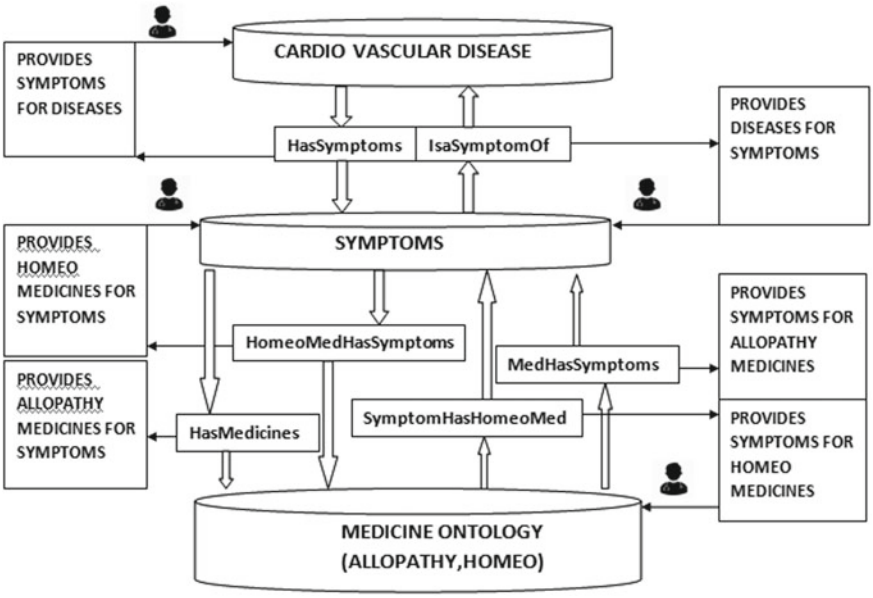


Fig. 4 Architecture Design of KBCA

SPARQL is used to retrieve relevant data, and it forms the integral process of retrieval system and it is explained in the next chapter.

5 Query Implementation

Knowledge retrieval process is realized using SPARQL [17] for querying domain ontologies. Domain resources of three ontologies are linked with property and class assertions. Suggestions of medicines for symptoms are made by considering the type of medicines, as every particular symptom is linked with the number of medicines in both allopathy and homeopathy treatments. Similarly, symptoms are retrieved for medicines or diseases in process of learning and knowledge acquisition by doctors to be used in treatments. Normally, details of the medicines are provided to doctors by medical representatives. Our system acts as an agent to render this task without any human effort. Thus, SPARQL query helps medicine practitioners to easily fetch required knowledge information and suggestions, which could be approved and prescribed to patients. Execution of sample queries for medicine recommendation, symptoms listing for given medicine and symptom-disease mapping are explained as follows.

All SPARQL query executed uses a set of standard prefixes which is mapped to Uniform Resource Locators (URL) of namespaces. Namespaces are web directory

or location where knowledge information is available in the web of world. It includes both standard namespaces and domain specific namespaces.

@PREFIX rdf: < http://www.w3.org/1999/02/22-rdf-syntax-ns# >

@PREFIX owl: < http://www.w3.org/2002/07/owl# >

@PREFIX rdfs: < http://www.w3.org/2000/01/rdf-schema# >

@PREFIX xsd: < http://www.w3.org/2001/XMLSchema# >

@PREFIXex: < http://www.semanticweb.org/devajith/ontologies/2021/1/untitled-ontology-33# >

5.1 Querying Knowledge Base for Medicines

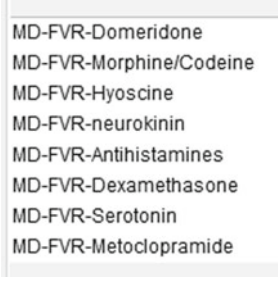
This task is initiated with symptom details of user, and appropriate query statements are executed to get medicines mapped with symptoms and its disease.

5.1.1 Medicine Recommendion

(a) Recommending allopathy medicines for symptom vomitting.

SELECT DISTINCT((?x) as ?name1).

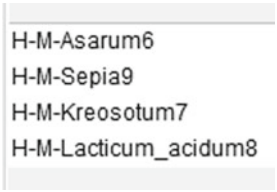
WHERE {ex:FVR-14Vomitting ex:hasMedicines ?x.}



(b) Recommending Homeo medicines for symptom Nausea.

SELECT DISTINCT((?x) as ?name1).

WHERE {ex:FVR-13Nausea ex:SymptomHasHomeoMed ?x.}



5.2 Querying Knowledge Base for Symptoms

5.2.1 Identifying the List of Symtoms to Medicines

This task is an inverse property of medicine recommendation. Medicine ontology acts as domain and symptom ontology act as range. Object property act as bridge connecting domain and range for symptom identification.

(a) Listing symptoms for Antihistamine medicine.

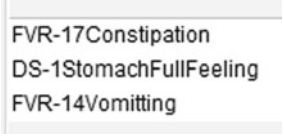
SELECT DISTINCT((?a) as ?name1).

WHERE { ex:MD-FVR-Antihistamines ex:MedhasSymptoms ?a.}

(b) Listing symptoms for homeo medicine Nux-Vomica.

SELECT DISTINCT((?a) as ?name1).

WHERE { ex:H-M-Nux_vomica12 ex:HomeoMedHasSymptoms ?a.}



5.3 Querying Disease Ontology

5.3.1 Identifying Symptoms for Disease and disease for Symptoms

In the linked data of disease and symptoms, symptoms act as domain and disease act as range for object property *hasSymptoms* and disease act as the domain and symptoms act as range for the inverse property *is Symptoms.*

(a) Listing symptoms for disease endocradilis.

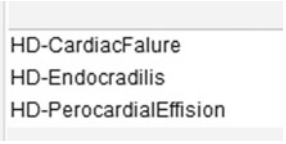
SELECT DISTINCT((?a) as ?name1).

WHERE {ex:HD-Endocradilis ex:hasSymptoms ?a.}

(b) Identifying diseases for symptom FeverAndChills.

SELECT DISTINCT((?a) as ?name1).

WHERE {ex:FVR-5FeverAndChills ex:isaSymptomof ?a.}



5.3.2 Identifying Symptoms Common for Combination of Two Diseases

(a) Perocardial effision and valvular disease.

SELECT Distinct ((?a) as ?name1).

WHERE { ex:HD-ValvularDisease ex:hasSymptoms ?a.

ex:HD-PerocardialEffision ex:hasSymptoms ?a.}

FVR-11Fatigue

5.3.3 Querying for Medicines used for Combination of Symptoms

(a) Vomitting and valvular Nausea.

SELECT DISTINCT ((?a) as ?name1).

WHERE {{ ex:FVR-14Vomitting ex:hasMedicines ?a.}

INTERSECTION.

{ ex:FVR-13Nausea ex:hasMedicines ?a.}}.

MD-FVR-Hyoscine
MD-FVR-neurokinin
MD-FVR-Metoclopramide
MD-FVR-Serotonin
MD-FVR-Domeridone
MD-FVR-Dexamethasone
MD-FVR-Antihistamines

Figure 5 depicts user interface design explaining top down approach from diseases to medicines as described in above queries. UI is designed as, first a particular heart disease is selected from the first drop down list box which populates list of heart

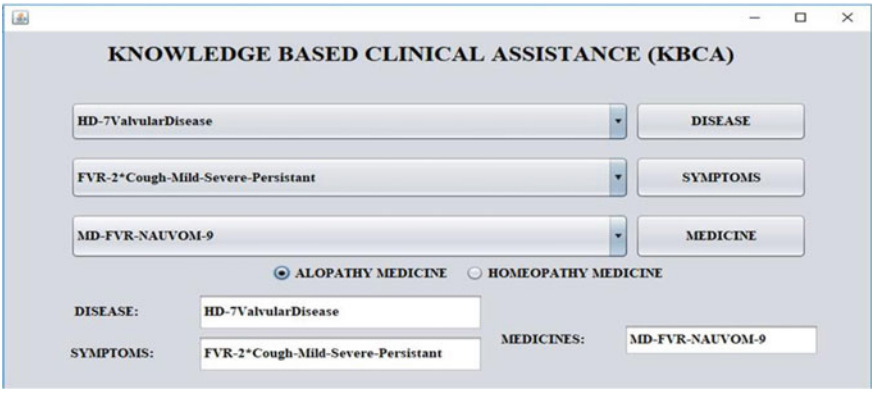


Fig. 5 User interface design for KBCA

diseases retrieved from disease ontology. Then, the second drop down list shows symptoms associated with selected disease.

Similarly, selections from symptoms can generate medicines associated to particular symptoms. Bottom up approach from medicines to disease is also implemented. UI is implemented in Net beans platform with java and its appropriate API to link ontologies and its querying statements [18] in the application. On execution of the UI all relevant knowledge information of medicines suggestion is retrieved from ontologies with minimum effort of selection of relevant symptoms or diseases. Updation of ontology could be automated linked data concept from different web resources. The system is proved to be a good system as it provides valuable suggestions to medical practitioners without much effort and minimum human interaction. It could also be used as a best clinical assistance during this COVID pandemic period where online doctor patient interaction is encouraged to avoid direct contacts.

6 Conclusion

The research work elaborates knowledge-based clinical assistance with semantic web technology. Initially, the knowledge base is created with information from various sources in the form of ontology classes, sub-classes, instances, assertions, logic and inference rules. It organizes common understanding of information relevant to the domain which is accessible by patients, medical practitioners and researchers. Ontologies are created for symptoms, heart disease and medicine domain bundled with a complex set of knowledge information. Medicines ontology includes knowledge base of both homeopathy and allopathy medicines. Clinical assistance helps medical practitioners to identify symptoms listing for a disease and medicines for the set of symptoms and diseases. It reveals mapping and linking of data between three domain information. SPARQL is used for querying knowledge information as per users requirements. System has also been designed with easy to use UI design which helps medical practitioners to get valuable suggestions with regard to medicines for diseases and to get details of medicines mapped with symptoms. Further, this research may focus on automatic linking of web data available in different web sources for creation of knowledge base ontologies. Limitation of the system is, it does not support dynamic updation of domain knowledge data into ontology, when new resources are available on the web. To overcome this issue, automatic linking of data will be implemented in future to keep up-to-date knowledge information in ontology.

References

1. S. Subbulakshmi, K. Ramar, A. Shaji, P. Prakash, Web Service recommendation based on semantic analysis of web service specification and enhanced collaborative filtering,in Intelligent Systems Technologies and Applications (Cham, 2018)
2. S. Kartik, R. Mohammed, R.F.H. Sailusha, R. Archanaa, Personalised tourism mobile application using semantic web, Int. J. Softw. Web Sci. **2**, (2014)
3. A.K. Mohan, D. Venkataraman, Forensic future of social media analysis using web ontology, in 2017 4th International Conference on Advanced Computing and Communication Systems, ICACCS (2017)
4. D. Venkataraman, R. Shaji, Improvising knowledge based acquisition on Zoo Management system, in 2017 4th International Conference on Advanced Computing and Communication Systems, ICACCS (2017)
5. S. Subbulakshmi, K. Ramar, R. Renjitha, T.U. Sreedevi, Implementation of adaptive framework and WS ontology for ımproving QoS in recommendation of WS, in The International Symposium on Intelligent Systems Technologies and Applications, Intelligent Systems Technologies and Applications 2016 (Cham, 2016)
6. I. Robu, V. Robu, B. Thirion, An introduction to the Semantic Web for health sciences librarians. J. Med. Libr. Assoc. **94**(2), 198 (2006)
7. B. Alkhatib, D. Briman, Building a herbal medicine ontology aligned with symptoms and diseases ontologies. J. Digit. Inf. Manag. **16**, 114 (2018). https://doi.org/10.6025/jdim/2018/16/3/114-126
8. T. Waheed, A.M. Enríquez, S. Amjad, A. Muhammad, Development of upper domain ontologies for knowledge preservation of unani medicines. Res. Comput. Sci. **68**, 115–125 (2013)
9. P.C. Sherimon, R. Krishnan, OntoDiabetic: an ontology-based clinical decision support system for diabetic patients. Arab J Sci Eng **41**, 1145–1160 (2016). https://doi.org/10.1007/s13369-015-1959-4
10. O. Arsene, I. Dumitrache, I. Mihu, Medicine expert system dynamic Bayesian Network and ontology based. Expert Syst. Appl. **38**(12), 15253–15261 (2011)
11. C. Doulaverakis, G. Nikolaidis, A. Kleontas, I. Kompatsiaris, GalenOWL: ontology-based drug recommendations discovery. J. Biomed. Semant. **3**, 14 (2012). https://doi.org/10.1186/2041-1480-3-14
12. S. Hussain, S.R. Abidi, S.S.R. Abidi, Semantic web framework for knowledge-centric clinical decision support systems, Conference on Artificial Intelligence in Medicine in Europe (Springer, Berlin, Heidelberg, 2007)
13. S. L. Achour et al. A UMLS-based knowledge acquisition tool for rule-based clinical decision support system development. J. Am. Med. Inform. Assoc. **8**(4), 351–360 (2001)
14. Z. He, C. Tao, J. Bian, M. Dumontier, W.R.Hogan, Semantics-powered healthcare engineering and data analytics (2017)
15. A. Armel et al. Towards a deep ontology for african traditional medicine. Intell. Inf. Manag. (2011)
16. R.A. Jenders, B. Dasgupta, Assessment of a knowledge-acquisition tool for writing medical logic modules in the arden syntax, in Proceedings of the AMIA Annual Fall Symposium. American Medical Informatics Association (1996)
17. G. Cima et al., On the SPARQL metamodeling semantics entailment regime for OWL 2 QL ontologies, in Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics (2017)
18. G. Veena et al., A concept-based model for query management in service desks. Innovations in Computer Science and Engineering (Springer, Singapore, 2016), 255–265

Deep Learning-Based Smart Mask for Social Distancing

L. V. Rajani Kumari, Mohammad Aatif Jaffery, K. Saketh Sai Nigam, G. Manaswi, and P. Tharangini

Abstract The COVID-19 is having a significant impact on our everyday lives. One of the measures put in place to slow the spread of the disease is social distancing. Because of the abrupt implementation of these laws, it become impossible for a person to follow this form of lifestyle. So we have developed a device that can assist both normal and visually disabled people in maintaining physical distance from other people by using deep learning to identify an individual and verify if the other person is six feet away from the user so that social distance can be sustained, and virus spread can be reduced. Our architecture includes a Webcam and a Raspberry Pi 4 model B to run an advanced image processing algorithm SSD MobileNet for human detection. SSD MobileNet is implemented to determine if the other individual is physically close to the user. If an individual does not adhere to social distancing, an automated reading aid is activated, and an audio warning is sent to the user. Since the whole setup is light and portable, it can be conveniently installed with any mask. The job is evaluated in real time. The results show that the proposed device improves mobility, convenience, and ease of navigation for both normal and visually disabled people in this pandemic situation.

Keywords COVID-19 · Social distancing · Smart mask · Raspberry Pi · SSD MobileNet · Object detection · Visually impaired · TensorFlow lite

L. V. Rajani Kumari (✉) · M. A. Jaffery · K. S. S. Nigam · G. Manaswi · P. Tharangini
Department of Electronics and Communication Engineering, VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad, TG, India
e-mail: rajanikumari_lv@vnrvjiet.in

M. A. Jaffery
e-mail: aatifjaffery_md17@vnrvjiet.in

K. S. S. Nigam
e-mail: sakethsainigam_n17@vnrvjiet.in

G. Manaswi
e-mail: manaswireddy_g17@vnrvjiet.in

P. Tharangini
e-mail: tharangini_p18@vnrvjiet.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_15

1 Introduction

The world is, as of now, battling COVID-19. Numerous nations have found primary ways to forestall this lethal infection like directing substantial screenings and 14 days required isolate for approaching global explorers. Also all visas, except for representatives and impermanent work, were wholly suspended. Because of the expanded danger of the infection spreading, numerous nations proclaimed a lockdown for the whole nation, restricting individuals from leaving their homes. Therefore, the infection's spread locally was seriously restricted because of this demonstration. In a new report, it has been shown that 60% of individuals are not after the standards any longer of wearing a veil or keeping up the close distance. The ramifications are much for those with vision disabilities. Visual hindrances diminish one's capacity to decide the distance to someone else, or it is amazingly troublesome in that situation. Our need is to help the client keep the Coronavirus anticipation administers cleverly to make the client aware of keeping up the legitimate social distance. Individuals can taint people before they understand they are wiped out. It is vital to keep at any rate six feet from others, whenever the situation allows, yet on the off chance that they are wiped out or not. Social removal is especially significant for individuals at a greater danger of creating extreme COVID-19 disease [1]. A model is fostered what might recognize the individual and check whether an individual is at six feet distance from the client with the goal that the client ought to be kept from the Coronavirus spread and the client ought to be completely secure when they go outside to satisfy their necessities, here the clients can be typical and outwardly disabled people. There is a parcel of related existing models that are accessible for recognizing items and estimating distance. A writing overview is finished by understanding what the existing model or framework is accessible in a word.

A comparative business related to our framework is created utilizing YOLOv3 and estimating the distance to get past friendly separating for an outwardly debilitated individual, however, the distance estimating method has some impediment, its relies upon the width and stature of the jumping box which may bomb when a little article comes in the edge of the camera, the procedure they utilized is the refraction of camera focal points. As we as a whole know, when a picture goes through a perspective, it is refracted because a beam of light can likewise enter the focal point, while the light consummately mirrors the picture in a mirror. When contrasted with faster R-CNN, YOLO, a technique utilized by different papers have a lower review and a higher restriction blunder. Since every network can propose two bounding boxes, it experiences issues identifying close by things and distinguishing little things, which are improved in this model. On account of the focal point, be that as it may, the portrayal is simply insignificantly developed. The outline underneath portrays how the picture shows up as anything but a focal point and its points; however, the procedure is bombed. We confirm its restriction in the exploratory outcomes segment that is segment three in this paper [2]. A framework utilizes a constant semantic division calculation on the RGB camera to perceive where individuals are and the profundity camera to decide how close they are. We then, at that point, convey aural input

through bone-leading earphones on the off chance that they are nearer than 1.5 m. The innovation educates the client if there is individuals around and does not respond to nonindividual things like dividers, trees, or entryways. The vision segment depends on the KR-vision scenes, which consolidate an RGB-D camera with bone-directing earphones for criticism [3]. The reference paper presents a profound learning framework for recognizing human-to-human distances to moderate the outcomes of the COVID pandemic. The identification device was made to caution clients when a video feed decides a protected separation from one another. This technique is generally used to distinguish social distance between people in CCTV or other fixed cameras [4]. The framework presents the visual social distance issue characterized as the mechanized evaluation of relational separation from a picture and the recognizable proof of related individual conglomerations. Visual social removing is needed to survey non-obtrusive if individuals hold fast to the social distancing (SD) limit and give security information in exact circumstances, when the impediment is penetrated. Singular congregations dispersed or potentially steady in friendly sign handling is often alluded to as "gatherings" or groups in PC vision [5].

In this paper [6], the creators fabricated a brilliant glass to recognize the article and visually impaired the thing by identifying and detailing the kind of thing to the client by the glass. The PC utilized a video feed to distinguish the item distance, just as the made item, utilizing an ultrasonic sensor. The ultrasonic sensor and microcontroller can distinguish the blockage and precisely gauge the distance. After social affair data from the environmental factors, it is shipped off the visually impaired individual using an earphone. The GSM/GPRS SIM900A module is utilized to obtain data from the Internet. A crisis gadget, like sending SMS with the time, temperature, and area to your watchman if individuals with low vision are at serious risk, is associated with a switch.

The topical audit is worried about deals with object recognition utilizing ebb and flow research in AI. Human recognition might be viewed as article discovery for arrangement in the PC vision assignment and video picture area of its shape. The profound examination uncovered an investigation pattern in object acknowledgment for various sorts of artificial reasoning discovery and was compelling extraordinary execution on troublesome datasets. Nguyen et al. documented an extensive, authoritative examination entitled recent human detection developments and challenges [7]. For the visual personality of a few picture ID benchmarks, innovation with the utilization of significant convolution neural network (CNN) organization has demonstrated better [8].

Profound CNN is a profound learning framework that utilization multi-facet neural perceptron networks with numerous convolution layers. Following that, loads in all organization layers are learned dependent on the dataset for everything arrangement. The CNN model was one of the profound investigation classifications managed by proper methods of adapting heartily to recognize the thing in different settings for object distinguishing proof in the picture. The firm has been especially fruitful in massive picture arrangement projects because of the most current elite PC innovations and enormous datasets, for example, ImageNet [9]. Many CNN models for object recognition with object area were proposed as far as organization engineering,

strategies, and one-of-a-kind ideas. CNN models like AlexNet [8], VGG16 [10], Inceptionv3 [11], and ResNet-50 [12] have been made lately to convey uncommon outcomes in object distinguishing proof. The achievement of broad information in object acknowledgment is refined because of its neural organization geography, which can self-develop the article descriptor and learning significant level properties not unequivocally indicated in the dataset.

A chronicled viewpoint on flow research, including an early investigation into electronic medical aid for movement, should be analyzed utilizing present-day fake vision models for BVIP routes; Many innovations are accessible for exploring BVIPs, including e-sticks canine aides, infrared sticks, laser-based walkers, and numerous others. Regardless of whether laser-based help is coordinated toward your own eyes or some other piece of your body, it could be unsafe to other people. These concessions are needed to convey this innovation. For exploring BVIPs, an expansive scope of innovations is accessible, including e-sticks or directing canines, infrared sticks, laser-based walkers, etc. While laser-based help is coordinated accurately toward their eye and different body pieces, it tends to be hazardous to other people. These concessions are needed to convey this innovation [13]. As our framework is valuable outwardly hindered people, a connected work is existing. However, using the additional sensors for identifying the items thus that the expense of their created item may expand as indicated by them, a few sensors are fitted on a material that could undoubtedly portable and it can help daze people [14]. The design of the paper is given as follows: Sect. 2 clarifies the new work and its curiosity. Segment 3 presents the discoveries, just as a correlation of existing methodologies. Segment 4 finishes up the paper.

The novelty of our made model is, as a matter of first importance, the framework's plan is flexible and straightforward by clients. Also, it is reversible and might be utilized with an assortment of covers. Third, it might recognize whether an individual is not sticking to social removing by estimating the distance between the client and the item. To wrap things up, it very well might be used by individuals who are both ordinary and outwardly tested.

2 Materials and Method

2.1 System Design

We have proposed a strategy for both normal and visually impaired people. OpenCV, Python, TensorFlow, and CNN were utilized as software tools, while Raspberry Pi, camera module, and earphones were used as hardware. The configuration may be adjusted to mask and only be attached to a camera that detects a human and checks whether a person has been detected at a distance of six feet from the user. When it comes to electricity, our project is battery-powered. It just needs 2.5–5 V to recharge in a short amount of time. Object detection is a computer vision approach

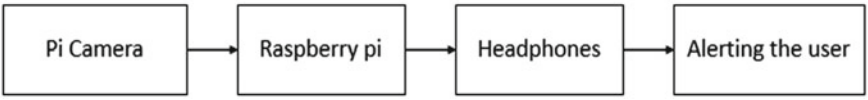


Fig. 1 Block diagram of the proposed system design

for identifying and locating things in images and videos. Object detection may be used to count items in a scene, determine and monitor their precise positions, and properly label them using this type of identification and localization. Convolutional neural networks (region-based convolutional neural networks (R-CNNs)), fast R-CNN, and YOLO are notable features object identification methods (You only look once). Figure 1 represents the block diagram of our proposed system design.

A social separation analysis methodology will be chosen and utilized to detect the interface, frameworks, and libraries such as TensorFlow object detection apps such as OpenCV, NumPy, and SSD. The label obtained from a visual object identification model may be sent into a speech synthesizer. eSpeak is a modest open-source (text-to-speech) voice synth for user auditory feedback. TensorFlow is an open-source advanced analytics framework that builds models using data flow graphs. It enables programmers to build large-scale neural networks with several layers. It is mostly employed in the areas of classification, perception, comprehension, discovery, prediction, and creation.

The working device is the Raspberry Pi 4 model B, which has the ability to multiprocessing. In the year 2019 of month June, the Raspberry Pi 4 model B was released. To identify obstacles and issue an alarm, a TensorFlow light object detection API was employed. The API was created using powerful, all-encompassing deep learning algorithms. The Raspberry Pi 4 model B has a 1.5 GHz quad-core Arm Cortex A72 CPU that can produce video at the full H.265 (4kp60 decoding) and H264 levels (1080p60 decode, 1080p30 encode).The Raspberry Pi 4 is capable of a stunning variety of tasks. For openers, Pi boards are used as media centers, file servers, vintage gaming consoles, routers, and system ad-blockers by amateur computer users. Figure 2 shows the proposed prototype of Raspberry Pi to which camera and headphone modules are integrated.

2.2 SSD MobileNet Architecture

In the race to identify the best method for object recognition, research teams are looking for a practical solution that is precise and has lower execution times for use in real-time apps. During the typical two-stage algorithms with regional proposals like the R-CNN [15] family, computationally demanding, and slow but in terms of accuracy using standard objects identification data sets such as MS COCO [16] and Pascal VOC2007&12 (usually measured by mean-average-precision or mAP). However, the fact that it employs small convolution filters to anticipate object classes

Fig. 2 Proposed prototype. Raspberry Pi with the camera module and headphones

and bound box positions for different aspect ratios are perhaps an essential component of the SSD [17] architecture. Also, it enables it to achieve high performance in terms of speed and precision. As a result of the utilization of several layers with relatively low input resolution and predictions at multiple scales, it may deliver high accuracy discoveries at faster detection rates. Figure 3 displays the brief overview of the MobileNet body architecture.

The MobileNet architecture is a convolution neural network [18, 19] that employs separable convolutions and has a lower parameter count and computational complexity. In addition, two additional global hyperparameters were used to build this network architecture. These are the width and resolution multipliers, which may

Type		Stride	Filter Shape	Size of input
Conv		S2	3X3X3X32	224X224X3
Conv dw		S1	3X3X32 dw	112X112X32
Conv		S1	1X1X32X64	112X112X32
Conv dw		S2	3X3X64 dw	112X112X64
Conv		S1	1X1X64X128	56X56X64
Conv dw		S1	3X3X128 dw	56X56X128
Conv		S1	1X1X128X128	56X56X128
Conv dw		S2	3X3X128 dw	56X56X128
Conv		S1	1X1X128X256	28X28X128
Conv dw		S1	3X3X256 dw	28X28X256
Conv		S1	1X1X256X256	28X28X256
Conv dw		S2	3X3X256 dw	28X28X256
Conv		S1	1X1X256X512	14X14X256
5X	Conv dw	S1	3X3X512 dw	14X14X512
	Conv	S1	1X1X512X1024	14X14X512
Conv dw		S2	3X3X512 dw	14X14X512
Conv		S1	1X1X512X1024	7X7X512
Conv dw		S2	3X3X1024 dw	7X7X1024
Conv		S1	1X1X1024X1024	7X7X1024
Avg Pool		S1	Pool 7X7	7X7X1024
FC		S1	1024 X 1000	1X1X1024
SoftMax		S1	Classifier	1X1X1000

Fig. 3 Table of MobileNet body architecture

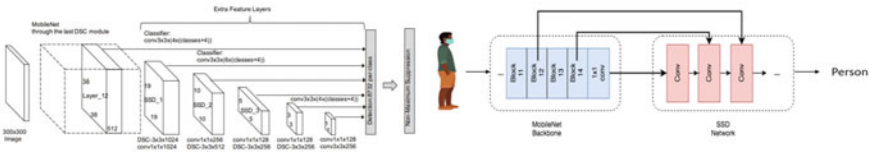


Fig. 4 SSD MobileNet architecture

be used to control the input/output channels of the layers and the input data resolution. These options can adjust the network's latency and accuracy according to the user's specific requirements.

The VGG had to be replaced with a network that reduced computation time while preserving accuracy. We found the ideal candidate in MobileNetv1. MobileNet is made up of Depthwise distinct innovative layers, which are much faster than traditional convolutional layers. The backbone network of the SSD architecture transforms incoming image pixels into features that characterize image contents and transmits them to the various layers of the SSD. As a result, it is used as a second neural network feature extractor.

Figure 4 shows connects MobileNet layers 12 and 14 of the Fig. 3 to SSD while replacing VGG16 with MobileNetv1.

The profoundly separable layer is coupled to a depth map 512 with a $1 \times 1 \times 512 \times 512$ filter (layer 12) for the table and image above (topmost in the above image).

The SSD layer is also connected to the final pointwise convolution layer with function map 1024. The MobileNet [20] has two hyperparameters that allow for a trade-off between accuracy and calculation. The width of the multiplier, which measures the input and output channels, and the resolution of the multiple which weights the input and output. It significantly decreases the computation's cost. We use the width multiplier extensively in our study because it influences the object detection mAP. However, we do not use the resolution multiplier since it interferes with the various size maps that object detection requires for the SSD. The width parameter indicates that the input and output channels have been reduced significantly, per MobileNetv1 [20]. This is exactly how we utilize the parameter with the number of channels acting as a single weight (0–1). The exact quantity of channels is weighted on every tier. The regression and classification SSD channel weighed with the same width factor linked to MobilNet, as illustrated in Fig. 2. We have to balance MobileNets with the same width factor and change the input and output channels of MobileNet.

2.3 Workflow of the System

Figure 5 depicts the whole process, including the proposed system's hardware and software interfaces. To generate the original picture or frame representation, each

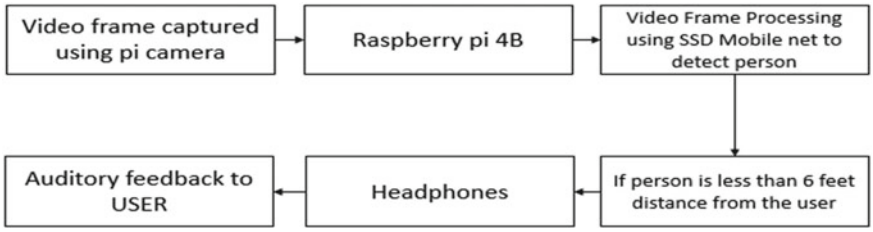


Fig. 5 Workflow of implementation of whole system

frame of the video is processed using a typical overlapping network. This backbone network is then pretrained using MobileNet as an image classification model in the SSD model to learn how to extract features from an image using SSD. The model then manually sets a set of bounding box aspect ratios at each grid cell place. For each border box, the offsets for the bounding box coordinates and size are predicted.

Let us take a quick look of the procedure from data collecting to the user's auditory feedback. The Raspberry Pi 4 model B is linked to other system components. The data were collected using a high-quality Sony IMX219 image sensor and a Raspberry Pi camera. The camera's sensor, with its fixed lens, is paired with the camera, which is designed to fit in the Raspberry Pi on board. It can capture a static image of 3280 pixels by 2464 pixels and a video of 1080p, 720p, and 640 pixels by 480 pixels. Small sockets link the specialized camera serial interface to the Pi module. Our program collects BGR data in real time and things from every previously known video frame may be recognized. The TensorFlow object detection API detects images from a live video feed by extracting features (objects). The TensorFlow object detection API is an open-source framework built on top of TensorFlow that can be readily integrated, trained, and constructed models that perform well in a variety of circumstances. TensorFlow, as the brains of object detection computations, represents deep learning networks. TensorFlow is the foundation of a visual object that includes a node network. The Protobuf library may generate visual objects to save network bandwidth. In the proposed architecture, dubbed the single-shot SSD-MobileNet detection, a pretrained model for TensorFlow recognition from the zoo was used. Google's model zoo is a collection of pretrained object-detection models that were trained on a variety of data sets, including the common objects in context (COCO) [16]. Formalized paraphrase the model was chosen specifically for the given prototype since it does not need high-end computer skills, which are commensurate with the Raspberry Pi's low processing capability. No more training is necessary to detect objects from the live video feed as the models have previously been trained in various object kinds. A picture can be tough since most of these potential places comprises background colors and not real subjects and spotting these items might be a challenge. SSD models often employ one-stage object sensing, which anticipates a picture to be bounded immediately by the object.

When the difference between the user base of the bounding box and the camera frame base (Eq. 1) exceeds a threshold value, our methodology, designed to identify if a person is sic feet away from the user, investigates it.

$$\text{Diff} = \text{Bottom of camera frame} - \text{Bottom of bounding box of the person} \quad (1)$$

It also determines the pixels covering the individual in the frame, when a particular threshold passes the height of the pixel and the width of the pixel, and a voice alarm is issued to the user so that the user bends its neck. Even if one criterion fails, additional conditions are monitored. There is therefore no opportunity for the user not to follow the social distance and when the user walks outside to satisfy his requirements, he may be at a safer distance from the other individual. Not just regular folks but even visually impaired folks are assisted by our designed alert system.

In Fig. 6, a man with the camera attached to the mask with an angle ***"a"*** in such a way that the bottom of that camera's camera frame lies precisely six feet from the person. The angle at which the camera is to be set so that the bottom frame of the camera lies at the six feet from the person is dependent upon the height of the person, and the person's distance from the user is more than six feet, so no alert is sent to the user.

Whereas in Fig. 7, the person's distance to the user is less than six feet so the system will alert the user as not following the social distancing. As we know, the six feet are constant and other variables such as angle and height are demonstrated in Eq. (2).

$$a = \tan^{-1}(D/h) \quad (2)$$

where,

a = angle at which camera to be arranged.

D = the distance at which physical distancing is observed. In this case ($D = 6$ feet).

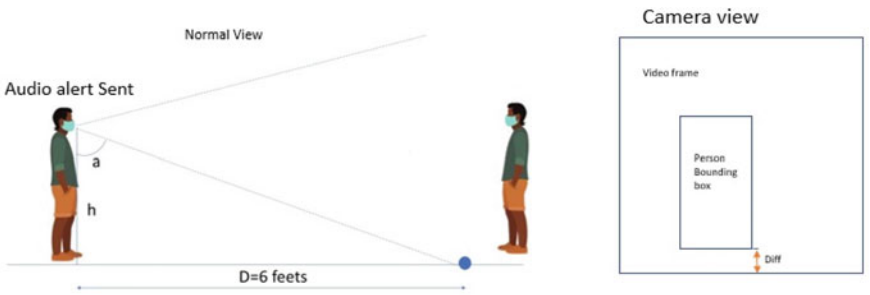


Fig. 6 Normal view and camera view of the detection model of person who is at social distance with other person

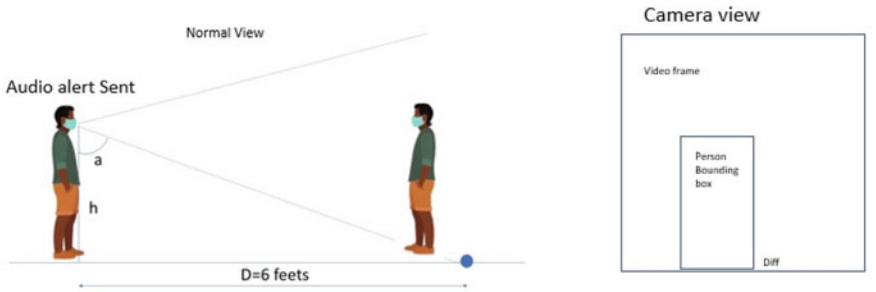


Fig. 7 Normal view and camera view of the detection model of person who is not at social distance with other person

Table 1 Angles according to the height of the user or person

Height (feet)	Angle ($a = \tan^{-1}(D = 6/h)$)
6–6.12	45°–40°
5–5.12	50°–45°
4–4.12	56°–50°

$h =$ height of the person or the user.

Angle at which camera should be arranged is calculated using Eq. (2). Table 1 shows the angles calculated with the help of Eq. (2) for different possible height of the person approximately.

In Table 1, only few standard heights are calculated as the variation in between them is less and it never impacts much on the system and the height are taken as an average person's height.

3 Experimental Results

3.1 Evaluating Results of Object Detection Models

Initially, before selecting the model, we tested the performance of the different pretrained model by implementing it with TensorFlow and selected the best performing pretrained model depending upon the FPS of the camera.

In Fig. 8a, the object detection model is pretrained with SSD VGG16 Net, as we can image the model's performance, which has an FPS of approximately 0.9. In contrast, in Fig. 8b, the object detection model is pretrained with SSD MobileNet v1, here, the FPS of it is approximately two.

Since the SSD MobileNet has better performance than the VGG16 Net, so SSD MobileNet is used as a detection model for our system.

Fig. 8 **a** Detection model implemented with SSD VGG16 Net **b** detection model implemented with SSD MobileNet

As the developed model is optimized according to the hardware while optimizing, the frozen inference graph is converted to the tflite. The whole model is modified and implemented in TensorFlow lite.

In Fig. 9, the detection model is optimized and implemented with TensorFlow lite in the Raspberry Pi [21]. We can observe that the FPS of our system is improved from two to five approximately, which more than two times better performance than two times better the TensorFlow model.

Performance of the SSD MobileNet object detection model can be determined by the mean average precision (mAP). mAP is a well-known assessment measure for object identification, localization, and classification. The location specifies where an instance is located. Object detection algorithms anticipate a class label as well as a bounding box.

For each bounding box, we measure the overlap between the predicted and actual bounding boxes. This is how intersection of union (IoU) is calculated.

$$\text{IntersectionofUnion(IoU)} = \frac{\text{Areaofoverlap}}{\text{AreaofUnion}} \tag{3}$$

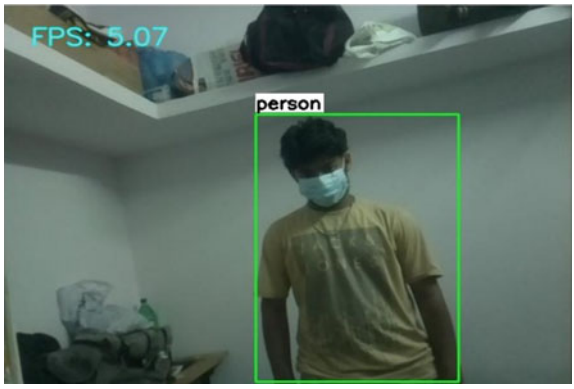


Fig. 9 Detection of person using optimized model deployed in Raspberry Pi

Table 2 Performances of SSD MobileNet v1

Performance parameter	Value
Mean average precision (mAP)	23.32
GFLOPs	2.494
Million parameters	6.807

For specific object detection tasks, we compute and return the precision and recall by utilizing an IoU value [22–24]. By calculating the mean AP for all classes or overall IoU thresholds, you compute the means average precision or mAP scores.

The performance is evaluated using precision and recall [25–27].

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{4}$$

$$\text{Recall} = \frac{T_p}{T_p + F_n} \tag{5}$$

We calculated the mAP by interpolating the 11 points.

$$\text{mAP} = \frac{1}{11} \sum_{\text{Recall}_j} \text{Precision}\left(\text{Recall}_j\right) \tag{6}$$

Table 2 shows performance of the SSD MobileNet v1 for detecting person the million parameters GFLOPs give the specifications of the model.

GFLOPs is defined as gigaflops is a measurement unit used for measuring the performance of the floating point unit of a computer, often known as the FPU. One gigaflop per second is one billion (1,000,000,000) FLOPS and million parameters considered to build the layers of the SSD MobileNet v1 object detection model.

3.2 Evaluating Results to Analyze Social Distance

The audio alert to the user is given when the user does not follow social distancing, that is, when another person's distance from the user is less than six feet. A method to calculate the distance is already available, but it fails in some instances.

As we can observe in Fig. 10, the person's distance depends on the width and height of the person that may fail when a person with less height will show as far distance and person more height will show less distance. In Fig. 7, as the two persons are standing at the same distance from the camera, the distances are different because the distance depends on the width and height of the ground truth box. According to the referenced paper [2], the equation used to measure the distance is Eq. (7).

Fig. 10 Measuring of distance according to reference paper

$$\text{dist(inches)} = \frac{2 \times 3.14 \times 180}{(w + h \times 360)} \times 1000 + 3 \tag{7}$$

Our advanced solution does not measure distance. However, as we know, the physical distance a person needs to follow from others is six feet, so by keeping this as default value and from Eq. (2) variable value will be the angle at which the camera has to be attached depending on the height of the user, as the motivation of that idea is obtained from the newly developed back camera for car's which will alert the driver when an obstacle reaches a point on it so in the same way we kept a threshold value for alerting the user when the other person distance from the user is less than six feet.

In Fig. 11, the person distance from the user is less than six feet, so that the system will alert the user through audio and we can observe the image there is a warning at the right bottom with the message "not following the social distancing." As a

Fig. 11 Detection of person and showing alert on the screen as person is not at the distance of six feet from the user

consequence of selecting and implementing a social separation analysis technique to detect the interface, frameworks, and libraries such as TensorFlow object detection applications such as OpenCV, NumPy, and SSD, a positive outcome is predicted. Finally, our system is ready after testing it in a real-time environment to be utilized. The developed is just a prototype and needed more optimized devices to make the user even more, ease to carry it out as the wireless Bluetooth earphones.

4 Conclusion

We proposed a prototype which can be integrated with wearable mask to protect the user from Coronavirus by giving alert to the user. This system can be helpful for both normal and visually impaired persons.

The following are the main features of the proposed system.

1. The complex algorithm processing with approximately5 FPS video stream using Raspberry Pi 4 model B.
2. Dual detection capabilities and a continuous analysis of social distance using a basic camera.
3. Audio-based feedback to the user which will be helpful to visually impaired persons.

Although there are lot of existing solutions to measure distance, but they all used either extra sensors or specialized cameras which will rather increase the cost of the product or increase the weight of the product. In addition to it, this model can be further optimized and deployed in a further small device which can be comfortable for a user to wear. The object detection model can be further increasing its detection of other objects for blind so that they can have a smooth navigation without any hurdles. Hence, our developed system helps a user wear a mask and reminds them to follow social distancing, which can help reduce the spread of the virus as far as possible for both regular and visually impaired persons.

References

1. H. Harapan, N. Itoh, A. Yufika, W. Winardi, S. Keam, H. Te, D. Megawati, Z. Hayati, A.L. Wagner, M. Mudatsir, Coronavirus disease 2019 (COVID-19): a literature review. J. Infect. Public Health **13**(5), 667–673 (2020). https://doi.org/10.1016/j.jiph.2020.03.019
2. M.A. Khan, P. Paul, M. Rashid, M. Hossain, M.A.R. Ahad, An AI-Based Visual Aid With Integrated Reading Assistant for the Completely Blind, in IEEE Transactions on Human-Machine Systems (published in 2020). https://doi.org/10.1109/THMS.2020.3027534.
3. M. Martinez, K. Yang, A. Constantinescu, R. Stiefelhagen, Helping the blind to get through COVID-19: social distancing assistant using real-time semantic segmentation on RGB-D video. Sensors **20**, 5202 (2020). https://doi.org/10.3390/s20185202

4. Y.C. Hou, M.Z. Baharuddin, S. Yussof, S. Dzulkifly, Social distancing detection with deep learning model, in 8th International Conference on Information Technology and Multimedia (ICIMU) (2020), pp. 334–338. https://doi.org/10.1109/ICIMU49871.2020.9243478
5. M. Cristani, A.D. Bue, V. Murino, F. Setti, A. Vinciarelli, The Visual Social Distancing Problem, arXiv preprint arXiv: 2005.04813v1 (2020)
6. M.R. Miah, M.S. Hussain, A unique smart eye glass for visually impaired people, in International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE) (2018). https://doi.org/10.1109/icaeee.2018.8643011
7. D.T. Nguyen, W. Li, P.O. Ogunbona, Human detection from images and videos: a survey. Pattern Recogn. **51**, 148–175 (2016). https://doi.org/10.1016/j.patcog.2015.08.027
8. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017). https://doi.org/10.1145/3065386
9. J. Deng, W. Dong, R. Socher, L. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, in 2009 IEEE Conference on Computer Vision and Pattern Recognition (2009), pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848
10. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014)
11. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) **2016**, 2818–2826 (2016). https://doi.org/10.1109/CVPR.2016.308
12. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) **2016**, 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
13. S. Khan, S. Nazir, H.U. Khan, Analysis of navigation assistants for blind and visually impaired people: a systematic review (published in 2021). https://doi.org/10.1109/ACCESS.2021.3052415
14. Rodríguez, Alberto; Yebes, J. Javier; Alcantarilla, Pablo F.; Bergasa, Luis M.; Almazán, Javier; Cela, Andrés, Assisting the visually impaired: obstacle detection and warning system by acoustic feedback. Sensors **12**(12), 17476–17496 (2012). https://doi.org/10.3390/s121217476
15. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6 (2017), pp. 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
16. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, C.L. Zitnick, Microsoft COCO: common objects in context. lecture notes in computer science (2014), pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48
17. W. Liu, D. Anguelov, D. Erhan, C Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, SSD: single shot multibox detector. Lecture Notes in Computer Science (2016), pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
18. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015), pp. 3431–3440
19. S. Albawi, T.A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network. International Conference on Engineering and Technology (ICET) **2017**, 1–6 (2017). https://doi.org/10.1109/ICEngTechnol.2017.8308186
20. A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv preprint arXiv:1704.04861v1 (2017)
21. J.W. Jolles, Broad-scale applications of the Raspberry Pi: a review and guide for biologists (published in 2021). https://doi.org/10.32942/osf.io/qh9sz,2021
22. V. Rajani Kumari, Y. Padma Sai, N. Balaji, Performance evaluation of neural networks and adaptive neuro fuzzy inference system for classification of cardiac arrhythmia. Int. J. Eng. Technol. [S.l.] **7**(4.22), 250–253 (Nov. 2018). ISSN 2227-524X
23. Mohebbanaaz, L.V. Rajani Kumari, Y. Padma Sai, Classification of arrhythmia beats using optimized K-nearest neighbor classifier, in *Intelligent Systems,* eds by S.K. Udgata, S. Sethi , S.N. Srirama. Lecture Notes in Networks and Systems, vol. 185 (Springer, Singapore, 2021)

24. R. Rohan, L.V.R. Kumari, Classification of sleep apneas using decision tree classifier, in 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (2021), pp. 1310–1316. https://doi.org/10.1109/ICICCS51141.2021.9432197
25. L.V. Rajani Kumari, Y. Padma Sai, N. Balaji, R-Peak identification in ECG signals usingp attern-Adapted Wavelet Technique, IETE J. Res. (2021). https://doi.org/10.1080/03772063.2021.1893229
26. L.V.R. Kumari, Y.P. Sai, N. Balaji, K. Viswada, FPGA Based Arrhythmia Detection, Procedia Computer Science **57**, 970–979 (2015)
27. Mohebbanaaz, Y.P. Sai, L.V.R. kumari, A Review on Arrhythmia Classification Using ECG Signals, in IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (Bhopal, India, 2020), pp. 1–6. https://doi.org/10.1109/SCEECS48394.2020.9

Dynamic Face Recognition System Using Histogram of Oriented Gradients and Deep Neural Network

L. V. Rajani Kumari, Syeeda Saher Fathima, G. Sai Praneeth, D. Mamatha, and B. Pranitha

Abstract Face recognition is a rapidly growing field. Face recognition is the ability to recognize and detect a person based on their visual features. Many methods are already proposed in which the face recognition system can recognize only a fixed set of people. This paper provides a revolutionary face recognition system implementation process that can recognize a large number of persons whose facial features may be derived from an input image. Face detection and recognition are the two phases of the human face recognition technique. In this model, face recognition is achieved by using histogram of oriented gradients and deep neural network. Histogram of oriented gradients (HOG) extracts all the important features from an image. HOG coupled with support vector machine (SVM) classifier detects the face from an input image. The localized face image is passed to the regression tree for facial landmark detection to obtain a centered face image which is fed as input to the deep neural network for face recognition. The proposed model gives an accuracy of 94.44% and is also implemented on video. This model can be used for biometric attendance systems.

Keywords Histogram of oriented gradients · Face detection · Convolutional neural network · Deep learning · Linear · SVM

1 Introduction

Humans know each other by looking at their faces, not by looking at their fingerprints or iris scans. As a result, facial recognition technology is less intrusive and easier to incorporate into our everyday lives and processes. In today's defense, administration, and business systems, biometric recognition software is becoming more and more important. Fingerprinting, retinal scanning, speech identification, and facial recognition are examples of biometrics. It is commonly used to authenticate users via ID authentication services.

L. V. Rajani Kumari (✉) · S. Saher Fathima · G. Sai Praneeth · D. Mamatha · B. Pranitha
Department of Electronics and Communication Engineering, VNR VJIET, Hyderabad, India
e-mail: rajanikumari_lv@vnrvjiet.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_16

Private and public CCTV camera feeds can be analyzed using this technology. The technology can be used for more than merely catching crooks. It could make it easier to find missing children and elder folks, for example, from 2020 to 2027, the global facial recognition market is projected to expand at a compound annual growth rate of 14.5%, reaching USD 9.9 billion. Face recognition technology was successfully utilized to recognize and follow a wanted fugitive at a performance attended by over 60,000 people [1].

Many statistical approaches like LDA and PCA were applied to develop a face recognition system. With the rise of deep learning, this approach proved to be a better methodology to achieve face recognition. Deep learning techniques can learn rich and compact representations of faces from very large datasets of faces, allowing modern models to perform on par with, if not better than human face recognition capabilities.

These traditional methodologies were trained on thousands of images of faces of a fixed set of people and therefore cannot recognize faces that were not included in the training data. This paper proposes a dynamic face recognition system that solves the above problem. This model is trained for the extraction of facial features from an input image per person. In this model, the face is detected from the images using the HOG and SVM and the output is fed to the convolutional neural network for face recognition after facial landmark detection. The proposed methodology efficiently recognizes the faces from the images and videos.

The structure of this paper is listed below. Section 2 discusses a study of related work in this field. Section 3 explains how the proposed face recognition methodology works. Section 4 presents the obtained results. Finally, in Section 5, the paper concludes with a suggestion for potential improvements and a list of applicable references.

2 Related Work

A substantial quantity of literature exists on the subject of face recognition. In this section, the paper focuses on the previous articles regarding face recognition. The previous articles on facial recognition systems were compared in this study. Different methodologies exist such as principal component analysis (PCA) and linear descriptive analysis (LDA) used by Bhale and Mankar. These algorithms are also used to reduce dimensionality [2]. The issue with this model is that it did not achieve good accuracy. Zhao Pei, Hang Ku, and Yanning Zhang presented a face recognition model based on orthogonal experiments combining deep learning and data augmentation [3]. A convolutional neural network (CNN) based on face detection was built for face identification using a convolutional neural network (CNN) based on TensorFlow, an open-source deep learning framework, as suggested by Liping Yuan and Zhiyi Qu [4].

Marko Arsenovic and Srdjan Sladojevic presented a deep neural network-based face recognition attendance system. The data augmentation was done on original

photos because the DNN can reach great accuracy on larger datasets. High accuracy was attained with a lower number of facial photos and the suggested approach of augmentation [5]. Muhmmad Zeeshan Khan, Saad Harous, Saleeth UL Hassn presented a convolutional neural network methodology and edge computing to perform face recognition efficiently. Edge computing has been used to process data at the nodes' edges to reduce data latency and improve real-time reaction times [6].

Taigman [7] and colleagues propose a deep face model based on CNNs architecture. 4.4 million face pictures were used to train this model. VGG-16 is a deep CNNs architecture proposed by S. Symonyan [8]. Meenakshi and M. Shiva Jyothi [9] demonstrated a face recognition system using deep neural networks across pose and illumination variations. Peter. J. B. Hancock, Rosyl S. Somal, and Victoria R. Mileva proposed a model titled convolutional neural net face recognition works in non-human-like ways [10]. The effort began with the observation of the CNN face recognition system. This model shows the performance of the CNNs and human participants on the four matching tasks. In the human data, there is a slight bias to say mismatch in the last three tests where there is an equal number of match and mismatch trails.

The framework naturally identifies the understudy face as he/she enters the class, and the framework is capable of marking the participation by remembering him/her, according to the Robotized attendance system utilizing face recognition [11]. Face detection is done with the Viola-Jones algorithm, which uses a course classifier to identify human faces, and PCA to include choice and SVM for grouping.

The framework perceives the face using Eigenface and Fisher face calculations by looking at receiver operating characteristics (ROC) bend, according to the execution of face acknowledgment calculation for biometrics-based time participation framework [12]. According to the results of the experiments in this research, ROC shows that Eigenface achieves the preferred outcome over Fisher's face.

Face recognition using LBPH face recognizer or anti-robbery and observation application using drone technology [13] provides a framework that uses local binary pattern histogram (LBPH) recognizer installed on drone technology. It is used as a reconnaissance drone that covers some areas; when the system recognizes a wanted person, it transmits the person's picture together with his or her location directions to the person in question. FaceNet is a face recognition system developed by Florian Schroff and Dmitry Kalenichenko. Rather than using an intermediate bottleneck layer, this approach uses a deep convolutional neural network to optimize the embedding. According to the authors, the most significant component of this strategy is the system's end-to-end learning [14].

The area of concern of the paper written by Gurlove Singh and Amit Kumar Goel is developing a face recognition system using digital image processing [15]. The proposed work did not achieve an accuracy greater than 90%. Further work where more effort is needed is in the field of completely automated frontal view face detection, which when exhibited virtually demonstrates faultless accuracy. The work previously done on face recognition systems are static models, hence the requirement of a dynamic system arises.

3 Proposed Methodology

The proposed model has two phases. Face detection and recognition are the two phases of facial recognition. Figure 1 represents the model architecture of this model. The histogram of oriented gradients (HOG) and linear support vector machine (SVM) are used to create a face detector. Although the HOG + SVM-based face detector [16] has been around for some time and has a sizable user base. Faces are detected even in cases where faces are not completely frontal. For a frontal face detector, this is excellent. An additional feature added to the model is facial landmark detection which is achieved using a regression tree [17]. The convolutional neural network receives the regression tree's output as input which is trained to generate face embeddings for every face present in an input image. The next step involves facial distance estimation which checks similarity between generated face embeddings from an input image with those which are present in the database. If there exists a similarity between two face embeddings, the face is recognized else the generated face embedding is appended to the database.

3.1 Face Detection Phase

Facial landmark detection, histogram of oriented gradients, and linear support vector machines are used to detect faces.

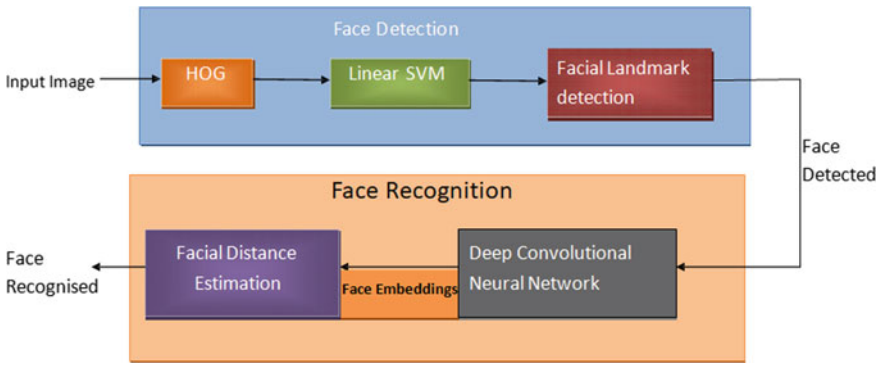


Fig. 1 Model architecture

3.1.1 Histogram of Oriented Gradients (HOG)

The block diagram of HOG is shown in Fig. 2. The image is initially sent into the model as input. Normalization is carried out on images and gradients are computed to reduce dimensionality. The gradient of the pixel $\nabla f(x, y)$ is defined as Eq. (1).

$$\begin{bmatrix} \frac{df}{dx} \\ \frac{df}{dy} \end{bmatrix} = \begin{bmatrix} f(x+1, y) - f(x-1, y) \\ f(x, y+1) - f(x, y-1) \end{bmatrix}$$

Here,

$$(g_x, g_y) = \left(\frac{df}{dx}, \frac{df}{dy} \right) \tag{1}$$

The df/dx term is the partial derivative on the x-axis, which is calculated as the color difference between adjacent pixels on the left and right of the target, $f(x + 1, y)$–$f(x, y-1)$, and the df/dy term is the partial derivative on the y-axis, which is calculated as $f(x, y + 1)-f(x,y-1)$, the color difference between adjacent pixels above and below the target.

An image gradient has two crucial characteristics:

1. Magnitude is defined as the vector's L2 norm, which is defined as Eq. (2),

$$g = \sqrt{g_x^2 + g_y^2}. \tag{2}$$

2. The arctangent of the ratio of partial derivatives in two directions, which is defined as is called direction.

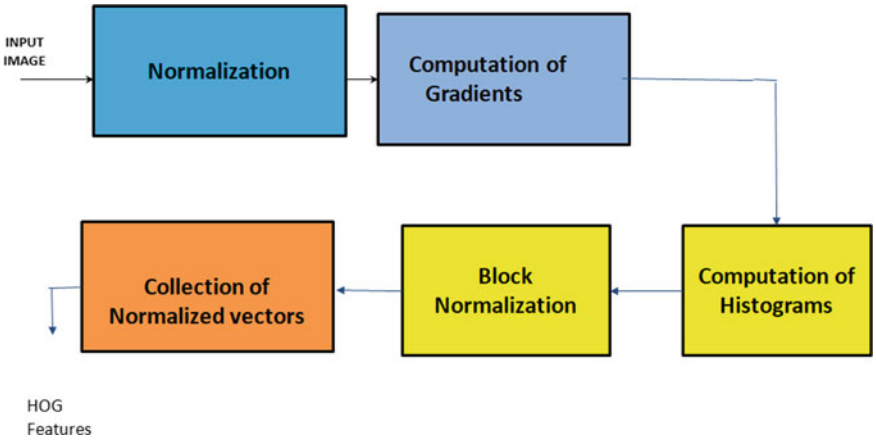


Fig. 2 Block diagram of HOG

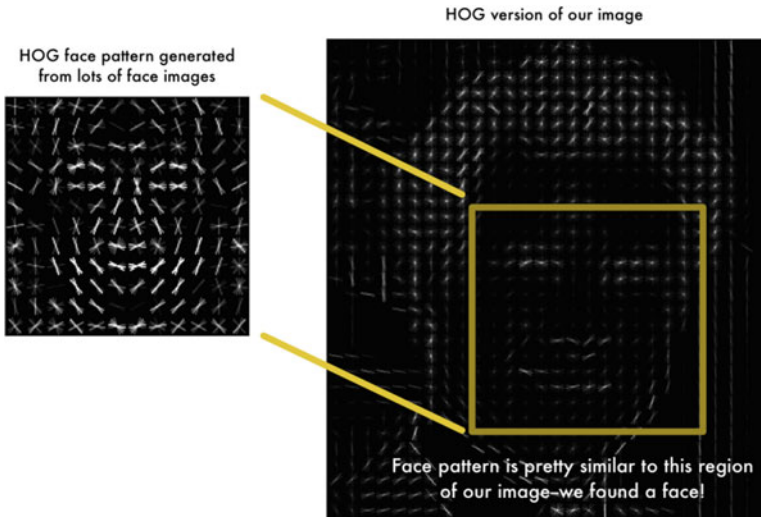


Fig. 3 Face localization

$$\theta = \arctan\left(g_y / g_x\right) \tag{3}$$

By splitting the picture window into tiny spatial segments throughout the experiment, each cell accumulates gradients over the cell's pixels [18]. The representation is made up of the combined histogram entries. On every four segments, block normalization is performed by tiling the detection window with a dense (often overlapping) grid of HOG descriptors and stacked onto a linear vector all over the image. The number of HOG features depends on the size of the image.

3.1.2 Support Vector Machines (SVMs)

The HOG features obtained above are input into a support vector machine that has been trained. The SVM was trained on HOG features of the face and HOG features that are not face. Therefore, the HOG features obtained are convoluted with those of face HOG features and fed to SVM. The obtained result is face localization. Figure 3 is a pictorial representation of face localization in an image.

3.1.3 Facial Landmark Detection

After the face is localized, the regression tree locates the facial landmarks on the obtained face. The purpose of locating the facial landmarks is to obtain a centered face image. The regression tree was trained on the images of faces of people with the location of 68 points on the face. This 68 point vector is called a facial landmark

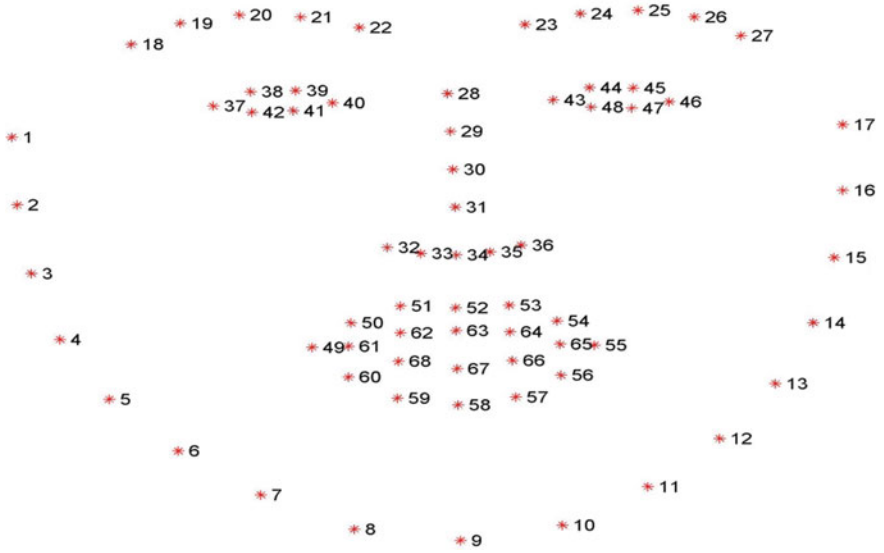


Fig. 4 Visualization of 68 facial landmarks

[19]. Then, the regression tree is used to detect landmarks on the face from an image in milliseconds. A regression tree is based on binary recursive partitioning. Binary recursive partitioning is an iterative procedure that divides data into partitions or branches, then separates each partition into smaller groups as the procedure continues along each branch. The pretrained landmark detector recognizes 68 points [20] in a human face ((x, y) coordinates). Figure 4 represents the visualization of facial landmark points. The area around the eyes, brows, nose, mouth, chin, and jaw is described by these points.

3.2 Face Recognition Phase

3.2.1 Deep Convolutional Neural Network

After the facial landmarks are located on the face, the deep convolutional neural network receives the centered image as input [21, 22]. In terms of algorithms, the convolution layer and the CNN layer share parameters. The structure of the CNN architecture VGG-16 [23] is shown below in Fig. 5. The number of layers in VGG-16 is less compared to those of other CNN architectures and also provides great accuracy.

The model's output is a 128-dimensional vector known as face embeddings which are unique for every face. At each epoch, the model was trained with two photographs of the same person and one image of a different person. The model was tweaked to

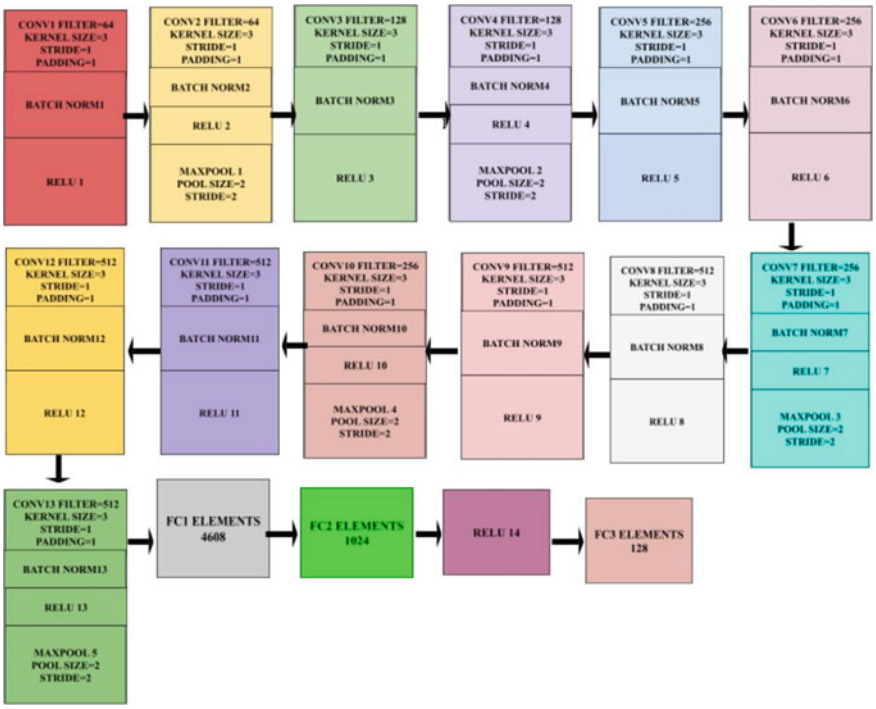


Fig. 5 VGG-16 architecture

generate similar face embeddings for the images of the same person and widely varied face embeddings for the image of a different person.

3.2.2 Facial Distance Estimation

Facial distance estimation is the measure of similarity between two faces. After the model is trained, facial distance is calculated between two images and the output is displayed. Figure 6 represents the flowchart for this process. One input image is considered for the process of face registration whose facial embeddings are stored in a facial database. The facial database consists of "known embeddings" which store facial embeddings and "known names" which stores names of corresponding embeddings. The image considered in the model is of RGB color space. After the face embeddings are generated for each face from the model, they are first compared to all the face embeddings present in the facial database. If there exists a similarity between two face embeddings, the face is recognized with the display of the name. The similarity between the two faces is checked using Euclidean distance. It is defined as Eq. 4.

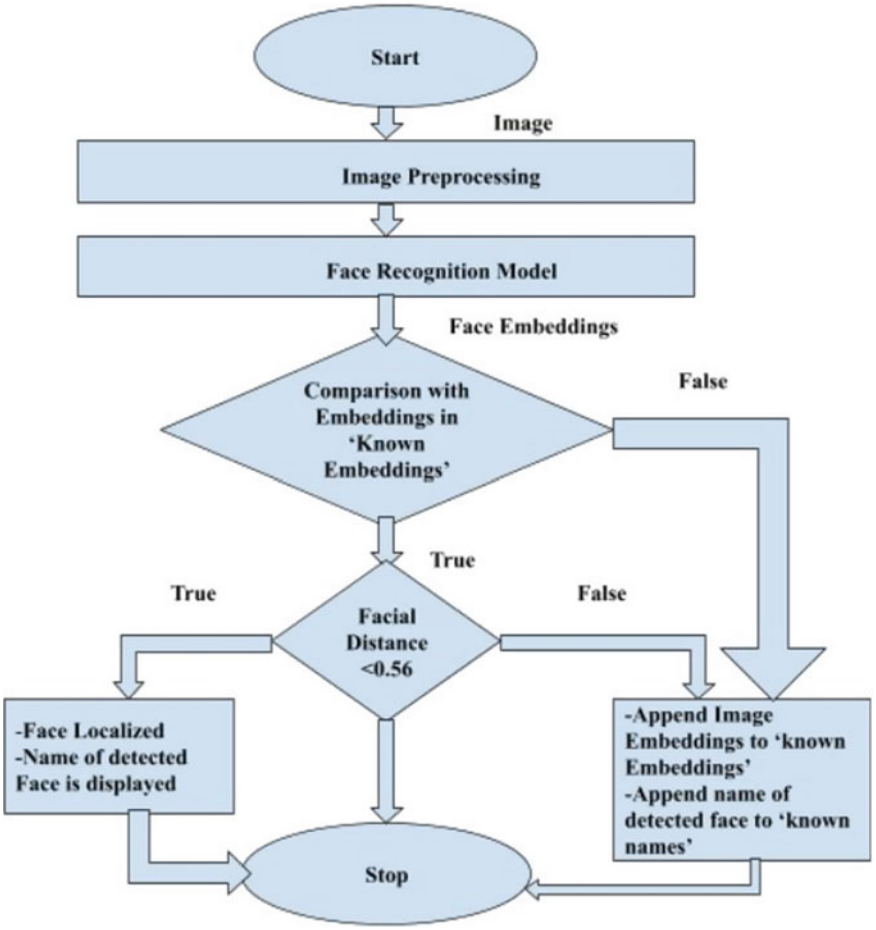


Fig. 6 Flowchart of facial distance estimation

$$d(p,q) = \sqrt{\sum_{i=1}^{128}(p_i + q_i)^2} \tag{4}$$

where p and q are defined as the two face embeddings, respectively. If the distance is greater than a threshold of 0.56, the face embeddings are appended to the facial database, else the corresponding name for which the minimum facial distance is obtained among all the face embeddings are displayed along with the face localization.

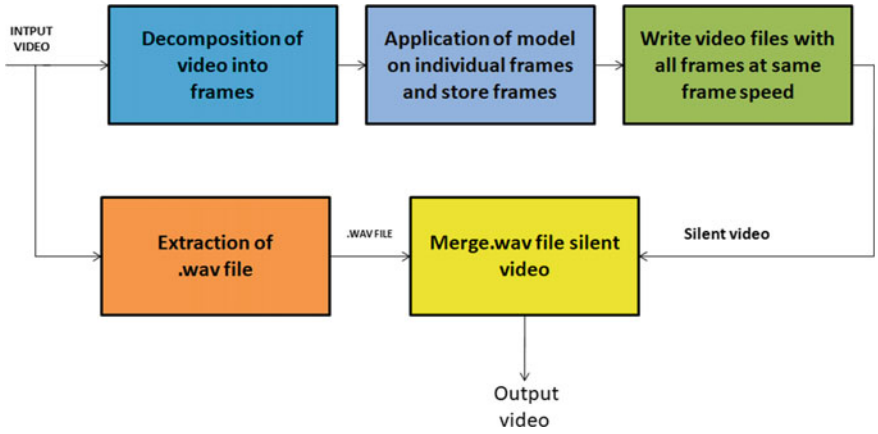


Fig. 7 Implementation of the model on video

3.3 Implementation of Face Recognition Model of Video

In recent years, there has been an increasing number of studies on video face recognition. Video-based recognition enables the integration of weak data from individual frames across time, perhaps leading to more accurate recognition despite constraints like pose variations and face expression. Zhang and Martinez [24] investigated whether approaches for recognizing faces from a single still image would perform better if they were extended to incorporate multiple photos or video sequences. Figure 7 represents the implementation of the model on video.

The implementation was done on both offline and YouTube videos. To detect the face from the video, initially, the video is split into frames, then the proposed model is applied on individual frames and these frames were clubbed together and written in video file to acquire silent video [25].

The audio of the video is lost when the video is split into frames. Here, the frame speed is the same as that of the original video. The audio for that video is extracted parallelly in the .wav extension, and the audio is merged with silent video with the same bit rate.

4 Results

This model was tested on 1000 images containing 1044 faces of different people. The testing was carried out on the model manually. Out of 1044 faces, the model detected 1030 faces and correctly recognized 986 faces. The performance of the model is evaluated in terms of accuracy [26–29]. The accuracy for any given model [29] is defined as Eq. (5)

Table 1 Accuracy of phases

Phase	Accuracy (%)
Face detection	98.65
Face recognition	94.44

$$\text{Accuracy} = \frac{(\text{No.of correctly recognized faces})}{(\text{Total No of detected faces})} \times 100 \tag{5}$$

Table 1, above, describes the accuracies of face detection and face recognition phases. Therefore, the accuracy achieved for the proposed model is 94.44%. This model can recognize people's [30] faces in various situations as shown in Fig. 8. It can recognize the person from his/her partial side view. And also it can recognize faces with or without a beard. Another observation is that a person's change in hairstyle does not affect the model. The model can identify people who transitioned from adolescence to adulthood or adulthood to old age. The proposed model can

Fig. 8 Model tested on various images

recognize a person's picture from ten years ago [31]. It can also recognize a person's picture, both with and without glasses. It also works on identifying faces from black and white photographs. Faces of people who have undergone plastic surgery can also be recognized by the model even when the model is given a photo of them before surgery. The model can recognize faces even when eyes are fully closed or half shut. The results are pickled and stored in a text file.

5 Conclusion

Various attendance and monitoring technologies are now in use in business, necessitating the usage of a dynamic face recognition system. A dynamic technique for facial recognition is provided in this paper. The proposed model extracts facial features from a single image and adds them to the database and therefore the model can identify many people compared to traditional face recognition systems. The model performs well in various scenarios and is also implemented on video. The accuracy obtained for the proposed model is 94.44%. This model can be used for a biometric attendance system. Further, improvements can be done in the model to recognize faces from images having bad illumination, low quality, rotation, and complete side view faces.

References

1. E. Drive, D. Casuarina, *Deep Learning for Face Recognition: A Critical Analysis* (Published in 2019)
2. S.G Bhale, V.H Mankar, *A Review on Face Recognition System* (2012)
3. Z. Pei, H. Ku, Y. Zhang, M. Guo, Y.H. Yang, Face recognition via deep learning using data augmentation based on orthogonal experiments (2019)
4. L. Yuan, Z. Qu, Y. Zhao, H. Zhang, Q. Nian, A convolutional neural network based on transfer flow for face recognition system (2017)
5. M. Srsenovic, S. Sladojevic, A. Anderla, D. Stefanovic, Face time-deep learning face recognition attendance system (2017)
6. M. Zeeshan Khan, S. Harous, S. UL Hassn, Deep unified model for face recognition based on convolutional neural network and edge computing (2017)
7. Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: closing the gap to human-level performance in face verification, in *Proceedings*, 23–28 June (2014)
8. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition (2014)
9. S. Meenakshi, M. Shiva Jyothi, D. Murugan, Face recognition system using deep neural network across variations in pose and illumination (2019)
10. P.J.B. Hancock, R.S. Somal, V.R. Mileva, Convolutional neural net face recognition works in non-human-like ways (2020)
11. I. Gupta, V. Patil, C. Kadam, S. Dumbre, Face detection and recognition using raspberry Pi, in *2016 IEEE International WIE Conference 185 on Electrical and Computer Engineering (WIECON-ECE)*(IEEE, Dec 2016), pp. 83–86. ISBN 978-1-5090-3745-2. https://doi.org/10.1109/WIECON-ECE.2016.8009092

12. A. Jadhav, A. Jadhav, T. Ladhe, K. Yeolekar, Automated attendance system using face recognition. IRJET **4**(1) (2017)
13. A.R.S.Siswanto, A.S. Nugroho, M. Galinium, Implementation of face recognition algorithm for biometrics-based time attendance system. ICISS. IEEE (2014)
14. F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: a unified embedding for face recognition and clustering
15. G. Singh, A.K. Goel, Face detection and recognition system using digital image processing, in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (2020)
16. N. Dalal, B. Triggs, Histograms of oriented gradients for human detection
17. V. Kazemi, J. Sullivan, One millisecond alignment with an ensemble of regression tree (2014)
18. F.Z. Salman, A. Madani, M. Kissi, Facial expression recognition using decision tree (2016)
19. Y. Wu, T. Hassner, K. Kim, G. Medioni, P. Natarajan, Facial landmark detection with tweaked convolutional neural networks (2018)
20. M. Sharma, J. Anuradha, H.K. Manne, G.S.C. Kashyap, Facial detection using deep learning (2017)
21. R.I. Bendjillali, M. Beladgham, K. Merit, A. Taleb-Ahmed, ILlumination robust face recognition based on deep convolutional neural networks architecture (2020)
22. S. Singh, D. Singh, V. Yadav, Face recognition using HOG feature extraction and SVM classifier (2020)
23. S. Liao, A.K. Jain, S.Z. Li, A fast and accurate unconstrained face detector (2016)
24. M. Zhang, A.M., from static to video, Face recognition Using a Parabolistic Approach, in *International Workshop On Face Processing in the video* (2004)
25. Z. Zhang, et al., Facial landmark detection by using deep multi-task learning (2014).
26. V. Rajani Kumari, L. Padma Sai, Y. Balaji, Performance evaluation of neural networks and adaptive neuro-fuzzy inference system for classification of cardiac arrhythmia. Int. J. Eng. Technol. [S.l.] **7**(4.22), 250–253 (2018). ISSN 2227-524X
27. L.V. Mohebbanaaz, Y. Rajani Kumari, P. Sai, Classification of arrhythmia beats using optimized K-nearest neighbor classifier, in S.K. Udgata, S. Sethi, S.N. Srirama (eds) Intelligent Systems. Lecture Notes in Networks and Systems, vol. 185 (Springer, Singapore, 2021)
28. R. Rohan, L.V.R. Kumari,Classification of sleep apneas using decision tree classifier, in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (2021), pp. 1310–1316. https://doi.org/10.1109/ICICCS51141.2021.9432197
29. L.V. Rajani Kumari, Y. Padma Sai, N. Balaji R-peak identification in ECG signals using pattern-adapted wavelet technique. IETE J. Res. (2021). https://doi.org/10.1080/03772063.2021.1893229
30. Y. Sun, X.X. Wang, X. tang, Deep learning face representation from predicting 10,000 classes (2014)
31. Y. Tagman, M. Yang, M. Ranzato, L. Wolf, Web scale training for face identification (2015)

Design and Implementation of Automatic Line Follower Robot for Assistance of COVID-19 Patients

Md. Alomgir Kabir, Md. Jakirul Sarker, Tomal Hossain, Mosa Israt Jahan Jerin, and Md. Hazrat Ali

Abstract Coronavirus disease (COVID-19) has caused unprecedented global health problems, and the disease's spread rate is extremely high. It spreads from infected people (COVID-19 positive) to others via droplets from the mouth or nose when they cough, sneeze, speak, sing, or take deep breaths. Frontline fighters of health-care organizations such as doctors, nurses, and other medical staff cannot have direct contact with COVID-19 patients in isolation room without personal protective equipment (PPE). Hence, hospital workers have to face different types of problems in distributing foods, medicines, and disposal of waste. An Automatic Line Follower Robot (ALFR) is designed and implemented for COVID-19 patients which is capable of serving infected patients in an isolation room. The main contribution of this paper is to serve essential medicines and foods from the hospital staff and serve it to the patients following the black line. The ALFR also proposes a system which maintains an emergency wireless communication protocol between doctors and patients. It also collects waste from a specified basket and damps it to a proper place. Finally, it can sanitize the isolated room with the help of a disinfectant machine which is assembled in ALFR. ALFR's performance has significantly improved, and it can successfully complete all tasks.

Keywords ALFR · COVID-19 · Digital audio amplifier (TAS5630) · Microcontroller (ATMEGA32A pro.) · Actuator (DC motor and servo motor) · Infrared ray (IR) sensor · Arduino Nano

1 Introduction

Technology is essential in our everyday lives. This is because life without technology is helpless in today's dynamic world. The need for automation in the healthcare sector is becoming increasingly important as the population grows. As a result, hospitals are experimenting with new approaches to boost services and reliability without

Md. A. Kabir (✉) · Md. J. Sarker · T. Hossain · M. I. J. Jerin · Md. H. Ali
Department of Electrical and Electronic Engineering, European University of Bangladesh, Dhaka, Bangladesh

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_17

spending more money. If automation is brought to healthcare centers, then it will be far more effective for hospitals. A coronavirus is one of many diseases that spread through indirect contact (COVID-19). In December 2019, COVID-19 was discovered in a food market in Wuhan, China [1]. In dealing with pandemic situations such as COVID 19, despite wearing all of their personal protective equipment and clothing when treating patients, doctors and nurses are influenced as well. Also, there is a problem of transporting goods such as medicines, food, linens, and other equipment in isolated rooms. Isolated rooms always need to be clean and tidy, but there is a high risk that medical staff will clean this up. Therefore, automatic guided vehicle (AGV) plays an important role in hospital functions like the supply of medicines, patient monitoring, and waste management [2]. Line follow is important for AGV because the majority of AGVs use the line follow technique to reach their destination [3]. Via Arduino software, some of the line follower robots are used to move the patient from the ambulance to the store room and walk along the lines mounted on the hospital floor [4]. Many restaurants have already adopted some methods of automation, such as conveyors for moving foods from one place to another. Some hospitals also use some automation techniques for carrying medicines from pharmacy to patients [5].

The creation of a system for medical science based on artificial intelligence, in which robots can navigate to desired destinations, diagnose an individual for COVID-19 using various parameters, and conduct a survey of a locality for the same [6]. The automated systems are primarily built for handling bulk supplies, pharmacy medications, laboratory tests, central supply, and transportation of food, dirty dishes, bed laundry, waste (biological, recyclable), biomedical equipment, and other products in healthcare facilities. Automating these supplies improves operational efficiency by allowing human resources to be transferred to other departments or activities [7]. Duplex technique is a very essential part of LFR. The wireless research community hopes to develop a full-duplex (FD) service, which would allow for simultaneous transmission and reception in a single time/frequency channel, doubling the achievable spectral performance [8]. Robotics have advanced to a new level as a result of the need to reduce physical interaction during the COVID-19 epidemic [10]. In [11], the authors design a double line sensor-following robot using a PID controller and a line reading algorithm. The ATMEGA32A Pro microcontroller can also be used to derive the line-following robots [12]. A simple and cost-effective circuit design of a black line follower robot is discussed in this study [13]. The mechanical and technical challenges of the line follower robot and its applications in numerous fields were examined in this study [14]. The first creation of a food serving robot utilizing a line-following approach is shown in this article [15]. Previously, this type of robot was designed for transportation automation in hospitals [2]. With advancement in technologies, this kind of robot is used in hospitals for medicine serving and waste collection. It is also a line follower robot but used only for single bed. There is also a paper which describes a line following robot that uses an Arduino to survey, inspect, and improve the transportation of essential materials inside healthcare facilities [9]. Review of the previous works on line follower robot is summarized in Table 1.

The primary goal of this project is to develop and implement an ALFR for isolation rooms which does not require any external command to accomplish its task.

Table 1 Comparison table of related work

References	Sensor	System	Application	COVID-19
[2]	IR and ultrasonic	Remote controlled	Serve medicine and monitoring patients	Yes
[4]	IR	Automatic	Transfer the patient from the ambulance to the stone room	Yes
[5]	IR and ultrasonic	Automatic	Carry the essentials and transport it to the patient's room	Yes
[9]	IR	Remote controlled	Carries components from desired source to destination	No
[11]	Line sensors	PID controller	Serve food to the customers	No

The proposed system will move by following a black line path that has been predefined. The proposed work is cost-effective and suitable for isolation rooms with large number of COVID-19 patients. Moreover, it can be more easily operated than other existing systems.

The rest of the work is organized as follows. Section 2 introduces the design approach of ALFR, Sect. 3 shows methodology, simulations are presented in Sect. 4, results and discussion are given in Sect. 5, and the work is concluded in Sect. 6.

2 Design Approach of ALFR

This section discussed the detailed design methods and analysis employed to implement an ALFR for COVID-19 patients.

Here, the internal circuit design of the ALFR is given in the "Fig. 1." There are three IR sensors, two motor drivers, two Arduino Nanos, one RF model, and an antenna fitted very neatly.

The drawing of an automatic robot is shown "Fig. 2." The whole structure of ALFR is 22 cm by 10 cm, which is mounted on six wheels.

The ALFR is subdivided into three major sections (communication, waste collection box, and vending machine).

Communication sections:

Direct contact of a doctor with a patient infected with COVID-19 is extremely dangerous, that is why we have a communication system in this project through which doctors and patients can communicate with each other very easily without getting too close to each other. The devices we used to communicate were called RF modules. In this project, we used two RF modules which is shown in "Fig. 5,"

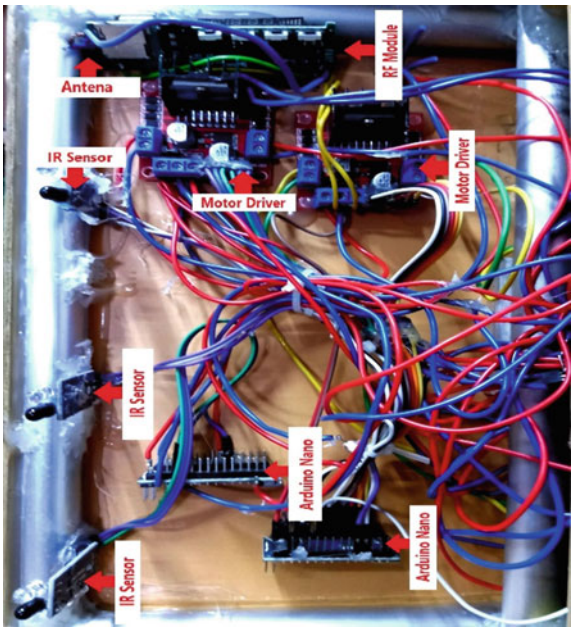


Fig. 1 System architecture of ALFR

Fig. 2 Side view of ALFR

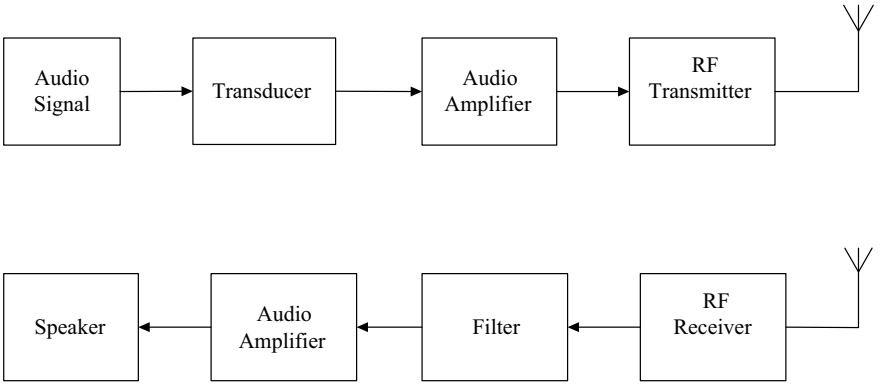


Fig. 3 Block diagram of transmitter and receiver part [8]

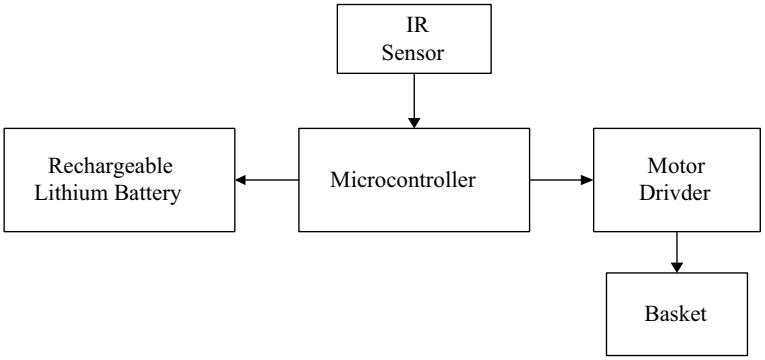


Fig. 4 Block diagram of smart dustbin [2]

one attached to the patient and the other to the ALFR. These RF modules have two parts. And our entire communication system is managed according to the duplex communication method. The duplex method has two sections which are shown in "Fig. 3."

In the transmitter section, the audio signal goes to a transducer, then it converts from the audio signal to an electrical signal. The electrical signal is very weak, that is why we use an audio amplifier. An audio amplifier coverts this weak signal and produces a strong signal whose frequency range is very high. Then, the RF transmitter transmits this signal with the help of an antenna.

In the receiver section, the RF receiver receives the sent signal from the RF transmitter with the help of an antenna. After receiving the signal, it goes to a filter to reduce the noise. Then, the signal goes to an audio amplifier to amplify the signal. After amplifying the signal, we can listen to the audio signal with the help of a speaker.

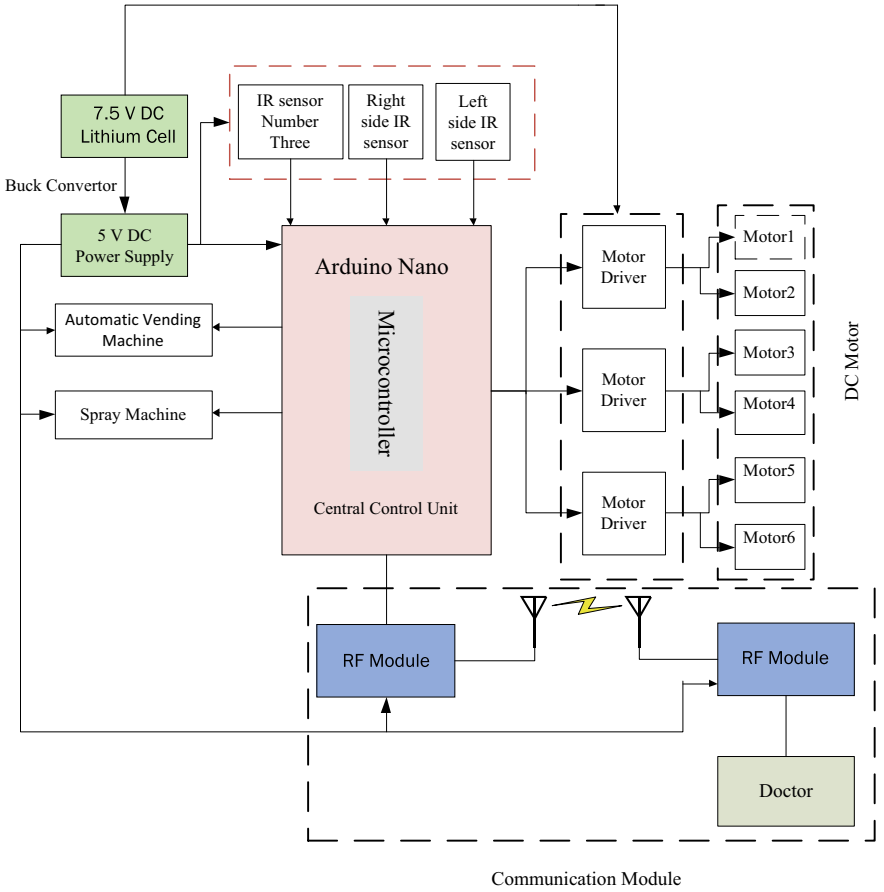


Fig. 5 Block diagram of ALFR

Waste Collection section:

Here, we have a waste collection section for cleaning the isolated room. Two things have been used in the ALFR to collect garbage. One is the smart dustbin, and the other is the basket that is attached to the robot. The smart dustbin is placed in a place very close to the black path that the robot follows, so that the robot can easily identify the dustbin. The smart dustbin consists of five items which are a microcontroller, a rechargeable lithium battery, a motor driver, a basket, and a IR sensor, which is shown in "Fig. 4." The smart dustbin is basically turned ON by the voltage received from the lithium battery. When the robot follows the black path and moves closer to the smart dustbin, the robot stands for 5 s. At the time, the IR sensor attached to the dustbin sensed the robot and instructed the microcontroller in the dustbin to dispose of the dirt. According to the microcontroller, with the help of the motor driver, it tilts

the dustbin a little and throws dirt in the basket attached to the robot. Then, after 5 s, the robot continues to follow the black path again.

Vending Machine section:

An important aspect of the robot is to provide medicine to patients. This work is mainly done with the help of a vending machine and this vending machine is installed in the robot. When the robot goes to the bed and stands for 5 s, the vending machine in the robot delivers food and medicine to the patient with the help of a plastic box as instructed by the microcontroller.

3 Methodology

All functions of ALFR are managed by the control center unit (CCU) which is shown in "Fig. 5." Here, CCU is basically made up of Arduino Nano where the microcontroller plays the main role. The CCU is mainly powered by a rechargeable lithium battery. There are three IR sensors which are connected with CCU. The main function of the IR sensors is identification of path and locating the patient's bed. When the sensors identify the desired path, they direct the microcontroller to work accordingly. Three motor drivers are linked to the CCU and turn on when given a DC 7 V supply. Six wheels are set with motor drivers. These wheels move back and forth with the help of motor drivers which are controlled by microcontroller. In addition, the automatic vending machine and the spray machine are used in the ALFR, which are also linked to the CCU. The automatic vending machine needs DC 5 V to turn on. After receiving the signal from the microcontroller, the automatic vending machine serves the food to the patient with the help of a small plastic box. Here, the spray machine turns ON after being given DC 5 V, and it is used to sanitize the ALFR. ALFR also performs the task of communication via full-duplex method. Here, the RF module is used for communication, which can be seen by looking at "Fig. 5." This RF module is basically composed of two parts, the transmitter and the receiver. Here, one RF module is attached to the ALFR, and the other is attached to the device by the doctor. These RF modules require 5 V to run, which we get from lithium batteries.

The ALFR mainly follows the line follower technique. There are three IR sensors that are connected on the front of the right side and left side of ALFR. Here, two IR sensors are mainly used to find the path for moving the robot, and the other sensor is used to identify the patient's bed. Here, the path that the robot follows is basically marked with a black line. When the robot is turned on, it follows the path by detecting the black line with the help of IR sensors. When both the sensors (left and right) do not detect the black line, the robot moves forward. If the sensor on the right side in front of the robot detects the white and the sensor on the left side detects the black line, then the right side's wheels on the robot will turn backwards and the wheels on the left side will turn in front, so that the robot will turn to the right side. Similarly, if the sensor on the right side in front of the robot detects the black line and the sensor

on the left side detects the white, then the left side's wheels on the robot will turn backwards and the wheels on the right side will turn in front, so that the robot will turn to the left side. In places where the patient's beds are placed, another small black line is used horizontally next to the black line, so that as soon as the robot approaches the bed, the IR sensor number three on the far right senses the black line and the robot stands at a short distance from the black line for 5 s. During those 5 s, the robot provides food to the patient, gives medicine, and sanitizes the room with the help of spray. When the 5 s time elapses, the sensor attached to the robot no longer finds the black line, so the robot moves forward again according to the program. When both sensors detect large black line, then Robot will stop. A large thick black line has been used to stop the robot. As a result, when the robot approaches the big black line, the number three IR sensor on the far right senses the black line and the robot stops for 5 s. However, as the black line grew larger, the IR sensor sensed it again after 5 s, causing the robot to stop moving forward. Here is a flowchart of how the ALFR follows the black path in "Fig. 6."

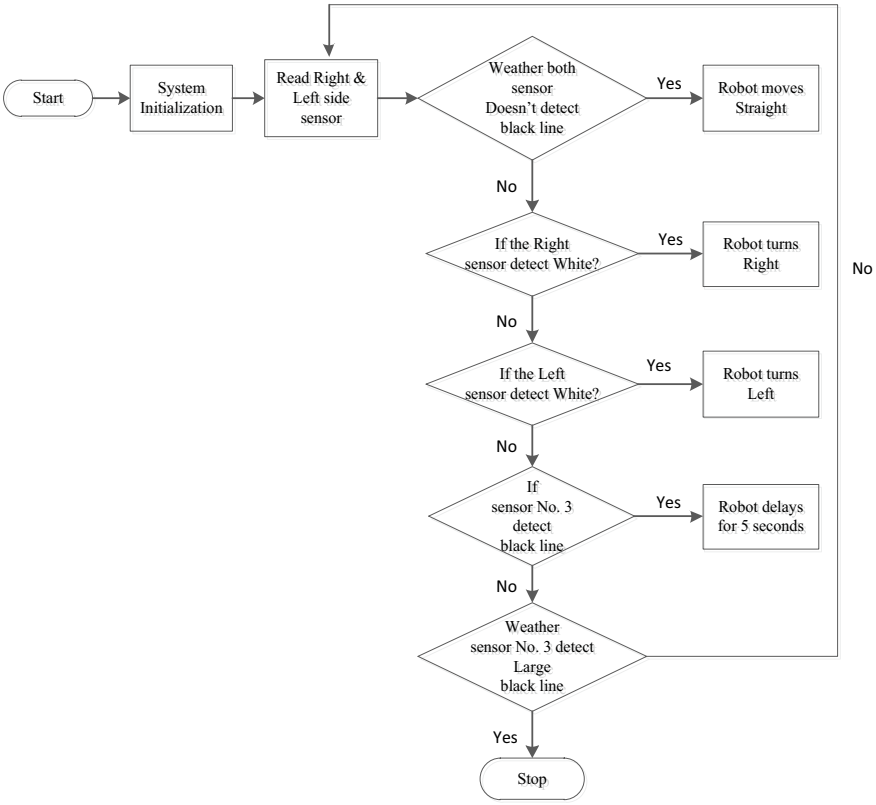


Fig. 6 Flowchart of ALFR

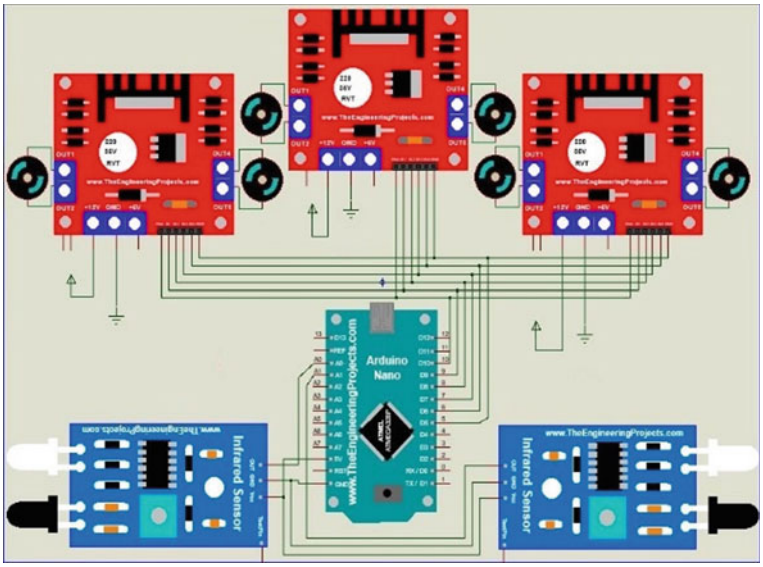


Fig. 7 System simulation circuit using proteus

4 Simulation

It is clear from the methodology's structure that a simulation part is required in order to assess this function in the implementation phase. Proteus, a well-known electronic circuit design and simulator software, which is used as the simulator. It also supports models for Arduino, IR sensors, motor drivers, and servo motor in "Fig. 7" which shows a simulation for ALFR circuit design.

5 Result and Discussions

This section shows the numerical results of all experimented data of ALFR in various distance as well as time. The numerical results are evaluated for the performance of ALFR, and it is simulated using MATLAB. The recommended model is used to work automatically. The prototype is designed to accommodate four beds with a single dustbin for isolation room. The optimal values of system parameters are used to determine the results. Tables 2 and 3 show the parameters used in the numerical computations.

In straight lines, the ALFR is shown to be faster than the other robots at passing distance and in less time, which is shown in case 1.

Table 2 Table for straight path

Sl. No	Distance covered	ALFR (Required time)	[9] (Required time)	[4] (Required time)
1	1.2 m	12 s	16 s	19 s
2	2.4 m	18 s	22 s	25 s
3	3.6 m	25 s	31 s	33 s

Table 3 Table for curved path

Sl. No	Distance covered	ALFR (Required time)	[9] (Required time)	[4] (Required time)
1	1.2 m	15 s	19 s	22 s
2	2.4 m	23 s	27 s	30 s
3	3.6 m	32 s	36 s	38 s

Case 1: Straight Path

On a curved path, the ALFR takes a little longer on a curved path than on a straight path. The ALFR moves faster than other robots on a curved path, which is shown in case 2.

Case 2: Curved Path

Observing the simulation results, it is found that better accuracy has been investigated in the new proposed system shown in Figs. 8 and 9.

We set up a black line with the help of black tape on the floor. There are four beds and a smart dustbin to keep a certain distance consecutively. The result is that the ALFR follows the path very accurately, which is shown in "Fig. 10."

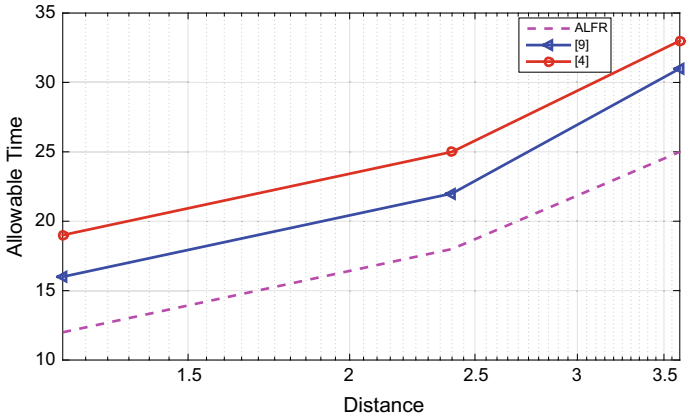


Fig. 8 Distance versus time curve for straight path

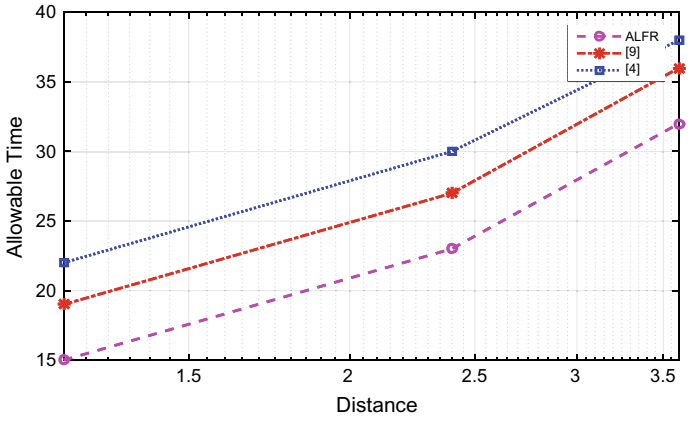


Fig. 9 Distance versus time curve for curved path

Fig. 10 Final project prototype

6 Conclusion

In the mid of this pandemic, the ALFR is capable of providing assistance to hospital workers. In this work, the ALFR has used a black line path to follow the proper direction of the isolation room, which is also proved by the simulation. In this project, the prototype is effectively capable of providing health care to COVID-19 patients such as serving foods and medicines which will enable the medical staff to deal with thousands of patients at the same time in an isolation room. The work is also expanded to include collecting waste from specially built baskets stored in the isolation room and properly disposing of it in a specified area. It also performs a communication link

between doctors and patients within 200 m and finally disinfects the whole isolation room. This is a tremendous help to frontline medical staff and is indispensable to controlling the global epidemic.

Looking at the other line follower robot, it can be seen that they only follow the line but do not delay anywhere. But when our robot moves, it stops for 5 s as soon as it gets close to the bed and smart dustbin and after 5 s, the robot starts moving again. Also, other robots can do only one job at a time, like giving food or giving medicine or collecting waste, but our robot can provide food, medicine, and collect waste at the same time. The robot can also spray the isolated room. And for all these reasons, this robot is different from other robots.

In addition, a number of sensors can be added to further increase the system performance of ALFR, which can measure the temperature of a patient's body and also oxygen level. Further work can be carried out to add a camera, so that doctors can monitor the COVID-19 patients from anywhere.

References

1. N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, W. Tan, A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. **382**(8), 727–733 (2020). https://doi.org/10.1056/nejmoa2001017
2. M. Antony, M. Parameswaran, N. Mathew, S. V.S, J. Joseph, C.M. Jacob, Design and Implementation of Automatic Guided Vehicle for Hospital Application. in 5th International Conference on Communication and Electronics Systems (ICCES) (2020). https://doi.org/10.1109/icces48766.2020.91378
3. I. Haider, K.B. Khan, M.A. Haider, A. Saeed, K. Nisar, Automated robotic system for assistance of isolated patients of coronavirus (COVID-19). in 2020 IEEE 23rd International Multitopic Conference (INMIC) (2020). https://doi.org/10.1109/inmic50486.2020.93181
4. H.A. Hadi, Line follower robot arduino (using robot to control Patient bed who was infected with Covid-19 Virus). in 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (2020). https://doi.org/10.1109/ismsit50672.2020.9254
5. M. Prabhakar, V. Paulraj, J.A. Dhanraj, S. Nagarajan, D.A.K. Kannappan, A. Hariharan, Design and simulation of an automated guided vehicle through webots for isolated COVID-19 patients in hospitals. in IEEE 4th Conference on Information & Communication Technology (CICT) (2020). https://doi.org/10.1109/cict51604.2020.931206
6. S. Karmore, R. Bodhe, F. Al-Turjman, R.L. Kumar, S. Pillai, IoT based humanoid software for identification and diagnosis of Covid-19 suspects. IEEE Sens. J. 1–1 (2020). https://doi.org/10.1109/jsen.2020.3030905
7. M. Pedan, M. Gregor, D. Plinta, Implementation of automated guided vehicle system in healthcare facility. Procedia Eng. **192**, 665–670 (2017). https://doi.org/10.1016/j.proeng.2017.06.115
8. Z. Zhang, K. Long, A.V. Vasilakos, L. Hanzo, Full-duplex wireless communications: challenges, solutions, and future research directions. Proceedings of the IEEE **104**(7), 1369–1409 (2016). https://doi.org/10.1109/jproc.2015.2497203
9. J. Chaudhari, A. Desai, S. Gavarskar, Line following robot using arduino for hospitals. in 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT) (2019). https://doi.org/10.1109/icct46177.2019.8969022
10. X.V. Wang, L. Wang, A literature survey of the robotic technologies during the COVID-19 pandemic. J. Manufact. Syst. **60**, 823–836 (2021). https://doi.org/10.1016/j.jmsy.2021.02.005

11. V.N. Thanh, D.P. Vinh, N.T. Nghi, L.H. Nam, D.L.H. Toan, Restaurant serving robot with double line sensors following approach. in IEEE International Conference on Mechatronics and Automation (ICMA) (2019). https://doi.org/10.1109/icma.2019.8816404
12. Y. Tian, G. Du, Infrared line following and ultrasonic navigating robot with ATMEGA328 Pro. in IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (2019). https://doi.org/10.1109/imcec46724.2019.89841
13. S. Tayal, H.P.G. Rao, S. Bhardwaj, H. Aggarwal, Line follower robot: design and hardware application. in 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) (2020). https://doi.org/10.1109/icrito48877.2020.9197
14. A.S. Vamsi, Design to ımplementation of a line follower robot using 5 sensors, Vol. 3 (2019), pp. 42–47
15. T. Vo, Development of restaurant serving robot using line following approach. J. Sci Technol. Issue Inf. Commun. Technol. **17**(1), (2019). https://doi.org/10.31130/JST-UD2019-087

Performance and Error Estimation Analysis of QAM with MRC Receiver for L-TAS/SC Over α–μ Fading Channels

K. S. Balamurugan, T. J. V. Subnahmanyeswara Rao, and G. Srihari

Abstract The diversity combining technique reduces the effect of fading. However, in practice, without error in a channel and imperfect channel estimation are unattainable, therefore, channel analysis is required when considering the imperfect channel estimation (ICE). In present work, the performance and error estimation analysis of Quadrature amplitude modulation with ICE is examining over α–μ fading channels for different combining technique like selection combining (SC), Transmit Antenna Selection (TAS/SC), Maximal Ratio Combining (MRC), and TAS/MRC. It has been detected that the performance of ABER improves for both 8-QAM and 16-QAM systems as compared to 32-QAM with the increase in fading parameter μ for both TAS/MRC and TAS/SC systems. The values of both transmit antenna and the user increases for TAS/SC systems and the presence of ABER can also be improved. The numerical values are obtained by using Monte-Carlo simulation (MC) method.

Keywords Selection combining · Transmit antenna selection · α–μ fading channels

1 Introduction

Channel State Information (CSI) is mainly determine the diversity combining technique occurs in the communications receiver. Selection combining (SC) is very simple compounding system where the transmitted and received antenna so that the total received selected signal power is maximized. Maximal Ratio Combining (MRC) is the optimal combining scheme of co-phased and multiplied to a weighted factor of received signals, and it is proportional to the signal amplitude added algebraically. Always the communication channel error is a common issue that is related with a real-world arrangement. So far, there is a matchless outcome, for that the receiver must have an efficient estimation of the received signal phase and envelope information [1].

K. S. Balamurugan (✉) · T. J. V. Subnahmanyeswara Rao · G. Srihari
Department of Electronics and Communication Engineering, Sasi Institute of Technology and Engineering, Tadepalligudem, Andhra Pradesh 534101, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_18

For the practical aspect, the selection of perfect channel estimation is a challenging one, which is important to show better performance of coherent receivers because of bringing together phase and envelope of the transmitted & received signals [2]. The short term α–μ fading channel consists of Rayleigh fading, One-sided Gaussian, Nakagami-*m*, Weibull, by replacing certain values for α and μ. The mathematical expression for the α–μ fading in communication channel with considering the moment generating function (MGF) with coherent modulation of bit error rate (BER) is specified in [3]. In terms of two-way relay networks (TWRNs) and importance of channel estimation, error was studied by taking account of separate and cascaded information schemes [4, 5]. The bit error level of gray-coded MPSK with random phase error is presented in [5]. From ref. [5], it found that fading channel bit error level converge the Additive-White-Gaussian noise (AWGN) channels bit error level. The bit error probability (BEP) of the rectangular QAM(R-QAM) or M-ary quadrature amplitude modulation (M-QAM) mathematical expressions with and without given spatial diversity, concatenated double μ, μ-μ or μ-fading are calculated. Process of Cho and Yoon in additive white Gaussian noise (AWGN) are used for BEP of digital QAM modulation expressed in [6]. The important and challenging issue are in the independent issues in the non-identically distributed (i.n.i.d.) double α–μ random variables (RVs) are studied in [7].

The Meijer's *G*-function, energy detection (ED), was derived for wireless networks like wireless sensor network (WSN) and cognitive radio network (CRN) in α–μ fading channels [8]. The performance of a κ–μ extreme fading conditions in the communication system were studied and the fresh systematic expressions for the normalized, average error rate, asymptotic fault analysis and normal channel capacity are examined in [9].

The arrangement of $\alpha - \mu$ fading channels and co-channel interferences in the error of AF relaying dual-hop network is investigated and analyze other fading representations like as Rayleigh model and Nakagami-*m* with different value of α and μ are studied in [10]. The Hexagonal Quadrature Amplitude Modulation (HQAM) along with SC combining receiver is investigated in detail. The Chiani, Prony, Chernoff, Trapezoidal approximations were used for assessing the error probability (EP) for Gaussian *Q*-function with η–μ fading channels in the ref. [11]. In the prior work, *K*-*G* Fading channels in the L-TAS/SC handset of the Performance and Estimation Error analysis is investigated in [12]. For the α–μ fading channel consideration of ICE, ABER and QAM modulation of the transmitting antenna is examined. The complete model of the system was discussed in Sect. 2, and the evaluation of the system using ABER details are given in section topic 3. The obtained numerical results and its detailed analysis are described in Sect. 5.

2 System Model

Arrange the structure containing of *L* and *R* no of transmitting and receiving antennas, respectively. Among all, the finest transmitting feeler is one in which can make full

advantage of the post-processing SNR of the selected MRC receiver. MRC receiver signals are co-phased and multiplied by their weighted factor. The instant MRC receiver SNR output is indicated by γ can be written as $\gamma = \sum_{m=1}^{R} \gamma_m$, where γ_m is the current SNR of the mth division of each MRC receiver. The instant PDF of SNR for the α–μ fading channel is specified by [3] as

$$f(\gamma) = \frac{\alpha \mu^{\mu} \gamma^{\frac{\alpha\mu}{2}-1}}{2\Gamma(\mu)\overline{\gamma}^{\frac{\alpha\mu}{2}}} e^{-\mu\left(\frac{\gamma}{\overline{\gamma}}\right)^{\frac{\alpha}{2}}} \tag{1}$$

The received SNR CDF is given as [6],

$$F(\gamma) = \frac{1}{\Gamma(\mu)} a\left(\mu, \mu\left(\frac{\gamma}{\overline{\gamma}}\right)^{\frac{\alpha}{2}}\right) \tag{2}$$

where lower incomplete gamma function $a(g, h) = \int_0^h x^{g-1} e^{-x} \mathrm{d}x$. An estimation of communication channel frequency vector $\alpha_l(t)$ is measured at the handset which is specified as $\alpha_l(t) = [\alpha_l(t), \ldots, \alpha_L(t)]^T$ wherever $\alpha_l(t)$ is first subdivision element of the vector and $[.]^T$ is the transpose of the vector. Each level of communication channel assessment can be given as $e_l(t) = \widehat{\alpha_l}(t) - \alpha_l(t)$ where $\alpha_l(t)$—channel coefficient-vector for L branches.

It can be written as $\alpha_l(t) = [\alpha_l(t), \ldots, \alpha_L(t)]^T$. The communication channel estimate error functional model and its analyzes of BER performance was given [2] as $\alpha_{i,l}(t) = \rho_l \widehat{\alpha}_{i,l}(t) + w_{i,l}(t)$ wherever, $(l = 1, 2, 3, \ldots, L)$ and the diffusion element indicated as 'i'.$\left\{w_{i,l}(t)\right\}_{l=1}^{L}$ is the second zero mean equivalent estimation error and its variance (σ_w^2). The complex number ρ_l is given as $\rho_l = |\rho_l| e^{j\Delta\theta_l}$. $\Delta\theta_l$ is the phase offset of ρ_l[2]. If we are considering ICE, the value either $|\rho_l|<1$ or $\Delta\theta_l \neq 0$. For MRC, the finding of diffused representation is $x(t)$. The composite decision variable (DV) can be represented [7] as $\widetilde{G} = \sum_{l=1}^{L} \widehat{a}_l(t) y_l(t)$ with attention of ICE. The new DV can be modified by using the half-plane decision method [8] by turning the complex decision variable $\widetilde{G}$ with angle of plane $\beta = \pm\left(\frac{\pi}{2} + \frac{\pi}{A}\right)$, A is the pattern dimensions of the modulation system used. The novel look achieved is

$$\widetilde{G}(\beta) = \Re\left(\widetilde{G} e^{-j\beta}\right) = \sum_{l=1}^{L} G_l(\beta) \tag{3}$$

$G_l(\beta)$ is the conclusion portion at individual division. As half-plane decision method [2],

$$\gamma_{\text{ICE}}^{\text{MRC}} = B(\beta) \sum_{l=1}^{L} \widehat{\gamma}_l$$

where

$$B(\beta) = \frac{|\rho_l|^2 \cos^2(\Delta\theta_l - \beta)}{\left[\left(1 - |\rho_l|^2\right)\overline{\gamma}_l + 1\right]}, \widehat{\gamma}_l = \frac{E_b}{N_0}\widehat{\alpha}_l^2$$

Regenerated value of instant SNR PDF using ICE in (1) as

$$f(\gamma) = \frac{\alpha\mu^{\mu}\gamma^{\frac{\alpha\mu}{2}-1}B^{\frac{\alpha\mu}{2}-1}}{2\Gamma(\mu)\overline{\gamma}^{\frac{\alpha\mu}{2}}} e^{-\mu\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha}{2}}} B^{-1} \tag{4}$$

The expression of CDF can be specified as

$$F(\gamma) = \int_0^{\gamma} \frac{\alpha\mu^{\mu}\gamma^{\frac{\alpha\mu}{2}-1}B^{\frac{\alpha\mu}{2}-1}}{2\Gamma(\mu)\overline{\gamma}^{\frac{\alpha\mu}{2}}} e^{-\mu\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha}{2}}} B^{-1} d\gamma \tag{5}$$

The above integral can be solved by using [9, 3.381]

$$F(\gamma) = \frac{1}{\Gamma(\mu)} g\left(\mu, \mu\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha}{2}}\right) \tag{6}$$

where, lower incomplete gamma function is $a(g,h) = \int_0^h x^{g-1}e^{-x}\mathrm{d}x$ by taking of $\Gamma(\mu)$ as gamma function. The highest transmitting antenna for TAS/MRC system can be selected by corresponding highest received SNR, and kind of MRC method is implemented at the receiver handset. The PDF of characteristic SNR in such a arrangement assuming all $\left(\left|x_{m,k}\right|\right)$'s are to be iid and can be given by [3] as

$$f_{\gamma_{(L)}}(\gamma) = LF_{\gamma}(\gamma)^{L-1} f_{\gamma}(\gamma) \tag{7}$$

Equation 7 holds for both TAS/SC and TAS/MRC arrangements. For TAS/MRC systems, L is transmitting antenna, whereas TAS/SC, $L = \text{TR}$. Where T is the transmitting antenna and R is receiving antenna. Putting the values of Eqs. 4 and 6 in Eq. 7, the SNR of PDF represented as

$$f_{\gamma_{(L)}}(\gamma) = L\left[\frac{g\left(\mu, \mu\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha}{2}}\right)}{\Gamma(\mu)}\right]^{L-1} \times \left(\frac{\alpha\mu^{\mu}\gamma^{\frac{\alpha\mu}{2}-1}B^{-\frac{\alpha\mu}{2}}}{2\Gamma(\mu)\overline{\gamma}^{\frac{\alpha\mu}{2}}}e^{-\mu\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha}{2}}}\right) \tag{8}$$

On making mathematical rearrangements, we obtain

$$f_{\gamma_{(L)}}(\gamma) = \frac{L\alpha\mu^{\mu}}{2\Gamma(\mu)^{L}\overline{\gamma}^{\frac{\alpha\mu}{2}}}\gamma^{\frac{\alpha\mu}{2}-1}B^{-\frac{\alpha\mu}{2}}e^{-\mu\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha}{2}}} \times \left\{g\left(\mu, \mu\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)\right)\right\}^{L-1} \tag{9}$$

Writing the incomplete gamma function in series using [10] as

$$f_{\gamma_{(L)}}(\gamma) = \frac{L\alpha\mu^{\mu}}{2\Gamma(\mu)^{L}\overline{\gamma}^{\frac{\alpha\mu}{2}}}\gamma^{\frac{\alpha\mu}{2}-1}B^{-\frac{\alpha\mu}{2}}e^{-\mu\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha}{2}}} \times \left[\Gamma(\mu)\left(\mu\left\{\frac{\gamma B^{-1}}{\overline{\gamma}}\right\}^{\frac{\alpha}{2}}\right)^{\mu}e^{-\mu\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha}{2}}}\right]^{L-1} \tag{10}$$

On making mathematical rearrangement, the above equation is obtained sequentially

$$f_{\gamma_{(L)}}(\gamma) = \frac{L\alpha\mu^{\mu L}}{2\Gamma(\mu)\overline{\gamma}^{\frac{\alpha\mu L}{2}}}\gamma^{\frac{\alpha\mu L}{2}-1}B^{-\frac{\alpha\mu L}{2}}e^{-\mu L\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha L}{2}}} \tag{11}$$

$$f_{\gamma_{(L)}}(\gamma) = \frac{L\alpha\mu^{\mu L}}{2\Gamma(\mu)\overline{\gamma}^{\frac{\alpha\mu L}{2}}}\gamma^{\frac{\alpha\mu L}{2}-1}B^{-\frac{\alpha\mu L}{2}}e^{-\mu L\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha L}{2}}} \times \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\cdots\sum_{h_{L-1}=0}^{\infty}\frac{\mu^{\sum_{i=1}^{L-1}h_i}\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha}{2}\sum_{i=1}^{L-1}h_i}}{\left\{\prod_{i=1}^{L-1}\Gamma(\mu+h_i+1)\right\}} \tag{12}$$

$$f_{\gamma_{(L)}}(\gamma) = \frac{L\alpha\mu^{\mu L}}{2\Gamma(\mu)\overline{\gamma}^{\frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1}h_i}}e^{-\mu L\left(\frac{\gamma B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha L}{2}}} \times \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\cdots\sum_{h_{L-1}=0}^{\infty}\frac{\mu^{\sum_{i=1}^{L-1}h_i}}{\left\{\prod_{i=1}^{L-1}\Gamma(\mu+h_i+1)\right\}} \tag{13}$$

3 Consideration of ICE-Average Bit Error Rate

Bit Error Rate (BER) of a particular system depends on both fading distribution and type of modulation techniques that have been used. By considering the characteristic of the uncertain bit error rate, the M-array modulations in digital communication system and the SNR PDF receiver output of ICE-ABER can be accomplished. Expression of the Hoyt fading imperfect channel ABER for M-QAM modulation with L-MRC is discussed in detail [13]. It can be utilized in this work.

$$P_e = \int_0^\infty P\left(e\middle|\widehat{\gamma}_{\mathrm{MRC}}\right) f_{\widehat{\gamma}_{\mathrm{MRC}}}\left(\widehat{\gamma}_{\mathrm{MRC}}\right) \mathrm{d}\widehat{\gamma}_{\mathrm{MRC}} \tag{14}$$

anywhere, conditional BER modulation scheme-$P\left(e\middle|\widehat{\gamma}_{\mathrm{MRC}}\right)$. Conditional BER expression for coherent modulation can be written as

$$P\left(e\middle|\widehat{\gamma}_{\mathrm{MRC}}\right) = aQ\left(\sqrt{b\widehat{\gamma}_{\mathrm{MRC}}}\right) \tag{15}$$

wherever, a and b are the variable parameters used for different modulation schemes.

The incomplete gamma function can be stated as $Q(.)$, and it can be written as by using Eq. (15),

$$P\left(e\middle|\widehat{\gamma}_{\mathrm{MRC}}\right) = \frac{a}{2\sqrt{\pi}}\Gamma\left(\frac{1}{2}, \frac{b\widehat{\gamma}_{\mathrm{MRC}}}{2}\right) \tag{16}$$

Putting the values of Eqs. 13 and 16 in Eq. 14, the ABER expression can be derived in terms of γ,

$$\begin{aligned} P_e = & \int_0^\infty \frac{a}{2\sqrt{\pi}}\Gamma\left(\frac{1}{2}, \frac{b\widehat{\gamma}}{2}\right) \frac{L\alpha\mu^{\mu L}}{2\Gamma(\mu)\overline{\gamma}^{\frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1}h_i}} e^{-\mu L\left(\frac{\widehat{\gamma}_{B^{-1}}}{\overline{\gamma}}\right)^{\frac{\alpha L}{2}}} \\ & \times \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\cdots\sum_{h_{L-1}=0}^{\infty} \frac{\mu^{\sum_{i=1}^{L-1}h_i}}{\left\{\prod_{i=1}^{L-1}\Gamma(\mu+h_i+1)\right\}} \\ & \times \widehat{\gamma}^{\frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1}h_i-1} B^{-\frac{\alpha\mu L}{2}-\frac{\alpha}{2}\sum_{i=1}^{L-1}h_i} \mathrm{d}\widehat{\gamma} \end{aligned} \tag{17}$$

$$P_e = \frac{a}{2\sqrt{\pi}} B^{-\frac{\alpha\mu L}{2}-\frac{\alpha}{2}\sum_{i=1}^{L-1} h_i} \frac{L\alpha\mu^{\mu L}}{2\Gamma(\mu)\overline{\gamma}^{\frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1} h_i}} \times \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\cdots\sum_{h_{L-1}=0}^{\infty} \frac{\mu^{\sum_{i=1}^{L-1} h_i}}{\left\{\prod_{i=1}^{L-1}\Gamma(\mu+h_i+1)\right\}} \times \int_0^{\infty} \widehat{\gamma}^{\frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1} h_i-1}\Gamma\left(\frac{1}{2},\frac{b\widehat{\gamma}}{2}\right) e^{-\mu L\left(\frac{\widehat{\gamma}_{B^{-1}}}{\overline{\gamma}}\right)^{\frac{\alpha L}{2}}} \mathrm{d}\widehat{\gamma} \tag{18}$$

Solving the integral using [9, (6.455.1)],

$$P_e = \frac{a}{2\sqrt{\pi}} B^{-\frac{\alpha\mu L}{2}-\frac{\alpha}{2}\sum_{i=1}^{L-1} h_i} \frac{L\alpha\mu^{\mu L}}{2\Gamma(\mu)\overline{\gamma}^{\frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1} h_i}} \times \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\cdots\sum_{h_{L-1}=0}^{\infty} \frac{\mu^{\sum_{i=1}^{L-1} h_i}}{\left\{\prod_{i=1}^{L-1}\Gamma(\mu+h_i+1)\right\}} \times \frac{\left(\frac{b}{2}\right)^{0.5}\Gamma\left(\frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1} h_i+0.5\right)}{\left(\frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1} h_i\right)\left(\frac{b}{2}+\left(\frac{\mu L B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1} h_i+0.5}\right)} \times {}_2F_1\left(\begin{matrix} 1;\ \frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1} h_i+0.5;\ \frac{\alpha\mu L}{2}+\frac{\alpha}{2}\sum_{i=1}^{L-1} h_i+1; \\ \frac{\mu L\left(\frac{B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha L}{2}}}{\frac{b}{2}+\mu L\left(\frac{B^{-1}}{\overline{\gamma}}\right)^{\frac{\alpha L}{2}}} \end{matrix}\right) \tag{19}$$

This is considered as the expression for the ABER of TAS-MIMO system under $\alpha-\mu$ fading channel for M-QAM modulation with ICE.

4 Results and Discussion

The numerical evaluation results are discussed in this section by chosen of two fading parameters such as α and μ. Figure 1 shows the BER performance of 8-QAM, 16-QAM and 32-QAM concerning different μ values, while keeping α, $\Delta\theta$, and ρ are constant. The values have been calculated for $L = 3$.

The numerical results have been evaluated and analyzed for ABER with error estimation. In Fig. 1, the QAM system with ICE for a TAS/MRC system graph was plotted for various fading parameters (μ). In Fig. 2, the values have been plotted for a different number of transmit antenna of QAM system keeping all the other parameters constant for TAS/MRC system. In Fig. 3, varying the values of fading parameter μ, the results have been plotted for a QAM system with ICE for TAS/SC system considering transmit antenna $T = 2$ and the number of users $K = 2$ keeping constant the number of receiver antenna. Figure 4 shows the comparison of ABER of QAM system with ICE for TAS/SC and TAS/MRC system with $T = 1$ and $K = 2$ and $\mu = 1$ and 2. It has been observed that the performance of the system can be enhanced for TAS/MRC as compared to TAS/SC system.

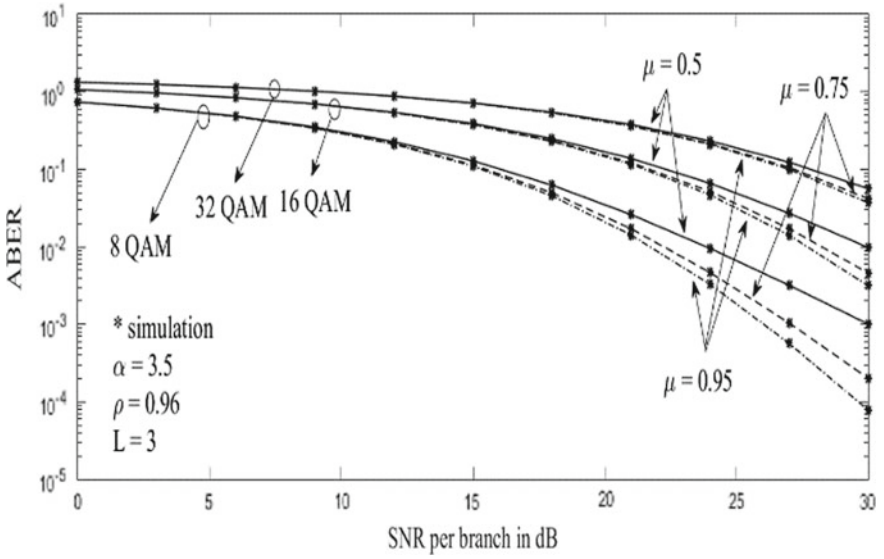


Fig. 1 QAM system with ICE & ABER performance for TAS/MRC system

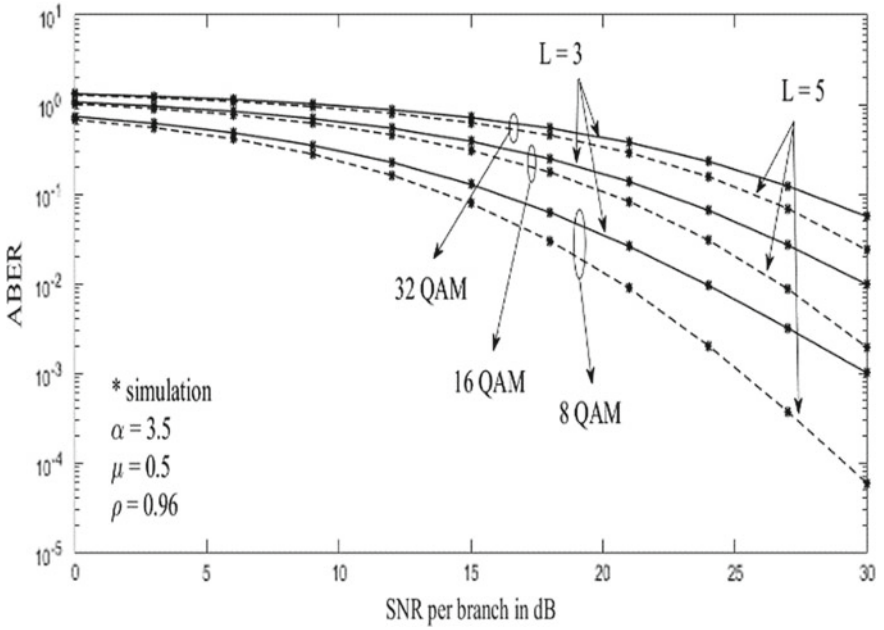


Fig. 2 QAM system with ICE & ABER performance for TAS/MRC system

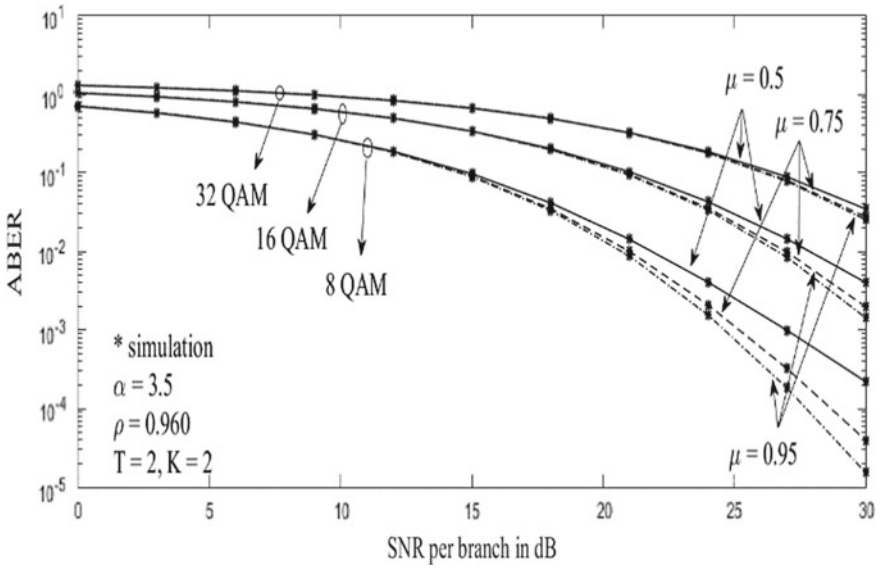


Fig. 3 QAM system with ICE & ABER performance for TAS/SC system

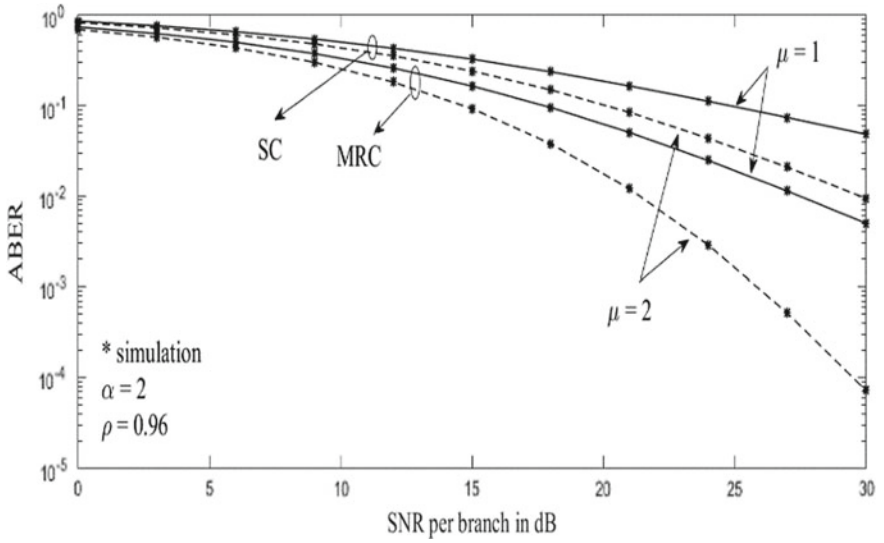


Fig. 4 TAS/SC and TAS/MRC system ABER comparison with ICE for 8-QAM system

5 Conclusion

The ICE-ABER performance of two systems TAS/SC, TAS/MRC over $\alpha - \mu$ fading channels were discussed. The expressions have been derived for the closed confluent hypergeometric functions. It has been observed that the performance of ABER improves for both 8-QAM and 16-QAM systems as compared to 32-QAM with the increase in fading parameter μ for both the TAS/SC and TAS/MRC arrangements. The values of both transmit antenna and the user increases for TAS/SC structures and the results of ABER can be improved. MC simulation was used for verification of the model through the numerical results generated.

References

1. M.K. Simon, M.S. Alouini, A unified approach to the performance analysis of digital communication over generalized fading channel. Proc IEEE **86**, 256–264 (1998)
2. Y. Ma, R. Schober, S. Pasupathy, Effect of channel estimation error on MRC diversity in Rician fading channels. IEEE Trans. On Vehicular Tech. **54**, 2137–2214 (2005)
3. M. Magableh, M.M. Matalgah, Moment generating function of the generalized α–μ distribution with applications. IEEE Commun. Lett. **13**, 411–413 (2009)
4. J.-W. Pu, T.-Y. Wang, S.-H. Li, C.-P. Li, H.-J. Li, Performance analysis of relay selection in two way relay networks with channel estimation errors, IEEE Trans. Broadcast. **61**, 66–71 (2015)
5. Y. Jhang, J. Jeong, D. Yoon, Bit error floor of MPSK in the presence of phase error. IEEE Trans. Veh. Technol. **60**, 1967–1973 (2011)

6. J.L.Q. Wamberto, H.S. Silvab, D.B.T. Almeidaa, A.S.R. Oliveirab, F. Madeiroc ,On the BEP analysis of M-QAM and R-QAM under cascaded double η–μ, κ–μ or α–μ fading channels, *Digital Signal Processing* vol. 297, (Elsevier, 2020), pp. 1–13
7. O.S. Badarneh, D.B. da Costa, M. Benjillali, M.-S. Alouini, Selection combining over double α–μ fading channels. IEEE Trans. Veh. Technol. **1**, 1–5 (2020)
8. K. Cao, P. Qian, J. An, L. Wang, Accurate and practical energy detection over α–μ fading channels. Sensors ,MDPI **20**, 1–8 (2020)
9. J.M. Moualeu, D.B. da Costa, R.A.A. de Souza, W. Hamouda, U.S. Dias, Performance analysis of wireless communication systems subject to κ–μ extreme fading, UTC from IEEE Xplore (2020)
10. A.M.J. Jwaifel, I. Ghareeb, S. Shaltaf, Impact of co-channel interference on performance of dual-hop wireless ad hoc networks over $\alpha-\mu$ fading channels, vol. 1. Int. J. Commun. Syst., (Wiley, 2020), pp. 1–14
11. M. Bilim, *Dual-Branch SC Wireless Systems with HQAM for beyond 5G over H–μ Fading Channels* (Peer-to-Peer Networking and Applications, Springer, 2020), pp. 1–14
12. K.S. Balamurugan, Performance and estimation error analysis for L-TAS/SC receiver over KG fading channels. Int. J. Adv. Sci. Technol. **29**, 5061–5067 (2020)
13. D. Das, R. Subadar, Performance analysis of QAM for L-MRC receiver with estimation error over independent Hoyt fading channels. Int. J. Electron. Commun. **107**, 15–20 (2019)
14. R.M. HubhaSaikia, Performance analysis of TAS/SC as well as TAS/MRC MIMO techniques under α–μ fading channels. IJRTE **8**, 1–6 (2019)
15. W. Jeong, J. Lee, D. Yoon, New BER Expression of MPSK. IEEE Trans. Veh. Technol. **60**, 1916–1924 (2011)
16. C. Toufik, E.B Faissal, B. Hussain, Performance Analysis of TAS/MRC based MIMO systems over Weibull Fading Channels, Proceedings of the International Conference and Advanced Communication systems and Information Security (ACOSIS) (2016), pp. 1–6

RETRACTED CHAPTER: Optimized Lower Part Constant-OR Adder for Multimedia Applications

Mahendra Vucha and A. L. Siridhara

Abstract Power consumption and speed of computing systems depend on their arithmetic modules such as adder, subtractor, and multiplier. So, the need for high speed, error tolerance, and power efficiency nature of few applications has been improved by developing approximate adders. Increasing the effectiveness of integrated circuits by making the trade-off between accuracy and cost has got significant importance. A systematic methodology for optimizing the architecture of approximate adders has been proposed and called optimized lower part constant-OR adder (LOCA). In this article, the approximate adders are designed by redesigning its logic circuit, implemented on reconfigurable architectures, and then compared with traditional adder architectures. The proposed architecture outperforms its contemporary architectures in terms of hardware and accuracy.

Keywords Approximation · Stochastic computing · Error metrics · Hardware trade-off

1 Introduction

VLSI systems rely on three major parameters, namely power consumption, delay, and space occupied (area). All these parameters must be optimized and kept controlled while designing the system. Computing architectures may face problems if these parameters are not maintained properly. In general, optimizing all these three parameters is not applicable for every system architecture, but the designers could balance them based on application requirements. For example, the design of ATMs strictly adheres to response to the inputs and transaction speed, where optimization required to be delay compared to power and area. For efficient computations, designers need

The original version of this chapter was retracted: The retraction note to this chapter is available at https://doi.org/10.1007/978-981-16-6605-6_67

M. Vucha (✉) · A. L. Siridhara
Department of Electrnonics and Communication Engineering, MLR Institute of Technology, Hyderabad, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_19

to minimize the complexity while optimizing the size of a system. Adder is one of the main components of arithmetic circuits, and analysts in the domain of approximate computing have paid attention to adders. There are two methods to approach adders called stochastic computing and approximate computing. Stochastic computing uses binary bitstream where the value of bitstream referred to as a stochastic number (SN) is encoded as 0s and 1s. The major disadvantage of stochastic computing is that it assumes bitstreams are independent, but this assumption does not hold if it fails. Approximate computing is a low-power means for digital signal processing applications and brings a trade-off between performance and accuracy. Approximate techniques may have some errors where no individual errors are recognized but only average errors can systematically predict the impact of error in the output. In the past decade, approximate computing is chosen for adders both at the software level and hardware level. Adders have significant importance in digital operations and signal processing. The approximation adders are the segmented adders, where the *m*-bit adder can partition into *k*-bit sub-adders. Carry select adders—where the multiple sub adders used, carry look-ahead adder, equal segmentation adder, Exact adder approximate full adders—where full adder is approximated. The design of all these approximate adders is to limit the carry generated by the adders. The length of carrying propagation in an *N*-bit conventional adder is similar to Log 2 *N*.

The (lower part OR adder (LOA) presented in literature [1] is shown in Fig. 1, and it is called OR adder, because it is observed that the adders are divided into sub-adders which are said to be *m*-bit one-half adders and the remaining part has OR gates [2–5]. So, the half adder is named as a higher sub-adder which consists of an (n_h–1)-bit-exact adder, and the 'n_l' of OR gates is represented as lower part sub-adder which consists of (0–n_l–1) bits [6–9]. A carry signal for an accurate adder is generated using an extra AND gate [10–14]. Since the approximation is restricted to the least significant bits, the magnitude of errors is limited [15–17]. This is the major advantage of LOA when compared with other architectures such as equal segmentation adder (ESA) [18, 19].

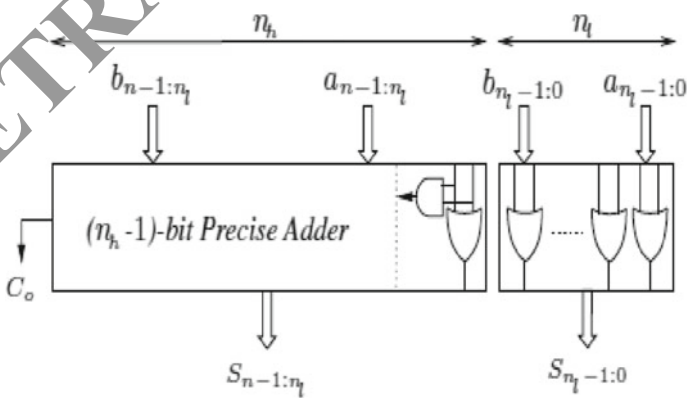


Fig. 1 Existing structure of LOA

The presented LOA is the slowest, but it is highly efficient at its computations. In this article, a method is presented to improve the LOA systematically by considering architectural templates from [1] and then implement all possible combinations to study its efficiency and propose an optimized lower part constant-OR adder (OLOCA) as reduced hardware architecture**.**

2 Optimized Lower Part Constant-OR Adder Architecture

The proposed optimal architecture can be obtained through the incorporation of the following sequential steps.

Step 1: Analyze the error metrics to value hardware quantifying the standard architecture.

Step 2: Consider LOA as a hardware template where the number of OR gates depends on the number of inputs.

Step 3: Implement mean square error (MSE) which is very minimum for OLOCA compared with any other error metric.

2.1 Error Metrics

Since an approximation technique has been adopted for this architecture, this approximation may generate errors in the output of a system which is not desired. In order to reduce the error, error metrics are preferred, and they play a major role in evaluation of different architecture in different fields. The quality of the approximate adders can be evaluated using these error metrics and shows the balance between error and cost of hardware. Error magnitude can be quantified with several metrics. Some of the error metrics are average error (μ), standard deviation (σ), mean square error (MSE), mean absolute error (MAE), root mean square (RMS), mean absolute percentage error (MAPE), and symmetric mean absolute percentage error (SMAPE). Error is observed as difference between approximate outcome and actual outcome, that is $\varepsilon = \tilde{S} - S$, where $\tilde{s}$ = Approximate outcome and S = Actual outcome. Error metrics can be calculated using the formulas below.

$$\text{Average Error}(\mu) = E[\varepsilon]$$

$$\text{Standard Deviation}(\sigma) = \sqrt{E\left[(\varepsilon - \mu)^2\right]}$$

$$\text{Mean Square Error(MSE)} = E\left[\varepsilon^2\right] = \mu^2 + \sigma^2$$

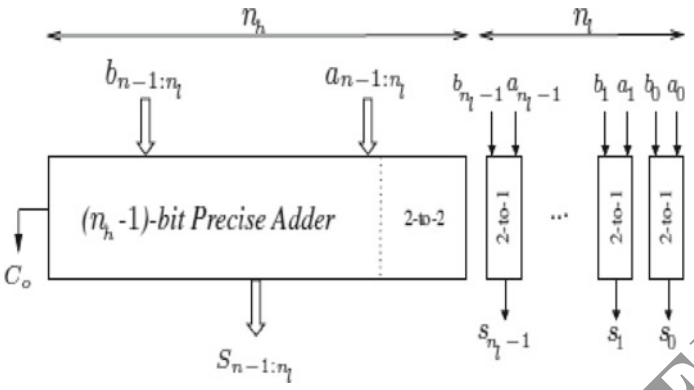


Fig. 2 Internal hardware of LOA

$$\text{Mean Absolute Error(MAE)} = E[|\varepsilon|],$$

where E is the expectation operator.

2.2 Architecture

The template architecture is considered from literature and shown in Fig. 2. The architecture template should figure out the efficiency of hardware architecture and delay, where 'A' and 'D' denote the area and delay of the architecture, respectively. In the approximation technique, many samples have been analyzed to consider the best one. Using the unit gate model, not only OR gates, more gates like AND, OR, and NAND are placed, and the combinations of their value are also noted to state that non-similar two input gates XOR and XNOR have more area and delay.

As literature and discussions state that error versus hardware cost trade-off is very efficient in LOA and found to be the best architecture among the existing approximate adders. Evaluation of the general template of LOA allows division of sub-adder into 'n_l' 2-to-1 logic blocks as shown in Fig. 2 and single 2-to-2 logic block which generates the carry for the accurate adder using AND gate by receiving the inputs of an adder at bit position n_l. The higher sub-adder is an accurate adder (exact adder), where XOR results in sum, and AND results in carry. The error metrics and unit gate characteristics of 2-to-1 blocks and 2-to-2 blocks of the architecture are stated in Tables1 and 2, respectively.

From Table 1, it is clear that consideration of MSE error metric and replacing OR gates in 2-to-1 blocks and OR-AND in the first bit of higher sub-adder is the best selection. MSE has strictly positive values (non-negative).

Table 1 Error metrics and unit gate characteristics of 2-to-1 blocks

	μ	σ^2	MSE	A	D
AND	–3/4	3/16	3/4	1	1
OR	–1/4	3/16	1/4	1	1
Buffer	–1/2	1/4	1/2	0	0
Cte-0	–1	1/2	3/2	0	0
Cte-1	0	1/2	1/2	0	0

Table 2 Error metrics and unit gate characteristics of 2-to-2 blocks

	μ	σ^2	MSE	A	D
Half adder	0	0	0	3	2(1)
OR_AND	¼	3/16	1/4	2	1(1)
Cte-1_AND	½	1/4	1/2	1	0(1)
Buffer_AND	0	1/2	1/2	1	0(1)

Table 2 represents all possible combinations of 2-to-2 blocks in the higher sub-adder block which eliminate maximum error values. In this case, half adder is used while it is having standard deviation value as zero, average error and MSE as zero, area as 3, and delay 2.

The data which is distributed parallelly, every bit is uncorrelated along with this error metrics, is to be analyzed and measured as a function of error metrics of each block. The block which contains better parameters can be chosen as optimal architecture. So, the overall error is concluded as the combination of error of each block with corresponding weight,

$$\varepsilon_T = \sum_{i=0}^{n_l} \varepsilon_i 2^i$$

$$\mu_T = \sum_{0}^{n_l} \mu_i 2^i$$

$$\sigma_T^2 = \sum_{0}^{n_l} \sigma_i^2 2^{2i}$$

$$\text{MSE}_T = \sum_{0}^{n_l} \sigma_i^2 2^{2i} + \left(\sum_{0}^{n_l} \mu_i 2^i \right)^2$$

where μ_i and σ_i^2 are average error and variance of error associated with block in bit position i.

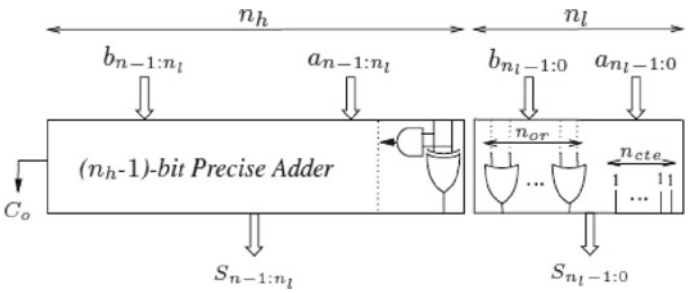


Fig. 3 Proposed system of LOCA: $n_l = n_{cte} + n_{or}$

3 Optimized LOCA

The LOCA can have various optimization methods based on the metrics values shown in Tables 1 and 2. In data processing and image processing applications, MSE is considered as one of the important error metrics because it measures the average of the errors that is the average difference between the approximate results to the accurate result. So, the proposed optimized architecture based on the mean square error metric surely brings the optimum and errorless computations to real-time applications. The various combinations of optimized lower part constant-OR adder architecture can be evaluated by including both lower sub-adder block and higher sub-adder block.

The upper bits (higher sub-adder) produce a high error rate as compared to lower bits (lower sub-adder). So, this article has concentrated on higher sub-adder rather than lower sub-adder. As seen in Table 2, the best higher sub-adder is OR_AND and half adder. Although it does not improve the delay by replacing the OR_AND with half adder, it improves the area. By fixing the higher sub-adder to a half adder, it is observed that the average error is considered as zero or positive, coming to the lower sub-adder block having a zero or negative average error. So, a higher sub-adder block is used with small u (C–1) or small (OR). Therefore, the optimal architecture of the lower sub-adder consists of OR gates followed by 1's blocks in the lower bits where hardware complexity of structural design is reduced while optimizing area and delay and hence called OLOCA (Fig. 3).

The proposed LOCA architecture is verified with an optimal number of OR gates and results in optimal value at nor = log2 (8/3). The integer numbers nor = 1 and nor = 2 produce the same MSE, but nor = 2 gives a better STD, and hence, this architecture is named OLOCA. The various error formulas in terms of architecture parameters are as shown (Table 3).

4 Result and Discussion

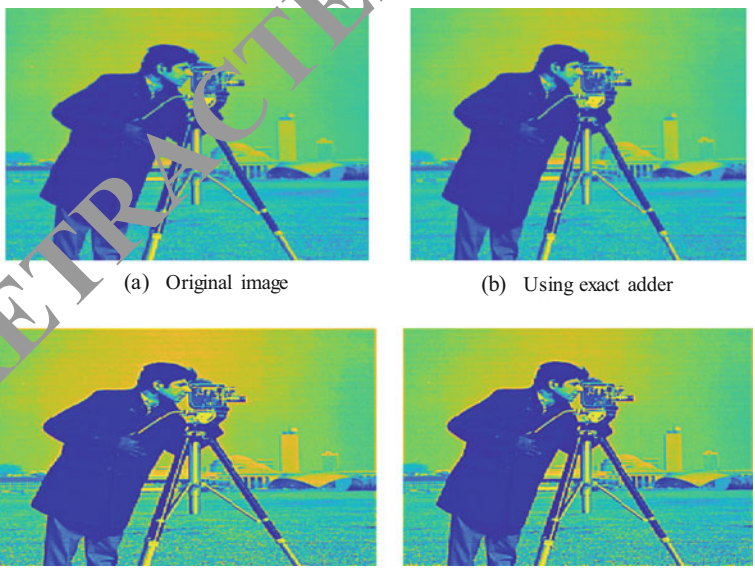
The LOCA found very significant image processing applications to improve the sharpness of an image. Structural similarity of an image can be the quality metric

Table 3 Formulas of error metrics, area, and delay

Parameter	LOA	OLOCA
μ	1/4	$-3/16\ 2^{n}{}_{1}$
σ^2	$1/4^{n}{}_{1} - 1/16$	$53/7684^{n}{}_{1}$
MSE	$1/4\ 4^{n}{}_{1}$	$5/48\ 4^{n}{}_{1} - 1/6$
MAE	$3/8\ 2^{n}{}_{1} - 3/8$	$15/64\ 2^{n}{}_{1} - 3/42^{-n}{}_{1}$
A	$(n_h - 1).A_{FA} + A_{AND} + (n_l + 1).A_{OR}$	$(n_h - 1).A_{FA} + A_{HA} + (n_1 - n_{cte}).A_{OR}$
D	$(n_h - 1).t_c + T_{AND}$	$(n_h - 1).t_c + T_{AND}$

used to measure the similarity between two images. Multimedia applications like animation programs where pixels are added in a picture can utilize the OLOCA addition operator and reduce the error rate. JPEG compression is used for saving storage space and transmission bandwidth for digital images. Reducing the data correlation by converting it from the time domain to the frequency domain is the strategy behind JPEG compression. The human eye is less sensitive to high frequencies. The LOCA techniques are implemented on a MATLAB simulation environment, and the results are shown in Fig. 4. The error metrics for JPEG image having 8-bit data size are summarized in Table 4.

The proposed OLOCA and ripple carry adder architectures are also implemented using Verilog HDL with targeted FPGA device xc7a100t-3-csg324, and the design parameters like area and speed of the architectures are tabulated in Table 5. From



(a) Original image (b) Using exact adder

(c) Image with two OR gates with n_l=8 (d) Image with two OR gates with n_l=7

Fig. 4 Simulated images processed using OLOCA

Table 4 Simulation results of 8-bit

Error technique		$n_l = 2$	$n_l = 3$	$n_l = 4$	$n_l = 5$
MAE	LOA	1.38 s	2.88	5.87	11.87
	OLOCA	0.75	1.78	3.70	7.48
MSE	LOA	4.00	16.00	63.93	255.90
	OLOCA	1.50	6.53	26.50	106.57
STD	LOA	1.99	3.99	7.99	16.00
	OLOCA	0.97	2.06	4.18	8.40
ADP	LOA	26.82	19.19	13.19	7.95
	OLOCA	27.00	18.89	12.24	6.74

Table 5 Design parameters of 8-bit adder architectures

Parameter	RCA	OLOCA
Number of XOR gates	16	7
Delay	2.523 ns	0.761 ns
Speed	0.4 GHz	1.3 GHz

Table 5, it has been proved that the proposed architecture has presented optimum performance both in terms of area (number of gates) and speed of computations.

5 Conclusion

The hardware architectures and their performance of adders have got significant importance in most of the computing architectures. In this article, an architecture called optimized lower part constant-Or adder architecture has improved the computing speed of addition operations compared to traditional adder architectures. The proposed architecture would balance the cost of hardware and accuracy while reducing the hardware complexity of existing architecture and proved that the proposed architecture showed significant optimization in both area and speed.

References

1. D. Ayad, N. Ardalan, G.O. Alberto, Systematic design of an approximate adder: the optimized lower-part constant-OR adder (2018)
2. J. Satori, R. Kumar, Stochastic computing. Found. Trends Electron. Design Autom. **5**(3), 153–210 (2011)
3. S. Mittal, A survey of techniques for approximate computing. ACM Comput. Surv. **48**(4), 62-1–62-33 (2016)

4. A.B. Kahng, S. Kang, Accuracy-configurable adder for approximate arithmetic designs, in Proceedings of the 49th Annual Design Automation Conference (DAC) (2012), pp. 820–825
5. D. Mohapatra, V.K. Chippa, A. Raghunathan, K. Roy, Design of voltage-scalable meta-functions for approximate computing, in Proceedings of the Design, Automation and Test in Europe (2011), pp. 1–6
6. S. Ajmera, M. Vucha, A. Kokkula, High speed architecture for orthogonal code convolution**,** in Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS, 2017)
7. N. Zhu, W.L. Goh, K.S. Yeo, An enhanced low-power high-speed adder for error tolerant application, in Proceedings of the 12th International Symposium Integerates Circuits (2009), pp. 69–72
8. M. Vucha, A.L. Siridhara, High speed cryptography architecture for health ınformation exchange. Int. J. Adv. Trends Comput. Sci. Eng. **8**(4), (2019)
9. T. Kalyani, S. Monika, B. Naresh, M. Vucha, Accident detection and alert system. Int. J. ... Technol. Exploring Eng. (IJITEE) **8**(4S2), (2019). ISSN: 2278–3075
10. S.-L. Lu, Speeding up processing approximate circuits, Computer **37**(3), L 67– (2004)
11. T. Anuradha, K.A. Manjusha, R. Karthik, M. Vucha, A.L. Siridhara, Design and ımplementation of an audio parser and player. J. Eng. Appl. Sci. **12**(20), 5301–5306 (2017)
12. D. Esposito, D. De Caro, A.G.M. Strollo, variable latency speculative parallel prefix adders for unsigned and signed operands. IEEE Trans. Circuits Syst. I, Reg. Papers **63**(8), 1200–1209 (2016)
13. H. Jiang, J. Han, F. Lombardi, A comparative review and evaluation of approximate adders, in Proceedings of the 25th Edition Great Lakes Symp VLSI (GLSVLSI) (2015), pp. 343–348
14. A.L. Siridhara, M. Vucha, T. Ravinder, Performance evaluation of parallel multipliers. J. Eng. Appl. Sci. **12**, 5186–5189 (2017)
15. A. Najafi, M. Weiβbrich, G.P. Vaya, A. Garica-Ortiz, A fair comparison of adders in stochastic regime, in Proceedings of the 27th International Symposium Power Timing Modeling, Optim. Simulation(PATMOS) (2017), pp. 1–6
16. M. Vucha, A. Rajawat, Design and VLSI ımplementation of systolic array architecture for matrix multiplications. Int. J. Comput. Appl. **26**(3), 18–22 (2011)
17. M. Vucha, L.S. Varghese, Design space exploration of DSP techniques for hardware software co-design: an OFDM transmitter case study. Int. J. Comput. Appl. **116**(20), 29–33 (2015)
18. P. Gurjar, R. Solanki, P. Kanslival, M. Vucha, VLSI ımplementation of adders for high speed ALU. Int. J. Comput. Appl. **29**(10), 11–15 (2011)
19. M.C. Sudeep, M. Sharath Bimba, M. Vucha, Design and FPGA implementation of high speed vedic multiplier. Int. J. Comput. Appl. **116**(20), 6–9 (2014)

GenSQL—NLP-Based SQL Generation

M. Sri Geetha, R. Yashwanthika, M. Sanjana Sri, and M. Sudiksa

Abstract Relational databases saved the world's facts in a considerable amount. The inability of insight of query languages limits the ability of the users to recapture the adequate knowledge. Here, natural language processing is enforced to relational databases. An NLP-based model is proposed to convert natural language utterances to SQL queries. Semantic parsing models find it difficult to derive to imaginary database schemas while convertion, which then retrieves corresponding results from the database. To handle schema encoding, schema look up and feature illustration within an encoder, the word uttered is taken as the input, and then sequence-sequence modelling takes place to show consideration to schema encoding based on the relation-aware self-attention system. The framework is evaluated on the two models based on the specific coordination and hardness norms of inquiry. The expected model upgrades the results of text-to-SQL (Yu et al. in Cosql: A conversational text-to-sql challenge towards cross-domain natural language interfaces to databases, pp. 1962–1979, 2019) task that is displayed on results of the Spider dataset.

Keywords Natural language processing (NLP) · SQL statements

1 Introduction

In this rapid technologically progressing world, the interaction among human and machine is the most vital and fundamental part that provides assistance to many

M. Sri Geetha (✉) · R. Yashwanthika · M. Sanjana Sri · M. Sudiksa
Sri Ramakrishna Engineering College, Coimbatore 641022, India
e-mail: srigeetha.m@srec.ac.in

R. Yashwanthika
e-mail: yashwanthika.1701259@srec.ac.in

M. Sanjana Sri
e-mail: sanjana.1701210@srec.ac.in

M. Sudiksa
e-mail: sudiksa.1701244@srec.ac.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_20

categories. In such cases, the amount of data keeps multiplying each day hence making it difficult to write queries for complex logics. In relational databases, a large part of the world's facts are stored. Earlier, many large companies were concerned about the data, whose ability is to query using structured query languages (SQL).

For the retrieval of valuable information, one of the sought-after language SQL is used. The amount of personal data getting stored keeps increasing each day along with the unavoidable growth of usage of cell phones and, thus, increases the need for a large number of people from different fields to query and use data on their own. Many people have not acquired the required knowledge to write SQL and query the information even though there is a great increase in the popularity and demand for data science. Not only them, even SQL experts find writing these queries again and again a monotonous process. Due to this fact, a great amount of data available today is not efficiently accessed. This pushes us to the verge of finding an effective solution to make SQL queries easy and simpler. Here is where the NLP comes to aid. Retrieval of the required information from the database can be a tedious process.

The AI model's function is to get parsed into SQL commands after the extraction of facts from various multiple information. Text language to SQL translation is smoothed with the support of machine learning and knowledge-based resources. By this, the course of data retrieval can be made simple and efficient for common man. Natural language processing (NLP) techniques are used to translate text into SQL queries. Natural language processing sees application in various fields like tourism, healthcare, financial services and sales industries. Text-to-SQL is a task that automatically translates a natural language query into SQL. This is a workaround that bridges the gap between non-technical users and database systems. It also demonstrates that individuals are not required to learn the database schema and semantics.

We must have seen a big improvement in NLP downside assignments, due to reconceptualized bearings on transfer learning over the past few years. A compelling need for an effective solution is required here due to this advancement. The goal is to allow the data talk directly using human language. Thus, a large number of data is analysed, and querying is made easy for users of any field with the help of NLP. It is necessary for the system to understand the question given by the user and translate them into SQL queries spontaneously to build this type of a natural language interface. From the start of NLP, the operation of natural language interface to access database has been a keen area of exploration and still continuous to be.

2 Related Works

2.1 IRNet

A neural approach called intermediate representation (IRNet) for broad and cross-domain text-to-SQL [1] addresses two fundamental challenges in Text2SQL [2] tasks. The misalignment of expressed intentions in simple language, as well as the

difficulty in predicting columns caused by a greater number of out-of-domain words, are among the issues. Instead of creating SQL queries end-to-end, IRNet decomposes natural language into three categories. The schema linking method is carried out over a database schema and a question in the initial step. IRNet tends to make use of SemQL, a domain-specific language that acts as the interface between SQL and plain language. Natural language (NL) encoder, schema encoder, and decoder are all part of the model. For the completion of the Text2SQL job, each component of the framework is said to have a unique role. The NL encoder takes natural language input and converts it to an embedding vector is obtained by converting the natural language input that is taken from an NL encoder.

To create hidden states, the embedding vectors are employed in a bi-directional LSTM. The schema encoder produces column and table representations by taking the database structures as input. In the end, using a context-free language, the decoder is utilized to generate SemQL queries. On the SPIDER dataset, IRNet achieves 46.7% accuracy, which is a considerable gain of 19.5% over prior reference models. When IRNet is combined with BERT, performance improves significantly and accuracy rises to 54.7%.

2.2 EditSQL

The main job of the cross-domain text2SQL that is context-dependent has been the main focus of the EditSQL. The neighbouring natural language inquiries are inter-dependent and the SQL queries coincide with them, and this fact is exploited by the authors. They take advantage of this reliance by amending the already anticipated query to enhance the effectiveness of the generation. At the token level, the editing mechanism receives SQL input as a series and recreates generating results. To deal with complicated tables in various domains, an utterance-table encoder [3] and a table-aware decoder are employed, which incorporate the context of the spoken language and the schema. User input and database schema are encoded using the utterance-table encoder. To encode tokens of utterances, a bi-LSTM is utilized. To choose the most important columns, an average of column header embedding is utilized for each token. The link between the table schema and the speech is captured via an attention layer. A communication decoder is developed on top of the speech encoder to capture data from different statements. Using an LSTM decoder, the conversion process concludes with the production of SQL queries based on the interaction history, database architecture and user voice. The SParC dataset, generated from SPIDER, is a large-scale cross-domain context-dependent conceptual parsing dataset, is used to test the model. The method achieves the previous state-of-the-art model, IRNet, on SPIDER and SParC, but not on SParC. In inter domain text2SQL generation, the model obtains a 32.9% accuracy. Furthermore, using BERT embedding results in a considerable increase in accuracy, with a score of 53.4%. According to our research, there are three problems with broad system language models when they are implemented in text-to-SQL semantic parsers: 1. The encrypting database

relations in an accessible way for the semantic parser, 2. Modelling integration between domain columns and their discusses in a specific input. This fails to recognize column names in utterances, infer column references from cell values and create sophisticated SQL queries.

3 Dataset Description

Spider dataset [4] is used to evaluate our proposed system. We use the same data split as Yu et al. [4]; 206 databases are split into 146 train, 20 dev and 40 tests. All questions for the same database are in the same split; there are 8659 questions for train, 1034 for dev and 2147 for test.

Spider dataset emphasizes on complicated SQL statements with several keywords and the ability to connect numerous tables. To deal with such intricacies, presented semantics SQLNet, a syntax tree network with modular decoders that creates a SQL statements by recursively invoking a module that conforms to the SQL semantics.

Table 1 Singer table

Singer_ID	Name	Birth_Year	Net_Worth_Millions	Citizenship
1	Liliane Bettencourt	1944	30	France
2	Christy Walton	1948	28.8	United States
3	Alice Walton	1949	26.3	United States
4	Iris Fontbona	1942	17.4	Chile
5	Jacqueline Mars	1940	17.8	United States
6	Gina Rinehart	1953	17	Australia
7	Susanne Klatten	1962	14.3	Germany
8	Abigail Johnson	1961	12.7	United States

Table 2 Song table

Song_ID	Title	Singer_ID	Sales	Highest_Position
1	Do They Know It's Christmas	1	1,094,000	1
2	I Don't Want You Back	1	552,407	1
3	Cha Cha Slide	2	351,421	1
4	Call on Me	4	335,000	1
5	Yeah	2	300,000	1
6	All This Time	6	292,000	1
7	Left Outside Alone	5	275,000	3
8	Mysterious Girl	7	261,000	1

The two above tables (Table 1 Singer table and Table 2 Song table) are the part of singer database in the spider dataset.

To do well on the database split, the model needs to learn to compose various SQL operators and generalize to new schemas, as all databases will be unseen at test time. To compare with prior work, we report the exact component matching score. The predicted query is decomposed by the SELECT, WHERE, GROUP BY, ORDER BY and KEYWORDS. Each component in the predicted query and the ground truth are then decomposed into subcomponents and checked if the sets of the components match exactly. The predicted query is correct when all components match.

4 Proposed Model

The model initially takes the input natural language query and sends it to the encoder. The encoder then retrieves a pre-trained glove [5] embedding for each word and processes the embeddings in each multi-word label using a bi-directional LSTM to produce the representations of each node. It additionally applies a second Bi-LSTM to the input natural language query to generate preliminary word representations. The model is helped by schema look up relations in aligning column/table references in the input query to the appropriate schema columns/tables. The keywords and attributes are involved to construct the SQL query with the corresponding attributes and keywords. The tree-structured design is used by the LSTM decoder. It constructs the SQL in depth-first traversal order as an abstract syntax tree (Fig. 1).

4.1 Attribute Selection

In Fig. 2, The attribute selection, by predicting whether or not a column is utilized in the statement, this learning objective assists the model to learn the common components between the utterance and the schema. Furthermore, a two-layer MLP is applied to each column representation produced from the transformer encoder, which is collected from the output of an average pooling layer that gathers all sub-tokens of the relevant column. The probability that the associated column is expressed in the utterance is then calculated using a sigmoid activation function.

For the input query “Name of the singers who is from France” is given as the input to the model, the input is compared with the database table and the direct attributes are retrieved in the attribute selection. Thus, the attribute “singer.name” is retrieved from the singer table from the database.

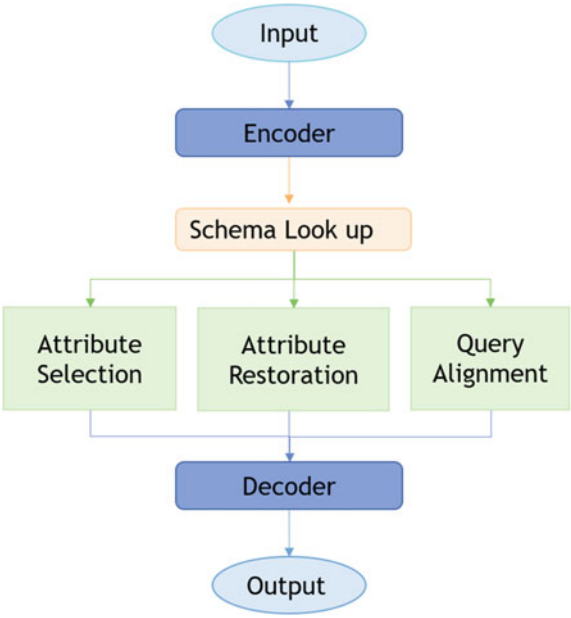


Fig. 1 Architecture of GenSQL

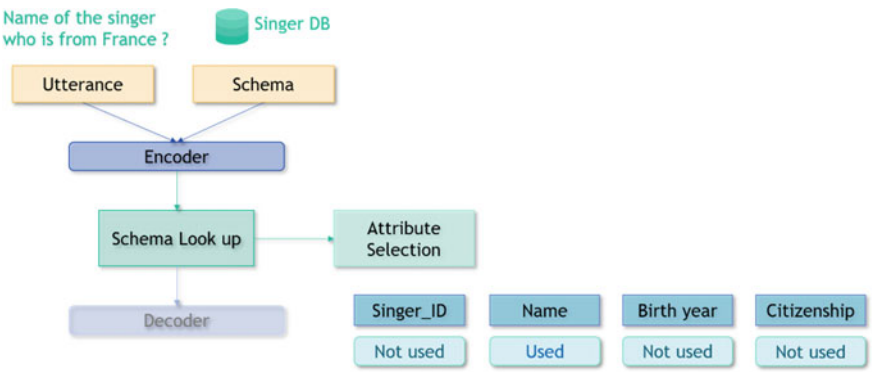


Fig. 2 Attribute selection description

4.2 Attribute Restoration

In Fig. 3, The attribute restoration, by restoring the attribute name based on a sampled entity value, this learning strategy increases the model's potential to expose the links between the entity values and the attribute names. For example, the model recovers the attribute "Citizenship" based on the input statement that states "France". In general, the transformer decoder obtains attribute names from two sources: the actual entity value and the attribute names mentioned in the statement.

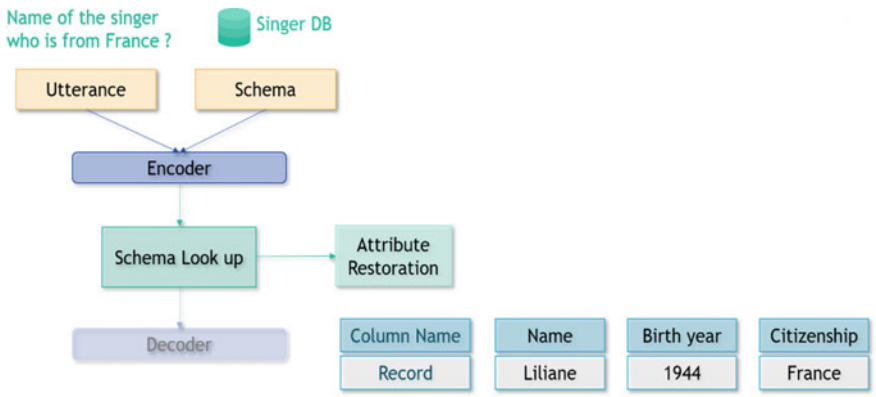


Fig. 3 Attribute restoration description

For the attribute restoration, we devise the enforcing procedures:

- If an attribute name is not specified in the input statement, the attribute name will be replaced with its entity value with a probability of 0.5. In this scenario, the attribute name will be retrieved from the entity value without any further context.
- If an attribute name is stated, the attribute name will be replaced directly with its entity value. In this situation, the model may use the contextual information from the statement and the entity value to retrieve the attribute name.

4.3 GenSQL

The proposed model is based on the mechanism of relation-aware self-attention, which facilitates in schema encoding, feature representation and schema lookup in the encoder. This learning goal is directly connected to the subsequent task [6]. The decoder optimizes probability based on the representation from the transformer encoder. Given the enormous number of complicated SQLs in crawled data, this learning goal enables the model to learn to build complicated SQL that involves logical thinking (Fig. 4).

For instance, instead of just predicting whether or not a column is utilized, the column must be created in the appropriate places by the model's decoder, such as in the ORDER BY clause or the WHERE clause. The decoder, in particular, using a small vocabulary database that is made up of keywords and column names the decoder emits the target SQL token by token.

During the well before phase, the SQL keyword embeddings are randomly initialized and trained. By averaging the column's sub-tokens representations, the column representations are obtained in the same method as that was utilized in the column prediction learning goal. The decoder constructs a value of the other variable at each

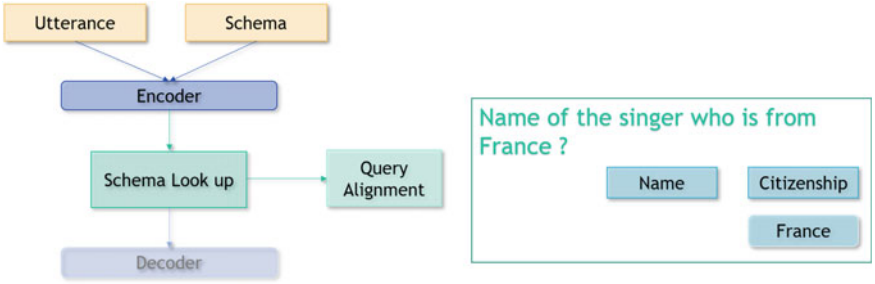


Fig. 4 Query alignment description

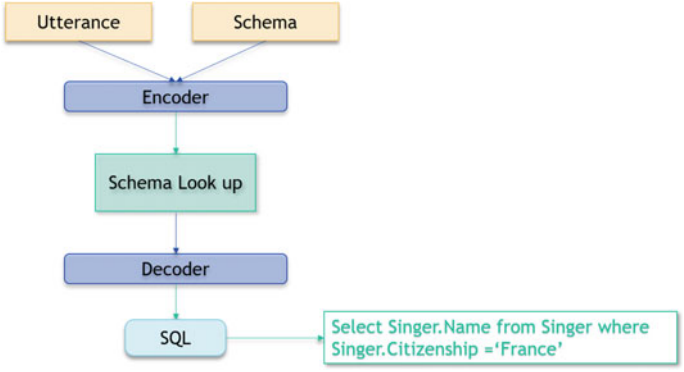


Fig. 5 Query generation description

decoding step, then applies an input feature operation to it and the target vocabulary presentations to construct a probability distribution from across vocabulary set (Fig. 5).

5 Results

The GenSQL model generates the SQL statements based on the available database schema and the input query. The output of the generated SQL (Fig. 6) is also obtained appropriately. During training, the encoder was kept constant while the probe parameters were modified to fine tune with Adam optimizer with cross-entropy loss function. The model is trained for 100 epochs on SPIDER dataset. The model achieved accuracy of 0.72.

```
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Natural Language:which country has more singers
Generated SQL:
SELECT singer.Citizenship, Count(*) FROM singer GROUP BY singer.Citizenship ORDER BY Count(*) Desc LIMIT 1
Output:
[('United States', 4)]


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Natural Language:the song which sold above 1000000
Generated SQL:
SELECT song.Title FROM song WHERE song.Sales > '1000000'
Output:
[("Do They Know It's Christmas",)]


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Natural Language:who is the richest singer
Generated SQL:
SELECT singer.Name FROM singer ORDER BY singer.Net_Worth_Millions Desc LIMIT 1
Output:
[('Liliane Bettencourt',)]


++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
Natural Language:Singer who is from France
Generated SQL:
SELECT * FROM singer WHERE singer.Citizenship = 'France'
Output:
[(1, 'Liliane Bettencourt', 1944.0, 30.0, 'France')]
```

Fig. 6 Generated query and its results

6 Conclusion

It is obvious that from the start of a new millennia, the number of information produced keeps increasing rampantly. Despite the ephemeral growth in the demand of data science, writing SQL and querying the information have become difficult for most of the people. Natural language processing is seen to have set a new model of all the other approaches that are used to retrieve data from a database. This displays a clear idea on the procedures that takes place in the natural language processing. The efforts of finding the appropriate data from the databases need an integrative skillset and a combination of the elemental difficulties of querying the disorganized text information.

Most present-day models find difficult to adapt to good portrayals to accordingly link column/table allusions in the question even though the text-to-SQL research is done earnestly. Encoding and usage of data from the architecture, role of a model in a question are a few hurdles related to it.

GenSQL motive is to give a solution by parsing the information into reasonable models by strengthening the relationship of written queries in an architecture and then querying the data into chatty plain text. The main work of the model is to allow bring in a significant improvement on text-to-SQL parsing. In conclusion, in terms of database management, NLP is said to be an enriching experience for any common man. To be precise, natural language processing is seen to be a reliable tool in many fields like education, business, tourism, medicine and so on for their betterment in the future.

References

1. T.S. Anisha, P.C. Rafeeque, M. Reena, Text to SQL query conversion using deep learning: a comparative analysis (August 15, 2019), in *Proceedings of the International Conference on Systems, Energy and Environment (ICSEE)* (GCE Kannur, Kerala, 2019)
2. T. Yu, M. Yasunaga, K. Yang, R. Zhang, D. Wang, Z. Li, D. Radev, SyntaxSQLNet: syntax tree networks for complex and cross-domain text-to-SQL task, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (2018)
3. R. Cai, B. Xu, Z. Zhang, X. Yang, Z. Li, Z. Liang, An encoderdecoder framework translating natural lang. to database queries, in *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (2018), pp. 3977–3983
4. T. Yu, R. Zhang, H. Er, S. Li, E. Xue, B. Pang, X.V. Lin, Y.C. Tan, T. Shi, Z. Li, et al., Cosql: a conversational text-to-sql challenge towards cross-domain natural language interfaces to databases, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Lang. Process. and the 9th International Joint Conference on Natural Lang. Processing* (2019), pp. 1962–1979
5. J. Pennington, R. Socher, C. Manning, Glove: global vectors for word representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Association for Computational Linguistics, Doha, Qatar, 2014), pp. 1532–1543
6. E. Levin, R. Pieraccini, W. Eckert, Learning dialogue strategies within the Markov decision process framework, in *Automatic Speech Recognition and Understanding IEEE Proceedings* (1997), pp. 72–79
7. J. Guo, Z. Zhan, Y. Gao, Y. Xiao, J.-G. Lou, T. Liu, D. Zhang, Towards complex text-to-SQL in cross-domain database with ıntermediate representation, (2019). arXiv: 1905.08205v2 [cs.CL], 29 May 2019
8. H.M. Anil Kumar, A.R. Kumar, P. Harshitha, S.D.N. Mahadevaswamy, Providing natural language ınterface to database using artificial ıntelligence. Int. J. Sci. Technol. Res. **8**(10) (2019). ISSN 2277-8616
9. J. Ma, Z. Yan, S. Pang, Y. Zhang, J. Shen, Mention extraction and linking for SQL query generation, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2020), pp. 6936–6942
10. E. ILevin, R. Pieraccini, A. Di Carlo, in *User Initiated Mixed Initiative Dialogue for Database Information Retrieval*, ed by S. LuperFoy. Automated Spoken Dialogue Systems (MIT Press)
11. P. He, Y. Mao, K. Chakrabarti, W. Chen, X-SQL: reinforce schema representation with context (2019). arXiv preprint arXiv:1908.08113
12. C. Finegan-Dollak, J.K. Kummerfeld, L. Zhang, K. Ramanathan, S. Sadasivam, R. Zhang, D. Radev, Improving text-to-SQL evaluation methodology, in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, vol 1 (Long Papers, 2018), pp. 351–360

An Unique 'Assessment Framework' for Agility in Software Development Projects

A. V. Ranjitha, M. Suresh, and S. Lakshmi Priyadarsini

Abstract The purpose of this paper is to develop unique assessment framework for evaluating agility in Software Development Project (SDP). This study makes an original contribution to the body of knowledge related to agility in SDPs. The conceptual framework is developed by using four enablers, twenty three criteria and one hundred and twenty three attributes for agility level assessment in SDP. The multi-grade fuzzy approach is used for development of this assessment framework. The identified major enablers are team agility, process agility, engineering agility, and organizational agility. With the help of this framework, project managers can identify weaker attributes in their current projects, thereby improving overall SDP agility. It is a generalized and continuous assessment tool for evaluating the agility level of SDP.

Keywords Organization agility · Project agility · Agility assessment · People agility · Multi-grade fuzzy · Software development projects

1 Introduction

Unpredictable situations, changes in organizational climate and budget cuts have dramatic effects on the project. The key to successful project management is to break the project into stages and deliver the finished parts as and when needed. This enables the concept of going agile project management. The word agility is derived from the Latin, agere, means 'to drive, act,' implying a sense of ownership, and the ability to drive something forward. Additionally, the definition refers to 'having a quick resourceful and adaptable character.' Schön et al. [1] state that agile software development was developed to provide on-time delivery and customer satisfaction

A. V. Ranjitha · M. Suresh (✉)
Amrita School of Business, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: m_suresh@cb.amrita.edu

S. Lakshmi Priyadarsini
Govt Victoria College, University of Calicut, Palakkad 678001, Kerala, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_21

hand in hand. Thus, the development process is carried out incrementally and empirically. Highsmith [2] states agility has three characteristics in accordance with project management. They are the sense of ownership and authority, quick and easy changes of direction, and resourceful and adaptable climate. In this paper proposes the unique framework for assessment of agility level in Software Development Project (SDP). The proposed generalized framework is applicable for any SDP irrespective of their agile methodologies like XP, SCRUM, DSDM, etc. This is the first time that the agility level assessment framework has been developed using multi-grade fuzzy approach in the SDP. The objective of this paper is converted in to the following Research Question (RQ)s:

RQ-1: What all are the criteria involved in SDP agility assessment framework?
RQ-2: What are the attributes that influence the agility of SDP?
RQ-3: How to measure the agility of SDP?

2 Literature Review

Moe et al. [3] conducted a study for a better understanding of the nature of self-managing agile teams and the teamwork challenges that arise when introducing such teams. Strode [4] through various literature review states that the central tenet of the agile philosophy is effective teamwork. Strode et al. [5] conceptualized the coordination strategy and its effects on coordination effectiveness. Fontana et al. [6] used process versus people approach to define the maturity of agile teams. Sheffield and Lemétayer [7] proposed a model with 20 factors for successful agile software development. Curcio et al. [8] discussed about the results of software projects, five of the eight top projects cancelation factors are related to requirements. Medeiros et al. [9] analyzed a cross-case of six companies to understand the quality of software requirements specification in agile projects. Tessem [10] studied how empowerment is enabled in software development teams both agile and non-agile; identified differences in empowering practices and levels of involvement. Lindsjørn et al. [11] studied whether learning and work satisfaction of agile software teams differ from that of traditional software teams. Rigby et al. [12] discussed six organizational practices to capitalize on agile potential. They are learning how agile works, understanding where agile does or does not work, starting small and letting the word spread, allowing master teams to customize their practices, practicing agile at the top level, destroying the barriers to agile behavior. Gandomani and Nafchi [13] constructed a model based on grounded theory of different aspects of human-related challenges throughout the agile transition process. Nurdiani et al. [14] carried a tertiary study on agile and Lean practices in software projects. Shastri et al. [15] studied the role of project manager in the agile software development projects.

3 Methodology

3.1 Identification of Agile Enablers, Criteria, and Attributes

The multi-grade fuzzy approach is widely used for assessment in manufacturing and service industries [16–21]. As far as none of the research studies have applied the multi-grade fuzzy to assess agility index of SDP. Multi-grade fuzzy approach is adopted from [22] to evaluate the agility of a SDP. In this study four enablers, twenty three criteria and one hundred and twenty three attributes have been captured through interview with various agile experts and literature review. Table 1 depicts the conceptual framework of agile performance indicators of SDP.

3.2 Case Study

Information Technology in India consists of IT services and BPO. According to the NASSCOM report in 2017, this sector contributed about 7.7% of India's GDP. One of the major functions of the IT sector is software development and maintenance. For the assessment, one of the fortune 500 companies was chosen. In that company, a project (case SDP) which is using the Scrum framework was chosen for assessment of the agility.

Let the agility level be represented as I. It is the product of rating R and weight, W.
Thus, the agility level is $I = R * W$.
Scale of agility level, $I = \{10, 8, 6, 4, 2\}$
8–10 represents 'Extremely Agile'
6–8 represents 'Agile'
4–6 represents 'Generally Agile'
2–4 represents 'Not Agile' and
<2 represents 'Extremely not Agile'.

The likert scale of 0–10 for rating R and weight W. The reliability of the scale has been proved in many cases [23, 24]. The attributes, criteria, enablers weightages are collected from agile experts from various SDP. Normalized weights of enablers (W_i), criteria (W_{ij}) and attributes (W_{ijk}) were obtained in Table 2. Ratings for attributes were obtained from five members (E_1, …, E_5) from case SDP at different levels like manager, developer, scrum master, team lead, tester and it's shown in Table 2.

Sample Calculation

The calculation comprises of three levels, namely primary level assessment, secondary level assessment, and tertiary level assessment. Through this process, we can obtain the agility index of the SDP.

Table 1 Conceptual framework for agility assessment in SDP

Enablers	Criteria	Attributes
1. Team agility	1.1. Team orientation	1.1.1. The propensity to take others behavior in to account during group interaction
		1.1.2. The belief in the importance of team goals over individual's goals
	1.2. Team Composition	1.2.1. Effective team size: 5–10 people
		1.2.2. Being multidisciplinary
		1.2.3. Check overlap of roles
		1.2.4. Roles are arranged relevantly
	1.3. Communications	1.3.1. Right conversation with the right level of details
		1.3.2. Relevant ideas and information relating to the team and its work is shared openly by all team members
		1.3.3. Timeliness in receiving information from other team members
		1.3.4. Define and communicate release contents to all
		1.3.5. 10–15 min daily call with all team members covering what they did yesterday? What they are planning to do today? Is there any issue in proceeding with the task?
		1.3.6. Sharing information about Ad-hoc activity
		1.3.7. The mode in which distributed teams communicate
		1.3.8. The work done on subtasks within the team is closely harmonized
		1.3.9. Goals are fully and clearly communicated within the team
	1.4. Decision making	1.4.1. Autonomy for fast decision making reduces waiting time for the superior to make decisions. 'Be autonomous but not sub-optimize.' Self-organized team

(continued)

Table 1 (continued)

Enablers	Criteria	Attributes
		1.4.2. Coach says what to solve and why to solve; Team decides how to solve?
		1.4.3. Roles and the decision making power are specified explicitly
		1.4.4. Take decisions based on the position in performance plan (forming, storming, norming, performing)
	1.5. Team leadership	Ability to direct and coordinate the activities of other team members, assess team performance, assign tasks, develop team knowledge, skills, and abilities, motivate team members, plan and organize, and establish a positive atmosphere 1.5.1. Structure of quick flow of ideas: lesser levels of hierarchy
		1.5.2. Closely connected leadership (technical lead, product lead, design lead are tightly connected)
		1.5.3. Clear definition of personnel's responsibility and authority
		1.5.4. Rotation of authority
	1.6. Cohesion	The action or fact of forming a united whole 1.6.1. Mutual trust: the shared belief that team members will perform their roles and protect the interests of their teammates
		1.6.2. All departments of projects are fully integrated into the team. (No more separate Dev, testing, ops, etc. Instead all in a team or if large team then representatives of each department must be in the other team)
		1.6.3. Every team member feels responsible for maintaining and protecting the team

(continued)

Table 1 (continued)

Enablers	Criteria	Attributes
	1.7.Collective code ownership	The code base is owned by the entire team and anyone may make changes anywhere Anyone can edit the code when a defect arises but it must be reviewed by SME. This will reduce the time of dependency of the person 1.7.1. Ability to anticipate other team members' needs through accurate knowledge about their responsibilities
		1.7.2. Ability to shift workload among members to achieve balance during high periods of workload or pressure
		1.7.3. The team is able to reach consensus regarding important issues
		1.7.4. Discussions and controversies are conducted constructively
	1.8. Knowledge	1.8.1. Keep lessons learned
		1.8.2. Knowledge of the customer's business culture
		1.8.3. Knowledge of the project
		1.8.4. Knowledge of the technology
	1.9. Adoptability	1.9.1. Ability to adjust strategies based on information gathered from the environment through the use of backup behavior and reallocation of intra-team resources
		1.9.2. Altering a course of action or team repertoire in response to changing conditions (internal or external)
	1.10. Involvement	1.10.1. Team members participate in ceremonies (planning meetings, retrospectives, review meetings) together with peers and other stakeholders
		1.10.2. Delegating the tasks based on the knowledge of the person by knowing the team in and out

(continued)

Table 1 (continued)

Enablers	Criteria	Attributes
	1.11. Empowerment	1.11.1. The team is responsible for estimates and distribution of work within the team. The team member may participate in activities (estimates, task assignment)
		1.11.2. To give suggestion and propose parts that can be taken for execution
		1.11.3. Employee is given much freedom to initiate activities where they find relevant knowledge for their project, and they may often be asked to review one or a few technologies for potential use and report on that
	1.12. Training	1.12.1. Time agile training
		1.12.2. Frequent technical training
		1.12.3. Knowledge transfer before coding
2. Process agility	2.1. Feedback	2.1.1. 360° feedback
		2.1.2. Retrospective meetings
		2.1.3. Continuous feedback
		2.1.4. Demo before the release
	2.2. Quality metrics	2.2.1. Number of defects/bugs/trouble reports, defects per KLOC/defect density, number/percentage of acceptance/black-box/external tests passed, number/percentage of unit tests passed, quality mark given by the client
		2.2.2. Mean time assertion per method, total assertions (passed/failed), number of test cases passed, number of passed acceptance tests or the total number of defects or number of defects/KLOC (defect density). The number of defects found before release or defects reported by customers
	2.3. Requirements	Using stories to define requirements 2.3.1. Delivering working software continuously

(continued)

Table 1 (continued)

Enablers	Criteria	Attributes
		2.3.2. Strong input—narrative, have a minimum set of rules that are very important
		2.3.3. User stories must be small, testable, independent, negotiable, valuable, the estimate will have the narrative
		2.3.4. Incremental design
		2.3.5. Measurements: min, hrs, days, months/T-shirt size, poker cards (Fibonacci series, take numbers with a difference to keenly know the difference in importance of stories)
		2.3.6. The effort required from the development team to understand the software requirements specification, with or without the presence of the client must be less
	2.4. Releasing project in iterations	2.4.1. Doing things when they have to be done, not before
		2.4.2. Doing iterative and incremental development
		2.4.3. Making short software releases
		2.4.4. Delivering working software continuously
	2.5. Process metrics	2.5.1. Less hedging, less blaming, less interruption
		2.5.2. Balanced team productivity
		2.5.3. Fast execution of ideas
		2.5.4. Max real value added
		2.5.5. Min non-value added
		2.5.6. Less defect
		2.5.7. How many features being released
	2.6. Before	2.6.1. Estimate bug fixes and add these estimates into the backlogs
		2.6.2. YAGNI—add functions as little pieces, which are necessary. Check how much to plan and implement now and are we implementing it?

(continued)

Table 1 (continued)

Enablers	Criteria	Attributes
		2.6.3. Building a story from the perspective of problems
		2.6.4. Vertical slicing which can be done within a day or two
		2.6.5. Define ready
		2.6.6. Define done
	2.7. During	2.7.1. Measure the continuous improvement
		2.7.2. Periodic checks of what is working and what is not
		2.7.3. Unit testing before functional testing
		2.7.4. Limit WIP size
		2.7.5. Give more focus on areas where you are predicting defect or is a heavy code continuous integration (versioning, merging of code, automation, a batch process)
	2.8. After	2.8.1. After launch functions: check with the user
		2.8.2. Zero day-usability check
		2.8.3. 30th day: are they using it?
		2.8.4. 90th day: are they getting what they asked for?
		2.8.5. Five-why analysis identifying causes of problems and taking actions to prevent them in the future
		2.8.6. Sprint review meeting
		2.8.7. Retrospection
3. Engineering agility	3.1. Information radiator	3.1.1. User story mapping board: to have the top line narrative and priorities in the timeline
		3.1.2. Burn-down charts: have granular details and keep work visible and understandable throughout the process. It can be used for a huge task, but using it for exploratory tasks is not that good
		3.1.3. Improvement board: have improvement theme: now, next target and benchmark
	3.2. Agile coding	3.2.1. Doing code refactoring

(continued)

Table 1 (continued)

Enablers	Criteria	Attributes
		3.2.2. Make code light by removing dysfunctional elements in the code
		3.2.3. Reusability
		3.2.4. Self-encapsulated codes
		3.2.5. Modular codes
		3.2.6. Limited blast radius—decoupled architecture; error means will affect only a specific spot
		3.2.7. Managing software configuration (version control)
		3.2.8. A close check on technical debt
	3.3. Agile testing	3.3.1. Automated test suites
		3.3.2. Make sure the code is tested well with all sample possible data
		3.3.3. Unit testing
		3.3.4. Functional testing
		3.3.5. Usability testing
	3.4. Tools	3.4.1. Managing projects with measures and analytic techniques
		3.4.2. Agile life cycle managing tool
		3.4.3. Automation tool
		3.4.4. Tracking tool
		3.4.5. Continuous integration tool
		3.4.6. Source control tools
4. Business agility	4.1. Company culture	4.1.1. Agility supported by top management (from plan-driven to agile)
		4.1.2. Criticality of project
		4.1.3. Close customer collaboration
		4.1.4. Agility supported by the customer
		4.1.5. Developing people's agility skills
		4.1.6. Align agility measure to performance and compensation
		4.1.7. Level of risk-taking willingness

(continued)

Table 1 (continued)

Enablers	Criteria	Attributes
		4.1.8. Responding to change positively
		4.1.9. Self-organizing culture
		4.1.10. Changeable organization, changeable architecture, changeable process. A healthy culture that heals the broken process
		4.1.11. Experimenting culture
	4.2. Core values	4.2.1. Focus on the innovation more than the prediction
		4.2.2. Focus on team trust more than team control
		4.2.3. Failure recovery is preferred than failure avoidance
		4.2.4. Scope for continuous learning for team members

Primary level assessment

Let's take the criteria team orientation, it consists of two attributes.

Weights corresponding to the criterion:

$$\mathrm{W11} = [0.476, 0.524]$$

Rating vector corresponding the criteria team orientation as given:

$$R_{11} = \begin{bmatrix} 8 & 7 & 9 & 6 & 9 \\ 9 & 8 & 9 & 9 & 9 \end{bmatrix}$$

The corresponding criterion agility index is [25, 26]

$$\boldsymbol{W}_{11} \times \boldsymbol{R}_{11} = \boldsymbol{I}_{11}$$

$I_{11} = [8.52, 7.52, 9, 7.57, 9]$

Similarly, the agility index corresponding to the remaining criteria have been calculated:

$I_{12} = [7.45, 8.05, 9.02, 8.05, 8.05]$
$I_{13} = [8.52, 7.7, 8.11, 8.95, 8.88]$
$I_{14} = [7.25, 7.25, 7.73, 7.96, 6.23]$
$I_{15} = [8.51, 7.55, 7.36, 6.21, 5.73]$
$I_{16} = [8.22, 7.66, 8.66, 7.57, 6.95]$
$I_{17} = [8.6, 8, 8.22, 9.04, 7.26]$

Table 2 Weightage and ratings of enablers, criteria and attributes

I_i	I_{ij}	I_{ijk}	E_1	E_2	E_3	E_4	E_5	W_{ijk}	W_{ij}	W_i
1	1.1	1.1.1	8	7	9	6	9	0.4761	0.0731	0.252
		1.1.2	9	8	9	9	9	0.5238		
	1.2	1.2.1	6	9	10	10	10	0.2765	0.0871	
		1.2.2	8	7	9	8	8	0.2234		
		1.2.3	9	8	8	7	7	0.2553		
		1.2.4	7	8	9	7	7	0.2446		
	1.3	1.3.1	7	8	7	9	8	0.1157	0.0836	
		1.3.2	9	8	7	10	10	0.1157		
		1.3.3	7	7	7	8	8	0.1157		
		1.3.4	9	7	8	10	10	0.0972		
		1.3.5	8	9	10	10	10	0.1296		
		1.3.6	9	8	9	6	9	0.0972		
		1.3.7	9	7	9	9	9	0.1064		
		1.3.8	9	7	8	8	8	0.1018		
		1.3.9	10	8	8	10	8	0.1203		
	1.4	1.4.1	7	7	7	7	6	0.2692	0.0836	
		1.4.2	8	8	8	8	6	0.2596		
		1.4.3	8	7	8	9	7	0.2307		
		1.4.4	6	7	8	8	6	0.2403		
	1.5	1.5.1	9	7	7	5	5	0.25	0.0836	
		1.5.2	10	8	8	7	7	0.2738		
		1.5.3	9	8	8	8	5	0.2857		
		1.5.4	5	7	6	4	6	0.1904		
	1.6	1.6.1	8	7	8	5	6	0.3382	0.0836	
		1.6.2	10	8	9	10	8	0.2941		
		1.6.3	7	8	9	8	7	0.3676		
	1.7	1.7.1	8	8	10	10	8	0.2047	0.0801	
		1.7.2	9	8	8	9	7	0.1968		
		1.7.3	9	8	8	10	9	0.2204		
		1.7.4	8	8	8	8	5	0.1889		
		1.7.5	9	8	7	8	7	0.1889		
	1.8	1.8.1	9	8	8	8	6	0.2475	0.0836	
		1.8.2	8	8	8	9	7	0.2079		
		1.8.3	8	8	9	9	8	0.2772		
		1.8.4	9	8	8	7	9	0.2673		
	1.9	1.9.1	5	8	7	8	5	0.4893	0.0871	

(continued)

Table 2 (continued)

I_i	I_{ij}	I_{ijk}	E_1	E_2	E_3	E_4	E_5	W_{ijk}	W_{ij}	W_i
		1.9.2	9	9	8	7	5	0.5106		
	1.10	1.10.1	10	9	9	10	10	0.5	0.0871	
		1.10.2	9	8	8	10	7	0.5		
	1.11	1.11.1	9	9	9	10	7	0.3235	0.0871	
		1.11.2	9	9	8	8	5	0.3382		
		1.11.3	8	9	8	8	5	0.3382		
	1.12	1.12.1	10	8	8	10	5	0.3417	0.0801	
		1.12.2	7	6	7	4	2	0.3291		
		1.12.3	9	9	7	8	5	0.3291		
2	2.1	2.1.1	8	8	7	7	5	0.28	0.1354	0.252
		2.1.2	10	9	9	10	10	0.24		
		2.1.3	9	8	10	10	8	0.22		
		2.1.4	10	9	9	10	9	0.26		
	2.2	2.2.1	5	8	9	10	8	0.4883	0.125	
		2.2.2	9	8	9	7	7	0.5116		
	2.3	2.3.1	9	8	8	10	8	0.1780	0.1302	
		2.3.2	8	8	8	8	6	0.1780		
		2.3.3	9	8	9	10	8	0.1643		
		2.3.4	10	8	9	8	8	0.1712		
		2.3.5	9	8	9	10	7	0.1438		
		2.3.6	9	8	8	9	4	0.1643		
	2.4	2.4.1	8	8	9	7	8	0.2209	0.125	
		2.4.2	8	8	8	8	9	0.2674		
		2.4.3	8	8	7	4	5	0.2325		
		2.4.4	9	9	9	8	9	0.2790		
	2.5	2.5.1	7	9	8	6	8	0.1474	0.125	
		2.5.2	9	9	8	7	8	0.1538		
		2.5.3	8	9	8	4	5	0.1474		
		2.5.4	8	7	8	5	5	0.1538		
		2.5.5	8	7	7	5	5	0.1282		
		2.5.6	7	7	6	8	3	0.1217		
		2.5.7	8	9	8	8	8	0.1474		
	2.6	2.6.1	9	9	9	10	10	0.1691	0.1302	
		2.6.2	9	9	9	5	8	0.1617		
		2.6.3	9	8	8	8	9	0.1691		
		2.6.4	8	8	6	4	5	0.1470		

(continued)

Table 2 (continued)

I_i	I_{ij}	I_{ijk}	E_1	E_2	E_3	E_4	E_5	W_{ijk}	W_{ij}	W_i
		2.6.5	10	9	8	9	7	0.1764		
		2.6.6	10	9	8	9	8	0.1764		
	2.7	2.7.1	8	8	8	8	8	0.2051	0.125	
		2.7.2	8	8	9	9	8	0.2051		
		2.7.3	8	8	8	8	10	0.2136		
		2.7.4	8	8	6	5	7	0.1794		
		2.7.5	9	8	9	6	7	0.1965		
	2.8	2.8.1	9	7	6	1	6	0.1558	0.1041	
		2.8.2	10	7	9	1	6	0.1558		
		2.8.3	9	7	8	1	5	0.1298		
		2.8.4	9	7	9	1	8	0.1233		
		2.8.5	8	7	10	1	8	0.1558		
		2.8.6	10	9	9	8	9	0.1428		
		2.8.7	10	9	9	8	9	0.1363		
3	3.1	3.1.1	8	9	9	7	8	0.3709	0.2421	0.242
		3.1.2	9	9	9	8	8	0.3064		
		3.1.3	8	8	7	9	7	0.3225		
	3.2	3.2.1	8	8	8	8	5	0.12	0.2631	
		3.2.2	5	8	8	9	8	0.135		
		3.2.3	9	8	8	8	8	0.125		
		3.2.4	8	8	8	8	8	0.125		
		3.2.5	8	8	8	8	8	0.12		
		3.2.6	7	8	9	7	8	0.125		
		3.2.7	10	8	9	8	9	0.125		
		3.2.8	8	8	8	6	9	0.125		
	3.3	3.3.1	4	8	9	9	9	0.2	0.2631	
		3.3.2	8	8	9	9	4	0.2083		
		3.3.3	8	8	9	10	9	0.1916		
		3.3.4	9	10	9	10	9	0.2083		
		3.3.5	9	8	7	6	9	0.1916		
	3.4	3.4.1	8	8	8	5	8	0.1631	0.2315	
		3.4.2	7	8	7	1	5	0.1560		
		3.4.3	5	9	3	1	4	0.1773		
		3.4.4	9	8	5	8	4	0.1631		
		3.4.5	9	8	5	6	5	0.1702		
		3.4.6	9	8	7	9	5	0.1702		

(continued)

Table 2 (continued)

I_i	I_{ij}	I_{ijk}	E_1	E_2	E_3	E_4	E_5	W_{ijk}	W_{ij}	W_i
4	4.1	4.1.1	9	8	7	8	8	0.0915	0.5	0.252
		4.1.2	10	8	8	6	9	0.0952		
		4.1.3	9	9	9	8	9	0.0952		
		4.1.4	8	9	9	7	9	0.0952		
		4.1.5	8	8	9	8	5	0.0952		
		4.1.6	8	9	8	2	7	0.0952		
		4.1.7	6	9	7	4	3	0.0952		
		4.1.8	7	9	8	5	5	0.0805		
		4.1.9	8	9	8	6	5	0.0879		
		4.1.10	9	8	8	5	3	0.0879		
		4.1.11	8	9	7	1	5	0.0805		
	4.2	4.2.1	6	8	7	3	5	0.2688	0.5	
		4.2.2	8	9	7	6	9	0.2688		
		4.2.3	8	9	7	6	5	0.2150		
		4.2.4	8	9	8	8	4	0.2473		

$I_{18} = [8.51, 8, 8.27, 8.21, 7.56]$
$I_{19} = [7.04, 8.51, 7.51, 7.48, 5]$
$I_{110} = [9.5, 8.5, 8.5, 10, 8.5]$
$I_{111} = [8.66, 9, 8.32, 8.64, 5.64]$
$I_{112} = [8.68, 7.67, 7.34, 7.36, 4.01]$

Secondary level assessment

The calculation corresponding to the 'team agility' enabler is shown below.

Weights corresponding to the team agility enabler:

$$W_1 = [0.073, 0.087, 0.083, 0.083, 0.083, 0.083, 0.08, 0.083, 0.087, 0.087, 0.087, 0.08]$$

Rating vector corresponding to the team agility enabler:

$$R_1 = \begin{bmatrix} 8.52 & 7.52 & 9 & 7.57 & 9 \\ 7.45 & 8.05 & 9.02 & 8.05 & 8.05 \\ 8.52 & 7.7 & 8.11 & 8.95 & 8.88 \\ 7.25 & 7.25 & 7.73 & 7.96 & 6.23 \\ 8.51 & 7.55 & 7.36 & 6.21 & 5.73 \\ 8.22 & 7.66 & 8.66 & 7.57 & 6.95 \\ 8.6 & 8 & 8.22 & 9.04 & 7.26 \\ 8.51 & 8 & 8.27 & 8.21 & 7.56 \\ 7.04 & 8.51 & 7.51 & 7.48 & 5 \\ 9.5 & 8.5 & 8.5 & 10 & 8.5 \\ 8.66 & 9 & 8.32 & 8.64 & 5.64 \\ 8.68 & 7.67 & 7.34 & 7.36 & 4.01 \end{bmatrix}$$

The corresponding enabler agility index is

$$\boldsymbol{W}_1 \times \boldsymbol{R}_1 = \boldsymbol{I}_1$$

$I_{1=}$ [8.28, 7.96, 8.16, 8.10, 6.89]

Similarly, agility index corresponding to the remaining enablers have been calculated:

$I_{2=}$ [8.51, 8.16, 8.33, 7.31, 7.43]
$I_{3=}$ [7.88, 8.31, 7.80, 7.44, 7.22]
$I_{4=}$ [7.82, 8.68, 7.63, 5.60, 6.03]

Tertiary level assessment

The comprehensive score for the overall agility index is shown below:

$$\text{Overall weight} : W = [0.252, 0.252, 0.242, 0.252]$$

Overall rating vector: [27–29]

$$R = \begin{bmatrix} 8.28 & 7.96 & 8.16 & 8.10 & 6.89 \\ 8.51 & 8.16 & 8.33 & 7.31 & 7.43 \\ 7.88 & 8.31 & 7.80 & 7.44 & 7.22 \\ 7.82 & 8.68 & 7.63 & 5.60 & 6.03 \end{bmatrix}$$

$$\boldsymbol{W} \times \boldsymbol{R} = \boldsymbol{I}$$

Therefore, the overall agility index is.

$$I = [8.12, 8.28, 7.98, 7.11, 6.89]$$
$$I = (8.12 + 8.28 + 7.98 + 7.11 + 6.89)/5$$
$$I = 7.68$$

Agility index is 7.68 as, the agility index falls between 6 and 8, the SDP is said to be 'Agile.'

3.3 Identification of the Weaker Attributes

As per the management need, threshold level has to be set to find the attributes which need to be addressed to improve the agility level of SDP. In this case, the management threshold is 6.5. Now, we need to find the mean rating of attributes which are below 6.5. There are 11 attributes which are below 6.5 as per the average of ratings provided by the experts. They are frequent technical training, making short software releases, min non-value added, less defect, vertical slicing which can be done within a day or two, after launch functions: check with the user, 30th day: are they using it?, agile life cycle managing tool, automation tool, experimenting culture, focus on the innovation more than the prediction.

Suggestive measures can be provided on the basis of criticality of the attribute based on the industry need and the expert weight. In this case, for example, according to expert's review, frequent technical training has higher criticality. Agility means to move quickly and easily this is possible only when the team is prepared to adapt to the new requirements. Thus, the learning and development team must rate the competency of the team members and prepare action plans according to need. Automation helps in quick delivery of requirements; thus, in-house tools can be developed to automate non-value adding activities as a part of the ad-hoc activity in each sprint. This might reduce defects and limits WIP due to short span software development.

4 Results and Discussions

Agility level of the case SDP was accessed. Its agility turned out to be 7.68. Among the 123 attributes around 11 attributes were below the management threshold of 6.5. This assessment tool covers most of the attributes required for the high performance of an agile SDP. This study provides an assessment tool which is generalized for any SDP using agile methodology at any level. The managers can implement this assessment tool to check the agility level of any SDP. Depending upon the level in the agile adoption the needed level of threshold can be chosen by the managers.

5 Conclusion

According to [30], Masters of agile + DevOps have attained 60% higher revenue and profit growth. Delta Matrix study performed on 8000 samples states that agile teams are 25% more productive than non-agile teams [31]. The assessment of agility in SDP is gaining vital significance. In this study, a new unique framework has been developed for assessing the agility level in SDP. The agility was calculated for the SDP based on the observations by experts from case SDP. Then, further improvements have been were suggested for implementing agile practices in the SDP. The performance of a team not only depends on the team agility it is also interconnected to overall agility of the organization. Thus, this paper categories all the enablers contributing to the agility of SDP, viz. team agility, process agility, engineering agility, and business agility. This study provides a generalized assessment tool to assess the agility level of any SDP. Further, this method could be used for benchmarking and periodical evaluation of the agility assessment in SDP. The present study has some limitations; the proposed framework has been validated for only one case organization in India. Thus, in future, more case studies should be investigated in different SDP in various geographic regions.

References

1. E.M. Schön, J. Thomaschewski, M.J. Escalona, Agile requirements engineering: A systematic literature review. Comput. Stand. Interfaces **49**, 79–91 (2017)
2. J. Highsmith, *Agile Project Management: Creating Innovative Products* (Pearson education, 2009)
3. N.B. Moe, T. Dingsøyr, T. Dybå, A teamwork model for understanding an agile team: A case study of a Scrum project. Inf. Softw. Technol. **52**(5), 480–491 (2010)
4. D. Strode, Applying Adapted big five teamwork theory to agile software development, in *Australasian Conference on Information Systems, Adelaide* (2015). https://arxiv.org/pdf/1606.03549. Accessed 23 Mar 2019
5. D.E. Strode, S.L. Huff, B. Hope, S. Link, Coordination in co-located agile software development projects. J. Syst. Softw. **85**(6), 1222–1238 (2012)
6. R.M. Fontana, I.M. Fontana, P.A. da Rosa Garbuio, S. Reinehr, A. Malucelli, Processes versus people: How should agile software development maturity be defined? J. Syst. Softw. **97**, 140–155 (2014)
7. J. Sheffield, J. Lemétayer, Factors associated with the software development agility of successful projects. Int. J. Project Manage. **31**(3), 459–472 (2013)
8. K. Curcio, T. Navarro, A. Malucelli, S. Reinehr, Requirements engineering: A systematic mapping study in agile software development. J. Syst. Softw. **139**, 32–50 (2018)
9. J. Medeiros, A. Vasconcelos, C. Silva, M. Goulão, Quality of software requirements specification in agile projects: A cross-case analysis of six companies. J. Syst. Softw. **142**, 171–194 (2018)
10. B. Tessem, Individual empowerment of agile and non-agile software developers in small teams. Inf. Softw. Technol. **56**(8), 873–889 (2014)
11. Y. Lindsjørn, D.I. Sjøberg, T. Dingsøyr, G.R. Bergersen, T. Dybå, Teamwork quality and project success in software development: A survey of agile development teams. J. Syst. Softw. **122**, 274–286 (2016)

12. D.K. Rigby, J. Sutherland, H. Takeuchi, Embracing agile. Harv. Bus. Rev. **94**(5), 40–50 (2016)
13. T.J. Gandomani, M.Z. Nafchi, Agile transition and adoption human-related challenges and issues: A grounded theory approach. Comput. Hum. Behav. **62**, 257–266 (2016)
14. I. Nurdiani, J. Börstler, S.A. Fricker, The impacts of agile and lean practices on project constraints: A tertiary study. J. Syst. Softw. **119**, 162–183 (2016)
15. Y. Shastri, R. Hoda, R. Amor, The role of the project manager in agile software development projects. J. Syst. Softw. **173**, 110871 (2021)
16. S. Vinodh, S.R. Devadasan, B. Vasudeva Reddy, K. Ravichand, Agility index measurement using multi-grade fuzzy approach integrated in a 20 criteria agile model. Int. J. Prod. Res. **48**(23), 7159–7176 (2010)
17. S. Vinodh, S.K. Chintha, Leanness assessment using multi-grade fuzzy approach. Int. J. Prod. Res. **49**(2), 431–445 (2011)
18. S. Vinodh, U.R. Madhyasta, T. Praveen, Scoring and multi-grade fuzzy assessment of agility in an Indian electric automotive car manufacturing organisation. Int. J. Prod. Res. **50**(3), 647–660 (2012)
19. M. Suresh, K. Gopakumar, Multi-grade fuzzy assessment framework for software professionals in work-from-home mode during and post-COVID-19 era. Future Bus. J. **7**(1), 1–9 (2021)
20. A. Akhil, M. Suresh, Assessment of service quality in restaurant using multi-grade fuzzy and importance performance analysis. Mater. Today: Proc. (2021). https://doi.org/10.1016/j.matpr.2021.01.767
21. E. Chacko, M. Suresh, Assessment of start-up agility using multi-grade fuzzy and importance performance analysis, in *Advances in Materials Research* (Springer, Singapore, 2021), pp. 685–694
22. S. Vinodh, Assessment of sustainability using multi-grade fuzzy approach. Clean Technol. Environ. Policy **13**(3), 509–515 (2011)
23. K.E.K. Vimal, S. Vinodh, R. Muralidharan, An approach for evaluation of process sustainability using multi-grade fuzzy method. Int. J. Sustain. Eng. **8**(1), 40–54 (2015)
24. S. Vinodh, M. Prasanna, Evaluation of agility in supply chains using multi-grade fuzzy approach. Int. J. Prod. Res. **49**(17), 5263–5276 (2011)
25. V. Vaishnavi, M. Suresh, Assessment of leagility in healthcare organization using multi-grade fuzzy approach, in *Data Intelligence and Cognitive Informatics* (Springer, Singapore, 2021), pp. 409–421
26. E. Chacko, M. Suresh, S.L. Priyadarsini, Start-Up leagility assessment using multi-grade fuzzy and importance performance analysis, in *Data Intelligence and Cognitive Informatics* (Springer, Singapore, 2021), pp. 397–407
27. S. Sreedharshini, M. Suresh, S.L. Priyadarsini, Workplace stress assessment of software employees using multi-grade fuzzy and importance performance analysis, in *Data Intelligence and Cognitive Informatics* (Springer, Singapore, 2021), pp. 433–443
28. M. Anil, M. Suresh, Assessment of service agility in power distribution company, in *IOP Conference Series: Materials Science and Engineering* vol 954, No 1 (IOP Publishing, 2020), p. 012010
29. M. Sivaraman, M. Suresh, R. Ranganathan, Assessment of stress and safety practice level of heavy commercial passenger vehicle drivers using multi-grade fuzzy. Mater. Today: Proc. (2021). https://doi.org/10.1016/j.matpr.2021.01.859
30. S. Panditi, Survey data shows that many companies are still not truly agile. Harvard Bus. Rev. (2018). https://hbr.org/sponsored/2018/03/survey-data-shows-that-many-companies-are-still-not-truly-agile. Assessed on 23 April 2019
31. N. Kremic, Why are agile teams 25% more productive? (2017). https://www.deltamatrix.com/why-are-agile-teams-25-more-productive/. Accessed on 23 Mar 2019

Hard Exudates Detection for Diabetic Retinopathy Early Diagnosis Using Deep Learning

P. Leela Jancy, A. Lazha, R. Prabha, S. Sridevi, and T. Thenmozhi

Abstract Diabetic retinopathy is the complication of the eye caused due to diabetes mellitus. It is caused due to high glucose level in blood which damages the blood vessels at the back of the eye namely retina. The abnormal blood vessels swell and leak into retina. The microstructures such as microaneurysm, exudates (hard/soft) will occupy the retina area which leads to vision threatening. Nowadays, in medical field, computer-aided systems are performing a promising work in terms of object recognizing, localization, classification, segmentation and analysis of images with the help of deep neural networks. Since diabetic retinopathy is irreversible, detection of early stages of diabetic retinopathy is needed. If it left without proper diagnosis and treatment, it will lead to vision loss. Detection of hard exudates indicates the presence of diabetic retinopathy. The early sign of diabetic retinopathy is the formation of hard exudates in retina. In this method, the main aim is to detect the presence of hard exudates in the retinal image using deep convolution neural network.

Keywords Diabetes mellitus · Classification · Convolution neural networks (CNN) · Diabetic retinopathy (DR) · Hard exudates · Microaneurysm · Deep learning

P. Leela Jancy (✉) · R. Prabha · T. Thenmozhi
Sri SaiRam Institute of Technology, Chennai, TamilNadu, India
e-mail: leela.it@sairamit.edu.in

T. Thenmozhi
e-mail: thenmozhi.eee@sairamit.edu.in

A. Lazha
Sri SaiRam Siddha Medical College & Research Centre, Chennai, TamilNadu, India

S. Sridevi
Vels Institute of Science, Technology & Advanced Studies, Chennai, TamilNadu, India
e-mail: sridevis.se@velsuniv.ac.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_22

1 Introduction

Diabetes mellitus is caused due to the presence of high glucose level in blood. Saeedi et al. [1] estimated that diabetes-affected people as 463 million people which is expected to increase by 578 million by 2030 and by 700 million by 2045 worldwide. Diabetic patient needs to be aware of eye diseases. Diabetic eye diseases are a group of problems that are affecting the eyes of diabetic patients. Diabetic retinopathy is caused due to the presence of high blood sugar in patients with diabetics of type II. This high blood sugar damages the eye (retina). Retina detects the light and sends signal through optic nerves to the brain. Earlier studies focused mainly on fundus images, but early stages of DR seem to be a challenging task.

1.1 Diabetic Retinopathy

Diabetic retinopathy is caused due to the presence of high blood sugar in patients with diabetics of type II. This high blood sugar damages the eye (retina). Retina detects the light and send signal through optic nerves to the brain. In order to prevent the vision loss and to prescribe correct medicine, it is a must to classify the severity level of diabetic retinopathy accurately. If not, it will lead to vision loss which is not reversible. Diabetic retinopathy can be classified into two types namely (i) NPDR—non-proliferative diabetic retinopathy comprises three stages, namely mild, moderate and severe. (ii) PDR—proliferative diabetic retinopathy.

1.1.1 Mild-Level NPDR

The patients with mild NPDR have at the minimum one microaneurysm as in Fig.1. Hence, diabetic patients need routine check-up and precise monitoring because the mild level turns into moderate level within one year.

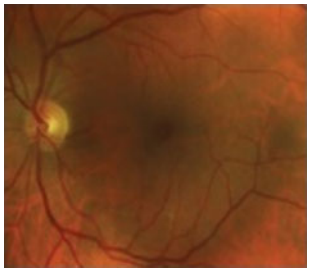


Fig. 1 Mild-level NPDR

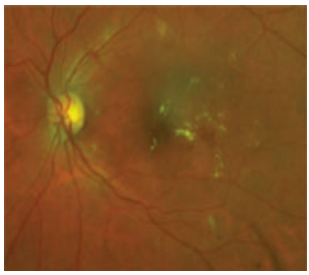


Fig. 2 Moderate-level NPDR

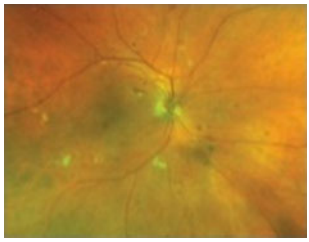


Fig. 3 Severe level NPDR

1.1.2 Moderate-Level NPDR

The patients have microaneurysms in more than one quadrant of retina. Along with/without the following signs—venous beading, hard exudates as in Fig.2.

1.1.3 Severe-Level NPDR

The patient will have more than 20 number of intraretinal bleeding and microaneurysms in each of four quadrants along with venous beading in more than 1 quadrant as in Fig.3.

1.1.4 Proliferative Diabetic Retinopathy

Many numbers of loop of new vessels are located in the optical disk. These vessels will leak which results in retinal edema which also bleeds as in Fig 4.

1.2 Hard Exudates

Small white deposits having sharp edges either yellow in color or whitish in color are called hard exudates. They can be detected at the outer layer of retina. They are formed by the leakage of extracellular lipid that flow from damaged capillaries.

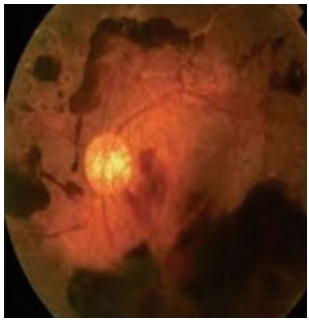


Fig. 4 Proliferative diabetic retinopathy

Accumulation of lipid in/and around retina is diagnosed as hard exudates. Fluids leak from the blood vessels, and it looks as like that of blood plasma in appearance. Hard exudates can be seen initially on the outer plexiform layer of the retina. The main cause of hard exudates is diabetic mellitus. Diabetic retinopathy leads to vision threatening problems and leads to vision loss. Hence, it is the important one to detect the hard exudates which indicates the presence of DR. Their size varies from tiny to larger patches and appears in shapes like rings, stars or plates.

1.3 Convolution Neural Network (CNN)

Convolution neural network (CNN) is a deep learning algorithm mainly used in applications like image processing, segmentation and classification. Manual feature extraction is not needed in convolution neural network. Each layer of CNN extract features using set of filters. The first layer of CNN extracts basic features like horizontal, vertical line from the input image. The second layers extract features like corners, edges from the input image. The hidden layers extract the features and pass the information to the next layer. The last layers of CNN extract the object/patterns from the input image. CNN is useful for detection, segmentation, classification and analysis of medical images. Convolution neural network consists of many layers like convolution layer, pooling layer, drop out layer, regularization layer, flattening layer and fully connected layers. The output from each layer (the activation map) can be input into the next layer. The accuracy of the model is defined as the ratio of total number of images predicted correctly to the total number of images predicted. For classification of images, fully connected network is in the last layer to predict the images either hard exudate image or normal image without hard exudate.

2 Related Works

Earlier works regarding hard exudates were proposed by many researchers. Ladkat et al. [2] proposed a method to differentiate between exudate and non-exudate pixels using intensity level and by tuning matched filters. Palavalasa and Sambaturu [3] proposed a method to detect hard exudates by using two methods, namely background subtraction method and decorrelation stretch-based method. Bharkad [4] proposed a method to detect hard exudates by removing the optical disk so as to avoid wrong classification and features are extracted from green component image. Samah et al. [5] proposed a method to classify three pathologies, namely exudates, hemorrhage, microaneurysm using image enhancement techniques and CNN. Rahi et al. [6] proposed an automatic system for the detection of skin cancer using deep neural networks. Abdar et al. [7] proposed an automatic detection of COVID-19-infected lungs using deep learning-based CNN. Ali and Kumar [8] proposed a model to classify X-ray images using deep CNN to detect pneumonia-affected persons using inception-V3 model. Zhang et al. [9] proposed attention residual learning convolutional neural network (ARL-CNN) to classify skin lesions. Ahmad et al. [10] proposed a method to classify skin diseases using deep CNN by fine-tuning layers of ResNet152 and InceptionResNet-V2 models using triplet loss function. Maya and Adarsh [11] proposed a method to classify the grading of DR severity by removing optical disk, blood vessels and by recognizing hard exudates. Albahar [12] proposed a method using deep CNN to classify the skin lesions. Yu et al. [13] proposed a method that uses four models to classify between normal cell samples and abnormal cells samples. Tiwari et al. [14] proposed the classification of living and non-living images using VGG-16 model. Peng Tang, Qiaokang Liang, Xintong Yan, Shao Xiang and Dan Zhang[15] classified skin lesions using Global-Part CNN. Prabha et al. [16] reviewed machine learning based classification algorithm for medical IOT. Deep learning out performs by extracting the features automatically. No manual interpretation is needed in deep learning algorithms.

3 Materials and Method

Deep learning is one of the promising areas in the field of medical imaging, image segmentation and analysis in research. Deep learning uses convolution neural network to extract patterns or objects from the input image. Convolution neural networks accept input and provide activation map which exhibits the relevant features of the image. The neurons in convolution neural network takes patch pixels as input and multiplies their color values with weights and sum them and executes them with activation function. Deep learning is one of the promising areas in the field of medical imaging, image segmentation and analysis in research.

Deep learning uses convolution neural network to extract patterns or objects from the input image. Convolution neural networks accept input and provide activation

map which exhibits the relevant features of the image. The neurons in convolution neural network takes patch pixels as input and multiplies their color values with weights and sum them and executes them with activation function. This implementation method make use of three stages: First stage describes about the dataset used. Second stage describes about methodology. Third stage describes about the image classification using CNN implementation, and the performance of the convolution neural network is evaluated using accuracy.

3.1 Dataset

The images are taken from public database DIARETDB1 which is used as a benchmark for diabetic retinopathy detection from digital images 89 color fundus are there. Samples from DIARETDB1 database is shown in Fig. 5.

The dataset consists of fundus retinal images, ground truth images and fundus mask as in Fig. 6.

The database contains 84 images that shows mild non-proliferative symptoms, and remaining five images do not have any symptoms and they are normal images as in Table 1.

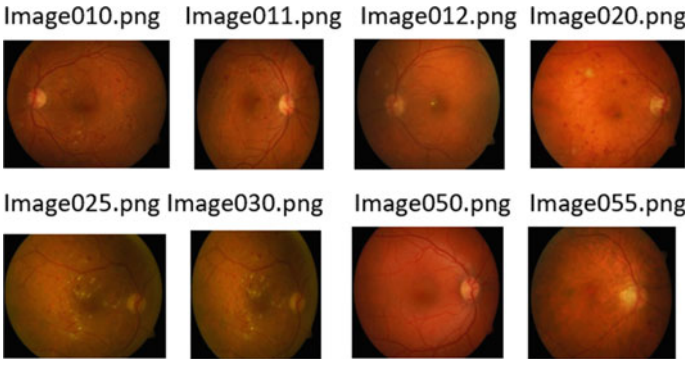


Fig. 5 Sample image of DIARETDB1database

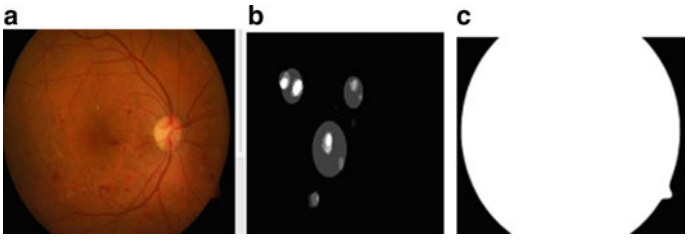


Fig. 6 **a** Original image, **b** hard exudates ground truth image, **c** fundus Mask

Table 1 Dataset details

DR diagnosis	Count	Total
Fundus retinal images with DR symptoms	84	89
Normal/Fundus retinal images without DR symptoms	5	

These color fundus images are also independently marked by four experts. The database consists of color fundus images showing the symptoms that occur for diabetic retinopathy.

3.2 Methodology

The data set itself consists of fundus images and ground truth images with a resolution of 1500 × 1152 pixels. The hard exudates images are identified using ground truth images. Pixels belonging to hard exudates images will have an intensity of 1 and remaining pixels will have an intensity 0. Two folders namely training set with 848 images, and test set with 132 images are created in the working directory. Training set is divided into two classes, namely class 1 (with hard exudate images) and class 0 (normal images) are created and each class consists of 424 images each. Likewise, test set is divided into two classes, namely class 1 (with hard exudate images) and class 0 (normal images) are created each class consists of 66 images each. Initially, input images are preprocessed. Since the dataset is an imbalance dataset (84 with mild DR symptoms and 5 normal images), data augmentation is needed. Data augmentation operations performs normalization, ZCA whitening and operations which resize and reshape the images including shift, rotation, zoom and flipping. The data augmentation creates more images. Data-augmented input images are fed into CNN model. Data augmentation improves the performance of the model.

3.3 CNN Implementation

The model used here is Keras API sequential model. The architecture of our model is shown in Fig. 7. The CNN model consists of six convolution layers. The first convolution layer uses 32 filters followed by 2 layers with 64 filters followed by layer with 128 filters, 256 filters and 512 filters for learning the features. These kernel filters are of 2D and are applied over the images. Rectified linear activation unit (ReLU) is used. ReLU return the input value if it receives a positive input, otherwise it returns zero; it can be written as $f(x) = \text{Max}(0, x)$ where $x > 0$. Pooling layer which performs down sampling follows the convolution layer. Here in this model, max pooling is used. Max pooling layer calculates the largest value of the image patch. Pooling layer summarizes the features of image patches. Convolution

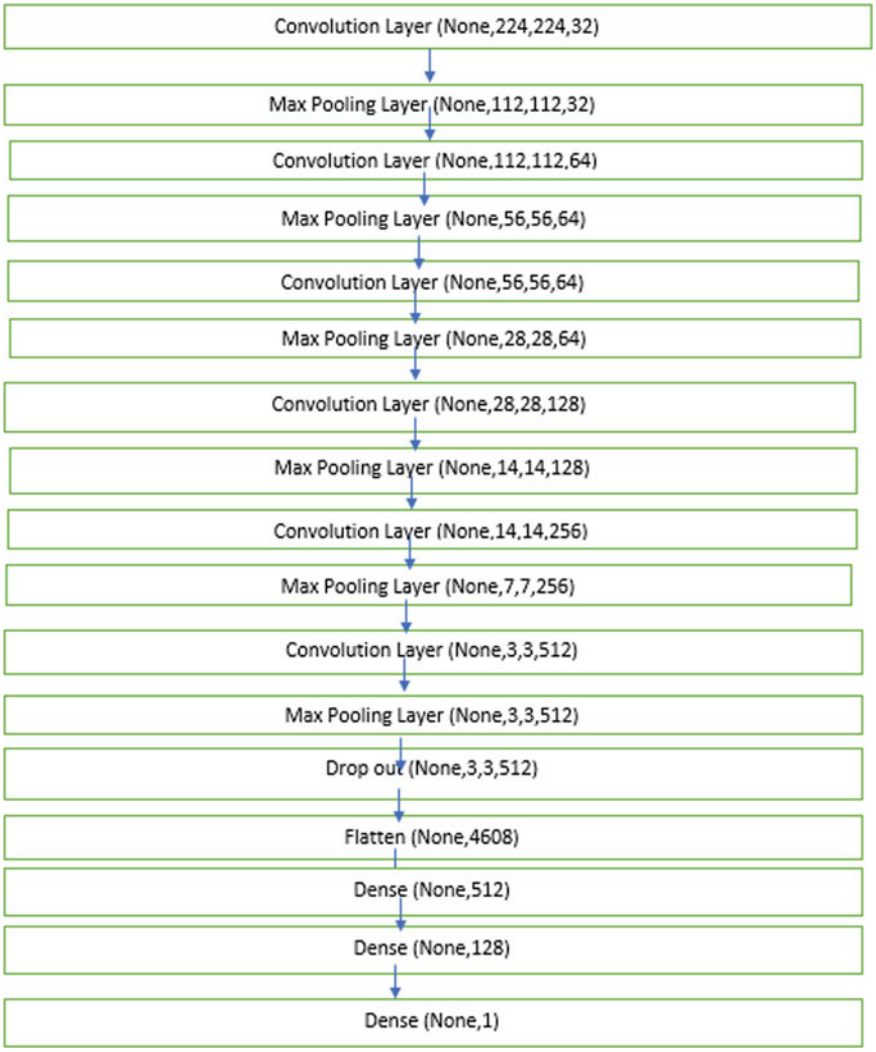


Fig. 7 CNN architecture model for hard exudates detection

and pooling layers learn the features from the input image patches. Drop-out layer is added at the end of the model. Drop-out layer prevent overfitting by removing some neurons in the training stage. Flattening layer converts 2D data into 1D. One-dimensional feature vector from the flattening layer is fed into fully connected layer. In this model, three fully connected layers are used. The sigmoid activation function $F(x) = 1/(1 + \exp(-x))$ is used in the final classification layer (fully connected layer) as in Fig. 7. The sigmoid activation function is used to classify the images into hard exudate image or normal image without hard exudates. The sigmoid activation

function is the best option for binary classification. The optimizer used is Adam optimizer with the learning rate of 0.001, and the loss function used in our model is binary cross entropy. It is a stochastic gradient descent method. Regularization and drop-out layers are used mainly added to improve the performance of the model in prediction. Regularization and drop-out layers are added to prevent overfitting.

Once the model is compiled, the model is used for training. Our model summary is shown in Fig. 8. Total trainable parameters are 4,031,105 and non-trainable parameters are 4,031,105.

The model is trained and training accuracy obtained is 99.1%. The test set is fed into the model and the model is evaluated for the performance in term of accuracy,

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_6 (Conv2D)            (None, 224, 224, 32)      896
_________________________________________________________________
max_pooling2d_6 (MaxPooling2 (None, 112, 112, 32)      0
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 112, 112, 64)      18496
_________________________________________________________________
max_pooling2d_7 (MaxPooling2 (None, 56, 56, 64)        0
_________________________________________________________________
conv2d_8 (Conv2D)            (None, 56, 56, 64)        36928
_________________________________________________________________
max_pooling2d_8 (MaxPooling2 (None, 28, 28, 64)        0
_________________________________________________________________
conv2d_9 (Conv2D)            (None, 28, 28, 128)       73856
_________________________________________________________________
max_pooling2d_9 (MaxPooling2 (None, 14, 14, 128)       0
_________________________________________________________________
conv2d_10 (Conv2D)           (None, 14, 14, 256)       295168
_________________________________________________________________
max_pooling2d_10 (MaxPooling (None, 7, 7, 256)         0
_________________________________________________________________
conv2d_11 (Conv2D)           (None, 7, 7, 512)         1180160
_________________________________________________________________
max_pooling2d_11 (MaxPooling (None, 3, 3, 512)         0
_________________________________________________________________
dropout_1 (Dropout)          (None, 3, 3, 512)         0
_________________________________________________________________
flatten_1 (Flatten)          (None, 4608)              0
_________________________________________________________________
dense_3 (Dense)              (None, 512)               2359808
_________________________________________________________________
dense_4 (Dense)              (None, 128)               65664
_________________________________________________________________
dense_5 (Dense)              (None, 1)                 129
=================================================================
Total params: 4,031,105
Trainable params: 4,031,105
Non-trainable params: 0
```

Fig. 8 CNN model summary

i.e., how accurate the model is classifying images into with hard exudates images and without hard exudates images.

3.4 Performance Evaluation of the Classification Model

The performance of the classification model is evaluated using accuracy, specificity and sensitivity. Let True Positive = PosT, True Negative = NegT, False Positive = PosF and False Negative = NegF.

Then

$$\text{Accuracy} = (\text{PosT} + \text{NegT})/(\text{PostT} + \text{NegT} + \text{PosF} + \text{NegF})$$

where PosT defines the number of images correctly classified as images with hard exudates. NegT defines the number of images correctly classified as images with no hard exudates. PosF defines the number of images incorrectly classified as images with hard exudates. NegF defines the number of images incorrectly classified as images with no hard exudates. Accuracy is defined as the ratio of correctly identified images (with hard exudates) and total images. Sensitivity measures the proportion of correctly classified images with hard exudates. Specificity measures the proportion of correctly classified images with no hard exudates.

4 Results and Future Scope

In this paper, the early sign of diabetic retinopathy, i.e., hard exudate is classified using deep CNN. The model starts training the images which are fed to convolution neural network. The model come up with a training accuracy of 99.1% and validation accuracy of 98%. Then, the test images are used for prediction. The model detected the presence of hard exudates with an accuracy of 98.94% for the test images. In future, we have planned to identify hard exudates classification for the early diagnosis of diabetic retinopathy using optical coherence tomography retinal (OCT) modality images. OCT retinal images are of high resolution. Hence, it will be more beneficial if OCT retinal images are explored for the identification of diabetic retinopathy.

References

1. P. Saeedi, I. Petersohn, P. Salpea, B. Malanda, S. Karuranga, N. Unwin, S. Colagiuri, L. Guariguata, A.A. Motala, K. Ogurtsova, J.E. Shaw, D. Bright, R. Williams, On behalf of the IDF Diabetes Atlas Committee, Global and regional diabetes prevalence estimates for 2019 and

projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition (2019)
2. A.S. Ladkat, S.S. Patankar, J.V. Kulkarni, Modified matched filter kernel for classification of hard exudate, in *2016 Internernational Conference on Inventive Computing Technolgy* (IEEE, 2017)
3. K.K. Palavalasa, B. Sambaturu, Automatic diabetic retinopathy detection using digital image processing, in *International Conference on Commnication and Signal Processing* (IEEE, 2018)
4. S. Bharkad, Morphological and neural network based approach for detection of exudates in fundus ımages, in *Proceedings of 2nd International Conference on Computer Methodology and Communication* (2018)
5. A.H.A. Samah, F. Ahmad, M.K. Osman, M. Idris, N.M. Tahir, N.A.A. Aziz, Classification of pathological signs for diabetic retinopathy diagnosis using ımage enhancement technique and convolution neural network, in *9th IEEE International Conference on Control Sysems Computing and Engineering* (2019)
6. M.M.I. Rahi, F.T. Khan, M.T. Mahtab, A.K.M. Amanat Ullah, M.G.R. Alam, M.A. Alam, Detection of skin cancer using deep neural networks, in *IEEE Asia-Pacific Conference on Computer Science and Data Engineering* (2019)
7. A.K. Abdar, S.M. Sadjadi, H. Soltanian-Zadeh, A. Bashirgonbadi, M. Naghibi, Automatic detection of coronavirus (COVID-19) from chest CT images using VGG16-based deep-learning (2020)
8. S.A. Ali, M.S. Kumar, Detecting pneumonia from X-ray images using transfer learning, in *IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation* (2020)
9. J. Zhang, Y. Xie, Y. Xia, C. Shen, Attention residual learning for skin lesion classification. IEEE Trans. Med. Imag. **038**(9), 2092–2103 (2019)
10. B. Ahmad, M. Usama, C.-M. Huang, K. Hwang, M.S. Hossain, G. Muhammad, Discriminative feature learning for skin disease classification using deep convolutional neural network. IEEE Access **008**, 39025–39033 (2020)
11. K.V. Maya, K.S. Adarsh, Detection of retinal lesions based on deep learning for diabetic retinopathy, in *5th International Conference on Electrical Energy Systems* (2019)
12. M.A. Albahar, Skin lesion classification using convolutional neural network with novel regularizer. IEEE Access **007**, 38306–38313 (2019)
13. S. Yu, X. Feng, B. Wang, H. Dun, S. Zhang, Automatic classification of cervical cells using deep learning method. IEEE Access **009**, 32559–32568 (2021)
14. V. Tiwari, C. Pandey, A. Dwivedi, V. Yadav, Image classification using deep neural network, in *2nd International Conference on Advanced in Computing, Communication Control and Network* (2020)
15. P. Tang, Q. Liang, X. Yan, S. Xiang, D. Zhang, GP-CNN-DTEL: Global-part CNN model with data-transformed ensemble learning for skin lesion classification. IEEE J. Bio. med. Health Inform. **024**(10), 2870–2882 (2020)
16. R. Prabha, S. Balakrishnan, S. Deivanayagi, V.K.G. Kalaiselvi, D.P. Rani, G. Aswin, A review of classification algorithms in machine leraning for medical IOT. Int. J. Pharm. Res. **013**(1) (2021). ISSN 0975-2366

ARFA-QR Code Based Furniture Assembly Using Augmented Reality

Ashok K. Chikaraddi, Suvarna G. Kanakaraddi, Shivanand V. Seeri, Jayalaxmi G. Naragund, and Shantala Giraddi

Abstract The current system of providing 2D instruction set and the products was a less efficient way to guide the user to assemble the product. Hence, the proposed application helps the user with an interactive augmented reality-based assembly instruction application. This application is mainly based on furniture models. The build and deployment of the model is performed on LabVIEW 2014 simulation environment. Unity GUI, Blender, Vuforia, Android Studio, and QR technology are used to create and deploy the model. The proposed model enables seamless integration of augmented reality into business applications. The outcome of this research is an Android application which shows different objects assembled using augmented reality.

Keywords Augmented reality · 2D · 3D · AR-Gaido · QR code · User interface

1 Introduction

User manuals help users to build a product they purchased, use the product, and become familiar with specifications of the product. Existing system uses a 2D catalog to know about the specifications or to know about how to build and use the product. This system may be helpful in knowing the specifications of the product, but when

A. K. Chikaraddi (✉) · S. G. Kanakaraddi · S. V. Seeri · J. G. Naragund · S. Giraddi
KLE Technological University, Hubballi, Karnataka, India
e-mail: chikaraddi@kletech.ac.in

S. G. Kanakaraddi
e-mail: suvarna_gk@kletech.ac.in

S. V. Seeri
e-mail: seer@kletech.ac.in

J. G. Naragund
e-mail: jaya_gn@kletech.ac.in

S. Giraddi
e-mail: shantala@kletech.ac.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_23

there comes a situation to build a complex product, the 2D manuals may not be helpful and the users may get confused as in what to do next. In the case of 2D manuals, the user would not have the privilege to know how the product may look in his surface when it is actually built or bought. The proposed application (AR-Gaido) provides the product to the users. The product features are built virtually using augmented reality [1]. The application is user interactive, i.e., the user has been given the option to control the building of the product. The user can also place the product virtually just to know how that product may look even before it was brought. There was no existing application that could help the user to build the complex products. The user manual guide books were not always sufficient. Hence, the proposed application has an advantage over the existing product. There was no guarantee if the product would look good when it was bought. Hence, there was a need of a system that would give the user the feature to virtually place the product were he/she wants to place it. The organizations may need this kind of application as there are a lot of complex products being released, where their customers may have the need for this application.

General Public: The general public or the buyers of the product are one of the main users of this application [2], as the application is mainly built for their ease to build the product. Not only the buyers can use this application, but also the users, who would want to know how the application may look like if they buy it in future, can use the application. The online shopping cart companies: These companies may want to use this product to advertise their shopping cart application. This would help them in their business to attract more users toward them. To help the user to build complex products.

- To let the users know how their product may look like, even before they purchase it.
- Useful for the company to advertise their products and to attract customers to buy the product.

QR code serves as a pathway to advertise the products through augmented reality. It redirects multimedia content. It enables users to understand the structure of the products they are buying by quickly scanning the QR code in their smart phone. Variety of products can be embedded in QR codes, making it easier to track the preference of the consumer. This enables the vendor to target specific advertisement toward consumers. This research focuses on building the images for chair and table using augmented reality embedded in QR technology.

2 Literature Survey

Here in [3], the author proposed developed augmented reality that shows real world overlaid by virtual pictures. Virtual monitors were freely positioned in user space to exemplify remote collaborators. By using shared virtual whiteboard, users cooperatively view and interact with virtual entities.

Using augmented reality technology virtual popup book system is introduced by the authors. Based on the posture and location, 3D virtual objects are shown on the real book by the machine [4]. Picture book with markers look vague despite many markers-based approaches being proposed over time. Markers are not used by the developed system. Virtual popup book exhibits the features such as 3D rendering assists the users in understanding the scenes [5]. By using motions, characters appear sprightly. Picture book uses the novel depiction which comprises the mix of 2D and 3D rendering. Finally, animation is used to display changes in time.

Authors have focused on the key technology necessary for developing mobile augmented reality application. Discussed the current glitches and a universal framework essential for the development. Discussed the applications for augmented reality in mobile devices [6].

Author's idea of object-based augmented reality [7] is to allow the users to view the virtual objects out in reality. Images from object pictures in front, back, top, bottom, left and right side can be provided by the users. They are placed onto the 3D cube, which will construct the whole virtual object. A prolonged environment is formed through the combination of real world and created object, and it appears as although the real-world object and virtual object coexist within the environment. Describes the effect of augmented reality and Unity 3D on learning and training of the future education.

Authors provide the information about the key extortions which affect the augmented reality in the near future and its existing and upcoming applications. Proposes the comparison of virtual and augmented reality. Also discussed outcome of augmented reality on the human life [8].

Author Balzerkiewitz and Stechert [9] suggested the use of virtual reality (VR) in distributed product distribution teams. This work focuses on the shared task and the associated problems. Analyzes the real-world technology. This study demonstrates the use of visible reality in product growth. Virtual software is adopted to support developmental distribution teams. Smooth data transfer between CAD and VR is achieved.

Togias et al. [10] have proposed a teleoperation-founded support system for the construction and control of industrial robots using the augmented reality [11]. The main purpose of this work is to reduce the time and effort required to restore the function of the robots without the presence of robotic operations on the floor of the store. The advanced robotic cell performing assembly functions can be quickly reproduced to handle different assembly variations. The virtual interface is designed as an offline simulation of the remote process. Provides real-time control of robotic resources to the user.

Hongyi's research project [12] author looks at the potential for using unpopular augmented reality that we see in the human-based robotic production support system. Personnel instructions and robotic commands can be made to the detriment of human labor readily and accurately.

The study of Ping et al. [13] explores the distinction between the virtual reality and augmented reality. The model compares the in-depth understanding of the AR and VR system using the head-mounted display. The result obtained demonstrates

the accuracy of the AR depth limitations higher than VR. Here a test is performed in the form of a shuffling board to meet the requirements for accurate depth judgment in sport.

Author Zhang et al. [14] reviewed the background of virtual reality and physical reality. We have proposed a coherent concept called symmetrical reality to describe the physical and real world in a coherent perspective. Under this reality of measuring cultural truth, the unpopularity of augmented reality, the real truth, the unchanging truth and the real contradictory truth can be explained using a combined demonstration. With two different viewing points, the features of the measurement truth are analyzed. All other physical and virtual forms are treated as special cases of symmetrical reality.

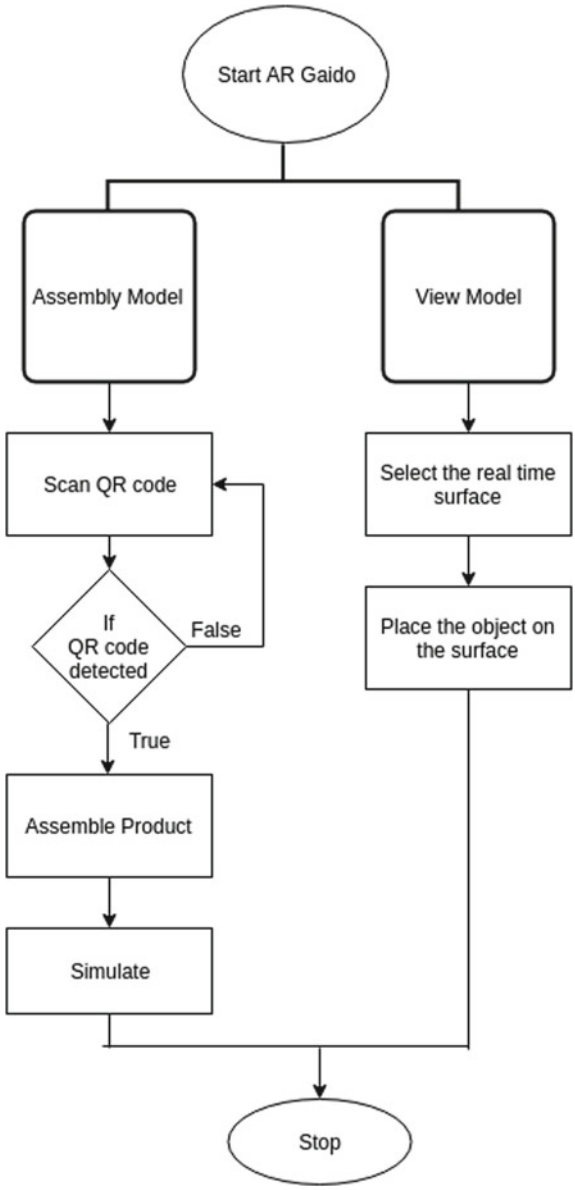
The proposed Thanyadit et al. [15] approach provides a solution for educators to see VR users on the scales using augmented reality. Augmented techniques are used to visualize the perception of real users in a visible environment and also enhances educator awareness.

Author Guo et al. [16] proposed a model to increase reality authenticity. Improved system real-time construction to facilitate closure management. The architecture is created in RGBD images and has three components, such as a real-time camera tracking system, a 3D reconstruction system and an AR fusion system. To implement the augmented reality program, two pass-through strategies have been used. The first camera for transit tracks arrives on time with video rate. The second pass occurs simultaneously to handle the closure between the visual object and the real scene such as the camera. Visual aids and color photographs are included to produce augmented reality contents.

3 Proposed System

The proposed system shown in Fig. 1 aims at building an augmented reality-based instruction set application. The application provides certain set of features to the user such as: The user can build his/her product in his mobile device virtually. The application is user interactive and hence the users have the control over building the product virtually. The users can also place the object on a real surface using his/her camera to know how it would look in reality. Whenever the user needs to know how the product may look, he/she downloads the application, goes to the 'view products' option. The "view products" option opens up the AR camera and the user will be given a list of products where he has to choose one product and place it anywhere he wants. Whenever the user purchases a product, the product will have a QR code attached to it. The user has to download the application and go to the assemble option where that option will again open the AR camera. The user then has to scan the QR code. Once detected, the user will have the option to build the product. The user can then build the product step by step and know how to build it in reality.

Fig. 1 Proposed system



Various games, construction kits and furniture come with build instructions. These are exemplified typically in 2D Print material or PDFs for ingesting via computers. 2D instructions are seldom inadequate contingent on the nature of the element.

This research focusses on the furniture product. Part of this product line is a set of models which the users build from wooden material. Users can be able to scan a

QR code on these products by means of phone or tablet app which will present user with augmented reality build and assembly instructions of all the models which are built. Next and previous buttons of the application will help the users to go back and forth among numerous intermediate steps of assembly. From cloud back end, the augmented reality data should be procured. Cloud backend will have the capability to upload the new 3D data for different models along with intermediate steps.

3.1 Architecture of ARFA

Graphical representation of the principles, concepts, elements and components of a system parts are represented using architectural diagram. Here in Fig. 2, we can see that there are two different sides, namely "Developer's Side" and "Customer's side." At the developer's side, the developer creates the 3D models and stores them in the cloud by assigning them a QR code each. The client on the other side uses the application to scan the QR code on the product. The QR code if matched downloads its respective model from the cloud, and the user can then start the assembly of the product.

Procedures that are discussed previously are developed by means of LabVIEW 2014 from TI. It provides robust environment for flow of data in a module and also it delivers user-friendly built-in functions. The simulation tool used to build and deploy this application is LabVIEW 2014 which is a programming environment to simplify hardware and software integration for engineering applications. This environment enables a consistent approach to acquire data from third-party hardware.

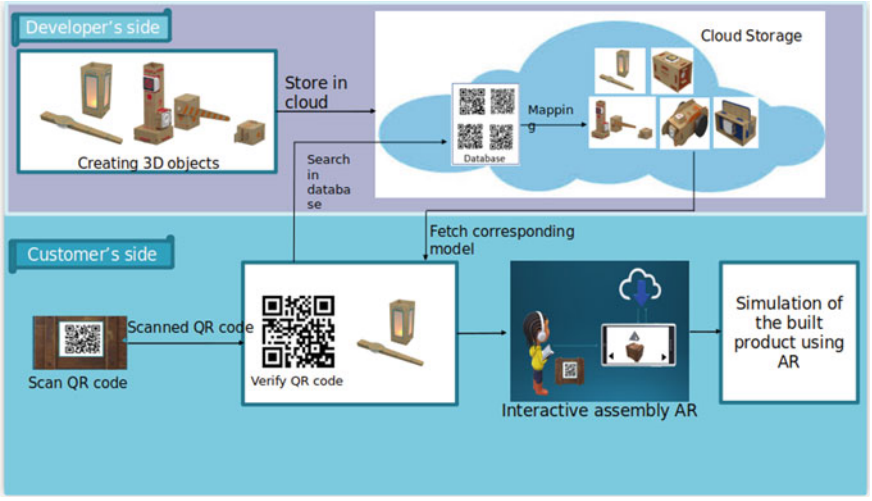


Fig. 2 Architecture of ARFA

4 Development

The development of the software process is discussed in this section. The methodology provides the complete procedure associated with the process depicted in proposed system shown in Fig. 1.

The prototype application provides a GUI-based 3D assembly instructions set manual in augmented reality. The application (AR-Guido) is developed using Unity engine. All the 3D models that are present in the database are built using the Blender. The next step is to deploy these models into Unity. Hence, Vuforia sdk is used to deploy these models. Vuforia has a database of the target images (QR codes) used for product identification. Later, the corresponding QR specific model is downloaded. The application reads the downloaded model and provides 3D assembly instructions in AR [17].

4.1 Assemble Product

The assemble product module has four different functions written in it. The first function is for the next button on the camera to work. Similarly, another function is for the back button to work and the third is for the home button to return to the home page. The update function is for each frame to run.

4.2 View Products

The view product module has two functions namely build NewTarget() and getSlider(). The build NewTarget() is used to detect a surface and place the product. The getSlider() is used to resize the placed product.

4.3 QR Code

The QR code is a unique ID where we add project name along with product category, e.g., (Arguido-table1) number; it is a unique ID so that we can differentiate different objects. The QR code is different for each model so that app will be able to understand which model to fetch when the user scans the QR code.

4.4 Blender

It is a tool that we used to create 3D models for our project. It provides all necessary GUI to create models to be exactly the way that they want it to be. Navigate the 3D scene in blender, press alt and right-click to move the scene. To create model, go to edit mode, consequently we can edit edges, vertices. Select vertex mode, right-click on a vertex (selection in blender uses right-click), and use manipulator to move it around. Then, we can extrude the model by selecting the vertices press the “e” key for extrude as we like. Now, add as many objects we need to finish the model, we can add texture in blender itself if we want, but we will add texture in Unity since Unity3D supports it, and we have already created it in Unity. Now, model switch is exported from edit mode to object mode. Make assured the model is selected and choose file/export/FBX. Provide name for model and location where to store the model select additional FBX exporting options from left pane of screen and select done.

4.5 Unity GUI

Select game object UI image from the top bar. This would add an image item to the scene. You should think of it as a descendant of the canvas from hierarchy. Canvas must be selected for all components. Unity will create one for you if you do not have one. Choose an image and move it to the desired location (X: 0, Y: 0). You have just gotten Rect Transform; in a second, you will specify the correct placement and size. In the meanwhile, there are three things you should be aware of. Three new articles have appeared on the scene in hierarchy: 1-Image; 2-Canvas; 3-EventSystem. This image is a non-interactive control that displays the sprite and provides numerous conversion possibilities. For example, you can colorize a photograph, assign it a material, regulate the extent to which the picture appears or motion it on the screen using a clockwise wipe. Canvas is the foundation for all of your UI components. Remember that Unity takes care of this for you when you add your first UI component. It contains a number of attributes that allow you to customize how your UI operates. Cycles and courses in the event system send events to objects in a scene. It is the same responsibility to supervise the beam's construction. The UI, like canvas, requires an event system, which Unity naturally adds.

- Constant pixel size: creates UI components, irrespective of screen size, preserves pixel size. This is the obliged worth of the canvas.
- Scale with screen size: size and acquisition of UI components as set adjustment. If the current adjustment is greater than the sample resolution, the canvas will retain the model adjustment, while increasing the elements to challenge the final resolution.
- Constant physical size: locations of the UI components are specified in physical measurements such as millimeters or points. It therefore requires an exact display of the screen's DPI.

To insert a button

- Click the mouse button on the right in Scene Hierarchy and navigate to the Create → UI → Button. If there is no existing canvas and an event system, Unity will build one for you, and insert a button inside canvas. Now, go to Button properties, and find the OnClick() property. Hit the + icon in the bottom tab, and a new entry should appear in the list. Drag and drop an empty game object, which contains the Button Manager script we created and which contains the OnButton Press method, onto the None (object) space.

Adding animation

- Create a new prospect.
- Create a text object and locate or change the perspective area of the object to be viewed by the camera.
- Choose the window menu and click on "Animation." This will open an animation window.
- In the animation window, click the "create" button in the middle of the window.

- A window opens and asks to specify a filename. Label with any name or with "TextAnim" can be used to name. If required, add other properties such as color and style.
- These steps are needed for base animation. It confirms that the text is white.
- To create another animation for the base color. Click on "Text Anim" header button and select "crate New Clip."
- One more dialog box will open and asks for file save. Name the file as "Base." Add the same color feature again by clicking "Add Property" and choose mesh Render → material color option. (Light color so that text is animated)
- Make sure that timeline should be set as 0.00 and make changes for *g* and *b* values to as you like.
- Now create the animation controller. On project window right-click and click Create → Animator controller and label it with suitable name.
- To open the animator controller editor, double-click on the animator controller. When it opens, drag and drop both images into the animator controller window.
- Create transition logic for the animations. Right-click on "Any State" and select "Make Transition" Move the mouse over the "Base" node and click on the left. It will connect and drag from any state into Base animation.
- It is now able to make changes from state base to "TextAnim" mode. Right-click on "Base" and select "Make Transition" and right-click on "TextAnim."
- To go back to the state base from "TextAnim," right-click on "TextAnim" and select "Make Transition" and connect it to the "Base" Node.
- In this regime, it is, in fact, what we are looking for. However, we still have to create a variable that we can use in our code in order to make the transition. In order to do so. Please click on the "+" icon in the top-left corner of the animator's driver window, and then select the "Enable" option from the drop-down list:
- Trigger is named as "Makechange" We refer to this as the trigger. However, we also need to address the transitions that we want to enable them in order to move around. On the left, click on the arrow that goes from the base to TextAnim. You may note that there is a transition that was released in the pop-up box to the right of the Unity. There is a section called "Conditions." Click on the "+" icon in this section, and it will add the "Makechange" trigger to the list of conditions.
- Setting animator controller on text. Drag and drop the animator controller made into a cube object in the inspector.
- In a project window left click on the animator. Check the "Loop Time" box in the inspector. It confirms that looping is turned on.

5 Results Analysis

Results achieved by the model are discussed in this section. The results obtained show four outcomes, which describes the home page, an assembly of a table, chair and shelf. The results depict usage of augmented reality embedded in QR technology to construct the desired images of the products.

Figure 3 shows the home page of the application AR-Gaido. The home page has four options, namely "Assemble Product," "View Product," "Support" and "Exit." This user interface is developed using Android studio. Each option performs specific tasks as per the assignment. Scanning the QR code prompts the user to choose between the above options from the home page. Assemble product crates the desired image of the object using augmented reality. Once it is assembled, product can be viewed using "View Product." Users can refer to "Support" option to know about the product details.

Figure 4 shows the "Assembly Product" option where the user scans the QR code and assembles the table using the interactive button.

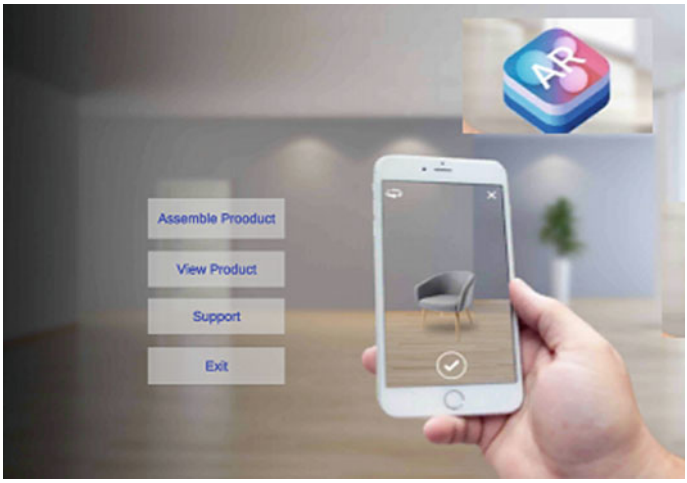


Fig. 3 Home page of the application AR-Gaido

Fig. 4 Assembly of table in the application AR-Gaido

Figure 5 shows the "Assembly Product" option where the user scans the QR code and assembles the chair using the interactive button.

Figure 6 shows the "Assemble Product" in the application. The option opens the camera with the "next," "back" and "home" buttons. There the user can scan the QR image and place the object.

The "View Product" in the application is a feature where the user can place the objects where they want using the camera just to know how it would look in reality. The object can be resized and rotated as the user wants. The feature has a camera button to place the product shown in Fig. 7.

Fig. 5 Assembly of chair in the application AR-Gaido

Fig. 6 An object of shelf in the application AR-Gaido

Fig. 7 View of a chair in the application AR-Gaido

6 Conclusions

The aim of the work is to create an android application to assemble the furniture products using augmented reality (AR). The requirement was to scan the QR code on the product and then assemble the product part by part. The application was required to be user friendly. Proposed model goal is achieved by creating an AR application that would scan the QR image on the product and that QR image was assigned to its respective model. The whole model was then assembled on the user commands. The different products assembled during this research are a table, a chair and a shelf.

The proposed model shows a sustainable mechanism, which is scalable as the technology advances. The build and deployment of different objects are possible by integrating varying technologies into the simulation environment. Hence, it can be said that the model provides a sustainable build mechanism.

The application can be expanded to different products other than furniture, such as cardboard. The application can be made available for the iPhone users. Other features such as, providing specifications of each product and how to use the product can be added for the existing model.

References

1. T. Hidayat, I.A. Astuti, Interactive augmented reality for the depth of an object using the model-based occlusion method, in *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE), Yogyakarta, Indonesia* (2020), pp. 382–387. https://doi.org/10.1109/IC2IE50715.2020.9274565
2. D. Melo Amaral, L. Chaves Dutra da Rocha, M. Carvalho Viana, M. de Paiva Guimarães, D.R. Colombo Dias, ARKLib: An augmented reality library for applications using kinect, in

2019 21st Symposium on Virtual and Augmented Reality (SVR), Rio de Janeiro, Brazil (2019) pp. 107–111. https://doi.org/10.1109/SVR.2019.00032
3. K. Hirokazu, M. Billinghurst, Marker tracking and hmd calibration for a video-based augmented reality conferencing system, in *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)* (IEEE, 1999)
4. N. Taketa et al., Virtual pop-up book based on augmented reality, in *Symposium on Human Interface and the Management of Information* (Springer, Berlin, Heidelberg, 2007)
5. F. Tang, Y. Wu, X. Hou, H. Ling, 3D mapping and 6D pose computation for real time augmented reality on cylindrical objects. IEEE Trans. Circuits Syst. Video Technol. **30**(9), 2887–2899 (2020). https://doi.org/10.1109/TCSVT.2019.2950449
6. H.-L. Chi, S.-C. Kang, X. Wang, Research trends and opportunities of augmented reality applications in architecture, engineering, and construction. Autom. Constr. **33**, 116–122 (2013)
7. D.R. Dhotre, Research on object based augmented reality using unity3d in education system. Student, MCA SEM VI, DES's NMITD, Mumbai, Maharashtra, India. IJARIIE-ISSN(O)-2395-4396 (2016)
8. R. Aggarwal, A. Singhal, Augmented reality and its effect on our life, in *2019 9th International Conference on Cloud Computing, Data Science and Engineering (Confluence), Noida, India* (2019), pp. 510–515. https://doi.org/10.1109/CONFLUENCE.2019.8776989.
9. H.-P. Balzerkiewitz, C. Stechert, Use of virtual reality in product development by distributed teams. Procedia CIRP **91**, 577–582 (2020)
10. T. Togias, C. Gkournelos, P. Angelakis, G. Michalos, S. Makris, Virtual reality environment for industrial robot control and path design. Procedia CIRP **100**, 133–138 (2021)
11. K. Pentenrieder et al.,Augmented reality-based factory planning-an application tailored to industrial needs, in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality* (IEEE, 2007)
12. H. Liu, L. Wang, An AR-based worker support system for human-robot collaboration. Procedia Manufacturing **11**, 22–30 (2017)
13. J. Ping, Y. Liu, D. Weng, Comparison in depth perception between virtual reality and augmented reality systems, in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces 23–27 March, Osaka, Japan*
14. Z. Zhang, C. Wang, D. Weng, Y. Liu, Y. Wang, Symmetrical reality: Toward a unified framework for physical and virtual reality, in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces 23–27 March, Osaka, Japan*
15. S. Thanyadit, P. Punpongsanon, T.-C. Pong, Investigating visualization techniques for observing a group of virtual reality users using augmented reality, in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces 23–27 March, Osaka, Japan*
16. X. Guo, C. Wang, Y. Qi, Real-time augmented reality with occlusion handling based on RGBD images, in *2017 International Conference on Virtual Reality and Visualization (ICVRV)*
17. J. Buchner, A. Jeghiazaryan, Work-in-progress–the ARI2VE model for augmented reality books, in *2020 6th International Conference of the Immersive Learning Research Network (iLRN), San Luis Obispo, CA, USA* (2020) pp. 287–290. https://doi.org/10.23919/iLRN47897.2020.9155110

Voice-Enabled Virtual Assistant

Ch. Lakshmi Chandana, V. Ashita, G. Neha, K. Sravan Kumar, D. Suresh Babu, G. Krishna Kishore, and Y. Vijaya Bharathi

Abstract Frameworks are being learned step by step nowadays and will assist people in their daily lives. Moreover, artificial intelligence [AI] techniques are now available in a broad number of sectors, ranging from ventures in manufacturing to medical innovation. As a result, within the completed framework, this research work has created a virtual assistant to solve college-related queries. The resulting framework is essentially a virtual assistant, who is meticulously college organized and resolve all college-related questions for students, teachers, and administration. Students confront a lack of critical information about college difficulties for a variety of reasons. The reason might be a result of system flaws. Such causes include a communication gap between the student and the college administration, a lack of student engagement, a lack of suitable guidance, and ignorance on the part of the student and/or the college administration. The student may be unaware of the class schedule, event timings, event location, vacations, examination schedule, permissions, and placement details. Here, the virtual assistant plays the key role in providing the necessary information. The advancements like natural language processing are utilized with the help of GTTS, Google API to change over from text to speech and the reverse way around. Here, the whole input is taken by the user and driven via graphical user interface (GUI). This application reduces the time and manual work of the user.

Keywords Natural language processing · GTTS · Artificial intelligence · Virtual assistant · Frameworks

Ch. Lakshmi Chandana (✉) · V. Ashita · G. Neha · K. Sravan Kumar · D. Suresh Babu · G. Krishna Kishore · Y. Vijaya Bharathi
Department of CSE, VR Siddhartha Engineering College, Vijayawada, India

G. Krishna Kishore
e-mail: gkk@vrsiddhartha.ac.in

Department of IT, VR Siddhartha Engineering College, Vijayawada, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_24

1 Introduction

Organizations today can give quick and customized reactions to clients with technology. In today's world, people are surrounded by smart watches, smart devices, and IoT devices in most cases. The IoT devices in today's world are capable of interacting with one another to provide useful services without interaction of humans in between. In this developing innovation, savvy conditions can be created utilizing artificial intelligence and machine learning to comprehend the climate needs. Knowing about the college before getting admission is very important. To be aware of the college and its schedules and its placement activities is necessary. To know more about the activities going on in our college without stress, we can use the latest technology. Technology nowadays can recognize our speech and text and can even reply back without human interference. We have seen voice assistants like Google Voice Assistants to know about nearby restaurants, schools, colleges, etc. Students going to college may forget the examination date and schedules; to help the students, the voice assistant is a great technology with the natural language processing technology in nowadays compared to printed ones. Understudies face absence of information on essential data with respect to school issues because of a few reasons. Such reasons incorporate communication holes among understudies and college organization. The understudy may have absence of data about the class timetable, occasion timings, occasions, examination timetable. Here, the virtual assistant assumes the critical part in giving the vital data. Today, the virtual bots are fueled by computerized reasoning; they are consequently, fit for helping through text or voice. Menial helpers give the best answers for both the students and administrators in different regions. Voice-actuated chatbots are the ones who can cooperate and convey through voice. They are equipped for tolerating the order in an oral or composed structure through natural language processing (NLP) innovation. They are customized to answer through voice. Here, this research work enhances the chatbot to furnish clients with the data in regard to college arrangement, class timetable, occasion timings, occasions, examination timetable, and fee structure. This voice-empowered chatbot is intended to furnish clients effortlessly gets to sites and applications.

Examples of our project are the Google Assistant and an Amazon Alexa which perform as according to our voice commands; similarly our application also works according to the user commands but with limitations, as these are restricted to the college website.

The objectives of the project are:

- To foster a remote helper for understudies to address college-based inquiries and give college-related data.
- To design a prototype with efficient performance.
- To develop/implement a working application for designed prototype
- To test the proposed model functioning and its performance on the college website.

Title justification:

The title to this project is "Voice enabled Virtual Assistant"; the virtual is having a meaning of not physically existing as such but made by software to appear to do the task. The virtual assistant is a non-physically existing human type support that is given to make things easy to the human. The title "voice enabled virtual assistant" is a physically non existing body done by software with voice recognition and talk back capability. The voice-enabled virtual assistant is built to help the students to know their schedule and timetable without much stress.

2 Related Work

Siddesh et al. [1] has done the work on "Artificial Intelligence-based Voice Assistant." In this work, the voice assistant performs mental tasks like turning on/off smart phone applications with the help of voice user interface (VUI) which is used to listen and process audio commands. They have installed GTTS (Google Text-to-Speech) engine package to make the voice assistant speak like a normal human being.

Gaglio et al. [2] proposed a method as "Artificial intelligent college oriented virtual assistant." In this work, an app for individual user with individual login and storage of data is developed. The artificial intelligence is used to create a virtual assistant, and the data related to specific user is sored into the application. Each and every individual can have the college data like timetable and examination timings, holidays etc. We can also get the information regarding location of the college and date and time etc.; here, technologies like artificial intelligence and Android studio are used.

Prajwal et al. [3] proposed "Universal Semantic Web Assistant based on Sequence to Sequence Model and Natural Language Understanding." The Web-based virtual assistant named as "Wisdom" is intended to build it is a complex system that includes several modules that are intended to perform specific tasks. These modules are later integrated to form the complex system as a whole. Each of our sub-systems interact to each other through cross module function calls similar to the request and response API's.

Pal et al. [4] proposed an user experience with smart voice assistants: The accent perspective answers the research questions regarding the usability, usage patterns, and the overall satisfaction received after using the VA's by the secondary English language speakers we wield a dual prong strategy: questionnaire survey via Google Forms and evaluating the serviceability of the VA's in a free-living condition. The questionnaire is created elicited from the System Usability Scale. The objective behind the questionnaire survey is of two-fold: (1) to let the users allocate their feedback towards their usability, usage pattern, and the usefulness of the VA's and (2) to act as an input to the second part of the experiment.

Atzeni et al. [5] proposed an "AskCO: A Multi-Language and Extensible Smart Virtual Assistant." It allows answering complex natural language queries by translating them into a proper formal meaning representation. The AskCO can be used to translate natural language utterances into a Java source code.

Bohouta et al. [6] proposed "Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon, Alexa and Google Home)." It process two or more combined user input modes, such as speech, image, video, touch, manual gestures, gaze, and head and body movement in order to design the next-generation of VPAs model. It can be used in other different areas of application, including education assistance, medical assistance, robotics and vehicles, disabilities systems, home automation, and security access control.

Novikov et al. [7] proposed "Development of Intelligent Virtual Assistant For Software Testing Team." In this work, the team builds an intelligent virtual assistant where one needs to combine in one project different technologies and tools from different areas. User input understanding: audio speech recognition, natural language understanding, user face recognition, sentiment analysis, opinion mining, etc. Intelligent reasoning: context understanding, dialogue management, social reasoning, domain specific knowledge, user model, etc., and next output generation, and software testing.

Iannizzotto et al. [8] proposed "A vision and Speech Enabled, Customizable, Virtual Assistant for Smart Environment." It is a graphical frontend and a coordinator that leverages on the services to offer to the user a multimodal and involving interconnection with the connected home automation and smart assistant systems.

Revathi et al. [9] has done the work on "Digital Revolution in Speech and Language Processing for Efficient Communication and Sustaining Knowledge Diversity." In this work, they have studied on the digital revolution of automatic speech recognition. In this paper, the user interface which are more suitable for people are discussed. The automatic speech recognition is subset of NLP. The latter is divided into natural language understanding (NLU) and natural language generation (NLG).

3 Proposed Work

In our proposed model, the various existing issues are covered. The disadvantages of the previous work are the user interaction with the application and the availability of the application. In this work, the application is available 24×7 and is easily used by the user through the website.

3.1 Creating GUI Interface for the Application

This project has a graphical user interface to interact with the user and prompts the user for the input via speech. This user interface (UI) acts as a way of interaction between the user and the application. This project contains a user interface useful for

these two types of people: those with physical disabilities and those of with lack of interaction. To create an effective interaction between the application and the user, we created the face of the application simply called as user interface. The HTML5 is used here to create interactive user interface to display the content in the webpage using browser. Webpage or website provides a fast and easy way for interacting with user, and it is supported in every system with Web browser and Internet connectivity. First, create the webpage with basic information and then create the style with tags to the webpage using CSS so that user feels easy to use by looking at the user interface. Add style to the webpage based on your necessity. Generally, the webpage supports buttons, labels, canvas, list box, images, etc. Here, the webpage uses button, images, and microphone. There are almost all types of tags the webpage and a browser support. Each style tag has some set of attributes width, height, bg, fg, etc. For button tags, we have to include "create button" tag which creates the button to be used in the webpage, and the JavaScript is used to create the function call for the button. The called or action to be performed after user clicks the button, and then the backend algorithms or techniques or function used will be executed and outputs the result through the webpage. At last, the script tag in html to take action against each event triggered by the user is created. The interface of the webpage of conversion has one button to trigger when the user speaks and generate the output speech to the user. In this way, the project used the webpage for prompting the user for the input and made it interactive. It is to summarize that in total, for each type of conversion, we used 1 button for the user to speak and stop listening when the user speaks and stops. The website is created to deploy the voice assistant.

3.2 Transforming Speech to Text

In the first transfiguration, i.e., to convert the speech to text, the implementation did here is by using webkitSpeechRecognition API. The webkitSpeechRecognition is a java script speech API. It makes easy to add voice assistant to the webpage. This API gives you complete control and flexibility over your Web browser's speech recognition features. In this project, we start by seeing if the browser supports the Web Speech API by looking for the webkitSpeechRecognition object. If this is the case, we recommend that the user upgrade their browser. (The API is now vendor prefixed because it is still experimental.) Finally, we set some of the characteristics and event handlers for the webkitSpeechRecognition object, which offers the speech interface. To begin, the user clicks the microphone button to input speech to system. We set the speech recognizer's spoken language "lang" to the BCP-47 value that the user has picked using the selection drop-down list, such as "en-US" for English-United States. It defaults to the lang of the HTML document root element and hierarchy if this is not defined. Chrome speech recognition supports a wide range of languages, as well as several right-to-left languages such as he-IL and ar-E. We call recognition after we have set the language. start() is used to turn on the speech recognizer. It calls the onstart event handler when it first starts recording audio and then the onresult

event handler for each new set of results. The onstart event sends the attribute speech to the speech recognition function and the recognizer sends attributes to the onresult event to transform the speech to text. The speech that is taken as input is measured with length and the transcript function is used to transcript from speech to text.

3.3 Transforming Text to Speech

In the case of the second transformation, i.e., converting the text to speech, the implementation is achieved with the speech synthesis module in HTML. The SpeechSynthesis interface of the Web Speech API is the speech service's controller interface; it can be used to receive information about the device's synthesis voices, start and stop speech, and other instructions. The SpeechSynthesis interface of the Web Speech API is the speech service's controller interface; it can be used to receive information about the device's synthesis voices, start and stop speech, and other instructions. The text in the code is generated as speech using speech synthesis. The language used is "en-US" in this application. The GTTs is a Speech Recognition API which is used to convert speech to text or text to speech.

3.4 Knowledge Abstraction

Gathering, manipulating, and augmenting knowledge are the three stages of knowledge abstraction. These phases derive lot for the virtual assistant's material.

Data gathering:
The first stage is to establish a knowledge foundation. This entails locating the course's major concepts and gathering information about them, with the course's core concepts organized in a hierarchical structure. The data is gathered from the user in this application.
Data manipulation:
The data stored in a form of output text to manipulate and categorize the data according to the data gathering.
Data Augmentation:
This process increases the number of natural language processing model training instances available in Dialog flow.

3.5 Response Generation

Contexts serve the function of overlaying the discussion in such a way that only particular intents can be activated when they are present. Because responses are the outputs of intents, they are triggered whenever a user entry matches intent. Using

precise questions, we can group data coming from student interaction with the bot. It is crucial to make a statement about the difficulties of drafting a response. We understand that how students view slide displays differs from how they should interact with a virtual assistant. This highlights the need of content adaptation or the process of identifying the media that best conveys the desired message. For example, the proposed project has added a location of a college by showing it in the Google Maps (Fig. 1).

The flowchart shows the flow of the data through the application. The user uses user interface to interact with the virtual assistant. The input speech is sent to Speech Recognition API. Giving the text as output to the system, the system checks whether the user is still speaking or not. If the user is speaking, it again changes speech to text and processes and gives the matched keyword result as speech output. Else it will terminate.

3.6 Algorithm

Step 1 Start Process.
Step 2 importing libraries and create an object to the classes.
Step 3 Get Audio from the user
Step 4 Convert speech to Text using GTTS.
Step 5 if the text with Keyword match
then return the corresponding result Response
Step 6 convert Text to speech using GTTS.
Step 7 return the Result.
Step 8 End Process

The process starts by opening the website, then the libraries are loaded, then the application get the audio from the user and convert that audio into speech to text using GTTS API. If the text that changes has a keyword that matches the keyword in the application, then the corresponding result will be outputted by the system. Then, the texts of response are converted to speech using GTTS and return the result to the user. This repeats throughout the processes and ends the processes.

4 Results Analysis

The functions are presented to the users through an interactive and attractive user interface, one for each module. The output and end results could be observed as below. Initially, when the application is made to run the website is opened, a UI as shown in Fig. 2 appears.

Click on the user at the right bottom of the page. As the steps in the user interface suggest, the user is prompted to speak, allow microphone and allow popups, the user

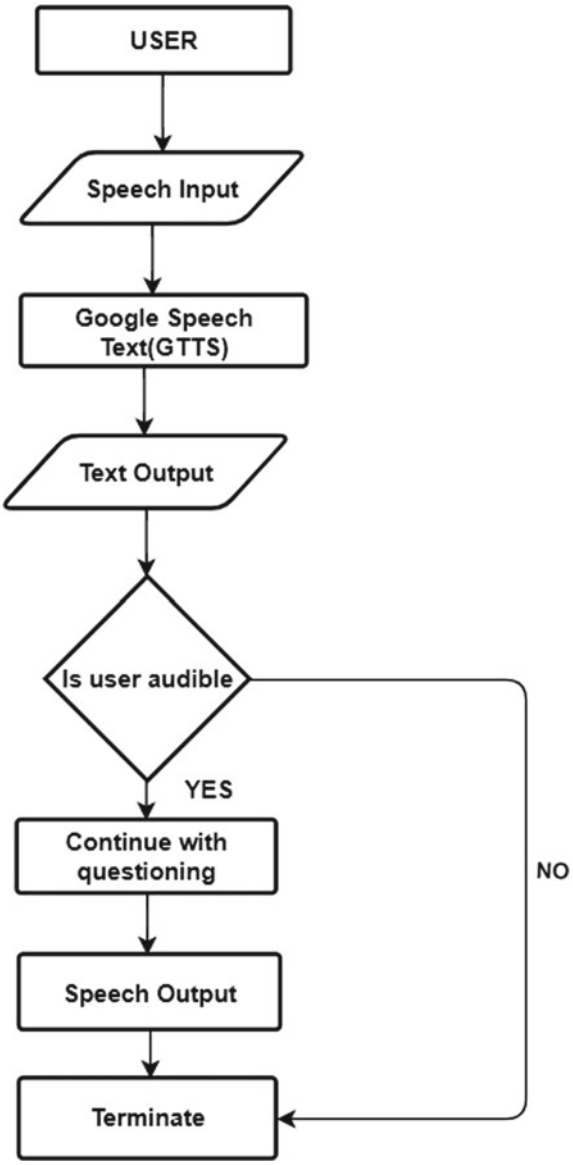


Fig. 1 Flow chart of voice assistant

clicks on the button to speak to the voice assistant, Fig. 3. Once the user speak to the assistant such as open the vr Siddhartha college website, then the college website will be opened as shown in Fig. 4. The user could go on and click on the button to start are stop the voice as shown in Fig. 3.

In Fig. 4, where the website that is commanded by the user to open is opened by virtual assistant is shown.

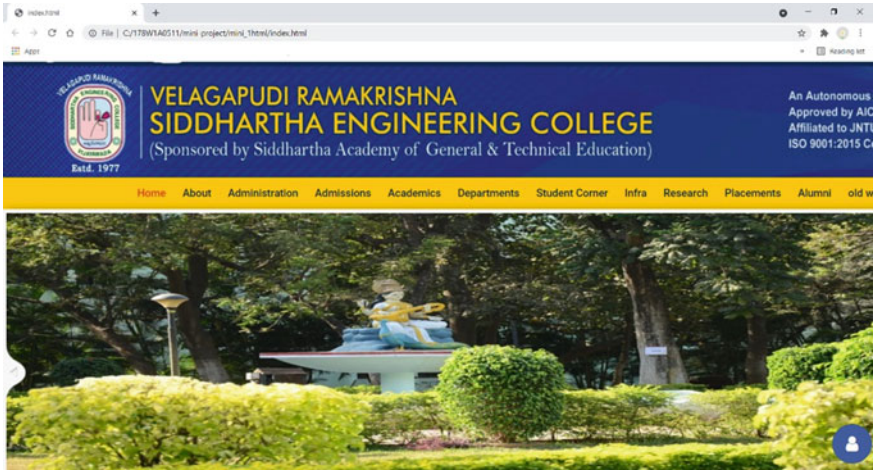


Fig. 2 UI of website to deploy the virtual assistant

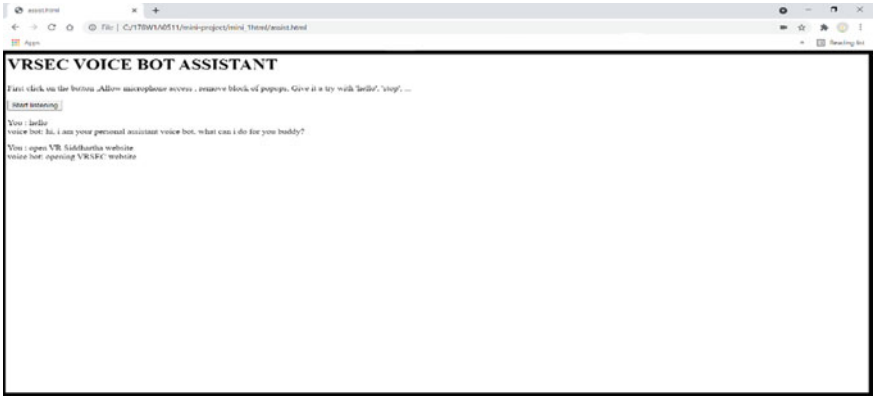


Fig. 3 User commanding the virtual assistant to open another website

So, we tested it out on the opening website, which is included in the code. If the user commands the website that is not included in the code, then the google will get the input of the user and return the results to the user. In Fig. 5, it is the user commanding more works to do by the virtual assistant.

The user has provided orders to the voice assistant to conduct tasks; therefore, the voice assistant opens one by one. We've expanded the use cases to include numerous websites as well. Figure 6 illustrates the commands being executed by opening a page that is not in the code.

We have extended the use cases to locations as well as well. In Fig. 7, it could be seen that a communicated location is also opened.

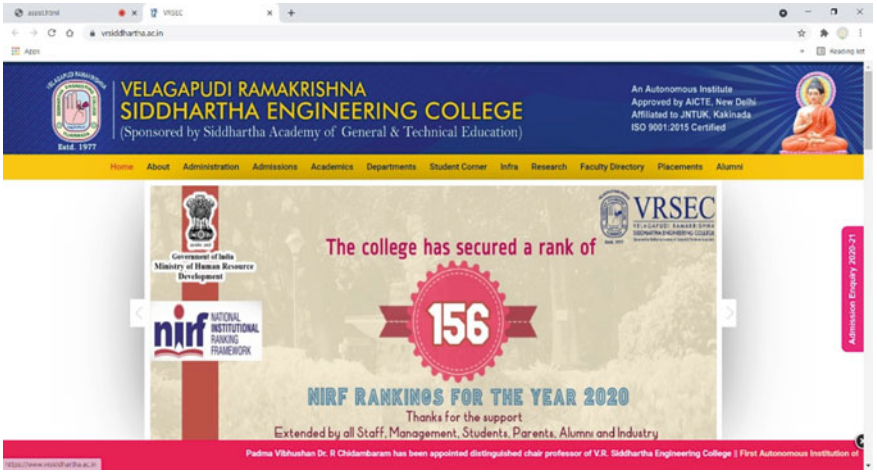


Fig. 4 Output of the website commanded by the user

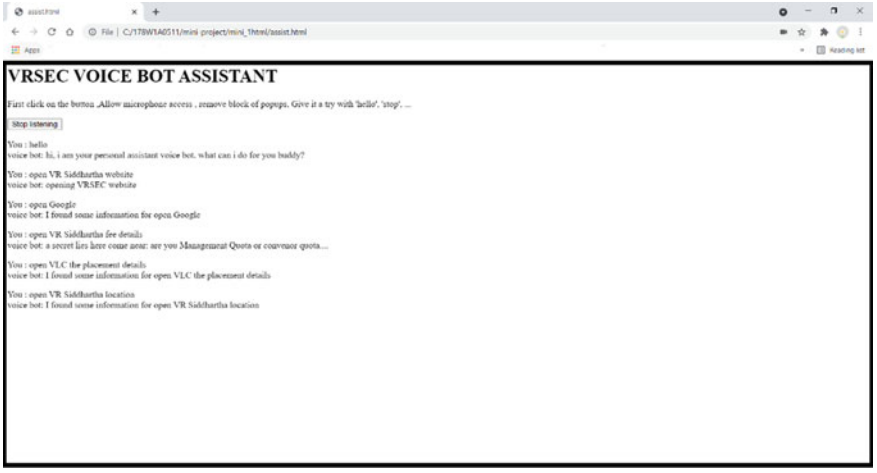


Fig. 5 Recognizing the speech and performing operations commanded by the user

The use cases that we have considered providing the system as an input is a speech as seen in Fig. 5. This is converted to text in the webpage. The resulting output successfully produced the text and speech from the given instructions by user.

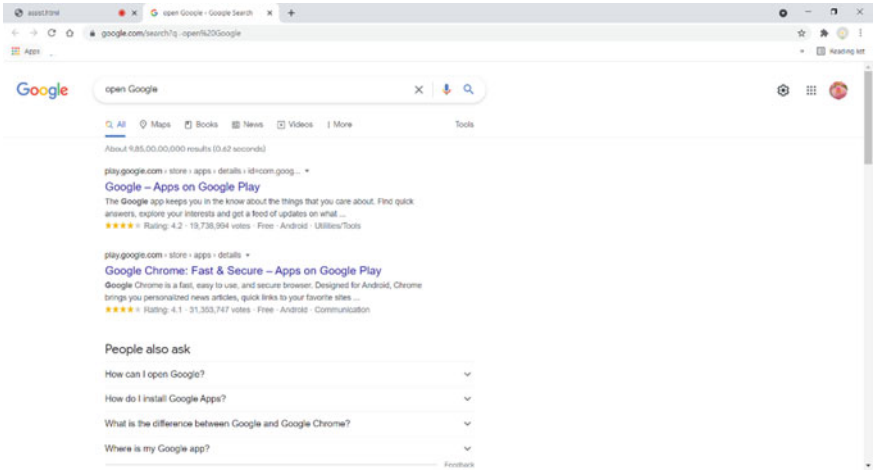


Fig. 6 Opening the page commanded by user

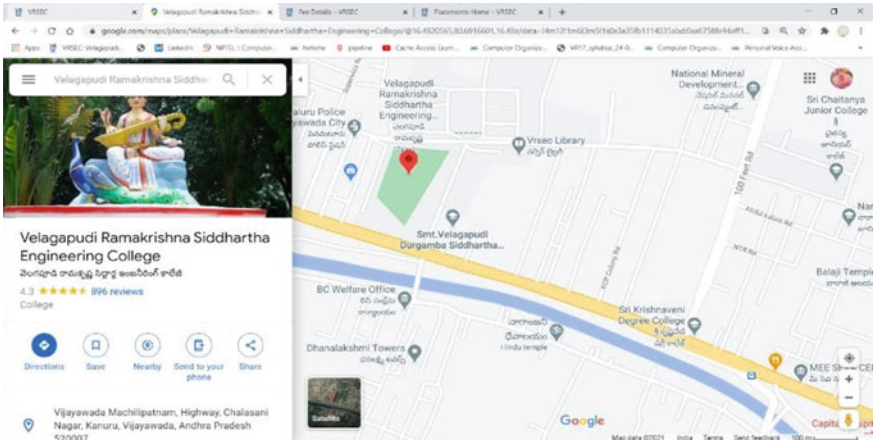


Fig. 7 Extracting the location that the user gave as speech

5 Conclusion

The deployed system is an AI-based college assistant that can help college students with their college-related questions, decreasing confusion and preventing opportunities from being lost. The established system can provide useful information such as college information, location, fee information, and a timetable. The disadvantages of the system are it is not self-learning system, and it is limited to only one college. As a future implementation of the system, we can use neural networks. Data compression and encryption can also be used to save information. Also, a machine

learning module may be added to AI so that trending topics for a specific branch can be recognized by AI based on priority, and users can be notified about the current top priority topics.

References

1. S. Siddesh, A. Ullas, B. Santosh, Aritificial intelligence based voice assistant, in *IEEE International Conference on Security and Sustainability (WorldS4)*, vol 978–1–7281–68234/20 (IEEE, 2020), pp. 593–596
2. S. Gaglio, V.Y. Shetty, S. Das, Artificial intelligent college oriented virtual assistant, in *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, vol 4, No 2 (2019), pp. 132–137
3. S.V. Prajwal, G. Mamatha, P. Ravi, D. Manoj, S.K. Joisa, Universal semantic web assistant based on sequence to sequence model and natural language understanding, in *International Conference on Advances in Computing and Communication (ICACC)* (IEEE, 2019)
4. D. Pal, C. Arpnikanondt, S. Funilkul, User experience with smart voice assistants: The accent perspective, in *IEEE conference, 10th ICCCNT* (2019)
5. M. Atzeni, M. Atzori, AskCO: A multilanguage and extensible smart virtual assistant, in *IEEE 2nd International Conference on Artificial Intelligence and, Knowledge Engineering* (2019)
6. G. Bohouta, Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon, Alexa and Google Home), in *IEEE 8th Annual Computing and Communication Workshop and Conference* (2018)
7. A. Novikov, I. Itkin, R. Yavorskiy, Development of intelligent virtual assistant for software testing team, in *IEEE, Nineteenth conference on software Quality, Reliability and Security Companion* (2020)
8. G. Iannizzotto, L.L. Bello, A. Nucita, A vision and speech enabled, customizable, virtual assistant for smart environment, in *11th International Conference On Human System Interaction(HSI)* (IEEE, 2018)
9. P.K. Deepthi, P.K. Vasanthi, Digital revolution in speech and language processing for efficient communication and sustaining knowledge diversity. CSI Commun. **41**(7), 6–9 (2020). ISSN 0970–647X
10. A.R. Revathi, P.K. Deepthi, P.K. Vasanthi, Digital revolution in speech and language processing for efficient communication and sustaining knowledge diversity. CSI Commun. **41**(7), 6–9 (2020). ISSN 0970–647X
11. D.S. Zwakman, D. Pal, C. Arpnikanondt, *Usability Evaluation of Artificial Intelligence-Based Voice Assistants: The Case of Amazon Alexa*, (Springers Nature, 2021), pp. 2–16
12. A. Poushneh, Humanizing voice assistant: The impact of voice assistant personality on consumers attitudes and behaviors. J. Retail. Consum. Serv. (2020)
13. G. Daniel, J. Cabot, L. Deruelle, M. Derras, Xatkit: A multimodel low-code chatbot development framework. IEEE Access (2020)
14. D. Carlander-Reuterfelt, A. Carrera, C.A. Iglesias, O. Araque, J.F. Sanchez Rada, S. Munoz, JAICOB: A data science chatbot. IEEE Access (2020)
15. A. Novikov, I. Itkin, R. Yavorskiy, Development of intelligent virtual assistant for software testing team, in *IEEE, 19th conference on software Quality, Reliability and Security Companion* (2020)
16. S. Manoharan, N. Ponraj, Analysis of complex non-linear environment exploration in speech recognition by hybrid learning technique. J. Innovative Image Process. (JIIP) **2**(04), 202–209 (2020)

Photonic-Based Front-Mid-Backhaul Access for 5G

Nihal Agarwal, Niranjan Kundap, Prajakta Joglekar, and Bharat S. Chaudhari

Abstract With the rapid growth in the deployments of Internet of things (IoT)-based smart devices and innovative real-life applications, there has been unceasing demand to offer the requisite data rates and quality of service. The 5G deployments are expected to fulfill such requirements at the access level. It would also serve the necessities of a wide and specialized range of applications that may need very high data rates or extremely low energy consumption. Furthermore, there would be potentials for the bottlenecks at the core and backhaul networks. Such constraints could be addressed by designing better transport networks at core and backhauls, and photonic networks are the best candidate. This paper presents the study on the integration of 5G and photonics networks, the proposed architectures, challenges, open issues, and silicon photonics' role in the near future to further enhance the end-to-end performance of the communication networks.

Keywords 5G · PON · Silicon photonics · Hybrid networks · Small cell · Backhaul

1 Introduction

In the fast-growing era, the evolution of technologies is playing an important role, and the Internet is one of them. Although initially the Internet was designed for connecting computers, however, due to explosive growth in mobile devices, there had been frequent evolution from 2.5G to 4G. This push is also driven by the innovative applications of the Internet, Internet of things (IoT) and its special access networks such as Bluetooth, Wi-Fi, ZigBee, low power wide area networks (LPWAN)-based LoRa, SIGFOX, and many others. According to the HIS Markit research, the number

N. Agarwal (✉) · N. Kundap · P. Joglekar · B. S. Chaudhari
School of Electronics and Communication Engineering, Dr. Vishwanath Karad MIT World Peace University, Pune 411038, India

B. S. Chaudhari
e-mail: bharat.chaudhari@mitwpu.edu.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_25

of connected IoT devices reaches 125 billion by 2030 [1]. The applications of IoT and LPWANs include health care, smart home, gaming, smart cities, smart environments, tracking, logistics, retails, automotive, smart manufacturing, and many more [2, 3].

With the imminent arrival of 5G, many challenges faced by today's world may be resolved because of its extremely high data rates, low latency, scalability, availability, and coverage. It is expected that 10 million user devices would be connected to the network per square kilometer. There would be a 1000x increase in the data traffic, so there are challenges for the bandwidth efficiency. Hence, the infrastructure needs to be developed, which is flexible to handle multiple devices with low latency and higher energy efficiency at the low cost [4].

The International Telecommunication Union Radio communication Sector (ITU-R) has identified the three basic services that 5G would support, viz. enhanced mobile broadband (eMBB), massive machine-type communication (mMTC), and ultra-reliable low-latency communication (URLLC), as shown in Fig. 1. It shows that these three services of 5G will touch each aspect of the ongoing evolution and advancements.

The existing wired networks at core and backhauls may be lossy and not immune to electromagnetic interference (EMI) and electromagnetic pulses (EMP). It may not be possible to achieve the 5G experience in terms of quality of service (QoS) because of higher bit error rates (BER) and bottlenecks due to large incoming and outgoing data. To address these challenges, photonic networks have to play an important role. As it transmits the data and control information in the form of light, huge bandwidth

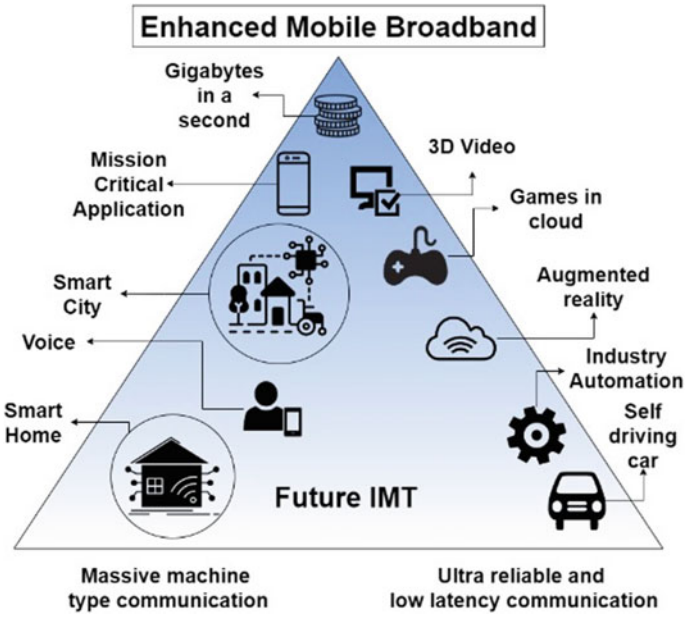


Fig. 1 5G usage scenarios by ITU

capability with immunity to EMI/EMP can be availed at a low cost. We can use passive optical networks (PON) to connect 5G base stations in the RAN. In several parts of the world, PONs are already deployed for fiber-to-the-home/business (FTTx), and the same networks can be used as core network for 5G. In such a case, the cost of setting up new PONs is reduced. Existing PONs in the developed world are vast and widespread, with enough line and port resources as well as the requisite optical equipment and power supplies.

This paper discusses the 5G and photonic network integration along existing architectures, challenges, and open issues. Section 2 presents the technology stack covering 5G and PON architectures. Section 3–6 focuses on the integration issues, the complementary technologies, and challenges. Section 7 concludes the paper.

2 Technology Stack

In this section, we present the 5G architecture and passive optical networks. Both technologies need to work in sync for the seamless performance of 5G.

2.1 5G Architecture

Currently, many wireless communication systems are working in the range below the 3 GHz range. However, to tackle evergrowing data or Internet traffic, very high bandwidth is required. Hence, for 5G, we are moving to a millimeter-wave (mmWave) spectrum, which ranges from 30 to 300 GHz. The typical 5G architecture is shown in Fig. 2.

The next-generation NodeB or UMTS Terrestrial Radio Access Network (UTRAN), also known as gNB, is directly connected to the core via the backhaul network. The backhaul network has to be energy efficient as there is a dynamic

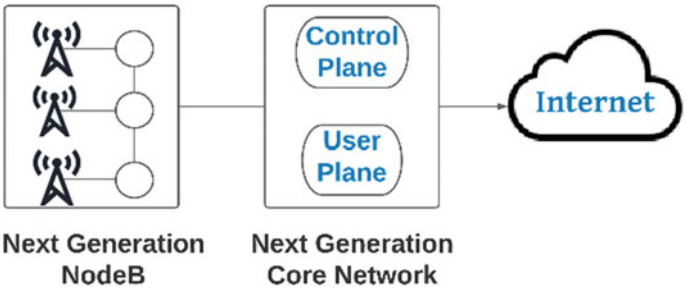


Fig. 2 Abstract architecture of 5G

requirement of the connection in dense 5G deployments. Although the wired backhaul networks are energy inefficient and not feasible economically [5], the researchers are looking to use pre-existing or new PON as a solution for the underlying transport network for 5G.

2.2 *Passive Optical Network*

As 5G networks have strict requirements of very high bandwidth capacity and very low latency, the current 3G/4G infrastructural supports such requirements. In such a case, utilization of pre-existing PON or deployment of new PONs will be the best solution. PON is a fiber-optic network that uses a point-to-multipoint architecture, with a single optical line divided into several optical splitters to serve numerous consumers [6]. The same splitters also link the numerous upstream channels from end-users to the OLT. The architecture of PON is as shown in Fig. 3. Primarily, it is comprised of three components, viz. optical line terminator (OLT), optical distribution network (ODN), and optical network unit (ONU). OLT is a core element connected to the fiber backbone, which also allocates bandwidth to ONU. ODN allows point to multipoint connection, involves the fibers, splitters. Its main function is to provide an optical communication channel between the OLT and ONU. The ONU terminates the PON at the customer's premises. PON is efficient, flexible, and reliable to accommodate fronthaul transport for cellular and fixed broadband in a single network.

There are different standards of PON services such as ATM PON (A-PON), broadband PON (B-PON), Ethernet PON (E-PON), Gigabit Ethernet (G-PON), XG-PON (10G PON), wavelength division multiplexing PON (WDM-PON), and next-generation PON (NG-PON). As we concentrate on the 5G network, thus we

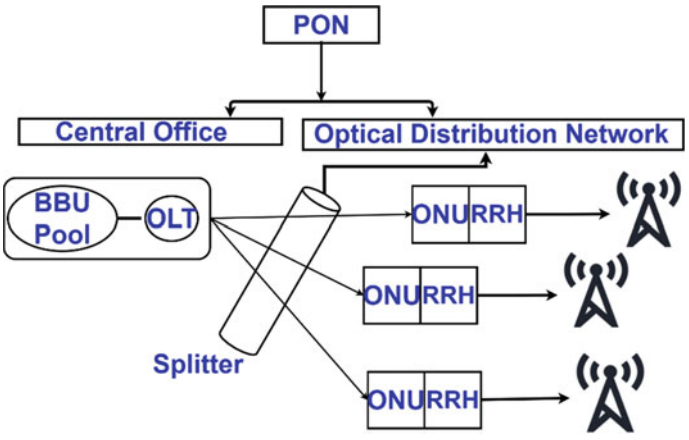


Fig. 3 Deployment of PON in fronthaul interface

will focus on high-speed standards of PON [7]. G-PON is a gigabit passive optical network from ITU-T. It provides a high-bandwidth downstream rate of up to 2.5 and 1.2 Gbps upstream rate.

XG-PON standard, also known as 10G-PON, is an enhancement to GPON, which has a downstream capacity of 10 Gbps and an upstream bandwidth of 2.5 Gbps. The downstream signal of an XG-PON user is specified in the wavelength range of 1575–1580 nm, while the upstream signal is specified in the range of 1260–1280 nm. XG-PON can coexist with G-PON (1310/1490 nm) and RF-video (1550 nm). The multiplexing scheme used is TDMA for uplink and TDM for downlink. In XG-PON, the spilt ratio is 1:64, and the fiber distance can be up to 20 km. 10G-PON supports dynamic bandwidth allocation, QoS, and traffic management.

WDM-PON is a bidirectional wavelength division multiplexing PON that enables operators to offer high bandwidth to many endpoints across long distances [8]. Due to WDM, the operator can overlay different types of PON on the same physical line. Using the same optical fiber, the operator can provide services like G-PON for FTTx and XG-PON for 5G mobile access networks on the same optical fiber line, i.e., four wavelengths on the same fiber. WDM-PON offers secure as well as scalable point-to-point wavelength links and is very important for the 5G fronthaul network.

NG-PON2 is a next-generation PON2 that provides a path for service providers to converged networks, in which numerous services coexist on the same fiber network infrastructure. The building block of NG-PON2 is a little different from the traditional PON system and comprises OLT, ONU, wavelength multiplexer (WM1), and coexistence element (CE), [9] as shown in Fig. 4.

NG-PON2, also known as time wavelength division multiplexing PON (TWDM-PON), can coexist with XG-PON, EPON, GPON, and RF Video. NG-PON2 is utilized for residential services, business services, small businesses, smart city

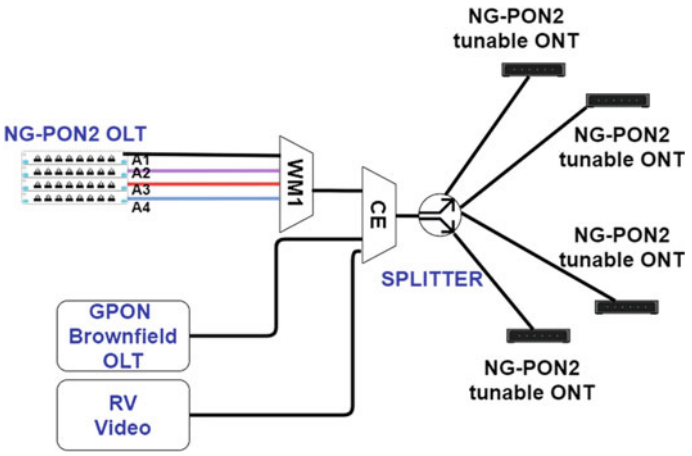


Fig. 4 NG-PON2 architecture

applications, smart home/building applications, mobile backhaul and fronthaul, and passive optical LANs.

3 Photonics for 5G

3.1 Radio Interfaces

A radio base station is used to transfer the data between user devices, and for that, a new interface called fronthaul interface [10]. In 5G architecture, the functions of base stations would be centralized so resources can be shared among different sites. This reduces the interference among the different remote radio heads. Different approaches for fronthaul connectivity of optical communication between the central office and antenna sites are proposed [11]. It includes digital radio over fiber (RoF), sigma-delta over fiber (SDoF), and analog RoF (ARoF). The wireless RF signal must be translated into optical frequencies since optical fiber operates at a different frequency range.

4 Data Centers

Data centers are the crucial physical part of the network storage in which all the applications and data is stored. The important components in the data center design are routers, switches, firewalls, storage systems, and servers. The current packet switching techniques of data centers consume more energy leading to high costs and high maintenance [10]. To tackle such challenges, using an optical switch instead of traditional packet switching can allow the packets to flow in the optical domain. It would also boost the processing speed and reduce OEO conversions. There can be intra data center communication through which flexibility is provided in managing the network resources with centralized control. This led to introduction of software-defined networks (SDN) control and virtualization of network function in the data center with a flexible photonic data plane. Even the emerging architectures of photonics integrated circuits would greatly impact small footprints and low-power chipsets.

5 Silicon Photonics

Due to the spectrum crunch, recently, the researchers are attracted to the mmWave spectrum. During the mmWave generation on the silicon photonic (SiP)-integrated circuit, two aspects have taken into account while designing the circuit: (i) which

band of frequencies is to be used and what would be the transmission distance of telecommunication transport network and the method used and (ii) the chip upon which mmWave is to be generated must be of low-cost, compact and energy-efficient [12, 13]. It is found that the energy efficiency of SiP circuits is far better than the energy efficiency of the electronic-driven circuit.

The photonic-integrated circuit would help in evolving the 5G radio access networks. Currently, on-chip circuits are connected with copper feed lines, but the photonic integrated circuits enable chip-to-chip interconnect so the bandwidth could be scaled up and demands can be met. They would also reduce the footprint and increase energy efficiency, and reducing costs. Higher data rates can be achieved by using PAM-4 modulation technique discussed in [14].

6 Hybrid Access Networks

To achieve the 5G milestone, advancement to new technologies in hybrid networks is not about removing and replacing; it is all about dynamic convergence, flexibility and coexistence of technologies. The term hybrid networking refers to integrating two or more technologies that operate together to build/design a single network design. For example, consider the combination of wired networking and wireless networking, which provide a single solution [15]. It might also refer to a network design that incorporates two or more fundamental physical topologies.

It may be challenging for operators to upgrade or replace their copper access networks installed in the old downtown areas or specific areas. On the other hand, customers' demands for greater bandwidth are increasing rapidly. Integrating various technologies such as optical, wireless, DSL, and others can be done to overcome this issue. Figure 5 shows the hybrid network of DSL and LTE [16, 17]. Due to the various

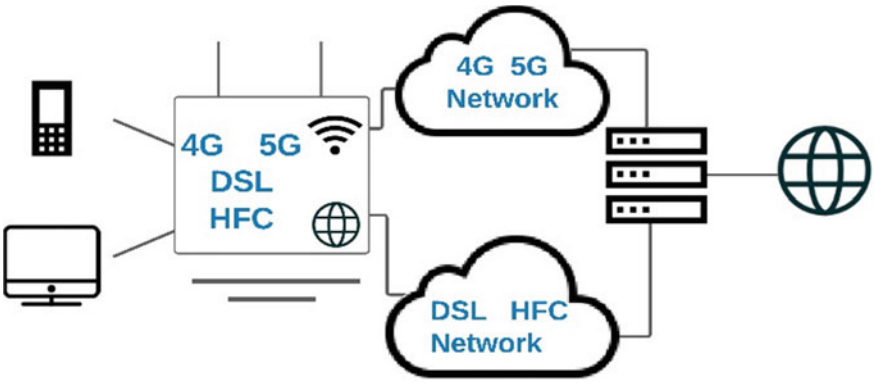


Fig. 5 DSL and LTE hybrid network framework

aspects contributing to the network's heterogeneity, hybrid networks have a way to go.

Multiple access interfaces shall continue to coexist to give operators with the coverage and capacity they require for consistent service throughout its footprint. As we prepare for widespread 5G deployments, 2G and 3G networks continue to run, and 4G will continue to handle the majority of traffic in most markets for many years. Operators may even develop hybrid networks for SD-WANs in which MPLS is tightly connected with LTE and, eventually, 5G. Wi-Fi, in addition, is designed to enhance cellular service by transferring the bulk of traffic to mobile devices, particularly in residential and business interior locations. Moreover, in low power and lossy IoT network, the connection has a high BER and a long inaccessibility time. In such a context, RPL (Routing Protocol for Low-Power and Lossy-Networks) will be quite beneficial. It is highly adaptive to dynamic network conditions as it provides alternate routes if default routes are inaccessible [18].

Considering core, network function virtualization (NFV) and software-defined networking (SDN) can build networks with both physical and virtualized components. The move to virtualized networks, however, some functions continue to rely on legacy technologies and may continue to do so for a long time. Furthermore, virtualization encourages the transition from centralized to distributed architectures that rely on edge and cloud infrastructure. The migration to virtualized, cloud RANs in the RAN will be progressive, with different kinds of RAN configurations and varied levels of functionality located both at the antenna and remotely. Wired and wireless networks will be continuing to coexist in the transportation network to fulfil traffic demand in a cost-effective manner [16].

5G and higher communications networks will deliver ultra-high quality of service. Such networks include both optical wireless networks and radio frequency (RF). Each network, taking into account both RF and optical wireless-based systems, has both amazing characteristics and limitations. Individual network restrictions/drawbacks are overcome by other system in the hybrid network. There are several optical wireless communication (OWC) technologies that can be considered for hybrid network such as visible light communication (VLC), free-space optical communication (FSOC), Li-Fi, optical camera communication (OCC) and for RF technologies, it includes 4G/5G, Wi-Fi, small cell, macrocell.

Optical-RF hybrid network will provide great significant benefits in aspects of handling very high-speed data rates, omnidirectional/directional communication, link reliability, support smooth handover, line-of-sight (LOS) and non-line-of-sight (NLOS) communication, traffic and load balancing, multi-tier networks. Let's take our point on very phenomenal optical-RF networks that would elevate the 5G scenario to new heights. Examples of such optical-RF hybrid systems are discussed below along with the diagram of optical-RF hybrid network system as shown in Fig. 6.

Hybrid VLC and Wi-Fi: In this network, VLC can enable high data rates, although Wi-Fi can provide a larger range of coverage for improved mobility. However, interference could be a problem with this hybrid system.

Hybrid Li-Fi and Wi-Fi: With low-to-medium mobility, the small cell network offers higher QoS, but it cannot handle extremely high data rate links. This hybrid

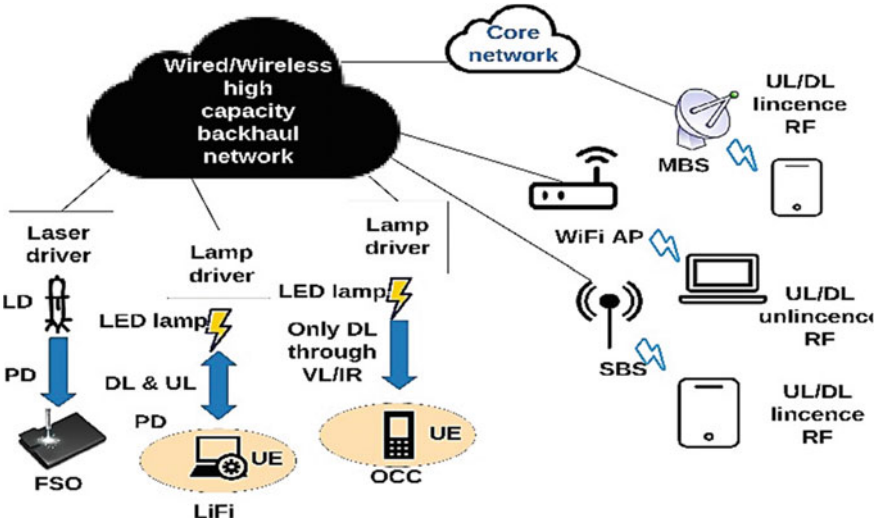


Fig. 6 Hybrid-optical-RF network

system is suitable for an indoor setting. Large amounts of traffic are offloaded from small cell networks to Li-Fi networks.

Hybrid FSOC and RF: In both indoor and outdoor environments, the FSOC system delivers a very high data rate point-to-point communication link. However, atmospheric conditions, notably fog and dust, can create disorder.

Hybrid OCC and Wi-Fi: Localization is done using both the OCC and Wi-Fi systems. However, OCC is performing better, whereas in NLoS situations, the Wi-Fi offers improved accuracy. Hence, this system can be used to improve localization performance.

Hybrid Li-Fi and Macrocell: In an indoor setting, low mobility, and proportionally lower QoS services are linked to Li-Fi networks, while greater mobility and proportionally higher QoS services are linked to macrocell networks.

6.1 Small Cells

5G networks are planned to take advantage of the cloud and virtualized network principles and support vast amounts of data communication between machines, in addition to quicker speeds and higher bandwidth than 4G/LTE networks. In a small dense wireless network with radio cells that are fewer than 10 m^2, a backbone network based on a fiber network design can be an appropriate option for serving these small cells. Backhauling will certainly gain convergence with other access networks, most notably next-generation optical networks, as small cell deployment densification pushes radio access antennas closer to end users.

In terms of capacity and latency, fiber optic backhaul is the greatest solution, particularly for centralized RAN (C-RAN), which employs the common public radio interface (CPRI) interface to transport multi-gigabit-per-second digital radio signals. In backhaul/fronthaul cases involving relatively long distances (>10 km), an optical transport network (OTN) with wavelength-division multiplexing (WDM) and a ring topology is used. For shorter links (<10 km), a point-to-multipoint unified passive optical network (Uni-PON), which uses an optical splitter to aggregate WDM signals from backhaul/fronthaul links of multiple cells, is suggested. As a complementary solution, point-to-point fiber, which yields low fiber resource utilization and weak protection, is also used. Point-to-point fiber, which has a low fiber resource use and weak protection, is also employed as an alternative option [19]. Selection of the new cell and hitless switching of the data stream to the new cell are the processes in the handover procedure. This necessitates synchronization between the central switches' wavelength and space switching. A simple point-to-point architecture with resources allocated per cell is ideal for scenarios with large user densities. This design will be expensive for low user densities.

Small cells are previously used to provide localized services such as hotspots to mitigate the dead spots. Commonly small cells are referred to as distributed antenna systems (DASs) [20]. To achieve the realization of small cells in 5G networks, a new approach of microwave photonics uses both optical and RF to achieve the purpose.

The transport schemes which are used are: (i) RF over fiber, (ii) Intermediate frequency (IF) over fiber, and (iii) Digital over fiber. A fiber–wireless (FiWi) architecture is presented in [21] which relies on efficient analog transport on intermediate frequency over fiber (IFoF). The optical streams are provided to individual RRH, and thus relaxing the burden on 5G mobile fronthaul and make possible to serve many user with high cell density. The transmission is done with the help of analog over fiber interface and give the data rates up to 4 Gbps with each link carry up to 1 Gbps and with the help of WDM technology.

7 Challenges

The 5G network will interconnect millions and billions of new smart devices, which will increase the requirement of backhaul capacity. It will add challenges like ultra-low latency and densification of a cell. As more densification of a small cell to support the huge capacity in 5G network, which will add more challenges for 5G backhaul network. Due to the denser 5G backhaul networks, it will limit the frequency reuse and forcing us to make better utilization of wireless spectrum communication. Currently, available backhaul situations vary as some places have dense fiber connections and some places do not have such. Also, good LoS connectivity is not available for all locations. As 5G will be using huge bandwidth, a denser fiber fronthaul, midhaul, and backhaul network is required. The resulting higher level of multiplexing increases the complexity. 5G promises to deliver heterogeneous services with a single network infrastructure that will support core networks and a radio access network (RAN) to

provide elastic and omnipresent connections. Due to such capability, it makes 5G architecture more complex than ever, and also it necessitates specialized knowledge to develop the architecture that can serve such requirements.

While transferring high data rate signals, there are some challenges associated with infrastructure such as prioritization/resource management, spectrum scarcity, filtering/encryption, internal fiber rollout. Because of the higher priority and real-time applications, if the network path's maximum capacity is reached, lower priority streams will encounter retransmission, packet loss, or disconnection. While filtering and encryption are employed to detect viruses, protect webservers from attacks, and protect users from malicious websites, such processing takes time, limiting data transmission speeds. As with 5G, more and more devices going to connect; but, because the RF spectrum is finite, spectrum scarcity will arise, impacting the connectivity and speed. Moreover, the internal fiber rollout and high-speed switches and routers are required for upgrading the backbone for 5G infrastructure [22].

7.1 Conclusion

5G will be the next major phase of cellular networks, aiming to meet the demands of rapidly expanding data traffic and spectrum scarcity. This paper presented the study of 5G and photonics networks, their integration challenges, and complementary technology. We discussed different technological aspects in the deployment of 5G and photonics fronthaul and backhaul networks. Passive optical networks to small cells would be the pillars of future 5G communication. Moving towards silicon photonics would be playing a crucial role in energy-efficient, ultra-high-speed communication to support IoT applications.

References

1. I.M. Research, https://cdn.ihs.com/www/pdf/IoT_ebook.pdf
2. S.N. Ghorpade, M. Zennaro, B.S. Chaudhari, GWO model for optimal localization of IoT-enabled sensor nodes in smart parking systems. IEEE Trans. Intell. Transp. Syst. **22**(2), 1217–1224 (2021). https://doi.org/10.1109/tits.2020.2964604
3. B.S. Chaudhari, M. Zennaro. (eds.), *LPWAN Technologies for IoT and M2M Applications* (2020)
21. S.N. Ghorpade, M. Zennaro, B.S. Chaudhari, R.A. Saeed, H. Alhumyani, S. Abdel-Khalek, Enhanced differential crossover and quantum particle swarm optimization for IoT applications. IEEE Access 993831-93846. https://doi.org/10.1109/ACCESS.2021.3093113
5. M. Agiwal, A. Roy, N. Saxena, Next generation 5G wireless networks: A comprehensive survey. IEEE Commun. Surv. Tutorials **18**(3), 1617–1655 (2016). https://doi.org/10.1109/comst.2016.2532458
6. J.S. Wey, J. Zhang, Passive optical networks for 5G transport: Technology and standards. J. Lightwave Technol. **37**(12), 2830–2837 (2019). https://doi.org/10.1109/jlt.2018.2856828
7. J.S. Wey, Y. Luo, T. Pfeiffer, 5G wireless transport in a PON context: An overview. IEEE Commun. Stand. Mag. **4**(1), 50–56 (2020). https://doi.org/10.1109/mcomstd.001.1900043

8. I.T. Xg-Pon, https://www.itu.int/en/ITU-T/studygroups/com15/Documents/tutorials/Optical_access_transmission.pdf
9. Z. Tayq, Fronthaul performance demonstration in a WDM-PON-based convergent network, in *2016 European Conference on Networks and Communications (EuCNC)* (2016), pp. 250–254
10. R. Sabella, Silicon photonics for 5G and future networks. IEEE J. Sel. Top. Quant. Electron. **26**(2), 1–11 (2020). https://doi.org/10.1109/jstqe.2019.2948501
11. K.V. Gasse, Analog radio-over-fiber transceivers based on III-V-on-silicon photonics. IEEE Photon. Technol. Lett. **30**(21), 1818–1821 (2018)
12. H. Mohammadhosseini, M.J.R. Heck, Silicon photonics to improve the energy-efficiency of millimeter wave communication systems, in *2017 International Topical Meeting on Microwave Photonics (MWP)* (2017), pp. 1–4
13. S. Bharat, S.S. Chaudhari, Patil (2020) Optimized designs of low loss non-blocking optical router for ONoC applications. Int. J. Inf. Technol. **12**(1), 91–96. https://doi.org/10.1007/s41870-019-00298-7
14. K.V. Gasse, Z. Wang, B.D. Deckere, G. Roelkens, Heterogeneous silicon photonic devices for wireless communication systems, in *2018 2nd URSI Atlantic Radio Science Meeting (AT-RASC), Meloneras* (2018), pp. 1–1
15. Ng-Pon2: https://www.broadband-forum.org/download/MU
16. A.O. Mufutau, F.P. Guiomar, M.A. Fernandes, A. Lorences-Riesgo, A. Oliveira, P.P. Monteiro, Demonstration of a hybrid optical fiber–wireless 5G fronthaul coexisting with end-to-end 4G networks. J. Opt. Commun. Networking **12**(3), 72–72 (2020). https://doi.org/10.1364/jocn.382654
17. M.Z. Chowdhury, M.K. Hasan, M. Shahjalal, T. Hossan, Y.M. Jang, Optical wireless hybrid networks for 5G and beyond communications, in, *2018 International Conference on Information and Communication Technology Convergence (ICTC)* (2018), pp. 709–712
18. J. Hariharakrishnan and B. N, "Adaptability Analysis of 6LoWPAN and RPL for Healthcare applications of Internet-of-Things," *JISMAC*, vol. 2, no. 2, pp. 69–81, May 2021, doi: https://doi.org/10.36548/jismac.2021.2.001.
19. N. Wang, E. Hossain, V.K. Bhargava, Backhauling 5G small cells: A radio resource management perspective. IEEE Wireless Commun. **22**(5), 41–49 (2015). https://doi.org/10.1109/mwc.2015.7306536
20. R. Waterhouse, D. Novack, Realizing 5G: Microwave photonics for 5G mobile wireless systems. IEEE Microwave **16**(8), 84–92 (2015). https://doi.org/10.1109/MMM.2015.2441593
21. E. Ruggeri, A. Tsakyridis, C. Vagionas, G. Kalfas, R.M. Oldenbeuving, P.W.L. van Dijk, C.G.H. Roeloffzen, Y. Leiba, N. Pleros, A. Miliou, A 5G fiber wireless 4Gb/s WDM fronthaul for flexible 360° coverage in V-band massive MIMO small cells. J Lightwave Technol. **39**(4), 1081–1088 (2021). https://doi.org/10.1109/JLT.2020.3029608
22. K. Cengiz, M. Aydemir, Next-generation infrastructure and technology issues in 5G systems. *JCOMSS* **14**(1) (2018). https://doi.org/10.24138/jcomss.v14i1.422

Sensorless Speed Control of BLDC Motor for EV Applications

R. Shanmugasundaram, C. Ganesh, B. Adhavan, A. Singaravelan, and B. Gunapriya

Abstract The limitation of the commonly used back-emf sensorless detection technique is that it is not possible to control the speed of the BLDC motor over a wide speed range as required in electric vehicles. In particular, it is not possible to control at low speeds due to the presence of noise signals near pulse width modulated (PWM) back-emf zero-crossing points at low speeds. Therefore, the duty-cycle of direct sensorless back-emf technique is limited to less than 100% since a minimum turn-off time is required to get the sample of back-emf and apply control action. In this paper, a microcontroller-based enhanced PWM back-emf zero-crossing detection method is proposed to control the speed of the BLDC motor over a wide speed range. The experimental results are presented in this paper for validation.

Keywords BLDC motor · Sensorless · PWM · Back-emf · Zero-crossing · Electric vehicle · Microcontroller

1 Introduction

The BLDC motors are used in electric vehicles, petrol pumps, etc., because of their sparkles operation and speed-torque characteristics that match the load requirements. The direct back-emf detection sensorless control techniques are used to control the

R. Shanmugasundaram (✉)
Sri Ramakrishna Engineering College, Coimbatore 641022, India
e-mail: rsseee@srec.ac.in

C. Ganesh
Rajalakshmi Institute of Technology, Chennai 600124, India

B. Adhavan
PSG Institute of Technology and Applied Research, Coimbatore 641062, India

A. Singaravelan · B. Gunapriya
New Horizon College of Engineering, Bangalore 560103, India
e-mail: singaravelan@newhorizonindia.edu

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_26

speed of the BLDC motor; however, these techniques have the limitation of controlling at low speeds. From the literatures, the current issues and challenges in implementing the sensorless control techniques are: (1) due to low signal–noise ratio in the back-emf, it is not possible to control the speed of BLDC motor at lower speeds, (2) difficult to extract third harmonic voltages, (3) need for virtual neutral point for interfacing with microcontroller, (4) filtering is required for reducing the signal–noise ratio in the back-emf, etc. In order to overcome the above shortcomings, a modified sensorless control scheme combining the back-emf zero-crossing detection during PWM off-time at lower speeds and starting, as well as back-emf zero-crossing detection during PWM on-time at higher speeds has been proposed in this paper. The sensorless control techniques that are reported in the literatures are as follows: In [1], implementation of indirect back-emf zero-crossing detection for BLDCM drive has been discussed and effectiveness in controlling speed and start-up is analyzed. In [2], a new sensorless control approach for BLDCM drive based on voltage difference between lines is discussed and results reveal that speed can be controlled and smooth commutation takes place. In [3], back-emf sensing technique for sensorless operation has been discussed. The effect of noise superimposed on back-emf and the difficulty in implementation this method is highlighted. In [4], sensorless position and speed control techniques has been compared and difficulties in implementation of each back-emf technique is highlighted. In [5], an optimal sensorless control technique for non-ideal back-emf of BLDCM drive fed by buck converter has been implemented and results reveal that speed can be controlled at higher speeds with smooth commutation. In [6], sensorless control technique for BLDM drive has been proposed for low speed applications. In this method, eliminates delay problems and use a amplifying and signal conditioning unit to sense the back-emf at lower speeds. In [7], new approach for the sensorless control of BLDC motor-based average voltage between lines has been discussed. This method eliminates phase shift and need for neutral point. In [8], a new method for sensorless speed control operation by extraction of third harmonic voltages and estimation of speed has been proposed. The results show that wide range speed control of BLDCM drive is possible. In [9–11], optimum design for high speed sensorless operation of BLDC motor using third harmonic voltage has been presented and the results reveal that improvement in the performance of PMBDM drive. In [12, 13], ASIC-based technique for sensorless high speed operation of BLDCM drive applications has been discussed. In this technique, third harmonic voltage is sensed instead of back-emfs and as a result smooth commutation can take place without retardation at high speeds due to conduction of freewheeling diodes. In [14], starting method for sensorless operation has been proposed. In this method, the voltage difference between terminals is amplified and used for the detection of initial rotor position and further zero-crossing of back-emf for triggering the switching devices to achieve desired acceleration characteristics of BLDC motor. In [15], sensorless technique is based on detection of back-emf has been proposed for BLDCM drive. This method requires virtual reference point to sense the back-emf during on-time of PWM and has a disadvantage that speed cannot be controlled over a wide range due to noise superimposed on the back-emfs. A sensorless speed control technique for BLDCM

drive by zero-crossing detection of back-emf has been discussed in [16]. The simulation results reveal that switching loss associated with the power stage is getting reduced by using 4-switch power inverter. In [17], a cost effective sensorless control technique for automotive applications and start-up method has been implemented. In [18], high frequency injection-based sensorless control of BLDC motor has been discussed. The disadvantage of this method is that the stator phase resistance and system delay has negative effect on the performance of BLDC motor drive. The literatures that discuss the modeling, development of low-cost BLDC motor drive and control techniques for performance improvement are as follows: The modeling of BLDC motor drive is discussed in [19, 20], and development of low-cost drive for BLDC motor is discussed in [21], performance improvement of BLDCM drive by fuzzy controller is discussed in [22, 23], performance improvement of BLDCM drive by ANN controller is discussed in [24, 25], and performance improvement of BLDCM drive by compensator is discussed in [26–29].

2 Direct Back-Emf Detection Scheme [6]

Normally, the power input to a three-phase BLDCM drive is controlled by power inverter powered by a battery source. In 120° conduction scheme, only two phases are energized at any time and one phase will be floating at any time. Generally, the upper switch in the legs of power inverter is PWM controlled to control the input power of BLDCM and the lower switch in the legs is switched on and off to perform commutation. This non-conducting phase voltage is sensed to detect the zero-crossing of this voltage during off-time of PWM.

Assuming that current 'i' is flowing through phase B and phase C while they are conducting, PWM switching takes place at phase B high-side switch and normal switching at low-side switch of phase A and phase C is floating as shown in Fig. 1, the equation for back-emf of non-conducting phase A during off-time of PWM can be derived as follows:

Assuming that current 'i' is flowing through phase B and phase C while they are conducting, PWM switching takes place at phase B high-side switch and normal switching at low-side switch of phase C and phase A is floating as shown in Fig. 1, the equation for back-emf of non-conducting phase A during off-time of PWM can be derived as follows:

Considering phase B, neglecting forward voltage drop of the diode, the expression for neutral voltage (V_n) can be written as,

$$v_n = -Ri - L\frac{di}{dt} - e_b \tag{1}$$

Also V_n can be written as,

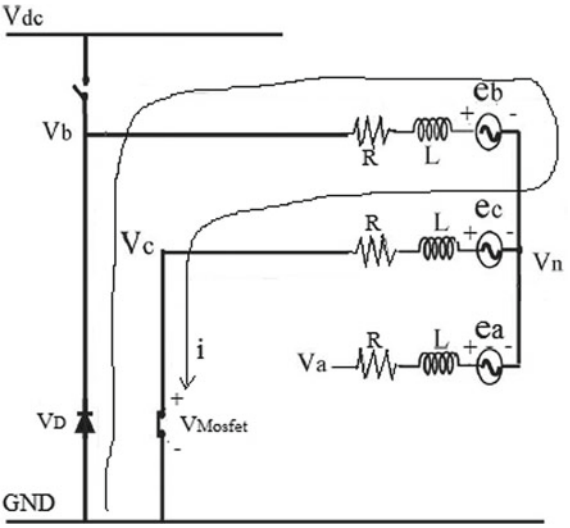


Fig. 1 Phase B and C are conducting current during PWM off-time

$$v_{\mathrm{n}} = Ri + L\frac{\mathrm{d}i}{\mathrm{d}t} - e_c \tag{2}$$

When Eqs. (1) and (2) are added, we get,

$$v_{\mathrm{n}} = \frac{-(e_b + e_c)}{2} \tag{3}$$

Assuming that the three-phase BLDCM is a balanced, we get,

$$e_a + e_b + e_c = 0 \quad \text{or} \quad e_b + e_c = -e_a \tag{4}$$

Substituting (4) in (3), we get

$$v_{\mathrm{n}} = \frac{e_a}{2} \tag{5}$$

$$e_a = 2v_{\mathrm{n}} \quad \text{or} \quad v_{\mathrm{n}} = \frac{e_a}{2} \tag{6}$$

The equation for phase voltage v_a is

$$v_a = e_a + v_{\mathrm{n}} \tag{7}$$

Substituting (6) in (7), we get,

$$v_a = \frac{3}{2}e_a \tag{8}$$

It is clear from the above equation that floating phase voltage is proportional to its back-emf and can be easily sampled during PWM off-time of high-side switch without the need for filtering and neutral point voltage (v_n). Due to switching noise near the end of PWM off-time, minimum time is required to sense the back-emf, and as a result the maximum duty-cycle of PWM is limited to 90% for the switching frequency of 20 kHz. Hence, by this method it is only possible to control the speed of the BLDCM upto rated speed and cannot utilize the full battery voltage. This problem can be overcome if detection of back-emf is carried out during PWM on-time as explained in Sect. 3.

3 Modified Direct Back-Emf Detection Scheme

Now let us assume that phase B and phase C are conducting current as shown in Fig. 2 and the derive the equation for back-emf [17] of non-conducting phase A during PWM on-time of high-side switch of phase B.

Considering phase B, neglecting forward voltage drop of the MOSFET, the expression for neutral voltage (v_n) can be written as,

$$v_n = V_{dc} - Ri - L\frac{di}{dt} - e_b \tag{9}$$

Also v_n can be written as,

$$v_n = Ri + L\frac{di}{dt} - e_c \tag{10}$$

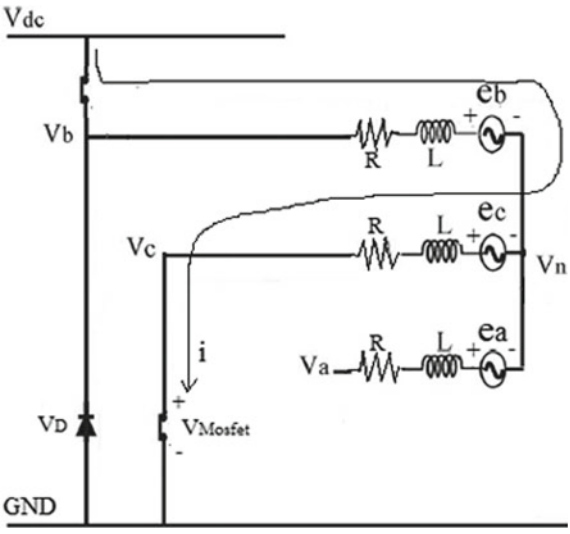


Fig. 2 Phase B and C are energized during PWM on-time

When Eqs. (9) and (10) are added, we get,

$$v_{\mathrm{n}} = \frac{v_{\mathrm{dc}}}{2} - \frac{(e_b + e_c)}{2} \tag{11}$$

Assuming that the three-phase BLDC motor is balanced, we get,

$$e_a + e_b + e_c = 0 \quad \text{or} \quad e_b + e_c = -e_a \tag{12}$$

Substituting (12) in (11), we get

$$v_{\mathrm{n}} = \frac{v_{\mathrm{dc}}}{2} + \frac{e_a}{2} \tag{13}$$

The equation for phase voltage v_a is

$$v_a = e_a + v_{\mathrm{n}} \tag{14}$$

Substituting (13) in (14), we get,

$$v_a = \frac{v_{\mathrm{dc}}}{2} + \frac{3}{2} e_a \tag{15}$$

It is clear from the above equation that it is required to use comparator for comparing the terminal voltage with the reference voltage ($V_{\mathrm{dc}}/2$) to detect the actual back-emf zero-crossing points. This technique can be used to control the BLDCM at higher speeds. So, for effective control the BLDCM over a wide speed range, the better choice would be to combine two methods together by employing back-emf detection during off-time of PWM at lower speeds and starting, and back-emf detection at high speeds during PWM on-time.

4 Hardware Implementation for the Proposed System

Figure 3 shows the hardware layout the combined PWM on-time and PWM off-time back-emf zero-crossing detection schemes. Figure 4 shows the hardware implementation for phase and reference voltage selection. The specifications of BLDC motor are: 3 Phase, 36 V, 5A, 4000 RPM, and 0.082 N m/A.

This system consists of multiplexer, PIC microcontroller and power inverter. The function of multiplexer is to connect the floating phase to the comparator input to compare the floating phase back-emf with reference voltage which is set to zero for zero-crossing detection during PWM off-time and then set to $V_{\mathrm{dc}}/2$ for zero-crossing detection during PWM on-time through the reference selection multiplexer. Based on the commutation sequence, the software connects the appropriate floating phase to the comparator for back-emf zero-crossing detection and makes the smooth

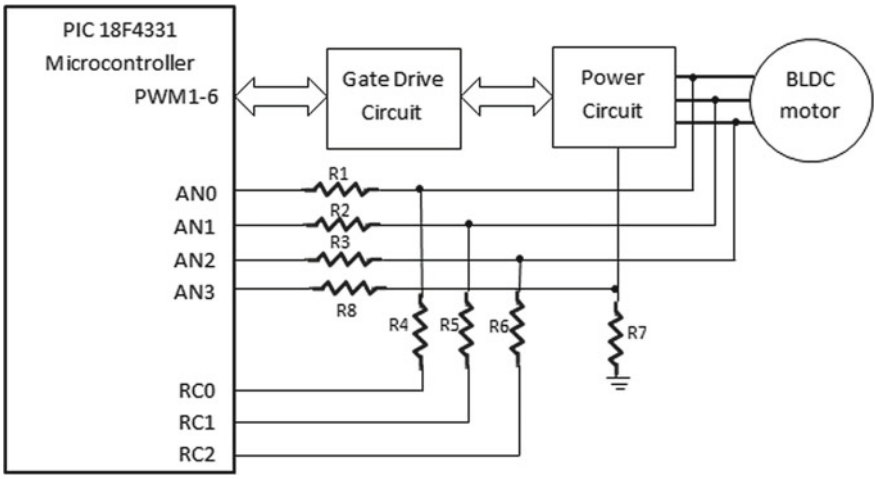


Fig. 3 Hardware layout for the combined zero-crossing detection scheme

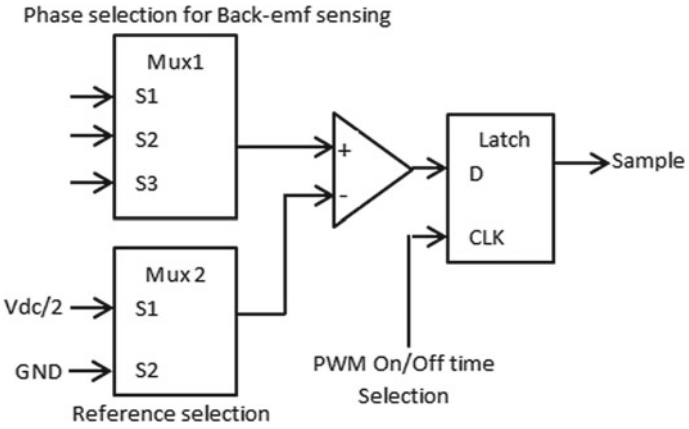


Fig. 4 Phase and reference selection for the proposed scheme

transition to take place from outgoing phase to incoming phase. The low-cost PIC 18F4331 microcontroller is used to implement the proposed method of zero-crossing detection during PWM on-time at lower speeds upto 60% of rated speed and at the time of starting, and PWM off-time for speeds above 60% of rated speed. The latch is enabled through software to latch the comparator output to collect samples either during on-time or off-time of PWM.

The phase voltages are applied to the high input impedance pins (AN0, AN1, and AN2) through the current limiting resistive network to detect the zero-crossing of back-emf without any attenuation during the starting and at lower speeds. The

comparator reference input is set near to zero. Similarly, during higher speeds, amplitude of back-emf is reduced and applied as one of the input to the comparator and the other input is set to $V_{dc}/2$ by the resistive potential divider for zero-crossing detection of back-emf.

Figures 5 and 6 show the floating phase back-emf, PWM gating pulses and PWM switched motor current during PWM on-time and PWM off-time switching, respectively. It is observed from Fig. 5 that the back-emf signal becomes low when the

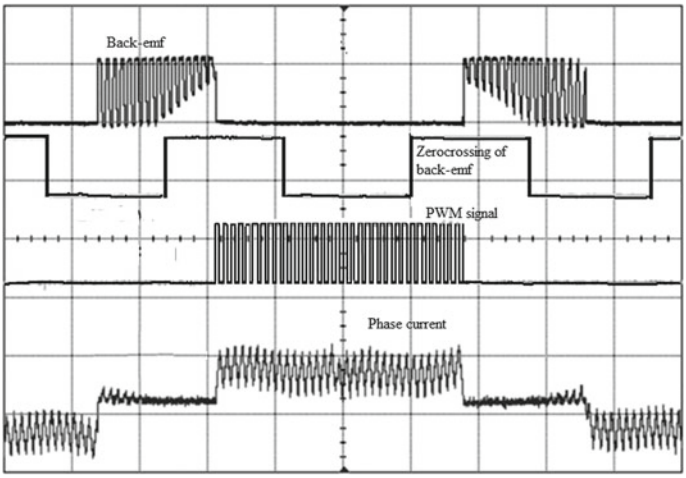


Fig. 5 Waveforms for back-emf sensing during PWM off-time (back-emf [30 V/div], back-emf zero-crossing [5 V/div], PWM signal [5 V/div], and phase current [2.5 A/div])

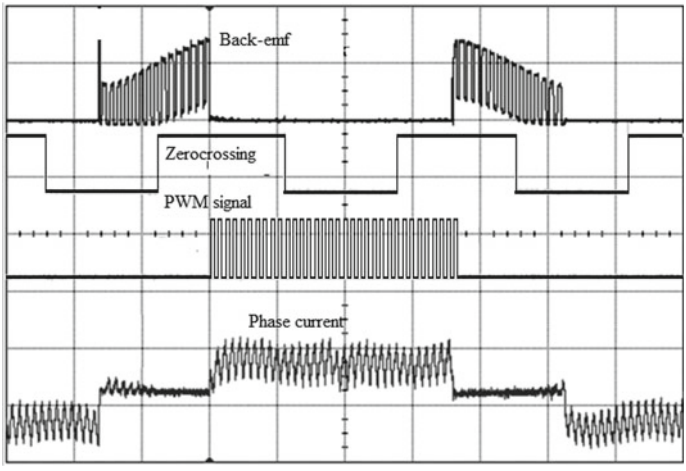


Fig. 6 Waveforms for back-emf sensing during PWM on-time (back-emf [30 V/div], back-emf zero-crossing [5 V/div], PWM signal [5 V/div], and phase current [2.5 A/div])

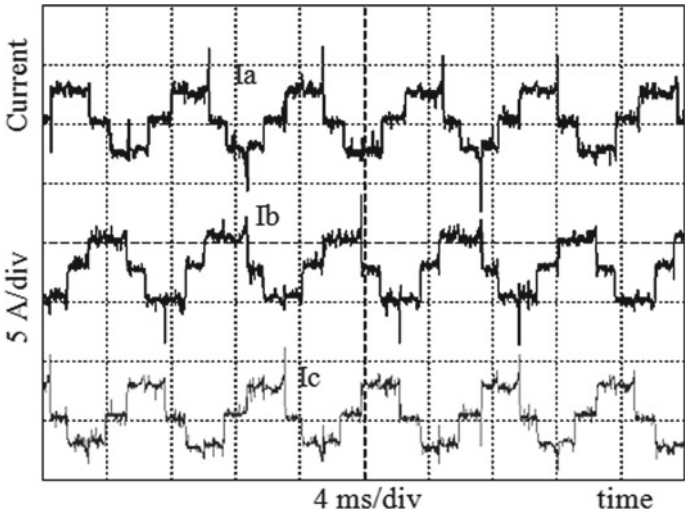


Fig. 7 Phase current waveforms

same phase is conducting current due to grounding of terminal through the resistor. Every transition of back-emf signal indicates the detection of zero-crossing. Figure 6 shows that attenuated back-emf signal during the high speed operation in the PWM on-time sensing scheme. It is clear from this waveform that it is possible to control the BLDC motor at maximum speed by varying duty-cycle up to 100%. The phase currents waveforms are shown in Fig. 7 and zero-crossing detection of back-emfs is shown in Fig. 8.

5 Starting Method and Current Sensing

The following start-up schemes are generally used for start-up of the motor: (i) starting in open-loop by energizing the stator winding at a predetermined increasing frequency profile, (ii) Starting the motor by an external motor. The first starting method is used in this proposed system. In this method, stator phase windings are energized in a particular sequence at the predetermined step size and increasing frequency profile. As the motor attains certain speed and generate sufficient back-emf, the controller is switched to the usual detection technique to get the information on position and turn on the appropriate MOSFETs in the power inverter to energize the phase windings for getting continuous rotation. In case, if the motor direction is opposite, then the microcontroller initiates braking mode operation to stop the motor by sensing the phase sequence. Then, the controller changes the phase sequence and start the motor by the same starting algorithm. As the microcontroller needs a reference point to sense the back-emf signals, a star connected resistive network is

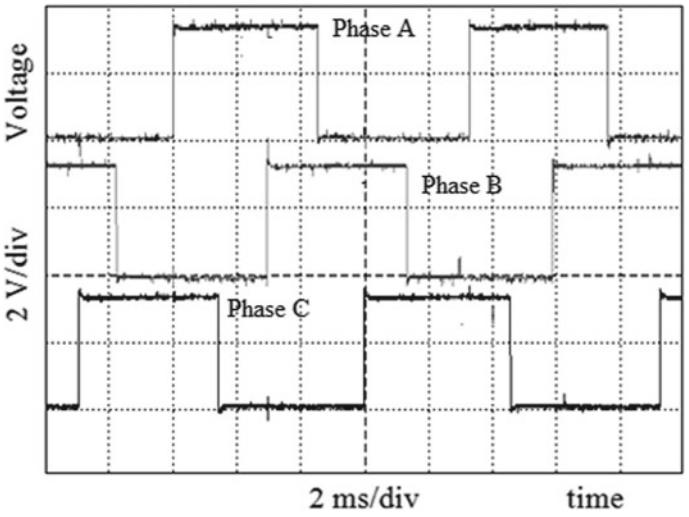


Fig. 8 Zero-crossing of back-emfs

formed as shown in Fig. 9. This method of starting can be applied to motors connected with small and medium sized loads by proper tuning of commutation time. In the second method, back-emf zero-crossing detection is carried out by using an external motor to rotate the motor till sufficient back-emf is generated.

The commonly used method to protect the motor under overloading conditions and stalled conditions is to sense the current flowing through the phase windings and turn-off the MOSFET switches under abnormal conditions. In this proposed system, a low resistance (Rsense) is placed in series with negative link terminal of power inverter. The voltage generated across this resistance is proportional to the current flowing through the phase windings. Therefore, the controllers sense this voltage and provide protection to the motor.

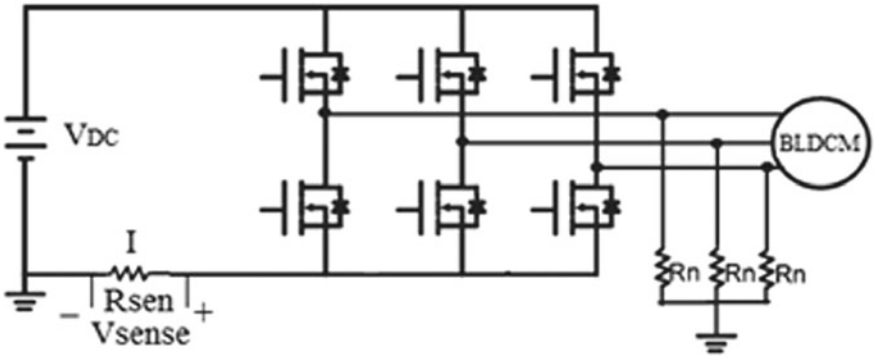


Fig. 9 Power stage of BLDC motor with reference point and current sensing resistor

6 Conclusion

The proposed optimized back-emf zero-crossing detection technique that combine detection of PWM off-time during starting and lower speeds, and PWM on-time during higher speeds has been implemented. The experimental results show that by combining the two methods, speed can be controlled over a wide range from zero to upto rated speed. By this proposed scheme, it is possible to overcome the limitation of existing back-emf zero-crossing detection method wherein minimum time required for sensing back-emf zero-detection during PWM off-time, hence limiting the maximum duty-cycle of MOSFET switches in the power inverter and the speed control range. The proposed starting method and protection schemes are found to be effective and enhance the performance of the BLDC motor drive system. Hence, this proposed is suitable for controlling the speed of the BLDC motor over a wide speed range in electric vehicle applications.

References

1. P. Damodharan, K. Vasudevan, Indirect back-EMF zero crossing detection for sensorless BLDC motor operation, in *Proceedings of IEEE Power Electronics and Drive Systems* (2005), pp. 1107–1111
2. T. Kim, C. Kim, J. Lyou, A new sensorless drive scheme for a BLDC motor based on the terminal voltage difference. IEEE Trans. Power Electron. **39**(8) (2011)
3. K. Kroics, J. Zakis, U. Sirmelis, Implementation of the back EMF zero-crossing detection for BLDC motor, in *IEEE 58th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON)* (Riga, 2017), pp. 1–4
4. R.M. Pindoriya, A.K. Mishra, B.S. Rajpurohit, R. Kumar, Analysis of position and speed control of sensorless BLDC motor using zero crossing back-EMF technique, in: *IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)* (Delhi, 2016), pp. 1–6
5. X. Chen, G. Liu, Sensorless optimal commutation steady speed control method for a nonideal back-EMF BLDC motor drive system including buck converter. IEEE Trans. Industr. Electron. **67**(7), 6147–6157 (2020)
6. J. Shao, D. Nolan, T. Hopkins, Improved direct back EMF detection for sensorless brushless DC (BLDC) motor drives, in *18th Annual IEEE Applied Power Electronics Conference and Exposition*, vol 1 (Miami Beach, FL, USA, 2003), pp. 300–305
7. C.-H. Chen, M.-Y. Cheng, New cost effective sensorless commutation method for brushless dc motors without phase shift circuit and neutral voltage. IEEE Trans. Power Electron. **22**(2), 644–653 (2007)
8. Y.H. Yoon, T.W. Lee, S.H. Park, B.K. Lee, C.Y. Won, New approach to rotor position detection and precision speed control of the BLDC motor. IEEE Ind. Electron. 1305–1310 (2006)
9. Z.Q. Zhu, J.D. Ede, D. Howe, Design criteria for high-speed brushless dc motors for sensorless operation. Int. J. Appl. Electromagnet. Mech. **15**, 79–87 (2002)
10. K. Wang, M.A. Rahman, J.X. Shen, *Control of High-Speed Sensorless PM Brushless DC Motors* (IEEE Industry Applications Society Annual Meeting, Houston, TX, USA, 2010)
11. K. Wang, J.X. Shen, F.Z. Zhou, W.Z. Fei, Design aspects of a high-speed sensorless brushless dc motor using third harmonic back-emf for sensorless control. J. Appl. Phys. **103**(7), 07F110-07F113 (2008)

12. J.X. She, S. Iwasaki, Sensorless control of ultrahigh-speed PM brushless motor using PLL and third harmonic back EMF. IEEE Trans. Ind. Electron. **53**(2), 421–428 (2006)
13. J.X. Shen, S. Iwasaki, Improvement of ASIC-based sensorless control for ultra high-speed brushless DC motor drive, in *Proceedings of the IEEE International Electric Machines and Drives Conference* (2003), pp. 1049–1054
14. P. Damodharan, R. Sandeep, K. Vasudevan, Simple position sensorless starting method for brushless DC motor. IEEE Electro. Power Appl. **2**(1), 49–55 (2008)
15. J. Shao, D. Nolan, T. Hopkins, A novel direct back EMF detection for sensorless brushless DC (BLOC) motor drives, in *Applied Power Electronic Conference* (2002), pp. 33–37
16. K.M.Reshma, T.B. Isha, A back-EMF based sensorless speed control of four switch BLDC motor drive, in *International Conference on Control, Power, Communication and Computing Technologies* (Kannur, India, 2018), pp. 283–287
17. J. Shao, An improved microcontroller-based sensorless brushless DC (BLDC) motor drive for automotive applications, in *14th IAS Industry Applications Conference. Hong*, vol. 4 (Kong, China, 2005), pp. 2512–2517
18. B. Tian, Q. An, M. Molinas, High-frequency ınjection-based sensorless control for a general five-phase BLDC motor ıncorporating system delay and phase resistance. IEEE Access **7**, 162862–162873 (2019)
19. R. Shanmugasundram, K.M. Zakaraiah, N. Yadaiah, Effect of parameter variations on the performance of direct current (DC) servomotor drives. J. Vib. Control **19**(10), 1575–1586 (2012)
20. R. Shanmugasundram, K.M. Zakariah, N. Yadaiah, Modeling, simulation and analysis of controllers for brushless direct current motor drives. J. Vib. Control **19**(8), 1250–1264 (2012)
21. R. Shanmugasundram, K.M. Zakariah, N. Yadaiah, Low-cost high performance brushless DC motor drive for speed control applications, in *International Conference on Advances in Recent Technologies in Communication and Computing* (Kottayam, India, 2009), pp. 456–460
22. R. Shanmugasundram, K.M. Zakariah, N. Yadaiah, Digital implementation of fuzzy logic controller for wide range speed control of brushless DC motor, in *IEEE International Conference on Vehicular Electronics and Safety (ICVES)* (Pune, India, 2009), pp. 119–124
23. R. Shanmugasundram, K. Muhammad Zakariah, N. Yadaiah, Implementation and performance analysis of digital controllers for brushless DC motor drives. IEEE/ASME Trans. Mechatron. **19**(1), 213–224 (2012)
24. N. Leena, R. Shanmugasundaram, Artificial neural network controller for improved performance of brushless DC motor, in *International Conference on Power Signals Control and Computations* (Thrissur, India, 2014), pp. 1–6
25. R. Shanmugasundaram, C. Ganesh, A. Singaravelan, ANN-based controllers for improved performance of BLDC motor drives, in *Proceedings of International Conference on Advances in Electrical Control and Signal Systems 2019, LNEE* vol 665 (Springer, Singapore, 2020), pp. 73–87
26. C. Ganesh, R. Shanmugasundaram, Design of non-iterative first order compensator for type-1 higher order systems, in *Proceedings of the 2nd International Conference on Communication, Devices and Computing 2020, LNEE* vol 602 (Springer, Singapore, 2020), pp. 355–367
27. C. Ganesh, S.K. Patnaik, Artificial neural network based proportional plus integral plus derivative controller for a brushless DC position control system. J. Vib. Control **18**(14), 2164–2175 (2012)
28. C. Ganesh, S.K. Patnaik, A simple first order compensator for brushless direct current drive based position control system. J. Vib. Control **21**(4), 647–661 (2015)
29. B. Tan, X. Wang, D. Zhao, K. Shen, J. Zhao, X. Ding, A lag angle compensation strategy of phase current for high-Speed BLDC motors. IEEE Access **7**, 9566–9574 (2019)

An IoT-Based Automated Smart Helmet

Shachi Sinha, Eesha Teli, and Washima Tasnin

Abstract Road accidents are on the rise as a result of motorists failing to wear helmets and drinking excessively. Wearing a helmet could minimise impact shock and potentially save a life. As a result, this paper is initiated to enhance the motorcycle rider's safety. The system comprises of an Arduino Uno microcontroller, a Node Microcontroller Unit (NodeMCU) ESP8266 Wireless Fidelity (Wi-Fi) module, Global Positioning System (GPS) sensor, Gyroscope, Infrared (IR) sensor, Alcohol Sensor and a Blynk Internet of Things (IoT) mobile application. This work is a comprehensive accident detection system that relies on a smart helmet that can sense whether or not the rider is wearing a helmet, track the rider's co-ordinates via real-time location tracking, recognise sudden severe jerks to deduce that an accident has occurred, assess whether or not the rider is drunk, and alert the rider's relatives and friends via a mobile application.

Keywords Smart helmet · Arduino Uno · Gyroscope · Blynk IoT app · IR sensor · Alcohol sensor · GPS module · Accident detection · Location tracking · Internet of Things · Node MCU

1 Introduction

Two-wheeler crashes are becoming more common by the day, resulting in the loss of many lives. According to a study conducted, in India, there are approximately 1214 deaths caused by road accidents each year. Of these, two-wheelers account for around 25% of these crashes [1]. There may be a variety of causes, including a lack of proper driving experience, bike fitness, fast driving, driving while inebriated, and so on. If crashes are one reason, casualties are also attributed to a lack of timely care. According to a report [2], almost half of all injured people die as a result of delayed diagnosis. Many factors contribute to this, including the late arrival of an ambulance and the absence of a personnel at the crash scene to notify the family or provide first

S. Sinha (✉) · E. Teli · W. Tasnin
Vellore Institute of Technology, Vellore, Tamil Nadu, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_27

aid to the injured. This is a problem that we encounter on a daily basis, and the need to find a solution to this issue prompted us to develop the concept of providing accident information as soon as possible. The following are the three main remedies that will help prevent accidents [3]: Firstly, making helmet use obligatory. Secondly, ensuring no driving when inebriated. Lastly, if an individual is involved in an accident and no one is there to assist him. In such a case, sending a text message to an ambulance or family members will help save his life. A considerable amount of work has been done in this domain. Rosenberg et al. [4] proposed a model of smart helmet which aimed at studying electrocardiography (ECG) and electroencephalogram (EEG) data using a motorbike helmet itself which could help to monitor the vitals of the rider when worn. This model implemented the installation of electrodes at different parts of the everyday helmet particularly near the jaws and forehead to monitor the vitals of the wearer continuously whereas at the same time, the model aimed at providing utmost comfort and convenience of usage to the helmet wearer. This was a novel idea aimed at tracking the rider's cardiac activity, his neural activities as well as his respiratory actions. However, the proposed model had no specific features to enable the bike rider's safety during driving or detecting an accident. Agarwal et al. [5]; Sreeja et al. [6] and Aravinda et al. [7] developed models of smart helmet which primarily focussed on the rider safety by ensuring that the bike ignition is not switched on until the helmet is mounted on the rider's head. A radio-frequency (RF) Transmitter is embedded in the helmets which is synchronised with an RF receiver installed in the bike unit. As the rider puts on the helmet, the RF transmitter senses it and irradiates an RF signal which is identified by the RF receiver using address matching which then enables the rider to switch on the bike ignition. Despite this, the proposed projects lacked any feature which could help in accident detection of the rider. Hegde et al. [8]; MohanRoopa et al. [9]; Rao et al. [10] and Eswari et al. [11] proposed a unique model which could help save thousands of lives via its alcohol detection and accident intimation framework. These smart helmet models incorporated alcohol sensors which could detect whether the driver was drunk or not, and they also had GPS and GSM modules, which could help in location tracking of the rider. They had a sensor to detect the tilt and sense whether or not the accident has occurred. The GSM module sent this message to the rider's relatives, friends and the nearest police station to alert about the rider's accident. However, these models lacked a mobile application as an alert system, because in present times people may or may not check an SMS, but a notification alert is a more seamless and infallible alerting process, which was incorporated in our model. Dubey et al. [12] introduced a smart helmet which provided advanced safety features for the riders. It was equipped with an ultrasonic sensor which could detect and warn the rider about an approaching vehicle, a pressure sensor to detect intense pressures in times of an accident, an alcohol sensor to detect if the rider is drunk, a GPS module for location tracking and a GSM module for alerting the ambulance and the riders' relatives in the unforeseen occurrence of an accident but most importantly the model had a rear view camera which alerted the rider when the rider changed lanes while driving. Despite such novel features, there wasn't any indication on whether or not the rider has worn the helmet as well as it did not have any provision of detecting accidents by sensing strong jerks.

Zhuang et al. [13] and Jurado et al. [14] presented a model of smart helmet equipped with a fibre optics sensor to detect blunt impacts on the helmet and alert that an accident has occurred. A single fibre Bragg grating (FBG)-embedded smart helmet prototype was introduced in this paper for rapid identification of blunt-force impact events. Real-time sensing was made possible by high-rate data processing at 5 kHz. For the precise estimation of blunt-force impact events, standalone and ensemble Machine Learning (ML) models were used. Using the ensemble ML models, high prediction accuracies were obtained for both the impact energy levels and directions. This model could be able to provide precise real-time guidelines for diagnoses of blunt-force effect cases. However, this model had no provision for alerting the near and dear ones or the ambulance regarding the occurrence of a mishap. It also could not detect accidents in case of sudden severe jerks, it as well did not provide location tracking of the driver neither did it indicate whether the rider was wearing the helmet or not. Sivagnam et al. [15] designed a helmet which was aimed at providing rider comfort and ease while riding. The helmet had a Bluetooth speaker controller, a speaker and a microphone to help the rider listen to music as well as communicate with the pillion rider without turning back. It also has an airflow compartment to enable the rider to breathe easily as well as it is powered by solar energy. However, the helmet does nothing to detect accidents or track the rider location neither does it detect whether the helmet is worn by the rider or not. Eldho et al. [16] also introduced a jacket cum helmet to enhance the wearers' safety. It enabled temperature regulation of the helmet and jacket by installing Peltier's module so as to perform temperature regulation so that the user does not feel stuffy and uncomfortable. Sathish et al. [17] introduced a smart bike system which could help detect bike as well as fuel theft. It also indicates the fuel quantity in the tank and alerts when fuel level is low. It also does auto speed monitoring and control as well as automatically turns on and off the headlights at night. It also does alcohol detection using helmet as well as does not switch on the bike ignition until the helmet is mounted. Despite all of these features, the model wasn't entirely based upon the helmet as the sensors like GPS and GSM which could help in tracking and locating the rider were mounted on the bike and not on the helmet. Moreover, there wasn't any provision for accident detection and alerting. While much work has been done from literatures [3–17] but none of them have a compact system which can indicate on whether or not the helmet is worn, track the rider location as well as detect sudden severe jerks to identify that an accident has occurred, indicate if the rider is drunk or not and also alert the riders' relatives and friends via a mobile application. Therefore, our work here aims at formulating such a complete accident detection system via a smart helmet and its main objective is to design a helmet.

- to automatically indicate whether or not the helmet is mounted on the head with the help of an IR sensor and
- to help detect accidents primarily using the intensity of movement of the helmet with the help of gyroscope and Blynk IoT app
- to equip the helmet with GPS and gyroscope to successfully track the driver's location and send an alert to rider's relatives if a mishap occurs.

2 Methodology

This IoT enabled smart helmet [18] is a unique, convenient and cost-effective proposition which makes use of low-cost easily available components to enhance the safety of a two-wheeler rider. The components used in our helmet are Arduino Uno which acts as a central microcontroller controlling all the major functions of the helmet, a Node MCU to connect the helmet with the mobile application over the Internet, a GPS and GSM sensor for continuous location tracking and IR sensor to detect if the helmet is worn or not.

2.1 Components Used

Arduino Uno. Arduino as shown in Fig. 1 is a free and open-source microcontroller that can be programmed, erased, and reprogrammed at any time. The Arduino platform, which debuted in 2005, was created to provide hobbyists, students, and professionals an affordable and simple way to build devices that communicate with their surroundings using sensors and actuators. It is an open-source computing framework for building and programming electronic devices that is based on basic microcontroller boards. It can also function as a mini computer, taking inputs and manipulating outputs for a number of electronic devices, same as most microcontrollers [19]. The Arduino board, which governs and handles all functions performed by the other components of the device, is the central element of the model. The instructions for the components are provided in the programming language of Arduino. Arduino consists of both a physical programmable circuit board and a piece of software or integrated development environment (IDE) running on a computer that we use to write computer code and upload it to the physical board. It runs on the ATMEGA328P microcontroller.

Node MCU. NodeMCU, depicted in Fig. 2 is an open-source framework with an open hardware architecture that anybody can edit, change, or create. The ESP8266 Wi-Fi enabled chip is included in the NodeMCU Dev Kit/board. The ESP8266 is a low-cost Wi-Fi chip with TCP/IP protocol developed by Espressif Systems [11]. The ESP8266 is a low-cost Wi-Fi module chip that can be programmed for Internet of Things (IoT)

Fig. 1 Arduino Uno. *Source* https://en.wikipedia.org/wiki/Arduino_Uno

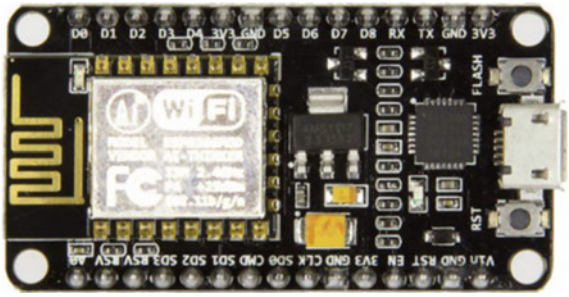


Fig. 2 Node MCU. *Source* https://www.elektor.com/nodemcu-microcontroller-board-with-esp8266-and-lua

and related hardware ventures to connect to the internet. It is fundamentally a System on Chip (SoC); an integrated circuit that incorporates all components of a computer or other electronic systems is a SoC [20]. It contains the TCP/IP protocol, which allows access to Wi-Fi networking. It can host an application or disable all Wi-Fi networking features on any application machine. A switch to the outside can trigger flash memory. Built-in cache memory reduces memory use and helps to improve device speed. Another situation is wireless internet connectivity, which recognises the Wi-Fi adapter's role and can be added to any microcontroller-based design with ease of communication, thanks to the SCI/SDIO framework. It's a sophisticated system with a powerful computing system and plenty of storage [21]. Basically, standard electrical and mechanical systems are unable to connect to the Internet on their own. This is because they don't have the inbuilt configuration. Thus, with these devices, we can configure ESP8266 to manage and track them. NodeMCU is an ESP82666 firmware module. It thus shows whether the helmet is worn or not through a smartphone application and provides the wearer's position by connecting the system over Wi-Fi thereby enabling easy monitoring of the helmet remotely.

GPS Module. The Global Positioning System (GPS) is a system made up of a network of 24 US satellites that were first used for military purposes before being made available for civilian use. The satellites send out brief radio signals to GPS receivers on a regular basis. To measure altitude, a GPS receiver collects signals from at least three satellites and uses a triangulation technique to calculate its two-dimensional (latitude and longitude) location. If a location has been calculated, the average speed and direction of travel can be calculated. As a result, GPS is an important technology for determining a device's location [22]. GPS modules include tiny processors and antennas that collect data directly from dedicated RF frequencies transmitted by satellites. From there, along with other pieces of evidence, it will collect timestamps from each visible satellite, used to locate the position of vehicle.

Gyroscope. Gyroscopes are instruments that are placed on a frame that can detect angular momentum as the frame rotates. There are several different types of gyroscopes, based on the operating physical theory and the technologies used. Gyroscopes may be used on their own or as part of more complex devices [23]. A Gyroscope may be interpreted as a gadget used in navigation, stabilisers, etc. to maintain a reference

position to provide stabilisation. It, therefore senses sudden jerks experienced by the helmet which in turn help in detecting a mishap.

Infrared Sensor. IR radiation is a form of electromagnetic radiation with a wave length that is longer than visible light but shorter than microwaves. IR radiation has wavelengths ranging from 750 nm to 1 mm, covering a five-order-of-magnitude range [24]. IR sensor makes use of the infrared radiations to detect an object in its proximity. This setup consists of two LED bulbs that emit an infrared signal continuously. This pair of leads is capable of receiving infrared signals. As a result, the two lead lamps are mounted within the helmet in a straight line. As a result, it continually transmits the signal, with the comparator noting any signal deviations [25]. Therefore, it detects whether or not the helmet is worn by the wearer and hence helps to give a warning if the helmet is not worn by the driver.

Alcohol Sensor. MQ3 sensor is a Metal Oxide Semiconductor also known as chemiresistors. Since sensing is based on the sensing material's resistance changing when exposed to alcohol, alcohol concentrations can be found by using it in a simple voltage divider network. The alcohol Sensor (MQ3) module can be used to detect gas leaks and can also detect alcohol, benzine, CH4, hexane, LPG, and CO. Measurements can be made as soon as possible because of its fast reaction time and high sensitivity. The potentiometer can be used to change the sensor's sensitivity. The MQ3 gas sensor is an alcohol sensor that detects the presence of alcohol in the breath. Dependent on the amount of alcohol in the system, this sensor produces an analogue resistive output [26]. The MQ3 alcohol sensor operates at 5 V DC and consumes about 800 mW. It can detect alcohol concentrations ranging from 25 to 500 parts per million (ppm) [27].

Blynk IoT Application. Blynk is an IoT platform whose screenshot is attached in Fig. 3; is used for iOS and Android devices that can monitor Arduino, Raspberry Pi, and NodeMCU over the Internet. By compiling and including the required address on the available widgets, this application is used to construct a graphical interface or human machine interface (HMI) [28]. Blynk is an IoT platform with white-label mobile apps, private clouds, device management, data analytics, and machine learning that is hardware agnostic. Blynk was created with the Internet of Things in mind. It has the ability to control hardware remotely, display sensor data, store data, and visualise it [29].

2.2 Procedure

The sensors like GPS module, MQ3 gas sensor, gyroscope as well as IR sensor are connected to the Arduino Uno microcontroller. The Arduino is provided with power supply. The Node MCU Wi-Fi module is also connected to the Arduino Uno and it interfaces with the Blynk mobile application via a cloud-based server.

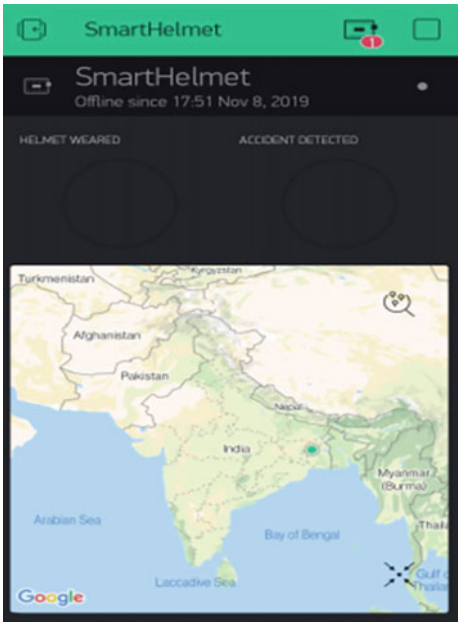


Fig. 3 Blynk IoT application screenshot

In our model which is depicted in Fig. 5, we make use of the above-mentioned hardware as well as Blynk IoT mobile application to successfully perform the accident detection and location tracking of a biker in case of a mishap. The Arduino Uno microcontroller performs the relaying of information to and from the sensors to the Wi-Fi module ESP-8266 which then transmits this information to the Blynk IoT app. As soon as the rider wearing the "smart helmet" meets with an accident, a notification is sent to the relatives of the rider about the rider's location. The helmet is equipped with an IR sensor which detects the heat from the rider's head and accordingly sends a notification to the Blynk application installed by the rider's close ones on whether the rider is wearing the helmet or not. If the rider wears helmet, the app button turns green and if he is not, the Arduino accordingly sends the signals it received from the IR sensor to the Wi-Fi module and it in turn alerts the riders near ones about the status of helmet worn. The gyroscope present in the helmet detects sudden heavy jerks experienced by the helmet. Thus, in case of a mishap, the gyroscope will sense the jerk experienced and accordingly signal the Arduino which will relay this message to the Node MCU which in turn will immediately send a notification to the rider's close ones notifying them about the rider's accident. The GPS Module will simultaneously help the close ones to locate the rider, thus ensuring timely medical treatment of the rider. Therefore, with the use of this novel model, several lives can be saved due to timely approach of medical help (Fig. 4).

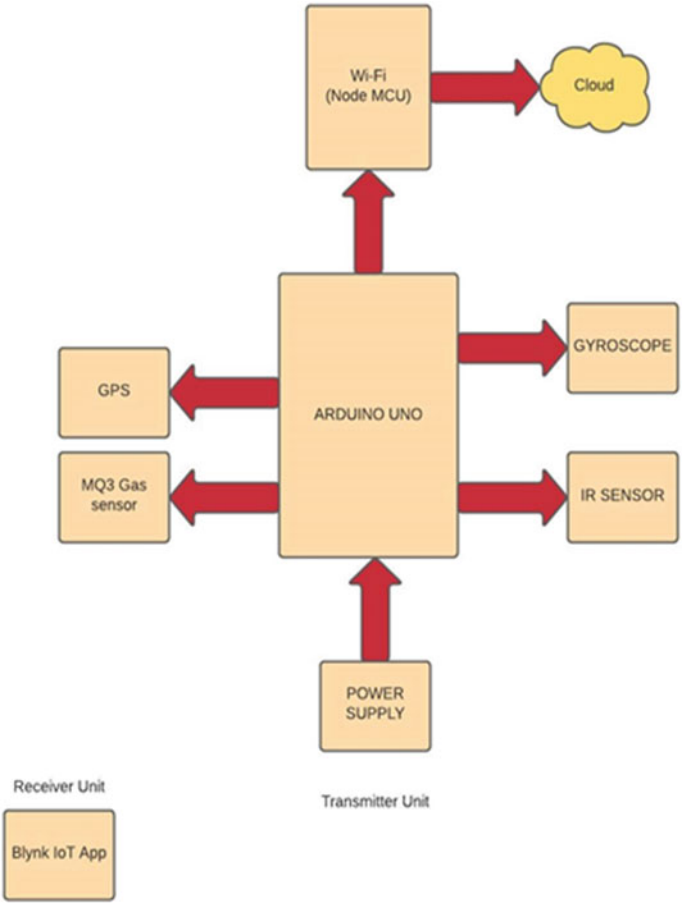


Fig. 4 Block diagram

2.3 Flowchart

The flowchart shown below in Fig. 6 highlights the functionalities of the smart helmet in a lucid and legible manner. It highlights the operation of the smart helmet.

- As soon as the helmet is powered on, the Node MCU tries to connect with the Blynk IoT application over the cloud as highlighted in the first process box. This initiates the live monitoring of the smart helmet.
- After this, the IR sensor indicates if the rider is wearing the helmet or not as indicated by the condition box.
- Similarly, the gyroscope detects the intensity of jerk to determine whether or not an accident has occurred as depicted by another condition box.

Fig. 5 Proposed helmet model

- The MQ3 gas sensor detects whether or not the person is drunk and transmits that message to the app viewer via the Arduino and Node MCU.
- The GPS sensor constantly tracks the location of the rider.
- All of this information is displayed on the Blynk IoT application as depicted by the final process block.

3 Results

A green LED starts glowing as shown in Fig. 9 in the Blynk smartphone application as soon as the wearer wears the helmet, signalling that the wearer is wearing the helmet. Any heavy unexpected jerk also triggers a red LED to flash in the app as depicted in Fig. 10, alerting the rider's relatives/friends that there has been an accident. In addition, the smartphone application continuously watches the rider's position providing information about the accident location which is shown in Fig. 8. This design as shown in Fig. 7, if it is adopted by the government, has a good real-life application. It will help to mitigate more two-wheeler road fatalities, as it is the leading cause of deaths worldwide. It will also reduce potential collisions from causing harm to the cars. In consideration of traffic, the traffic laws and even the welfare of individuals, this study is implemented here. The government's introduction of this sort of project saves the traffic police a lot of time and, most importantly, saves a person's precious life as he cannot drive an automobile while he is intoxicated and if there is no helmet. Family members will also be informed. While privacy is put at risk to a small extent with such a model but it is a small price to pay when one's life is at stake (Fig. 10).

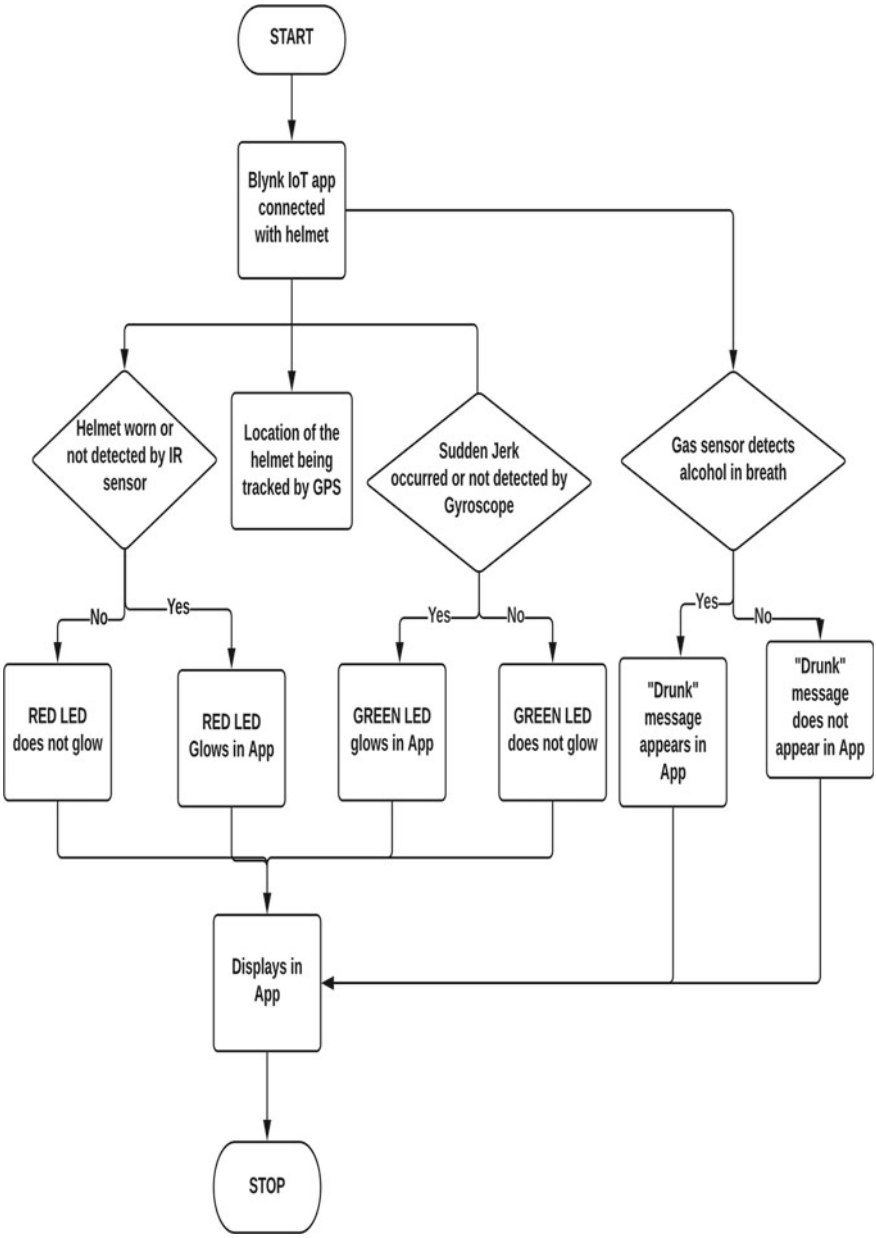


Fig. 6 Flow chart of smart helmet

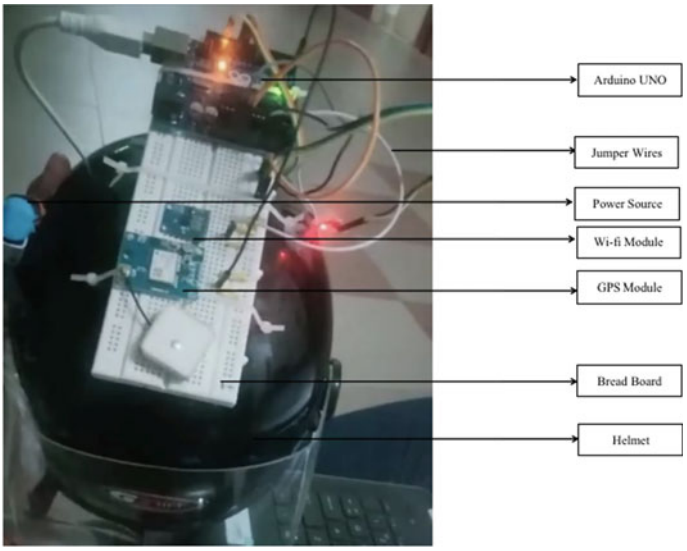


Fig. 7 Working model of smart helmet

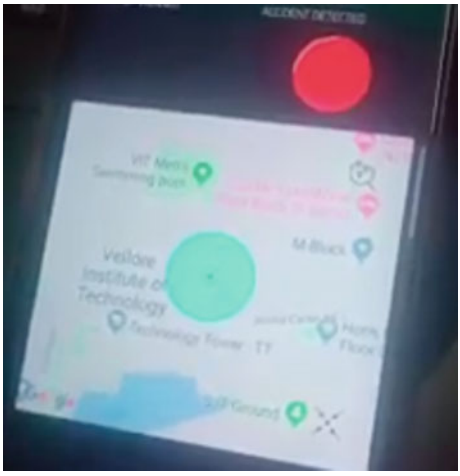


Fig. 8 Location tracking via Blynk app

4 Conclusion

This IoT-based smart helmet is a smart device which could help enhance the safety and security of the two-wheeler riders, particularly in a country like India where there are a high number of road crashes and numerous breaches of traffic laws. Wearing a helmet when riding a two-wheeler is mandatory for the rider's protection particularly from severe head injuries. This is where our helmet's IR sensor came into play as it

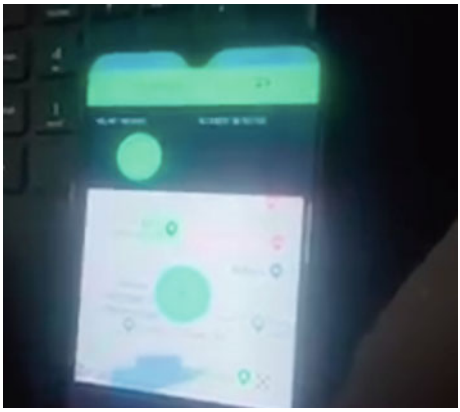


Fig. 9 Green light indicating helmet not worn

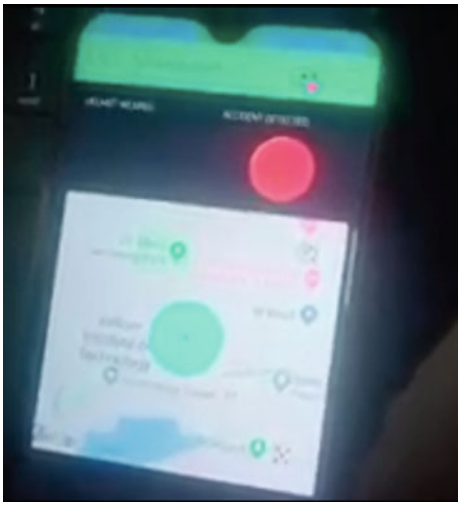


Fig. 10 Red light indicating accident detected

alerted the rider as well as its relatives whether or not the rider was wearing a helmet. Moreover, having immediate information about the rider's accident along with its location could help save lives even more. This was facilitated by the gyroscope, GPS module, and the IoT mobile application present in our model. Driving while drunk also increases the risk of accidents several manifolds and thus should be prevented. In this facet, the MQ3 gas sensor enabled alcohol detection and notified the rider as well as its relatives of its inebriated state. This work is an extensive accident detection system that uses a smart helmet that can detect whether the rider is wearing a helmet or not; track the co-ordinates of the rider; recognise the abrupt serious jerks in order to conclude that there is an accident. Therefore, this helmet proved to be an innovative and cost-effective solution for saving lives and increasing safety of two-wheeler riders. Few of the limitations of this model on which we intend to work

upon in future include, rapid draining of power, unavailability of power backup in case of sudden shut-down of the helmet, risk of privacy breaches and information theft due to constant location tracking. However, it may be noted that both of these disadvantages can be overcome by looking for alternate energy sources, such as solar energy as power source, and following proper privacy and safety protocols to prevent unnecessary privacy lapses.

Acknowledgements The authors are thankful to VIT University, Vellore for providing the necessary facility to successfully complete the concerned work.

References

1. Road Accident Statistics in India, https://sites.ndtv.com/roadsafety/important-feature-to-you-in-your-car-5/
2. W.J. Vles, E.J. Veen, J.A. Roukema, J.D. Meeuwis, L.P.H. Leenen, Consequences of delayed diagnoses in trauma patients: A prospective study. J. Am. Coll. Surg. **197**(4), 596–602 (2003)
3. S. Shalini, J. Muruganandham, S. Surya, Identification and prevention of accidents using smart helmet and GPS system. J. Phys.: Conf. Ser. **1717** (2021)
4. W. Rosenberg, T. Chanwimalueang, V. Goverdovsky, D. Looney, D. Sharp, D.P. Mandic, Smart helmet: Wearable multichannel ECG and EEG. IEEE J. Trans. Eng. Health Med. **4**(February), 1–11 (2016)
5. R.S.N. Agarwal, A.K. Singh, P.P. Singh, Smart helmet, in *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2018—Proceedings* (2018), pp. 2338–2341
6. B.P. Sreeja, S. Manoj Kumar, S. Sriram, Smart helmet for alcohol detection to prevent from accidents. Ann. Rom. Soc. Cell Biol. 2808–2815 (2021)
7. N.L. Aravinda, M. Jabirullah, D. Kirtana, An intelligent helmet system using IoT and Raspberry Pi. IOP Conf. Ser.: Mater. Sci. Eng. **981** (2020)
8. R.R. Hegde, L.M. Shaikh, R.S. Naik, A. Aminaerum, I. Shaikh, N.M. Bhomkar, Smart helmet for bikers. IOSR J. Mech. Civ. Eng. 49–52 (2020)
9. Y. MohanRoopa, N. Soujanya, V.S. Vaishnavi, U.V. Vardhan, An IOT based smart helmet for accident detection and notification. J. Interdisc. Cycle Res. **12**(6), 1–6 (2020)
10. P.K. Rao, P.T. Sai, N.V. Kumar, S.K.Y.V. Sagar, Design and implementation of smart helmet using. IoT: Int. Conf. Adv. Res. Innov. (ICARI) 323–325 (2020)
11. P. Eswari, S.M. Praveena, P. Gowtham, G. Vetrichelvi, H. Harsha, Intelligent helmet based on web communication and Iot technology for safe driving. Int. Res. J. Modernization Eng. Tech. Sci. **2**(12), 284–293 (2020)
12. S. Dubey, K. Meghana, M. Likhitha, K. Gupta, R. Balaji, An experimental study on advanced lane changing signal assist technology with smart helmet. Mater. Today Proc. **33**, 4771–4776 (2020)
13. Y. Zhuang, Q. Yang, T. Han, R. O'Malley, A. Kumar, R.E. Gerald, J. Huang, Fiber optic sensor embedded smart helmet for real-time impact sensing and analysis through machine learning. J. Neurosci. Methods **351**, 109073 (2021)
14. I. Campero-Jurado, S. Márquez-Sánchez, J. Quintanar-Gómez, S. Rodríguez, J.M. Corchado, Smart helmet 5.0 for industrial internet of things using artificial intelligence. Sensors **20**, 6241 (2020)
15. T. Sivagnam, S. Salama, H. Hajjaj, K. Rao, M. Thariq, H. Sultan, L. Seng, Materials today: Proceedings smart motorcycle helmet for enhanced Rider' s comfort and safety. Mater. Today: Proc. (2021)

16. A.S.T. Eldho, J.T. Paul, K. John, J. Jose, B. Paul, Design and analysis of a smart-attachment to jacket and helmet used by two-wheeler riders using Peltier-module. Mater. Today: Proc. (2021)
17. E. Sathish, G. Manikandan, G. Bhuvaneswari, Materials today: Proceedings design and development of multi controlled smart bike. Mater. Today: Proc. 1–5 (2021)
18. S. Shakya, L.N. Pulchowk, Sensor assisted incident alarm system for smart city applications. J. Trends Comput. Sci. Smart Technol. **2**(1), 37–45 (2020)
19. L. Louis, Working principle of Arduino and Using it as a Tool for Study and Research. Int. J. Control Autom. Commun. Syst. **1**(2), 21–29 (2016)
20. S.A. Jain, A. Bharadwaj, Characterizing WDT subsystem of a Wi-Fi controller in an automobile based on MIPS32 CPU platform across PVT. J. Ubiquit. Comput. Commun. Technol. (UCCT) **2**(04), 187–196 (2020)
21. Y.S. Parihar, Internet of Things and Nodemcu: A review of use of Nodemcu ESP8266 in IoT products. J. Emerg. Technol. Innovative Res. (JETIR) **6**(6), 1085–1086 (2019)
22. P. Srivastava, M. Bajaj, A.S. Rana, Irrigation system using IOT, *2018 Fourth International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)* (2018), pp. 1–5
23. N. Chadil, A. Russameesawang, P. Keeratiwintakorn, Real-time tracking management system using GPS, GPRS and Google Earth, in *5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2008* vol 1, (2008), pp. 393–396
24. V.M.N. Passaro, A. Cuccovillo, L. Vaiani, M. de Carlo, C.E. Campanella, Gyroscope technology and applications: A review in the industrial perspective. Sensors (Switzerland) **17**(10) (2017)
25. D. Vijayakumar, G. Ramesh, C. Jayabalan, S. Palani, M. Selvam, Micro controller based smart helmet by IR motion sensors. Int. J. Eng. Adv. Technol. **8**(6), 1850–1853 (2019)
26. K. Vidhya, M. Kasiselvanathan, Smart helmet and bike system. Int. J. Recent Technol. Eng. **7**(4), 1–4 (2018)
27. In-Depth: *How MQ3 Alcohol Sensor Works? and Interface it with Arduino.* https://lastminuteengineers.com/mq3-alcohol-sensor-arduino-tutorial
28. E. Media's, S. Syufrijal, M. Rif'an, BLYNK framework for smart home. KnE Soc. Sci. **3**(12), 579 (2019)
29. Blynk IoT platform: For businesses and developers. https://blynk.io

Manual (Wired) and Control (Wireless) Modes of Automation System with Multi-level Voice Strings

Venkatarao Dadi, Naresh Pathakamuri, Mohammed Ashik, M. Ramesh Patnaik, D. V. Rama Koti Reddy, and B. Ravichandra

Abstract Automation plays a significant role in making our lives easier and comfortable in day-to-day activities. Various wireless technologies like Bluetooth, IR remote, and Wi-Fi, etc., are used in the design of automation systems. For mobile applications, it is needed to perform the initial training process for user's specified (target) voice command. In this work, new user's initial training process of specified voice command is reduced by implementing the Initially Observed Plurality of Voice Commands (IOPVCs) data method in control unit. In automation systems, at the time of failures in network or control unit, it is required to include the manual mode of operation to ensure the reliability. It is achieved in this work by configuring the control unit with two-way electrical switches. For wireless voice connectivity, HC-05 Bluetooth module and for non-voice connectivity, IR (infrared) TSOP decoding sensor with IR remote is used. Graphical and virtual simulation of voice-controlled automation system is performed in the Proteus 8 professional simulation software, and test results are compared and validated with real-time experimental setup. Bluetooth is used to store the status of loads changes in customized excel file using PLXDAQ tool for monitoring.

Keywords Bluetooth · IR · Initially Observed Plurality of Voice Command (IOPVC) · Multi-voice string · PLXDAQ server · Two-way electrical switches

1 Introduction

The objective of automation is to design a low-cost and effective automation system for the people, who are unable to afford the high-end automation systems. With the

V. Dadi · N. Pathakamuri (✉) · M. Ashik · M. Ramesh Patnaik · D. V. R. K. Reddy · B. Ravichandra
Department of Instrument Technology, Andhra University College of Engineering, Visakhapatnam, Andhra Pradesh 530003, India

N. Pathakamuri
Department of Electrical and Electronics Engineering, Raghu Engineering College (A), Visakapatnam 531162, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_28

voice-activated automation system, one can easily control different loads like fans, lights, air conditioners, refrigerators, etc. Programmable control unit is the heart of the automation system to control the appliances or loads. The appliances are made to interface with control unit over a wireless network and relay driver modules.

Cristina Stolojescu-Crisan [1] designed a smart home automation system using qToggle application with basic API for user-friendly interface to incorporate the automation and security. Kunaraj [2] discussed the implementation of hardware for smart home automation to increase the user's comfort and security. Islam [3] implemented an IoT-based home automation system framework. The system is designed with NodeMCU controller and MQTT protocol to control and monitor appliances over the Internet in mobile application. Waleed [4] discussed the smart home incorporated with advanced intelligent technologies to the living environments to respond and control the appliances as per the home residents' requirements.

Yue [5] proposed a home automation system by imparting various wireless technologies like Zigbee, Bluetooth, wireless fidelity (Wi-Fi), radio frequency identification (RFID), and low-power wireless personal area network. Basanta [6] has implemented an automation system using voice signals and gestures to control the devices. Ebrahim Abidi [7] implemented a smart home automation system for home security with voice control and message alerting system connected via global system for mobile communication (GSM) SIM 900A. In the system using HC-SR04 ultrasonic sensor, the movement of unauthorized person is tested to detect for a distance of 20 m. Muhammad Asadullah [8] discussed remote controlled home automation system using Arduino controller, Bluetooth module, and smart phone to control up to 18 devices. Additional features are detection of water level and automatic plant irrigation. Mohammed Akour [9] implemented a speech-based remote control to operate the electrical appliances using an Android mobile phone. In the system, Voice to Text Google App is developed using the tasker application to convert the voice command to text command.

Al Mamun [10] proposed a smart home automation system using Bluetooth in a mobile application to maintain the required room temperature and water level in addition to control of light and door. Dhruva Kumar [11] implemented a smart home with security and automation using the Google Voice assisted mobile application designed based on Blynk application to send home notification security alerts to the user. Sowah [12] designed an authenticated wireless home automation system interconnected with OpenHAB server. OpenHAB server is configured on the Rasperry Pi, and Arduino controller is set to communicate with OpenHAB server to perform the automation. Singh [13] designed a smart home automation system controllable either with Internet or manual switches to use even in case of no Internet facility. In the design HTTP protocol is used for communication between the Raspberry Pi and the Web site.

Inam Ullah Khan [14] incorporated the Android mobile in-built voice recognizer in the automation system using the Arduino Uno controller and the Bluetooth module to control the loads and handles the client orders with the feature of controlled exchange of information among the gadgets. Sonali Sen [15] designed system controls the appliances with Speech-to-Text service (STT) conversion commands

using the built-in Android voice recognizer and Arduino Uno controller, Bluetooth module, and relay driver module. Depends on the received text commands, the relays are set to ON or OFF to operate the household appliances.

Shinde [16] implemented a voice-controlled home automation system for both voice and non-voice signals. Problems occurred in voice recognition module is rectified with keypad which is useful and accessible for old people and handicapped persons. Shaishav [17] implemented an automation system using speech processing unit (computer), microcontroller, and microphone. Vector quantification output is matched with the stored speech in computer, the output is calculated using Mel Frequency Cepstrum Coefficients extracted from voice signal. If matched signal is equal then corresponding loads are activated by controller. Dhawan Singh [18] designed a high-range Zigbee-based home automation system for older and disabled persons to avoid complications with wired systems.

From the above discussion of voice-based automation literature, authors did not discuss the initial voice recognition error during speaking specific voice command by a new person in Google assisted voice mobile application for voice-controlled automation. Some authors did not discuss the wireless automation implementation enabled with real-time manually operating of loads or appliance. However, some authors have discussed manual operating implemented for low-power (electronic) applications only but still there was a problem that forms a short circuit of phase line (malfunctioning due to physical connectivity of electrical switches) connection between isolated high-power output terminals of relay module and one-way electrical switches during the wireless controlling mode of operation by control unit.

In this work, home automation system (HAS) with IOPVC is implemented to control the appliances either in manual mode or wireless control mode. Failure of network or control unit in automation system requires manual operating of loads without changing the existing physical electrical wiring in home. Also, the short circuit of phase line in one-way manual switches in automation system is resolved using two-way electrical switches during operating in wireless controlling mode.

IOPVC method is implemented in control unit program to reduce the new person voice recognition training process in Google Voice assisted mobile application. Specific voice command in different pronunciations converted into text and observed in Google Voice assisted mobile application is considered as IOPVC text data. This IOPVC text data of specific voice command is used in control unit IDE program for reducing the training process.

2 Methodology

2.1 *Multi-string Commands to Minimize the Initial Training Process for a New Person in Mobile Voice Application*

Appliances (loads) are unable to control if the user's pronunciation is not mapping with the stored voice command in the control unit. Controlling a particular appliance using Google assisted voice command in mobile application requires a training time until a specific (target) voice signal is recognized correctly.

Initially observed that for a target voice command (e.g., "L1 ON") in the Google assisted voice mobile application with change in user pronunciation is recognized as various voice strings (IOPVC) as shown in columnwise in Fig. 5. Even the given target voice command (e.g., "L1 ON") is same, the variations are occurred until it is trained in Google assisted voice mobile application. To ensure the immediate response of appliances based on target voice command (e.g., "L1 ON") and reducing initial training process (i.e., wrong recognition of specific voice command in Google Voice assisted mobile application) for a new person is possible by considering IOPVCs data as a target voice command (e.g., "L1 ON") in control unit is shown in Fig. 1.

Control unit checks the user's received voice command (e.g., "L1 ON") with the already stored IOPVCs in control unit. If any one of IOPVCs is same as received voice command, then received voice commands assigned as target voice command in control unit and activated the corresponding appliance as shown in Fig. 2. By speaking "L1 ON" voice command in various pronunciation by user and its corresponding observed text error observed in Google Voice assisted mobile application is shown in Fig. 1. Specified voice command (e.g., "L1 ON") in Google Voice assisted mobile

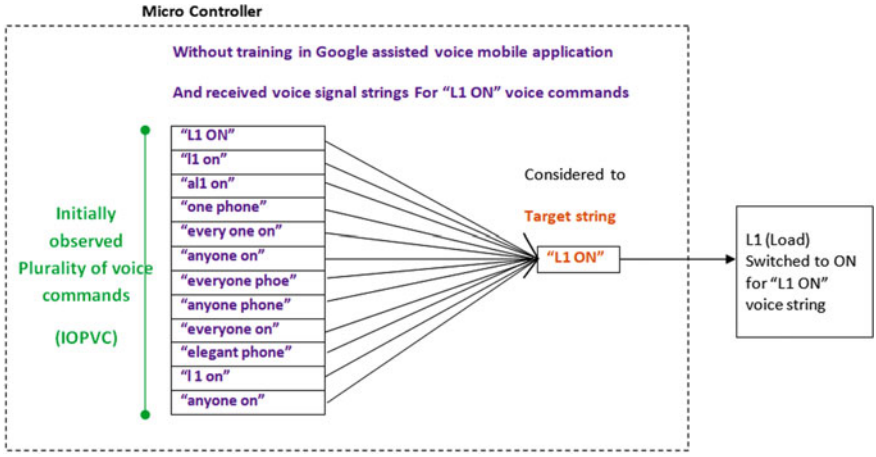


Fig. 1 Initially observed plurality of voice commands (IOPVC) for "L1 ON" speaking in Google Voice assisted mobile application

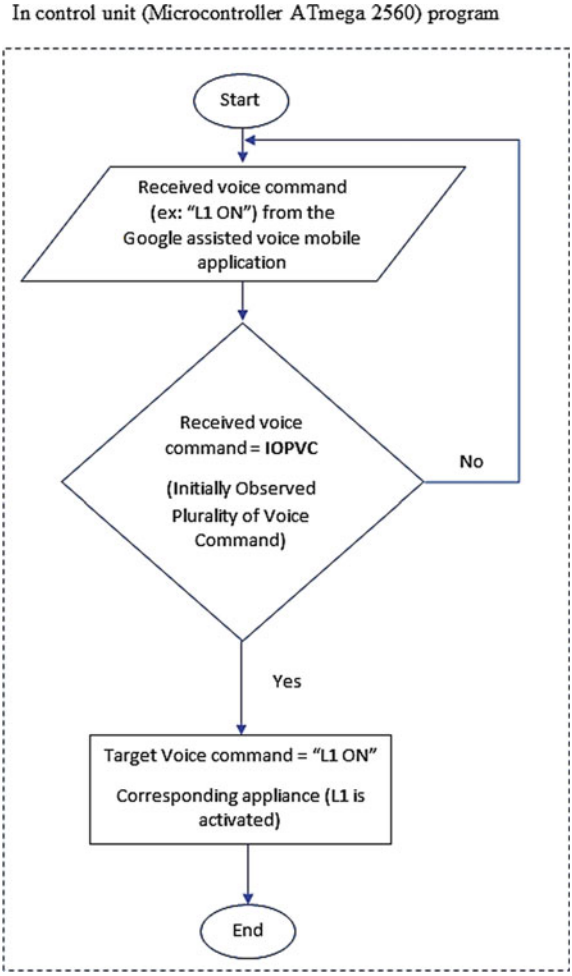


Fig. 2 Activate the appliances by control unit when received voice command is equal to anyone of the sample data of IOPVCs

application is correctly identified only when specified command is continuously speaking by user for some time then the specific command is trained in mobile voice application.

So, considering IOPVCs data in control unit makes the better chances for the control unit to identify the voice commands in short time and precise for a new person.

2.2 *Phase Line is Shorted During Wireless Control Mode with Physical One-way Switches*

Phase line is shorted during wireless control of appliances in automation system configured with one-way electrical switches as shown in Fig. 3a. So, in wireless mode of operation formation of closed phase line makes the loads are operative with manual one-way switches even the manual mode is in open condition. To avoid this confusion and for a desired mode of operation, two-way electrical switches are placed in the place of one-way electrical switches.

Phase line shorting is removed during wireless connectivity when the two-way electrical switches are properly connected with the isolated high-power output terminals of relay circuit as shown in Fig. 3.

Even manual mode is in open condition, the phase line of electrical wiring is shorted due to physical manual one-way switches that are switched upward and any one of relay (RL) is activated by digital signal from control unit as shown in Fig. 3a. When any one of relay is activated (i.e., relay ON condition means shorting is established between common (C) and no-operation (NO)) at the same time any one of one-way physical switch is switched upward, it leads to connectivity of phase line to other one-way manual switches. This problem is overcome by using the two-way electrical switches in place of one-way electrical switches, and their schematic connection is shown in Fig. 3b. Connection between relay isolated high-power output terminals and two-way switches connection of loads phase line helps to reduce the cost of automation system without modifying the existing electrical wiring of homes.

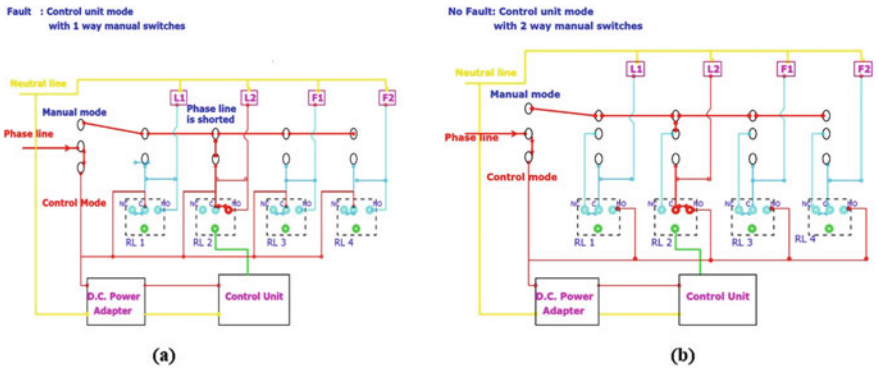


Fig. 3 **a** Two-way switch is used for switching between manual and control mode and loads are connected with one-way switches. **b** Two-way switches are used for change of mode and loads

3 Experimental Setup

The experimental setup comprises of an Arduino Mega2560 microcontroller, HC-05 Bluetooth Module for wireless transmission of voice signal, a 4-Channel Relay Driver to drive the four loads such as four bulbs named L1, L2, F1, and F2, and these are exemplary names given for the home appliances like lights1& 2, fans1& 2, respectively. Interconnection of experimental setup is shown in Fig. 4. Power supply is given to the Arduino controller board using 12 V, 1A DC power adapter. For the voice controlling of loads, the mobile Bluetooth is paired with the HC-05 Bluetooth. Non-voice wireless controlling of loads is implemented with IR remote and controller. The controller decodes the received IR signals from TSOP IR sensor via wireless remote. Voice-based commands are provided to the Arduino controller via mobile application which is developed with MIT application. Voice signal is transmitted via established wireless connection between mobile Bluetooth and HC-05 module which are connected to the AT Mega 2560 controller.

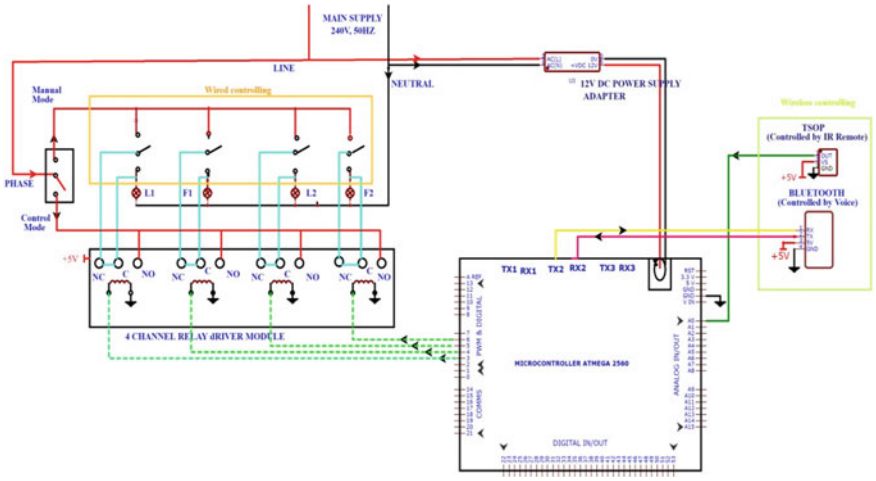


Fig. 4 Schematic diagram of wired and wireless automation system with two-way electrical switches

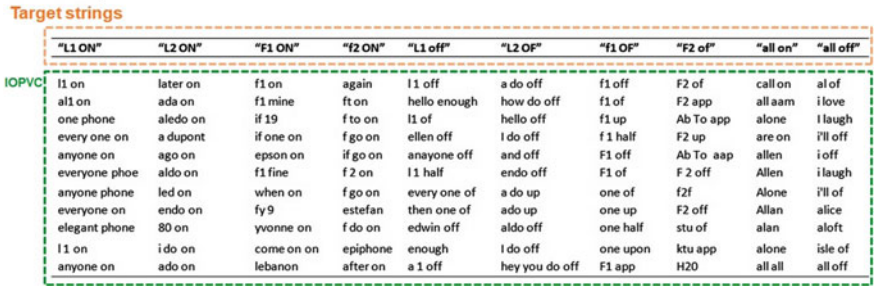


Fig. 5 Initially observed plurality voice commands (IOPVC) mentioned columnwise under various pronunciations of users for a specific (target) voice command

4 Results and Discussions

4.1 Initial Observed Plural Voice Commands (IOPVC) to Minimize the Training Process for a New Person Pronunciation for a Target Voice Command

Target voice instructions like "L1 ON," "L2 OF," "L2 ON," "L2 OF," "F1 ON," "f1 OF," "f2 ON," "F2 of," "all on," and "all off" with their corresponding 12 samples of IOPVCs (columnwise) are observed in Google Voice assisted mobile by speaking user in different pronunciations as shown in Fig. 5.

Controller unit checks the received voice command with the particular stored IOPVCs samples. If received voice command is matched with the any one of the particular stored instruction (column wise), then it is replaced with the specific (target) command. Controller unit activates the corresponding load based on target voice command.

4.2 Virtual and Graphical Simulation of Wireless Voice-Controlled Appliance

Virtual and graphical simulation of automation system with IOPVCs is done using PROTEUS 8.6 SP2 professional software. Bluetooth library is added in the Proteus software to use the virtual Bluetooth module. For simulation in PROTEUS 8.6 SP2 Professional application, all the virtual devices like Arduino Mega 2560, HC-05 Bluetooth module, four-channel relay driver modules and Lamps L1, L2, F1, and F2 are connected as shown in Fig. 6. Mobile Bluetooth and system Bluetooth are connected virtually. "L1 ON" and "F1 ON" are spoken by user then virtual relay pins of L1 and F1 appliances are switched to ON as indicated by arrows during virtual simulation.

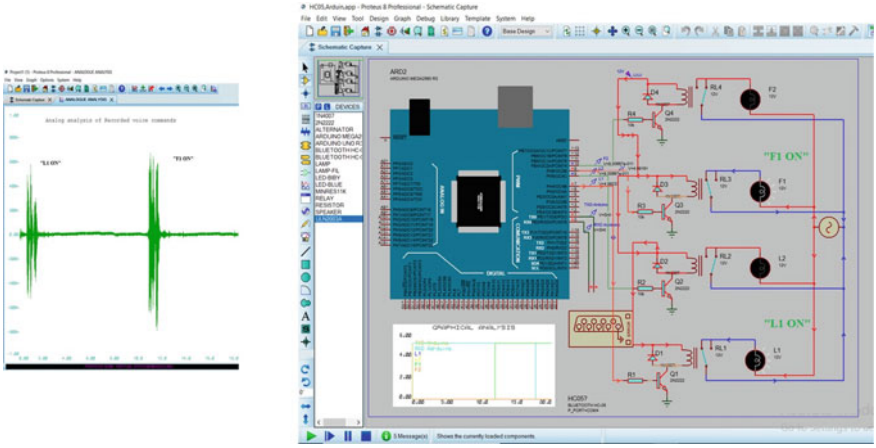


Fig. 6 Virtual simulation for voice signals "L1 ON" and "F1 ON"

Voltage probes are placed at TXD (Transmitted data), RXD (received data), and digital ports of Arduino controller to analyze the graphical analysis of change in switching states for a particular voice signals as shown in Fig. 7. Simulation time interval is 0–20 s. "F1 ON" voice command spoken by the user is received in control unit via virtually connected Bluetooth. With the received voice command either "F1 ON" or anyone of sample of IOPVCs (various texts observed by speaking "F1 ON" voice command in Google Voice assisted mobile application), the control unit considers the voice command as "F1 ON" and activated the appliance L1. At 0 s TXD pin of Arduino is activated to high (change the state from 0 to 5 V) represents virtual

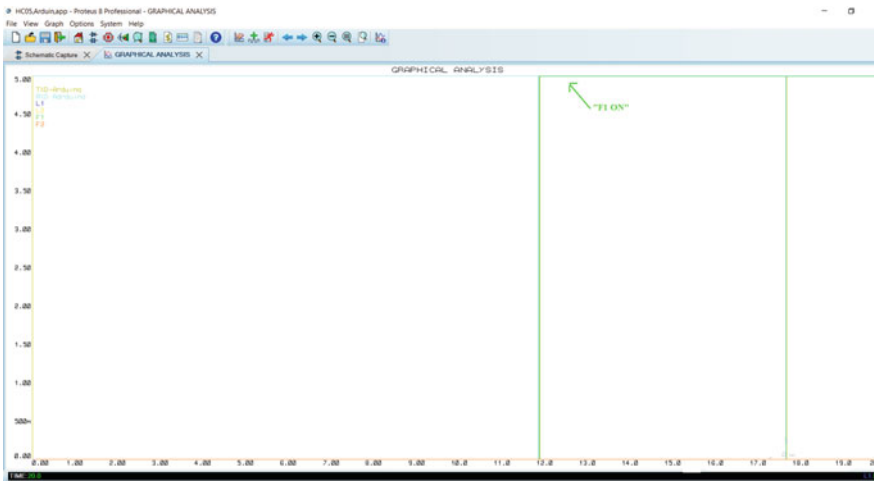


Fig. 7 Graphical simulation for "F1 ON" voice command

Bluetooth are connected at same time RXD pin of Arduino is activated means ready to receive the signals from mobile application. For "F1 ON" voice spoken by user, it is received by control unit and checks received command along with IOPVCs. So F1 appliance is activated (change the state from 0 to 5 V) at 11.9 s. At 16.8 s RXD deactivated to low (change it state from 5 to 0 V) means it available to receive next command.

4.3 Real-Time Wired and Wireless Implementation

The mobile Bluetooth is synchronized with the system Bluetooth HC-05 and established the wireless connection. During the transmission of voice data via HC-05 Bluetooth LEDs of RXD, TXD in Arduino prototype are blinked. The input pins of the relay driver module are interfaced and communicated with digital output ports of the control unit and control the corresponding appliances. TSOP IR receiving sensor receives control signals from remote wireless to control the appliances. Control unit decodes the received IR signals and changes the state of appliance. An application is designed with MIT inventor application for voice-controlling mode of operation as shown in Fig. 8. In addition to wireless connectivity, the circuit is designed with manual controlling of appliances using two-way switches.

The appliances can be controlled in manual mode even with failure in wireless control mode as the one-way switches are replaced with the two-way switches to eliminate the shorting of phase line operating in wireless control mode. In the real-time

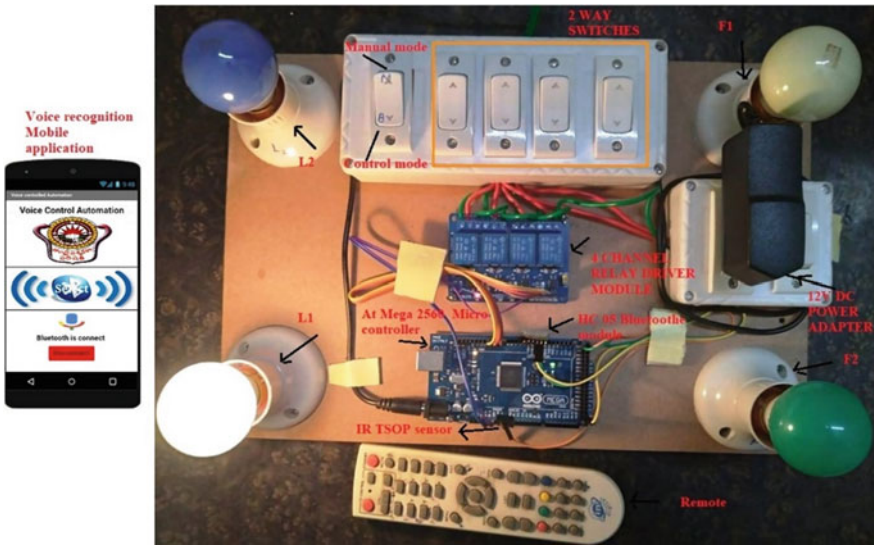


Fig. 8 Real-time implemented setup with two-way manual switches

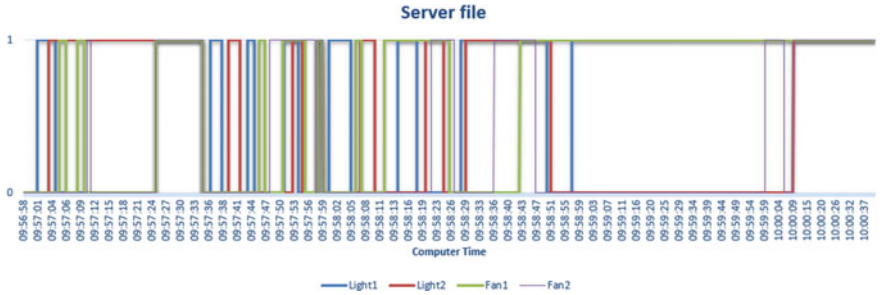


Fig. 9 Data stored and monitored in EXCEL using PLXDAQ tool

implementation, four-channel relay driver module isolates the high-power electrical connectivity with low power connectivity on input side of relay as shown in Fig. 4. This design does not require to modify the existing electrical phase line wiring of loads rather requires only modifying the phase line connecting point at electrical switches in the electrical switch board to make it affordable in ordinary homes.

The status of loads changes is monitored and saved in customized excel file using PLXDAQ tool via Bluetooth. Graphical analysis for status updating of loads with time interval is obtained which is shown in Fig. 9. This graphical analysis is particularly useful in automation industry in identifying the machinery operating time to ensure the proper maintenance.

5 Conclusion

In automation systems, voice and non-voice controlling modes are used for controlling the equipment remotely. In the control unit, the task of assigning a target voice string from the initially observed multi-voice string reduces the voice training of new person. Activated relays configured with one-way electrical switches creates a phase line shorting problem even in wireless control mode, which results into user confusion operating with any one of the wired or wireless mode. In this work, replacing the one-way electrical switches with the configured two-way electrical switches the user can operate safely in any one of the desired modes. This real-time implementation for automation makes it possible to operate the real-time electrical appliances in desired and safe manner. Secured home automation with voice command encryption using fast Fourier transformation is our future scope.

References

1. C. Stolojescu-Crisan et al., An IoT-based smart home automation system. Sensors **21**, 3784 (2021). https://doi.org/10.3390/s21113784

2. A. Kunaraj et al., Hardware implementation of smart home automation system using Aurdino UNO, in *Springer Lecture Notes in Electrical Engineering, Advances in Electrical and Computer Technologies,* pp. 547–558. https://doi.org/10.1007/978-981-15-9019-1_48
3. Sk.F. Islam, Md.I. Hasan, M. Akter, M.S. Uddin, Implementation and analysis of an IoT-based home automation framework. J. Comput. Commun. **9**, 143–157 (2021).https://doi.org/10.4236/jcc.2021.93011
4. J. Waleed et al., Smart home as a new trend, a simplicity led to revolution, in *1st International Scientific Conference of Engineering Sciences—3rd Scientific Conference of Engineering Sciences (ISCES).* (IEEE, 2018). https://doi.org/10.1109/ISCES.2018.8340523
5. C.Z. Yue, S. Ping, Voice activated smart home design and implementation, in *2nd International Conference on Frontiers of Sensors Technologies (ICFST).* (IEEE, 2017), pp. 489–492. https://doi.org/10.1109/ICFST.2017.8210563
6. H. Basanta, Y. Huang, T. Lee, Assistive design for elderly living ambient using voice and gesture recognition system, in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (2017), pp. 840–845
7. M.E. Abidi et al., Development of voice control and home security for smart home automation, in *2018 7th International Conference on Computer and Communication Engineering (ICCCE).* (IEEE, 2018), pp. 246–249. https://doi.org/10.1109/ICCCE.2018.8539258
8. M. Asadullah et al., Smart home automation system using Bluetooth technology, in *International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT).* (IEEE, 2017), pp. 1–6. https://doi.org/10.1109/ICIEECT.2017.7916544
9. M. Akour et al., Mobile voice recognition based for smart home automation control. Int. J. Adv. Trends Comput. Sci. Eng. **9**(3) (2020). https://doi.org/10.30534/ijatcse/2020/196932020
10. M.A.A. Mamun et al., Smart home automation system using Arduino and android application. J. Comput. Sci. Eng. Softw. Test. (6) (2020). https://doi.org/10.5281/zenodo.3816960
11. K. DhruvaKumar et al., IoT based home security and home automation using google assistant. High Technol. Lett. **26**(5) (2020). http://www.gjstx-e.cn/
12. R.A. Sowah et al., Design of a secure wireless home automation system with an open home automation bus (OpenHAB 2) framework. J. Sens. **2020**, 1–22 (2020).https://doi.org/10.1155/2020/8868602
13. A.P. Singh, A. Biswas, B. Singh, IoT based smart home automation enabled with manual mode switch, in *Control 2019 2nd International Conference on Intelligent Communication and Computational Techniques (2019978)*
14. I.U. Khan, Voice controlled home automation system. Int. J. Res. Comput. Commun. Technol. **6**(5) (2017). https://www.researchgate.net/publication/317386416_Voice_Controlled_Home_Automation_System
15. S. Sen et al., Design of an intelligent voice controlled home automation system. Int. J. Comput. Appl. **121**(15), 39–42 (2015). https://doi.org/10.5120/21619-4904
16. G.R. Shinde et al., Voice and non-voice control based wireless home automation system. Int. J. Innov. Res. Sci. Eng. Technol. (2015). https://doi.org/10.15680/IJIRSET.2015.0407005
17. Shaishav et al., Interactive home automation system using speech recognition. Int. J. Eng. Res. Technol. (IJERT) **3**(04), 2295–2299 (2014). https://www.ijert.org/research/interactive-home-automation-system-using-speech-recognitionIJERTV3IS042264.pdf
18. D. Singh et al., Voice recognition wireless home automation system based on Zigbee. J. Electron. Commun. Eng. **6**, 65–75 (2013).https://doi.org/10.9790/2834-616575

Smart Devices to Detect Health Abnormalities

K. Bhagavan and M. Kavitha

Abstract The smart devices gain a lot of attraction in the health care field. Medical sensors, wireless communications, and computational analytics advance medical services in healthcare field. Due to mobile technologies, smart services are becoming part of everyone's regular life. Even though we had medical advancements, day-to-day disease incidents and death rates are increasing for the people. Lifestyles, food habits, and other simple health concerns are causing major health issues that also lead to death. Continuous monitoring of vital symptoms may allow people to detect abnormal health conditions at an early stage of disease attack. It supports the patient to increase the survival rate through undergone better treatment. For this, smart devices are one of the solutions, to closely monitor human health. Smart devices and their role in the early detection of health abnormal condition are projected in this article by taking the heart disease paradigm.

Keywords Heart disease · Data analytics · Health monitor · Machine learning · Smart device · Sensor

1 Introduction

Now a day, the usage of smart devices increases in many areas including healthcare, industry, agriculture, and transportation. Healthcare and medical are the two most benefitted areas of these smart solutions. Smart devices can increase the healthcare applications like elderly people assistance, remote monitoring, fitness management, chronic disease identification, etc., [1]. The inbuilt sensors in these devices support continuous monitoring of people's health. These reduce healthcare costs and also help the healthcare provider to identify the disease at an early stage. It helps the

K. Bhagavan · M. Kavitha (✉)
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh 522502, India
e-mail: mkavita@kluniversity.in

K. Bhagavan
e-mail: bhagavan@kluniversity.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_29

doctors to increase the life span of a patient through better treatment steps. Smart solutions in the healthcare field are still an attractive research problem even though many innovative solutions available in this area. Many researchers discussed an internet of things (IoT) as a solution to predict health abnormalities at an early stage and diagnosis them. Smartwatches, smartphones, and other electronic gadgets are the basis of IoT devices. Figure 1 shows various smart devices using in the market to monitor people's health [2]. These devices help the users to connect with the medical IoT platforms. Sensors, connectivity, data processing, and user interface are the main components of this platform. The launch of cost-effective wireless sensors has drastically changed healthcare practices over the past two decades. A wireless biomedical sensor is a small, tiny device that closely monitors human health by attaching it directly or through some device. Many salable devices are now accessible, and exploration in wireless health appliances is proceeding at rapid grown. Smart and wireless biosensors have achieved significant force together in research and industrial regions. The smart solutions over patient healthcare include remote health monitors, surgical systems, and real-time diagnostic systems, etc., [3].

Section 1 discussed usage of smart solutions in healthcare. Section 2 discussed various commercial health monitor devices. Various authors contributions in smart healthcare is discussed in Sect. 3. Section 4 discussed smart solution to monitor heart patient. Machine learning approach to analyze health data is discussed in Sect. 5. Results part is discussed in Sect. 6 and finally paper is ended with conclusion and future work discussion.

Fig. 1 Various wearable devices in healthcare

2 Commercial Health Monitor Devices

In recent years, various sensors produced for remote sensor networks-based medical care. Healthcare services are advanced using these sensors. With the goal of medical care administrations, the industries are producing small and advanced featured sensors. Wearable gadgets can be controlled or observed through web applications or mobile applications. The remote patient observation using sensors or wearable gadgets is the principal advantage in medical care administrations [4, 5]. Figure 3 represents various wearable devices present in the healthcare domain [3].

Many industries came with smart products for healthcare based on market requirements with modern technological innovations like wireless technologies, IoT, AI, Bigdata [6, 7]. Wearable devices can provide enormous services with a variety of sensing features in healthcare. These devices can be controlled or monitored through a web application or mobile applications [9]. Various commercial wearable devices present in the healthcare domain are shown in Fig. 2. All these devices are app-supported, the user can self-monitor the health and these devices can support to detection of health abnormal conditions at an early stage. Wireless BP cuff supports to check the blood pressure levels of a human. The oximeter can measure the oxygen level in human blood for nearly 5 days continuously with a fully charged battery. ECG device supports continuous ECG monitor for nearly 24 h with a fully charged battery. Glucose monitor device support measuring the sugar level in the human blood. A wireless wristband device shown in Fig. 3 supports continuous monitor of COVID-19 vital symptoms for a patient, which supports treating the COVID patient at an early stage. FIT activity monitor device is designed using a temperature sensor,

Fig. 2 Commercial devices to monitor health vital conditions

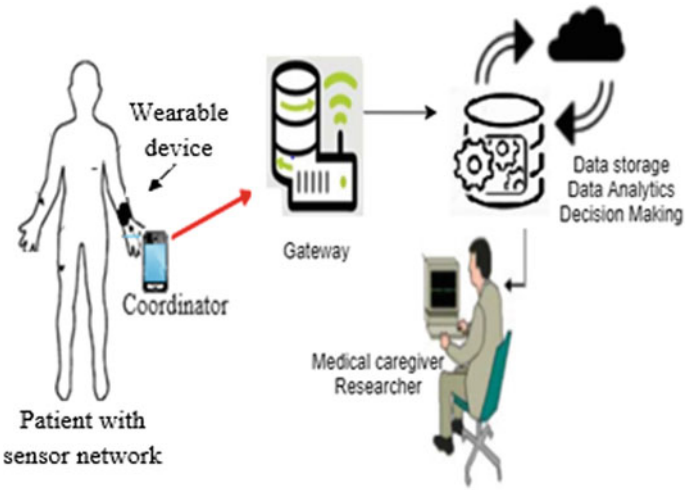


Fig. 3 Block diagram of smart health monitor

sweat sensor, and other biosensors to monitor the fitness of the body by closely monitoring the human body activities. These devices permit real-time observing, ambient assisted living, adverse drug reaction monitor, etc., [9]. All these devices allow mobility of the patients and support non-stop monitoring of patient health.

3 Literature Survey

Many industries are effectively engaged with the production of new kinds of biosensors. For instance, BIOTEX was a European Union subsidized task focused on the advancement of material sensors for the estimation of physiological biosignals. Sensors were created to gauge sweat rate, ECG, breath, blood oxygenation, sodium, conductivity, pH, and temperature [10].

Kalantarian et al. proposed an innovative nutrition monitoring wearable device to monitor the nutrition level of humans by monitoring food intake. The authors also proposed a framework to analyze the food habits in long run, volumes of meals using classification mechanisms [11].

Tong et al. recommended a sensor-enabled personalized health care targeting structure to keep track of patient health. ECG signals are monitored continuously through mobile app and web interfaces. This allows the caregivers to keep track of patient health continuously and allows them to give better treatment at an early stage [12].

Huang et al. discussed a wearable device to monitor physiological parameters of human health and it is set at Taiwan care institute from October 2018 to monitor patient health. The care notification system sends the notifications to the patient

and caregiver based on the collected physiological symptoms data processing. This system supports the doctors to identify abnormal incidents in a long time and provides more attractive healthcare services to the patients [13].

Azevedo et al. developed a wearable device to monitor children's health outdoor. The system gives the alert notification based on threshold values of body temperature, solar radiation, etc. This system supports continuous monitoring of children's health [14].

Yang et al. proposed a smartwatch to monitor the physiological conditions of a patient. This allows personal health monitoring to the chronic disease suffering patient. It is app-supported and provides a new way of patient care in the health domain [15].

Haolin et al. proposed a smartwatch to detect the abnormal movement of the body which leads to early diagnosis of Parkinson's disease. It supports the analysis of the abnormal behavior of human health by observing heart rate, steps count, abnormal hand movement closely [16].

Kavitha et al. proposed an IoT-based device to monitor breast health. The sensors in the device can monitor the patient's breast health measurable parameters data over a period. Based on the threshold value of said parameters the abnormality is noticed in the system and the alert notification sends to the user. The proposed system supports the patient to attend treatment at an early stage of disease and the measured data support the caregiver to provide better treatment [17].

Kolli et al. reviewed the internet of things mechanism to assist the physically challenges people. The authors discussed various IoT assisted solutions for making peoples life easier and comfort [18].

Pavani et al., discussed smart solution for visually challenged people for safe navigation system. The authors used sensor device to detect the vehicles which are coming toward the person and system will generate signal to the person to take the right direction [19].

Phani et al., discussed smart solution to monitor patient health. The authors identified few basic parameters which are more probable to measure the current health condition of health and that are measured using IoT device [20].

Ates et al. proposed a wearable device to monitor corona patients remotely. The device can monitor the activities of carona patient to detect the health abnormalities at early stage. The monitored data support the doctor to provide better treatment [21].

The literature study in wireless biosensors is centered around more modest and less power consumption biosensors which incorporate multiple biosensors together into one framework including sensors to monitor sweat rate, temperature, pH, and so forth. Table 1 shows summary report on smart solutions in healthcare. Research is also centered around the improvement of wireless sensor organizations to make smart gadgets composed of many sensor nodes.

Table 1 Smart healthcare solutions summary report

Author	Proposal	Symptoms monitored	Results
Coyle et al. 2010	Discussed textile embed supported bio sensors to monitor physiological changes in routine activities	Sweat rate, ECG, breath, blood oxygenation, sodium, conductivity, pH, and temperature	Proposal is more benefitted for the humans in sports field and helps to enhance the services in clinical field
Kalantarian et al. 2011	Proposed wearable device to monitor nutrition level of person based on food intake reading	Volume of mean, type of meal and meals timings	It helps to regulate diet habits by taking appropriate levels of food and it also helps to preserve healthy life style
Tong et al. 2017	Proposed sensor-enabled wearable device to monitor patient health	Continuous monitoring of ECG signals	This supports the caregivers to keep track of patient health continuously and allows them to give better treatment
Huang et al. 2019	Developed healthcare system to continuously monitor the patient health	Physiological changes of health	It leads to provide more attractive healthcare services based on long time health condition of a patient
Pavani et al. 2019	Discussed smart solution to monitor patient health	Basic health conditions of human	It supports the care giver to treat the patient at early stage of disease
Yang et al. 2019	Proposed a wearable device to monitor chronic disease patients	Physical and physiological conditions	It provides a new way of patient care in the health domain
Azevedo et al. 2020	Proposed a wearable model to monitor children health based on climatic conditions	Body temperature, solar radiation and other climatic changes	It supports to constantly monitor the children health at outdoor
Kavitha et al. 2020	Proposed a sensor-enabled device to monitor breast health of human	Temperature, skin color change, lumps, nipple discharge	It helps to detect the breast abnormal conditions at early stage
Haolin et al. 2020	Proposed a wearable device to notice abnormal movement of body of a human	Heart rate, steps count, hand movement	It helps to diagnose the Parkinson's disease patient at early stage
Ates et al. 2021	Proposed an activity tracker to measure corona patient health condition	Body temperature, oxygen level and other physical symptoms	This helps the care giver to track the health condition of corona patient

4 Smart Solution to Monitor Heart Patient

This section proposed the solution to monitor and predict heart abnormality at an early stage using smart device. Smart devices are having inbuilt sense, store, analyze and communication power. Heart disease is one of the familiar reasons for death among people. Early detection of asymmetry in heart health leads to an increased survival rate of a patient through better treatment practices and increases the survival rate of a patient. Continuous monitoring of vital signs of health supports the diagnosis of this disease at an early stage. The smart device capable of sensing heartbeat rate, oxygen level, body temperature, bp, cholesterol, weight, sugar level is suggested here. The smart device is a wearable with the integrated sensors which support to sense and record these parameters. Now a day, bio sensors for measuring health conditions are available in the market so the sensors and its details are not discussed too much here. The device can monitor and process these vital conditions of user. The variation in these measures helps to predict health abnormal conditions at an early stage by comparing with set threshold values of corresponding. The related sensors in a smart device can monitor these measures and the signals are transferred into the required format at the coordinator site. Recorded data directed to the cloud for permanent storage as well as further processing. The machine learning algorithms support predicting the health abnormality by processing measured data. These meaningful predictions support the users to make better decisions toward their health as well as support to provide better treatment at the doctor's end.

Figure 3 represents the block diagram of proposed smart solution for heart patient health monitor. This is four-tier architecture. Stage-1 is patient with sensor network where the sensors are the initial point of contact to humans which sense the health data. The sensed data is converted into the required format and transferred to storage using a communication medium in stage-2. Stage-3 is the medical server at cloud that allows storage of data in a variety of forms. The sensor system generates a big amount of data due to its continuous monitoring nature. Cloud is a feasible solution to store the big amount of generated data because of its elasticity and stable nature. The data processing techniques applied thereon collected data to provide meaningful insights. This result helps to make decisions over patient health. Here, the prediction of health abnormality is done using a machine learning classifier. Stage-4 is the user interface that allows users like caregivers, doctors, laboratory researchers, government agencies, etc., to access the information. Figure 4 represents workflow of proposed approach. Step-1 is data collection phase, where health data is collected through sensors, step-2 is data storage phase, where sensor generated data is stored, step-3 is data analytics phase, where data processing is done, and step-4 is user access phase, where user interaction is done.

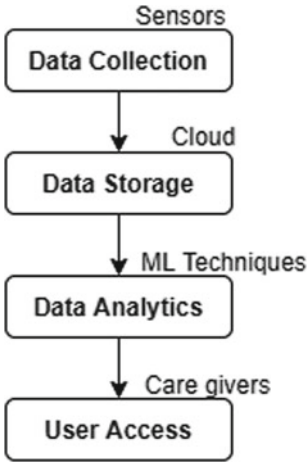


Fig. 4 Workflow of proposed approach

5 Machine Learning Techniques for Health Data Analysis

The data analysis using machine learning (ML) Techniques includes the steps of data preprocessing, data partition, classifier, and performance analysis steps [22]. Data preprocessing is the key step where the sensed data is processed to remove outliers and to transform the data into the required format [23]. In general, data is partitioned into train and test parts in the 70:30 ratio. Train part data is used to train the model and test part data is used to evaluate the system performance. The Naive Bayes, linear regression and random forest classification techniques are implemented in this work. Navie Bayes is a statistical technique constructed based on Bayes theorem. Linear regression also a statistical approach which builds the predictive model based on the relationship among the dependent and independent features in the dataset. Random forest technique is the most popular approach which takes several decision trees based on different subsets of features in the given dataset and takes the average and that average result improves the prediction accuracy. Classifier performance is measured in terms of accuracy, precision, recall, and f1-score.

Accuracy means the percentage of correct classifications. The higher value of accuracy refers the model is more accurate.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Number of total predictions}}$$

Precision refers the proportion of positive prediction results that are correct.

$$\text{Precision} = \frac{\text{True positive}}{(\text{True positive} + \text{False positive})}$$

Recall refers the probability of actual positives that predicts correctly.

$$\text{Recall} = \frac{\text{True positive}}{(\text{True positive} + \text{False negetive})}$$

F1 score is the harmonic mean of precision and recall. The best model gives highest f1-score as 1.

$$\text{F1 - score} = 2.\frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

6 Result Discussion

To discuss the actual results of proposed smart solution, the exact dataset with identified parameters is not available in the public datasets repository. For sample results discussion purpose, the UCI heart disease dataset is studied, which consists of 305 records and seventy-five features. Based on this dataset, a sample dataset with the parameters heartbeat rate, oxygen level, body temperature, bp, cholesterol, weight, sugar level is considered to predict heart abnormality. There are three categories of patients in the dataset which are normal, low risk, and high risk. The number of instances under each category is shown in Fig. 5.

In this discussion, the smart device attached to the patient can sense medical data and send it to the cloud for storage. The data is processed using the machine learning

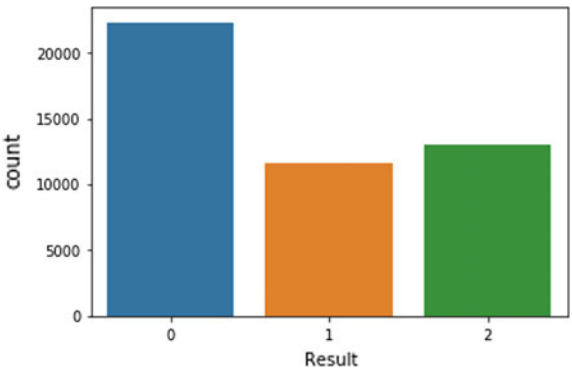


Fig. 5 Categories of patients in dataset

Table 2 Classifiers performance report

Classifier	Risk category	Precision	Recall	F1-Score	Accuracy (%)
Naïve Bayes	0	0.94	1.00	0.97	83.15
	1	0.80	0.49	0.61	
	2	0.72	0.89	0.80	
Logistic regression	0	1.00	1.00	1.00	87.13
	1	0.96	0.46	0.63	
	2	0.68	0.98	0.80	
Random forest	0	0.93	0.97	0.95	93.15
	1	0.77	0.87	0.81	
	2	0.97	0.85	0.91	

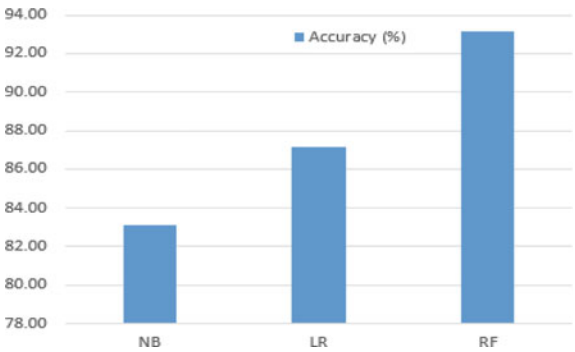


Fig. 6 Graphical representation of classifiers accuracy

technique. Python programming platform Jupiter notebook is used for implementation. Three machine learning techniques Naïve Bayes (NB), linear regression (LR) and random forest (RF) are applied on dataset. Among these the random forest machine learning classification technique given more accuracy in this heart health classification task. Its performance is analyzed using accuracy, precision, recall, and f1-score. The performance report is shown in Table 2. The graphical representation of classifiers accuracy is shown in Fig. 6. The classification results help the caregiver to provide an early stage treatment which enhances the survival rate of the patient.

7 Conclusion and Future Work

Many smart devices are available commercially and can measure vital symptoms of health. These devices are having inbuilt sensors which support continuous monitoring of people's health. These devices reduce the cost of healthcare. Smart devices help the healthcare provider to identify the disease at an early stage which makes them increase the life span of the patient through better treatment steps. In this study, heart

patient vital health parameters are identified and the smart device with supportable sensors is proposed to measure it. The measured data is processed using the machine learning technique and the notification is sent to the user based on the prediction report of the classifier. Early detection of heart abnormal conditions supports people to undergone better treatment. This smart healthcare solution reduces the incident and death rate of heart relates health problems.

In the future work, a smart device to measure all health-related vital conditions are supposed to design and implement using appropriate sensors. Machine learning techniques are widely using in processing medical data, so the performance of machine learning techniques is intended to assess with deep learning techniques on proposed model.

References

1. M. Kavitha, V. Saritha, K.P. Venkata, M.S. Obaidat, Wireless sensor enabled breast self-examination assistance to detect abnormality, in *2018 International Conference on Computer, Information and Telecommunication Systems (CITS).* (IEEE, 2018, July), pp. 1–5
2. http://hdl.handle.net/10603/314681
3. M.R. Mahfouz, M.J. Kuhn, G. To, Wireless medical devices: a review of current research and commercial systems, in *2013 IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems.* (IEEE, 2013), pp. 16–18
4. K.R.R.M. Rao, V.A. Reddy, C.H. Shankar, K.V.G. Reddy, Smart health care system using IoT. Int. J. Innov. Technol. Explor. Eng. **8**(6), 772–775 (2019)
5. K. Satwik, N.V.K. Ramesh, S.K. Reshma, Estimation and monitoring of vital signs in the human body by using smart device. Int. J. Innov. Technol. Explor. Eng. 1035–1038 (2019)
6. S. Srinivasulu, M. Kavitha, C.S. Kolli, Wireless technology over Internet of Things, in *Emerging Research in Data Engineering Systems and Computer Communications.* (Springer, Singapore, 2020), pp. 589–601
7. G. Akhila, N. Madhubhavana, N.V. Ramareddy, M. Hurshitha, N. Ravinder, A survey on health prediction using human activity patterns through smart devices. Int. J. Eng. Technol. **7**(1.1), 226–229 (2018)
8. M. Kavitha et al., Speech controlled home mechanization framework using android gadgets. Int. J. Eng. Technol. **7**(1.1), 655–659 (2018)
9. M. Kavitha, P. Krishna, Sensor enabled wearable technological tool to detect breast abnormality. Int. J. Eng. Technol. **7**(2.7), 589–593 (2018)
10. S. Coyle, K.-T. Lau, N. Moyna, D. O'Gorman, D. Diamond, F. Di Francesco, D. Costanzo et al., BIOTEX—biosensing textiles for personalised healthcare management. IEEE Trans. Inform. Technol. Biomed. **14**(2), 364–370 (2010)
11. H. Kalantarian, N. Alshurafa, M. Sarrafzadeh,A wearable nutrition monitoring system, in *2014 11th International Conference on Wearable and Implantable Body Sensor Networks.* (IEEE, 2014), pp. 75–80
12. K.-K. Tong, K.-S. Leung, Y. Leung, A system for personalized health care with ECG and EEG signals for analysis, in *2017 International Smart Cities Conference (ISC2).* (IEEE, 2017), pp. 1–6
13. P.-C. Huang, C.-C. Lin, Y.-H. Wang, H.-J. Hsieh, Development of health care system based on wearable devices, in *2019 Prognostics and System Health Management Conference (PHM-Paris).* (IEEE, 2019), pp. 249–252
14. D. Azevedo, A. Esteves, F. Ribeiro, L. Farinha, J. Metrôlho, A wearable device for monitoring health risks when children play outdoors, in *2020 15th Iberian Conference on Information Systems and Technologies (CISTI).* (IEEE, 2020), pp. 1–6

15. Y. Gu, J. Shen, Y. Chen,Know you better: a smart watch based health monitoring system, in *2019 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. (IEEE, 2019), pp. 7–8
16. H. Fei, M. Ur-Rehman,A wearable health monitoring system, in *2020 International Conference on UK-China Emerging Technologies (UCET)*. (IEEE, 2020), pp. 1–4
17. M. Kavitha, P.V. Krishna, IoT-cloud-based health care system framework to detect breast abnormality, in *Emerging Research in Data Engineering Systems and Computer Communications*. (Springer, Singapore, 2020), pp. 615–625
18. C.S. Kolli, V.V.K. Reddy, M. Kavitha, A critical review on internet of things to empower the living style of physically challenged people, in *Emerging Research in Data Engineering Systems and Computer Communications* (2020), pp. 603–614
19. M. Kavitha, D. Pavani, Y. Lahari, M. Prasanna, Safe navigation system framework for visually impaired users using a micro-digital camera. J. Adv. Res. Dyn. Control Syst. **11**(7), 631–637 (2019)
20. K.P. Babu, G.S.N. Rao, Smart healthcare system. Int. J. Innov. Technol. Explor. Eng. **8**(12) (2019). https://doi.org/10.35940/ijitee.L3667.1081219
21. Ates, H. Ceren, A.K. Yetisen, F. Güder, C. Dincer, Wearable devices for the detection of COVID-19. Nat. Electron. **4**(1), 13–14 (2021)
22. M. Kavitha, G. Gnaneswar, R. Dinesh, Y.R. Sai, R.S. Suraj, Heart disease prediction using hybrid machine learning model, in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. (IEEE, 2021, January), pp. 1329–1333
23. S.K. Jonnavithula, A.K. Jha, M. Kavitha, S. Srinivasulu, Role of machine learning algorithms over heart diseases prediction, in *AIP Conference Proceedings*, vol. 2292(1). (AIP Publishing LLC, 2020, October), p. 040013

Battery Management System Simulator

P. Uthara Balraj and P. Sivraj

Abstract Availability of diverse methods or algorithms to implement the basic battery management system (BMS) functionalities make it a challenging task to choose the best that suits a vehicular feature requirement. This paper focuses on the development of a simulator tool for a software BMS model which incorporates the functionality and implementation approach selected by the user via a user interface. The major functionalities such as state of charge and state of health estimation, cell balancing, and cell protection are implemented using basic algorithms. The model is interfaced to a user interface that enable parameter initialization from the user's end, along with the display of simulation results. MATLAB tools Simulink and App Designer are used for BMS modelling and user interface design, respectively. The user interface has the provision to export simulation results to excel and view the outputs as graph. The work can be extended to include more algorithms and functionalities, and hardware interfaces for development of the simulator into a full fledged teaching–learning and system-in-loop tool.

Keywords Battery management system · State of charge · State of health · Cell balancing · Cell protection · User interface · MATLAB/Simulink

1 Introduction

There has been a growing demand for clean or renewable power, in all sectors, since the past few decades and so the deployment of Electric Vehicles (EV) shot up worldwide. While fossil fuelled combustion engines account for the largest share of CO_2 emissions in the world, EVs with zero tank-to-wheel emissions has a huge chance for large market share in automotive industry [1]. Lithium-ion cells are extensively

P. U. Balraj (✉) · P. Sivraj
Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

P. Sivraj
e-mail: p_sivraj@cb.amrita.edu

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_30

used in EVs over other cell chemistries because of their better performance and low weight [2]. To improve their efficiency, lifespan and most importantly the safety associated with different operating conditions, all Li-ion batteries are accompanied by a battery management system (BMS) that monitors and controls their critical parameters [2]. Each battery pack configuration, cell distribution and cell chemistry demand different cooling strategies, distribution of monitoring resources and battery state estimations. An EV manufacturer must choose the right BMS architecture and software algorithm from the wide pool available to suit their system requirements and to optimise the vehicle as per the target users [3, 4].

Accurate estimation of state of charge (SoC) and state of health (SoH) is crucial in predicting the battery state [3]. It is a challenging task since they rely on multiple physical parameters, operating conditions and even cell chemistries [4, 5]. Baccouche et al. have categorised different estimation algorithms into direct measurement, model-based, impedance-based, adaptive and machine learning-based approaches [6]. Daowd et al. have done detailed comparison of sixteen cell balancing techniques, both passive and active based on cost, complexity, speed etc. [7]. All these approaches have their trade-offs and can distinctively satisfy the varied requirements in EV market. Considering the time, safety and possible equipment damage associated with hardware testing, it is a challenge to evaluate these algorithms for every possible operation scenario. A simulation model to properly validate the algorithms is hence a mandatory before hardware commitment [8]. A desktop simulation model for functional aspects of BMS can verify algorithms and validate them to industry standards. This gives a preliminary insight to battery pack's dynamic behaviour and the overall BMS performance [9, 10].

BMS Reference Design by Intel Corporation implements a user interface (UI)-based system-in-loop simulation where a MATLAB-Simulink model of BMS communicates with an FPGA-based real-time platform that computes SoC using Dual Extended Kalman Filter [11]. However, the model does not implement functionalities such as SoH estimation, cell balancing or cell protection. Liao developed a UI-based ad-hoc BMS learning tool by interfacing commercial BMS hardware and software modules to computer-controlled unit that emulate cell voltages. The voltages could be manually changed on the fly using UI [12]. Even though the tool could emulate and analyse overcharge, over-discharge and balancing condition of each cell, there was no provision to implement alternate algorithms apart from the ones used in the module. Chirag designed a reference model for Li-ion BMS using Simulink-MATLAB that implements SoC estimation, cell balancing, current and fault management [13]. The model lacks a user-friendly medium for selecting algorithms, viewing results, and configuring parameters.

Considering the unavailability of a flexible platform that support multiple modelling approaches for BMS, the work intends to build a UI-based simulation tool that implements major BMS functionalities using basic algorithms. Functionalities focussed by the work includes cell balancing, cell protection, SoC and SoH estimation. Multiple and complex algorithms can be added to the tool to realise these functionalities. User may select them through the UI from where model parameters can be initialised and results can be visualised as well.

The next section gives an insight of different cell balancing, SoC, SoH estimation methods and similar BMS simulation tools. Section 3 describes the system overview and methodology which is used for the system implementation discussed in Sect. 4. Section 5 discusses the result and Sect. 6 concludes the paper.

2 Related Work

A detailed survey is done on literature covering algorithms that implement BMS functionalities such as SoC, SoH, cell balancing; important terminologies associated with each, and similar available BMS development tools.

The SoC of a battery expresses available capacity, in Ampere hours, as a percentage of its rated capacity. Accurate SoC estimation is important to prevent unpredicted system interruption [14]. Coulomb counting is one of the widely used SoC estimation technique where charging/discharging current is integrated over the time to calculate the coulombs entered or left the cell [15]. However, factors like self-discharge, charging and discharging losses or inaccurate initial SoC value can add cumulative errors to this approach [14, 16]. Another known way to estimate SoC is by determining cell's open-circuit voltage (OCV) which depend on electrolyte concentration that varies with cell charge [17]. This method is only valid if the battery is in unloaded state. Improved coulomb counting method where initial SoC is found using piecewise linear SoC-OCV relationship is discussed in [6, 14, 20]. Tejaswini, and Sivraj [19] uses a purely data driven approach where SoC is estimates using artificial neural network (ANN) and support vector machine (SVM) models. Although it works for all battery chemistries, large datasets are necessary for accurate modelling. SoH on the other hand relates to loss of rated capacity in a cell due to cycling and calendar ageing. It indicates the state of the battery, in percentage, between the beginning of life and end of life of a cell beyond which it cannot meet minimum operational requirements [20]. The main indicators of SoH are battery internal resistance, battery impedance and capacity whose values change over time due to internal and external factors [21]. Landi and Gross propose two methods, fuzzy logic, and neural network for SoH estimation in vehicle-to-grid applications. In the former approach, a double exponential fitting curve for capacity decay is obtained for all the batteries of same family. From the curve, making use of fuzzy logic, a health index that is dependent on temperature, DoD and current is obtained. Latter method is computationally intensive, thus need powerful processors like graphics processing units for implementation [22].

Cells that constitute a battery pack may not have perfectly equal characteristics because of the differences in their manufacturing or environmental conditions. Over the time, due to charge–discharge process, the voltage of these cells can differ and deteriorate the performance and reliability of the battery pack. Cell equalisation is an important function of BMS, and the balancing topologies can be classified as passive and active balancing [23, 24]. A detailed comparison of different topologies that fall under the classification can be done based on cost, complexity, speed, and

intelligent control. Each of those models have numerous advantages and trade-offs [7]. Passive balancing uses a bleed resistor to connect each cell via a controlled switch so that any imbalance can be equalised by draining power from the high energy cells. This method is the least efficient because of heat loss and temperature rise [25]. Active balancing works by active transportation of energy within cells where no amount of energy is wasted [7]. Capacitor-based, inductor/transformer-based, and converter-based active cell balancing topologies are broadly discussed in [26], while [23] discusses a flyback transformer-based active balancing that takes energy from primary windings connected across overall battery pack and transfer it to its secondary connected to each cell, equalising SoC of all the cells.

Simulink serves a good platform for BMS desktop simulation using model-based design approach. It gives provision to develop supervisory control algorithms, for SoC and SoH, using state machines and flowcharts [11]. A Simulink BMS model with individual subsystems for SoC estimation, battery modelling and thermal management is discussed in [27]. Kumar et al. [28] presents a Neuro-Fuzzy model in Simulink that could be connected to HDL coder for BMS chip design. These systems implemented multiple algorithms and compared their performances using Simulink tool. Chirag [13] describes development of an extensive closed loop BMS model in Simulink with a harness dashboard that indicates fault conditions such as overvoltage, overcurrent etc. The model implements coulomb counting, extended Kalman filter and unscented Kalman filter for SoC estimation and uses Stateflow for passive cell balancing. Jacobitz et al. [29] highlights development of a continuous model-based verification-oriented design methodology for software-hardware BMS model. This is a scalable, low-cost design and performs Model-in-Loop simulation (MiL) for system validation. With coulomb counting, it implements cell balancing by means of charge transfer among adjacent cells so to maintain the mean SoC value of the battery pack. Shah et al. [30] develops a platform software for BMS in EVs to monitor cell voltages of 6 to 16 cells simultaneously and implements passive balancing.

SoC, SoH, cell balancing, and protection are identified as the main functionalities of BMS, and a range of methods that are available were discussed in this section. The BMS simulators in [13, 27–30] use multiple modelling approaches to realise these functionalities, but none support quick model reconfiguration using a friendly user interface or quick access to outputs for further detailed analysis and model comparisons. Therefore, a scalable comprehensive UI-based simulator that support multiple configurations and modelling approaches has large scope and can be developed using platforms like MATLAB that provide tools such as Simulink, Stateflow, App Designer etc. to implement different functional aspects and even features for system-in-loop simulations.

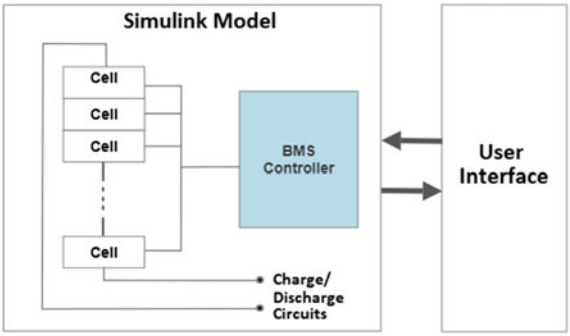


Fig. 1 System overview

3 System Overview and Methodology

3.1 System Overview

Two integral units of the overall system as shown in Fig. 1 are a Simulink BMS model that implements various functionalities, and a MATLAB UI integrated to it. These units exchange model parameters before and after the simulation through the MATLAB workspace.

Simulink Model This is a BMS model with cells in series connected to discharging/charging circuit. The cell model contains the circuitry for cell balancing and protection as well. These cells are then continuously monitored by a BMS controller block which take necessary inputs such as voltage, current, temperature, etc. from cells and constitute subblocks that perform each of different functionalities. The BMS controller block estimates SoC, SoH, and implement logics for cell balancing and model protection using defined methods.

User Interface UI support configuration of parameters, algorithm selection, and plotting and exporting simulation results. A protection scheme is implemented where out-of-bound values cannot be given as input when tool is configured. The UI incorporates features to run the simulation, obtain the results as graphs and export the results to an excel file.

3.2 Methodology

The methodology can be largely divided into two steps as shown in Fig. 2. First, a skeleton BMS framework is created where the initial objective is selection and configuration of a cell model. SoC and SoH algorithms are then implemented, and the model is analysed for different charge/discharge cycles. More identical cells are added to the framework; cell balancing, and protection logic are then implemented.

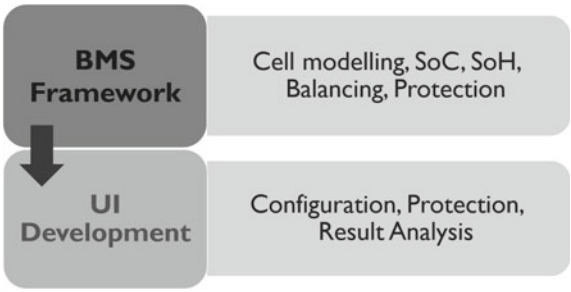


Fig. 2 Workflow of implementation

All implementations are grouped to subsystems—one for charge/discharge cycle generation, one for battery model and one for the BMS controller block that monitors the cell and perform functionalities such as state estimations, protection, and balancing. Each model parameter is defined as a variable in workspace to enable access and control from UI.

UI development is the next phase where a basic layout is initially built to display cell specifications, drive cycles, simulation settings and the algorithms used. These data fields are then connected to the previously defined model variables in workspace so that user may change them to reconfigure the simulation model through UI. A protection layer is then implemented in UI to avoid any data entry that may cause over current, deep-discharge, overcharge, and unsafe temperature conditions in the simulation. Finally, options to run the simulation, plot the outputs, and export the data to an excel file are provided in the interface.

4 Implementation

4.1 BMS Framework

A basic BMS framework in Simulink with three main subsystems, each for charge/discharge cycles, battery model and BMS controller are shown in Fig. 3. The load profile generated from the charge/discharge cycle block feed corresponding currents to the battery model which gets continuously monitored by the BMS controller block that implement algorithms for SoC, SoH, cell balancing and protection. The controller block measures necessary cell parameters from the battery model to implement these functionalities.

The cycle generation block, shown in Fig. 4 create continuous charge/discharge cycles based on five different timeseries inputs–charge/discharge current, DoD, current SoC and maximum SoC. The function logic switches from charge to discharge when current SoC reaches the defined maximum SoC level and vice versa when it reaches maximum DoD. Correspondingly charge/discharge currents are generated from this block.

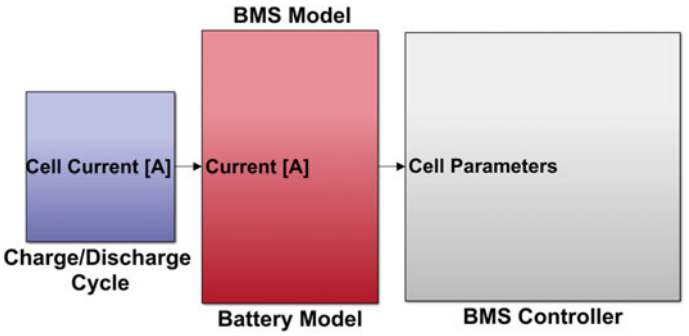


Fig. 3 Simulink framework

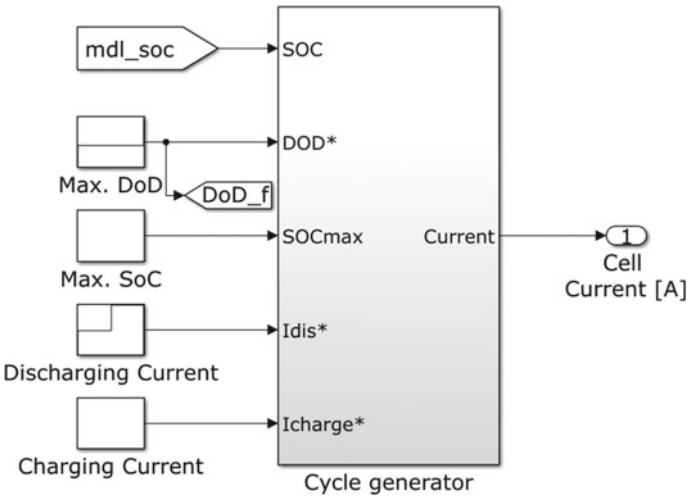


Fig. 4 Drive cycle generation block

Two battery models are implemented in the system based on their temperature dependency. One is a temperature dependant model and other an independent model. One of them can be selected by means of a switching circuit as shown in Fig. 5, where a variable T_en can be set/reset to enable/disable the dependency.

Three MATLAB generic battery blocks in series configuration with passive cell balancing circuit, shown in Fig. 6, is implemented for both temperature dependant and independent models. Parameters such as cell nominal voltage, capacity, initial SoC/SoH, and temperature dependency are set to default values through workspace variables that can be reconfigured, when required, through UI. A 3.6 V, 2.05 Ah $LiCoO_2$ model with temperature effects is chosen by default. A function logic for passive balancing is written for enabling and disabling control signals of MOSFET switches that connect each cell to the bleed resistor. During an imbalance, switches

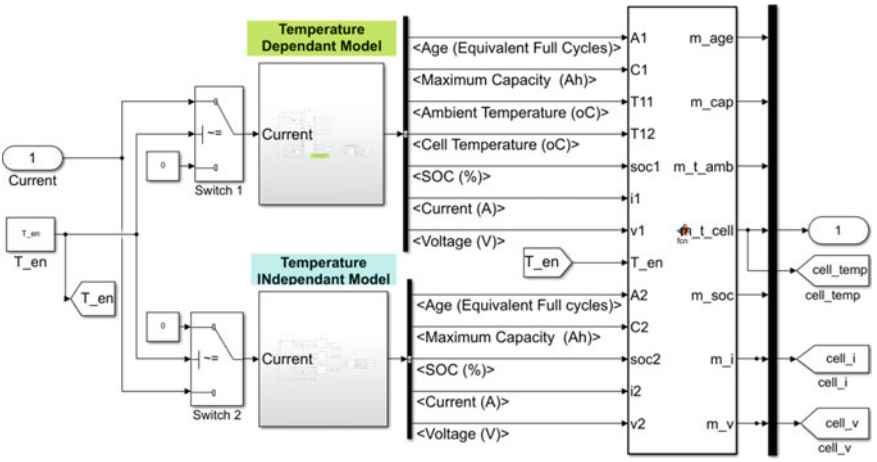


Fig. 5 Simulink implementation for battery model selection

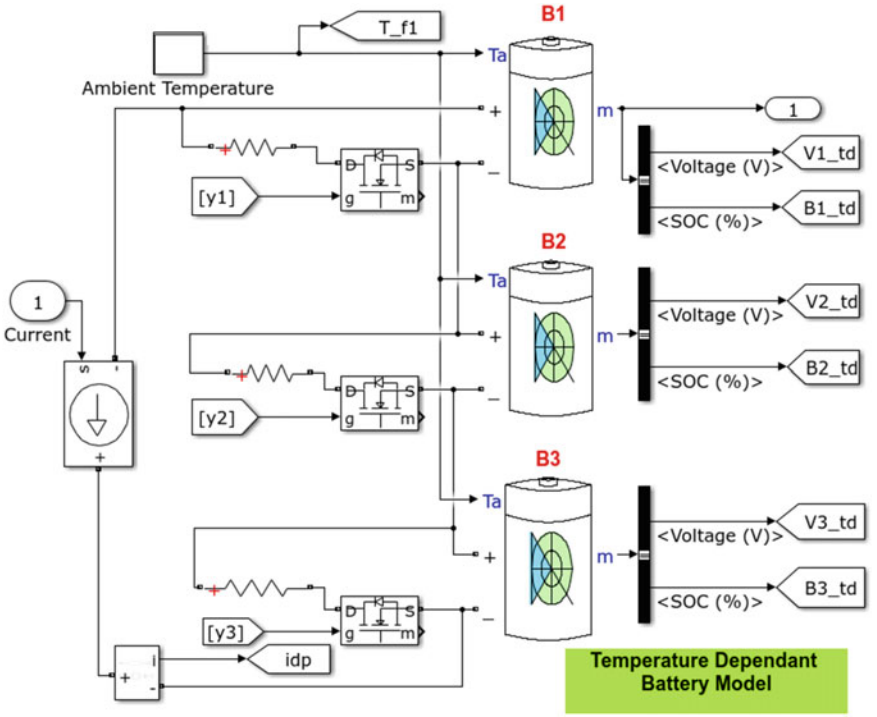


Fig. 6 Implementation of battery model with balancing circuitry

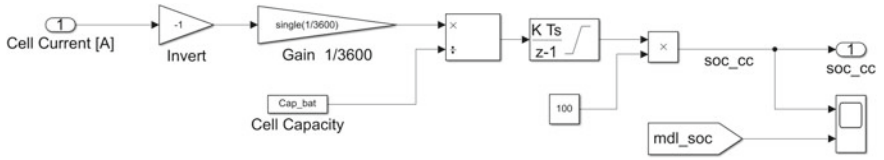


Fig. 7 Simulink implementation of SoC estimation using coulomb counting

for high energy cells are enabled until their SoC reach the same level as the cell with minimum SoC in the battery pack. Values such as cell voltage, current, temperature etc. are taken out from the battery model during the simulation and is fed to the BMS controller block for implementation of protection, SoC and SoH estimation functionalities.

Coulomb counting is used to estimate SoC in the simulator tool. State of charge of a cell at any point express the amount of releasable charge with respect to the cell's maximum charge capacity at that point. This can be represented as [16]:

$$\mathrm{SoC} = \frac{\mathrm{Q_{releasable}}}{\mathrm{Q_{rated}}} \tag{1}$$

Since charge is integral of current over time, amount of charge that moves in and out of the cell can be found out by integrating the current that goes in and out. If initial charge of the battery is known, the newly received or released charge can be added or subtracted to find the current state. Thus, SoC can be calculated as [14]:

$$\mathrm{SoC}(\tau + 1) = \mathrm{SoC}(\tau) + \frac{\int_{\tau}^{\tau+1} I \, \mathrm{d}t}{Q_{\mathrm{rated}}} \tag{2}$$

where $\mathrm{SoC}(\tau)$ represents the state of charge at time τ, I is charge/discharge current in amperes and Q_{rated} is the rated ampere hour capacity of the cell. The schematic of coulomb counting implementation is shown in Fig. 7.

Initial SoC and rated cell capacity values are already obtained from the configured battery model. Current going in and out of the battery is integrated using a discrete-time integrator block with initial SoC as initial condition. The integrated current is then multiplied by a gain value of 1/3600 to obtain received/released charge in Ah. The ratio of this Ah capacity to rated capacity calculates SoC for the time step.

Fuzzy logic is the method used to estimate SoH in the simulator tool. SoH can be expressed as:

$$\mathrm{SoH} = \frac{Q_M}{Q_{\mathrm{rated}}} \tag{3}$$

where Q_M is maximum releasable capacity of the cell in Ampere hours and Q_{rated} is rated Ampere hour capacity [20]. Q_M is largely affected by charge/discharge

current, temperature, DoD, and number of equivalent charge–discharge full cycle. Hence, based on [22], a SoH estimation model that is dependent on all the above factors is developed using Simulink fuzzy logic controller blocks.

The generic battery model with the mentioned specifications is run with well-defined DoD, temperature, and current values for over 100 h to obtain a capacity decay curve that can be approximated as equation:

$$y = ae^{-bx} \tag{4}$$

where y is the cell capacity and x is the equivalent number of full charge/discharge cycles. a and b are curve coefficients solved with equations of local approximation [22] implemented as function logic inside Simulink. The values obtained have dependency only on cycle number x, and hence, a new equation is considered [22]:

$$y = a_f e^{-b_f x} \tag{5}$$

Here,

$$a_f = a * ad \tag{6}$$

and,

$$b_f = b * bd \tag{7}$$

ad and bd are corrective coefficients that depend on temperature, discharge current and DoD and are determined using fuzzy algorithm.

Mamdani type fuzzy logic controllers are used to define relationships between the coefficients a and b with the three parameters. Membership functions are created, and rule base developed to mimic the decay curve obtained before. Coefficients are passed through each controller one after the other as shown in Fig. 8 to obtain ad and

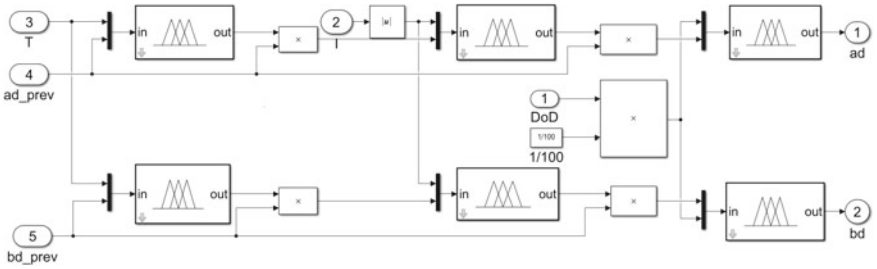


Fig. 8 Internal architecture of fuzzy blockset for SoH implementation

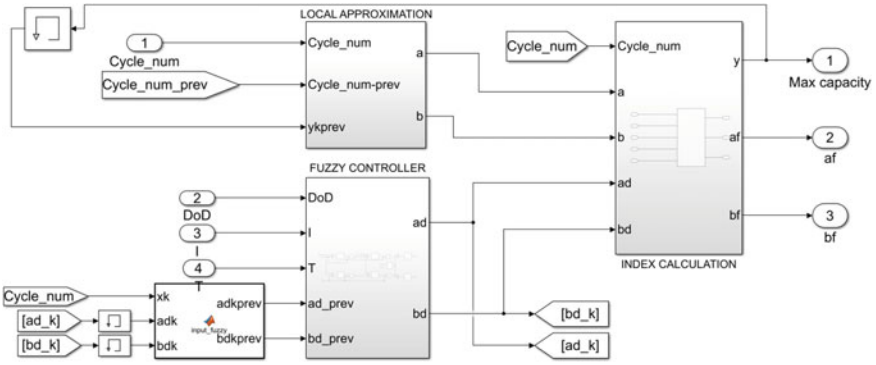


Fig. 9 Fuzzy logic implementation for SoH estimation in Simulink

bd. Finally, (5) is calculated using the index calculation subsystem in Fig. 9 where overall fuzzy implementation is shown.

The overall system predicts the new value of y on each time step based on the two previous values of x, a and b. The accuracy of the system is largely dependent on the type and number of membership functions, and the defined rule base.

A basic protection scheme as shown in Fig. 10 is implemented, where a function block continuously checks the charging/discharging currents, maximum SoC and DoD values, and maximum/minimum temperature values throughout the simulation. Simulation will be stopped if charging temperature exceeds the range of (0–45) °C or if the discharging temperature goes below −10 °C or above 60 °C. Similarly, the allowed ranges for charging current, discharging current, DoD and SoC are 0–1.3 A, 0–4 A, 0–99%, 0–95%, respectively. Once the simulation stops, a 'StopFcn' model call-back function is called. The function identifies the outlier and triggers the corresponding error dialogue that asks user to correct the model.

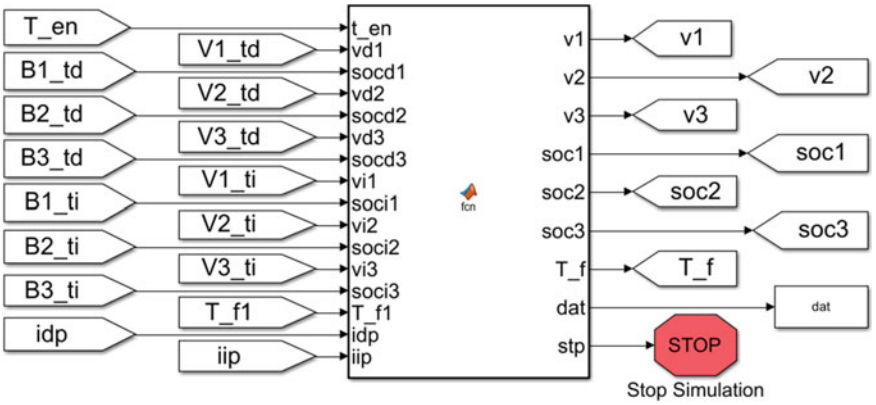


Fig. 10 Simulink implementation of basic protection scheme

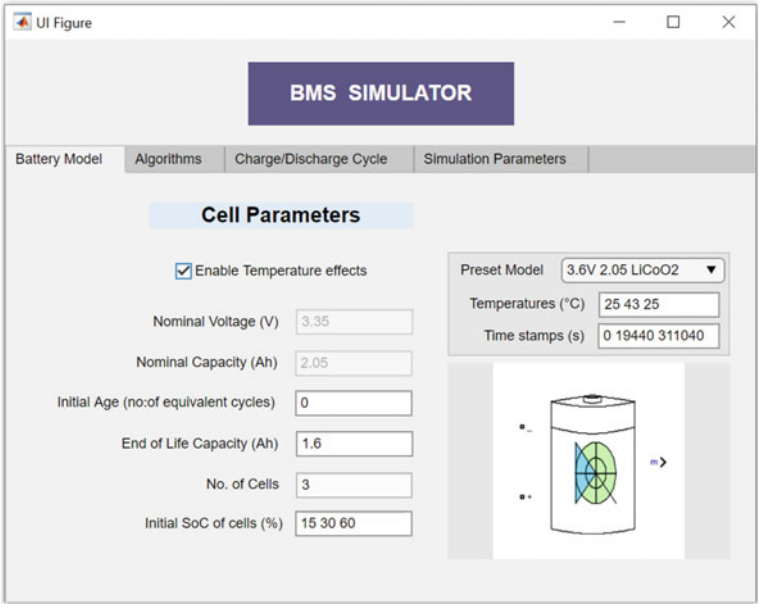


Fig. 11 UI window for the tool

4.2 UI Development

MATLAB's App Designer tool is used to implement the UI. UI is designed to have one single window with a tab each for cell configuration, algorithm identification, drive cycle definition, and result analysis. Only necessary model parameters are given so that user can edit them to modify the model. All fields are filled with default values upon opening of the tool, any edits trigger a function call back that executes a piece of code to save the changes to necessary workspace variables which will in turn reflect in the Simulink model.

A protection layer for data entry is implemented where any out-of-bound values entered by the user initiates pop-up with error message. Permissible values for temperature, charge/discharge current, DoD or maximum SoC are 0–45 °C, 0–1.3 A, 0–4 A, 0–99%, 0–95%, respectively. Entries are not saved unless they pass this layer. A push button is implemented, that initiates call-back to open and run the Simulink model with configured values. Similar buttons are created which can be used for opening the Simulink scope and for exporting values of cell voltage, current, temperature, SoC and SoH.

5 Results and Discussion

The simulator tool is opened by running the UI whose main window is as shown in Fig. 11. It has four main tabs with all parameter fields populated with default values. Each tab and their functions are explained with figures in this section.

The first tab configures the battery model based on nominal voltage, nominal capacity, initial age, and end of life capacity. Here the default model is a 100% healthy $LiCoO_2$ cell of 3.6 V, 2.05Ah with temperature dependency, all specified in Fig. 11. Initial SoCs of each cell is given as 15%, 30%, and 60%. Different ambient temperatures along with their timestamps can be provided, however, unchecking the temperature effects will disable these edit fields. The second tab, shown in Fig. 12, displays algorithms used for implementing the functionalities—SoC, SoH and cell balancing. The fields are given as dropdowns so that any future additions to the model can be displayed and required algorithms can be chosen from.

The third tab shown in Fig. 13 has fields to input charge/discharge currents, maximum SoC and DoD values along with their timestamps of occurrence so that a custom drive cycle with constant current charging/discharging can be defined. Charging current and maximum SoC values are fixed at 1 A and 99%. Discharge current and DoD values are changed at three different timestamps in the simulation. The fourth tab, shown in Fig. 14, provides option to set simulation stop time, run the simulation and view, or export the result. A default simulation time of 4.5 days is given so that there is considerable battery health degradation from the large number

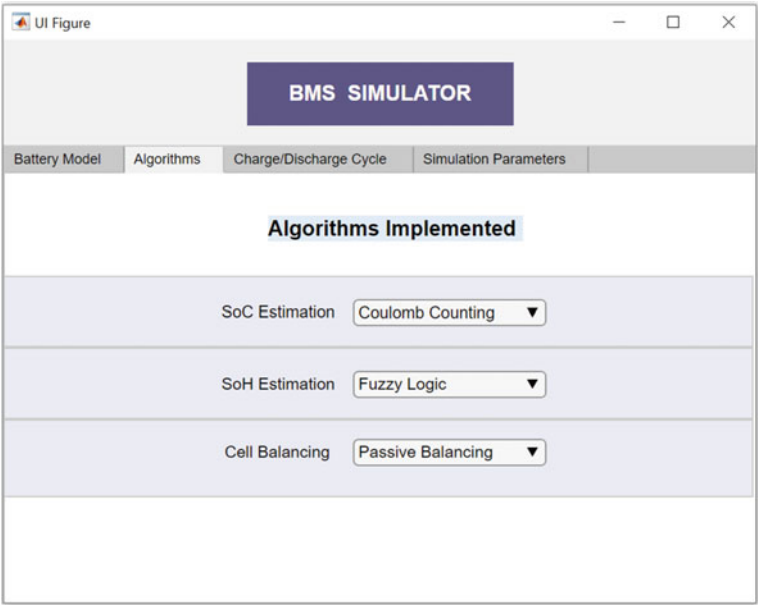


Fig. 12 UI tab for algorithm selection

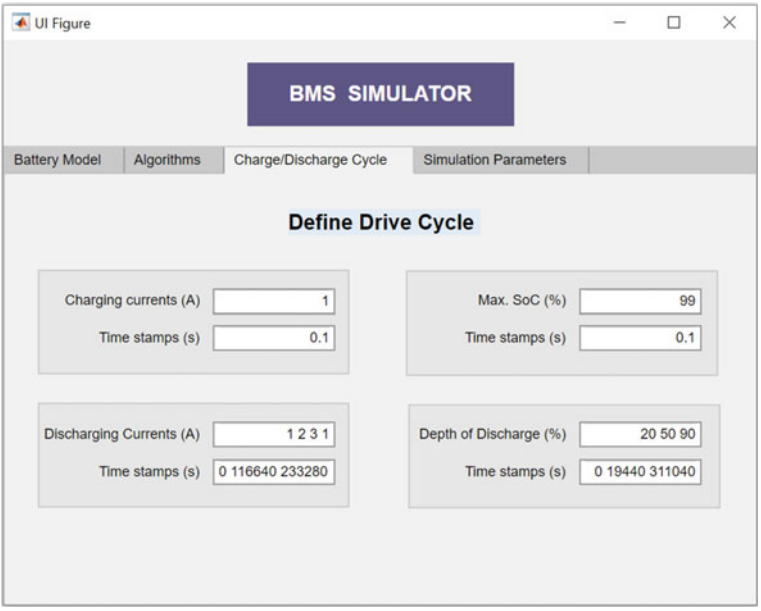


Fig. 13 UI tab to configure drive cycle

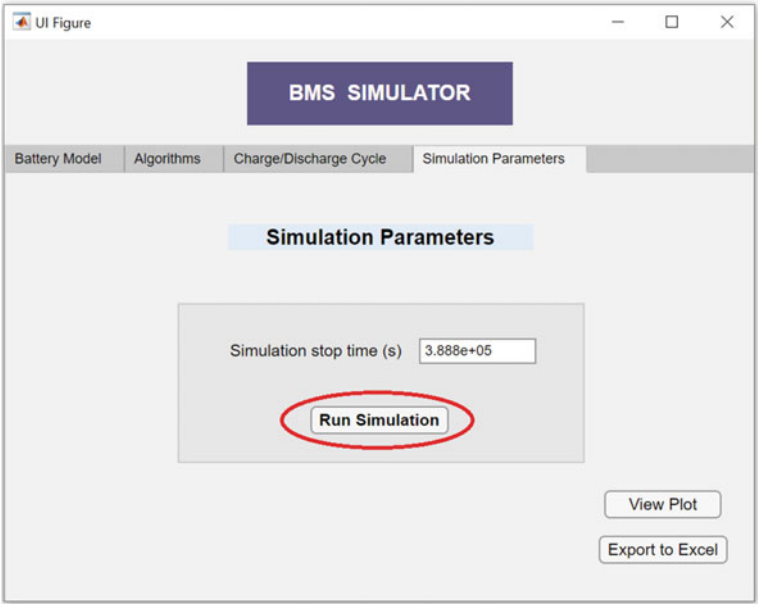


Fig. 14 UI tab to run simulation

of charge/discharge cycles. Simulation can then be run from the UI and the plots of critical cell parameters such as voltage, current, temperature, along with the estimated battery states are made available to be viewed under the button 'View Plot'. Also, there is a provision to export these values to an excel file for further analysis using the 'Export to Excel' button.

Figure 15 shows how an entry of −10 °C in the temperature field of the first tab trigger an error dialogue. Similar error messages were generated if the values for charge current, discharge current, maximum SoC or DoD exceed permissible ranges. The battery protection scheme implemented stops the charging or discharging operation of battery with warnings for abnormal values of battery and operational parameters like charging current, discharge current, temperature, DoD, and maximum SoC.

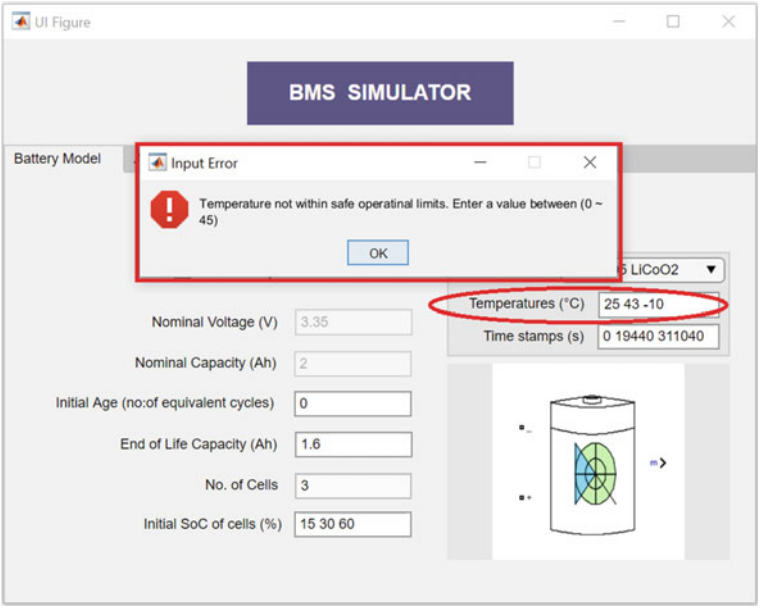


Fig. 15 Error dialogue from UI

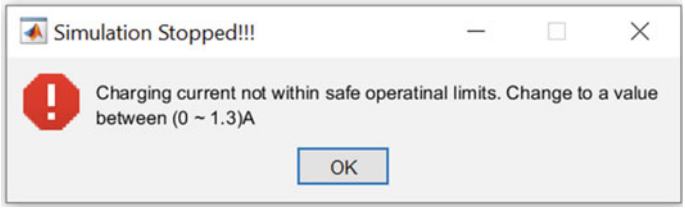


Fig. 16 Error dialogue while simulation

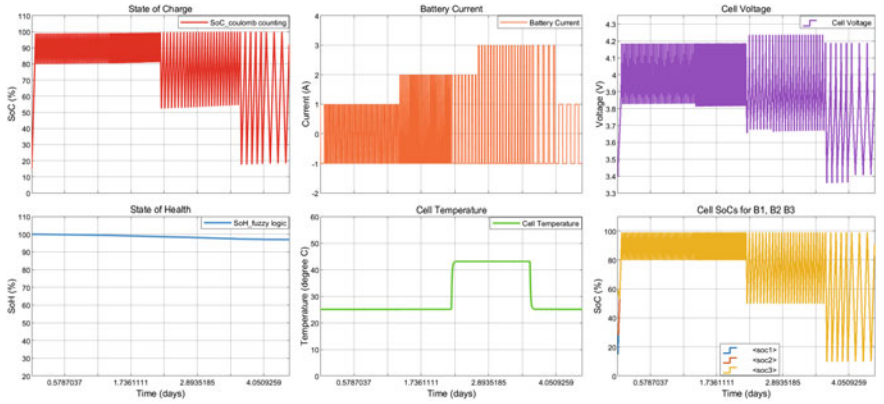


Fig. 17 Simulation results for the temperature dependent model

Figure 16 shows such a scenario where the charging process is abruptly stopped on charging current exceeding the set upper limit.

A sample result plot when the model is operated under acceptable conditions is as shown in Fig. 17. First five graphs plot SoC, SoH, charge/discharge current, cell temperature and cell voltage of the first battery, and the last one plots SoC of all three batteries. Simulation, carried out here, is divided into three stages based on the value of DoDs starting with 20%, which changes to 50% at 54 h and then to 90% at 87 h.

The SoC starts rising from the set value of 15% until 99% and then decreases till 80% and continues the pattern. The values are accurate until when discharge current rises from 1 to 2 A in the first stage. Lower side of SoC curve tends to shrink from this point onwards and do not hit the expected values throughout the simulation. This occurs due to capacity degradation that does not get corrected in the model. Analysing the SoH curve shows the cell degradation with increased load current, but the rate of degradation is much higher when temperature changes to 45 from 25 °C. Ageing is quicker with increased DoD and temperature, although has less slope variation with discharge current. Model can thus be corrected by improving the rule base and the corresponding membership functions. To check the balancing operation, initial SoCs for the three cells in series were given a 15, 30 and 60%. Model starts with charging the batteries at 1 A, and by the end of fifty minutes, SoCs of all three cells converge to 53%, thus performing the passive cell balancing. This can be viewed from the last graph of Fig. 17. Also, the export output option records the simulation outputs such as time (s), cell temperature (°C), cell current (A), cell voltage (V), SoC (%) and SoH (%) to an excel sheet till the last time step.

6 Conclusion

The availability of plenty of modelling approaches for BMS suggests a tool that can support multiple model implementations. A UI-based platform that simulates a BMS model as per user's configuration is implemented using MATLAB tools Simulink and App Designer. The model with temperature dependent/independent generic battery blocks in series implement passive cell balancing, model protection, SoC estimation using coulomb counting and SoH estimation using Fuzzy Logic. The model could successfully run through the UI for a pre-set configuration and obtain the important cell parameters which then can be properly analysed with generated plots and results that can be exported.

The simulator tool eases the effort to reconfigure the model from scratch in case of integration of alternate estimation algorithm or battery model or a new functionality. Parameters and algorithms can be mix-matched to analyse and compare their effects on critical battery states such as SoC and SoH. The tool can further be improved through addition of new battery models, drive cycles, test cases, and other BMS functionalities. Adding better UI features and improved algorithms can elevate the model to an advanced software tool that can later facilitate hardware interfacing and system-in-loop simulation.

References

1. Snehalika, S. Das, Emerging trends in electric vehicle in Indian market, in *3rd International Conference on Recent Developments in Control, Automation and Power Engineering* (2019), pp. 514–518
2. Y. Miao, P. Hynan, A. von Jouanne, A. Yokochi, Current li-ion battery technologies in electric vehicles and opportunities for advancements. Energies **12**, 1074–1094 (2019)
3. A. Petrova, Technology overview on software development for battery management systems (2020). https://www.integrasources.com/blog/battery-management-systems-software-development/. Accessed 22 Jun 2021
4. T. Lombardo, A look at BMS validation testing (2020). https://chargedevs.com/features/a-look-at-bms-validation-testing/. Accessed 22 Jun 2021
5. Cadex Electronics Inc, BU-1003: electric vehicle (EV) (2019). https://batteryuniversity.com/article/bu-1003-electric-vehicle-ev. Accessed 22 Jun 2021
6. I. Baccouche, S. Jemmali, A. Mlayah, B. Manai, N.E.B. Amara, Implementation of an improved coulomb-counting algorithm based on a piecewise SOC-OCV relationship for SOC estimation of li-ion battery. Int. J. Renew. Energ. Res. **8**(1), 178–187 (2018)
7. M. Daowd, N. Omar, P.V.D. Bossche, J.V. Mierlo, Passive and active battery balancing comparison based on MATLAB simulation, in *2011 IEEE Vehicle Power and Propulsion Conference* (2011), pp. 1–7
8. S. Hwang, Developing an HIL simulator for testing battery management system logic. https://www.ni.com/en-in/innovations/case-studies/19/developing-an-hil-simulator-for-testing-battery-management-system-logic.html. Accessed 22 Jun 2021
9. T. Lennon, Validating battery management systems with simulation models (2020). https://eepower.com/technical-articles/validating-battery-management-systems-with-simulation-models/#. Accessed 22 Jun 2021

10. The MathWorks Inc, Developing battery management systems with simulink and model-based design (2018). https://in.mathworks.com/content/dam/mathworks/white-paper/simulink-bms-development-white-paper.pdf. Accessed 22 Jun 2021
11. Intel Corporation, Battery management system reference design (2016). https://www.intel.com/content/www/us/en/programmable/documentation/dmi1457349952129.html. Accessed 22 Jun 2021
12. G.Y. Liao, BYOE: learning tool for lithium-ıon battery management system, in *ASEE Annual Conference & Exposition*, New Orleans, LA (2016)
13. Chirag, Design and test lithium-ıon battery management algorithms (2021). https://in.mathworks.com/matlabcentral/fileexchange/72865-design-and-test-lithium-ion-battery-management-algorithms. Accessed 22 Jun 2021
14. M. Murnane, A. Ghazel, A closer look at state of charge (SOC) and state of health (SOH) estimation techniques for batteries. Accessed 22 Jun 2021
15. S. Pai, M.R. Sindhu, Intelligent driving range predictor for green transport, in *First International Conference on Materials Science and Manufacturing Technology,* Coimbatore, TN, ICMSMT-vol. 561. (IOP Publishing Ltd, 2019), p. 012110
16. A. Harigopal, S. Nithin, Assessment of state of charge estimation techniques for li-ion battery pack, in *International Conference on Smart Electronics and Communication*, Trichy, TN (2020), pp. 988–991
17. Y.H. Chiang, W.Y. Sean, J.C. Ke, Online estimation of internal resistance and open-circuit voltage of lithium-ion batteries in electric vehicles. J. Power Sour. **196**(8), 3921–3932 (2011)
18. K.S. Ng, C.S. Moo, Y. Chen, Y. Hsieh, Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries. Appl. Energ. **86**(9), 1506–1511 (2009)
19. P. Tejaswini, P. Sivraj, Artificial intelligence-based state of charge estimation of li-ion battery for EV applications, in *5th International Conference on Communication and Electronics System,* Coimbatore, TN (2020), pp. 1356–1361
20. P.K.S. Venkatesh, B.P. Divakar, Real time estimation of SOC and SOH of batteries. Int. J. Renew. Energ. Resour. **8**(1), 44–55 (2018)
21. N. Noura, L. Boulon, S. Jemeï, A review of battery state of health estimation methods: hybrid electric vehicle challenges. World Electr. Veh. J. **11**(4), 66 (2020)
22. M. Landi, G. Gross, Measurement techniques for online battery state of health estimation in vehicle-to-grid applications. IEEE Trans. Instrum. Meas. **63**(5), 1224–1234 (2014)
23. A.F. Moghaddam, A.V. Bossche, Flyback converter balancing technique for lithium-based batteries, in *8th International Conference on Modern Circuits and Systems Technologies*, Thessaloniki, Greece (2019), pp. 1–4
24. S. Hemavathi, Overview of cell balancing methods for li-ion battery technology. Energ. Storage **3**(2) (2020)
25. K. Scott, S. Nork, Passive battery cell balancing. https://www.analog.com/en/technical-articles/passive-battery-cell-balancing.html. Accessed 22 Jun 2021
26. T. Duraisamy, D. Kaliyaperumal, Active cell balancing for electric vehicle battery management system. Int. J. Power Electron. Drive Syst. **11**(2), 571–579 (2020)
27. B.P. Divakar, H.J. Wu, J. Xu, H.B. Ma, W. Ting, K. Ding, W.F. Choi, B.F. Huang, C.H. Leung, Battery management system and control strategy for hybrid and electric vehicle, in *3rd International Conference on Power Electronics Systems and Applications*, Hong Kong (2009), pp. 1–6
28. B. Kumar, N. Khare, P.K. Chaturvedi, Advanced battery management system using MATLAB/Simulink, in *IEEE International Telecommunications Energy Conference*, Osaka (2015), pp. 1–6
29. S. Jacobitz, S. Scherler, X. Liu-Henke, Model-based design of a scalable battery management system by using a low-cost rapid control prototyping system, in *Thirteenth International Conference on Ecological Vehicles and Renewable Energies*, Monte Carlo, Monaco (2018), pp. 1–7

30. S. Shah, M. Murali, P. Gandhi, Platform software development for battery management system in electric vehicle, in *IEEE International Conference on Sustainable Energy Technologies and Systems*, Bhubaneswar (2019), pp. 262–267

Analysis of RNN Capsule Model for Multiclass Imbalanced Data

Shreekant Jere, Annapurna P. Patil, and Ananya Muralidhar

Abstract Sentiment analysis, often known as opinion mining, is the process of mining text for views, attitudes, and emotions. Existing models are a tedious process as they require a significant amount of effort to develop the quality of instance representations and linguistic knowledge, which they depend on. Real-world datasets frequently lack training input for regular classifiers with some classes being overrepresented. With little data to learn from, imbalanced data raises issues in machine learning classification and outcome prediction becomes more challenging. This research analyzes a no-linguistic-knowledge RNN-based capsule model for sentiment analysis of a multiclass imbalanced emotion dataset. The RNN capsule model allows each capsule to produce words with sentiment inclinations that match the sentiment category to which it has been assigned. The RNN's hidden vectors provide the instance representation of an input instance, from which each capsule outputs the state probability via its probability module and the reconstruction representation via its reconstruction module. Ensemble modeling is used to enhance the performance of the RNN capsule model, in which each classifier in the ensemble model evaluates a portion of the dataset comprising balanced data.

Keywords Emotion classification · Imbalanced data · RNN capsule

1 Introduction

RNN capsule [17] achieved state-of-the-art performance on sentiment classification tasks with balanced data (Stanford Sentiment Treebank, Movie Review, and Hospital Feedback datasets). The principle of the RNN capsule model is to create a basic capsule structure and utilize each capsule to focus on a single emotion category. Each capsule generates an active probability and a reconstruction representation. The

S. Jere (✉) · A. P. Patil · A. Muralidhar
M S Ramaiah Institute of Technology Affiliated to Visvesvaraya Technological University, Belagavi, Bengaluru, Karnataka, India
e-mail: shreekantjere@msrit.edu

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_31

aim of learning is to increase the capsule's state probability that matches the ground truth while minimizing the reconstruction representation with the provided instance representation. Simultaneously, the active probability of the other capsules must be minimized, with the distance between their reconstruction representations and the instance representation maximized. As a result, the RNN capsule model achieves state-of-the-art sentiment classification accuracy without carefully designed instance representations or linguistic knowledge and outputs the words that best reflect the sentiment category.

The data collected from many real-world NLP applications has a skewed distribution [6, 8, 18]: data from a few classes appear more frequently than data from other classes. Tweets about shootings or fires, for example, are generally less common than those about sports or entertainment. These data samples are usually the focus of attention since their scarcity may convey important and useful information. However, due to their disadvantage in the population, most learning algorithms employ them inefficiently [11]. As a result, creating discriminative models for imbalanced class distribution has become an important and difficult challenge for the machine learning community.

The solutions put forth in prior research may be generally divided into three groups [11]: (1) Data-level methods that utilize under-sampling or oversampling to balance class distributions [2, 13, 14, 20]. (2) Algorithm-level techniques for modifying present learners' to reduce their bias toward majority classes with the most common branch being cost-sensitive algorithms, which assign a higher cost to misclassifying minority class occurrences [5]. (3) Ensemble-based techniques combine the advantages of data-level and algorithm-level methods by integrating data-level solutions with classifier ensembles to produce resilient and efficient learners [9, 15].

This research describes an ensemble RNN capsule model for text classification that can handle imbalanced data. The emotions dataset is used to evaluate RNN capsule architecture for multiclass classification of imbalanced data. Each tweet is classified into one of 13 categories: "boredom," "anger," "empty," "fun," "enthusiasm," "happy," "love," "hate," "relief," "surprise," "neutral," "concern," and "sadness" in multiclass imbalanced data classification. To implement an ensemble model, the classes of emotion datasets are grouped such that each group has balanced data.

The primary contributions of this article are as follows:

1. Detailed literature on RNN capsule architecture and sentiment classification with imbalanced data is provided.
2. A methodology for implementing the ensemble RNN capsule model utilized to manage the imbalanced data is presented.
3. Experimental analysis is conducted on ensemble RNN capsule model with imbalanced emotion dataset.

The remainder of this article is structured as follows: Sect. 2 provides a brief overview of RNN capsule architecture and sentiment classification with imbalanced data. Section 3 describes the methodology. Next, Sect. 4 examines the RNN capsule model's experimental performance on the imbalanced emotion dataset. Finally, Sect. 5 presents the conclusions.

2 Related Work

2.1 RNN Capsule Architecture

Yequan Wang et al. [17] attempted to perform sentiment analysis using RNN capsule architecture. Each sentiment capsule has a state, an attribute, and three modules: representation, reconstruction, and probability. No prior linguistic knowledge is required for the RNN capsule. Furthermore, the capsule model is capable of attending opinion words that represent the dataset's domain knowledge. Figure 1 depicts architecture of RNN sentiment capsule model. Each capsule represents a single sentiment class.

RNNs can be realized using LSTM [10], GRU [4], or variants like biLSTM. RNN encodes an instance (e.g., a paragraph or a sentence) and outputs the hidden vectors given a dense vector representation of the instance. In a nutshell, memory cell c_t and the hidden states h_t in LSTM are functions of the previous c_{t-1} and h_{t-1}, and the input vector x_t, or formally:

$$h_t, c_t = LSTM(h_{t-1}, c_{t-1}, x_t) \tag{1}$$

Gated recurrent unit (GRU) [4] merges the input and forget gates into an update gate. Similarly, in GRU, the hidden state h_t indicates the representation of location t.

$$h_t = GRU(h_{t-1}, x_t) \tag{2}$$

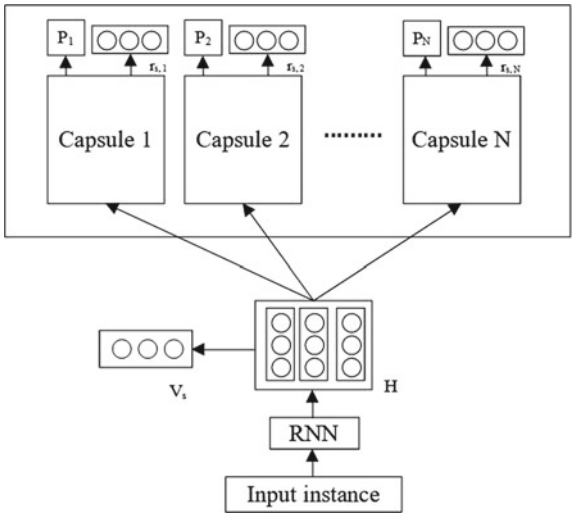


Fig. 1 RNN Capsule architecture. The number of sentiment capsules is equal to the number of sentiment classes

RNN encodes the instance representation for all capsules. Formally, v_s, the instance representation is the arithmetic mean of the RNN-derived hidden vectors.

$$v_s = \frac{1}{N_s} \sum_{j=1}^{N_s} h_j \tag{3}$$

where N_s denotes the instance's length, such as the number of words in a paragraph or sentence. A dense vector is generated for each word using word2vec or other similar techniques.

Each capsule in Fig. 1 outputs a reconstruction representation and a state probability, via its reconstruction and probability modules respectively. The capsule with the largest state probability becomes "active," while the others remain "inactive." One of the training goals is to increase the state probability of the sentiment capsule that corresponds to the specific sentiment while decreasing the state probability of the other capsules. Another goal is to reduce the difference between the reconstruction and the instance representation, as well as to maximize such differences for other capsules.

Figure 2 depicts the architecture of a single sentiment capsule. A capsule is composed of three modules: representation, probability, and reconstruction. The capsule representation $v_{c,j}$ is constructed by the representation module using the attention mechanism. The sigmoid function is used by the probability module to determine the active state probability p_j of the j-th capsule. The reconstruction module calculates a reconstruction representation for a given instance by multiplying $v_{c,j}$ and p_j.

Representation Module. The attention mechanism is used to build capsule representation within a capsule. The representation module uses the attention mechanism to determine the significance of words. Attention mechanism used in RNN capsule architecture is inspired by [1, 7, 16, 19] with a one parameter in each capsule:

$$e_{t,j} = h_t w_{a,j} \tag{4}$$

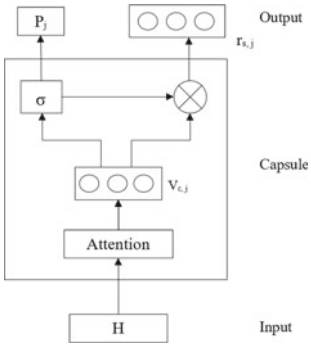


Fig. 2 The single capsule architecture. The hidden vectors $H = [h_1, h_2, \ldots, h_{N_s}]$ from recurrent neural network is the input to a capsule

$$\alpha_{t,j} = \frac{exp(e_{t,j})}{\sum_{k=1}^{N_s} exp(e_{k,j})} \tag{5}$$

$$v_{c,j} = \sum_{k=1}^{N_s} a_{t,j} h_t \tag{6}$$

where h_t denotes the representation of a word at location t and $w_{a,j}$ denotes the attention layer parameter of j^{th} capsule.

The attention score $\alpha_{t,j}$ for each position is computed by multiplying the representations with the weight matrix and then normalizing to a probability distribution over all words. $\alpha_j = [\alpha_{1,j}, \alpha_{2,j}, \ldots, \alpha_{N_s,j}]$. Finally, $v_{c,j}$, the capsule representation vector, is a weighted sum of all the positions, with the attention scores providing as weights.

Probability Module. The active state probability p_j is calculated after obtaining the capsule representation vector $v_{c,j}$ as shown below

$$p_j = \sigma(W_{p,j} v_{c,j} + b_{p,j}) \tag{7}$$

where $W_{p,j}$ and $b_{p,j}$ are the active probability parameters for the current capsule j.

Reconstruction Module. Multiplying probability p_j and $v_{c,j}$ of an input instance yields the reconstruction representation.

$$r_{s,j} = p_j v_{c,j} \tag{8}$$

2.2 Imbalance in the Sentiment Class

Many real-world datasets have an imbalanced class distribution, which means that the proportion of samples from various classes is not evenly distributed [12]. Standard classification methods are hampered by high imbalance because many algorithms anticipate a homogeneous class distribution and fail when faced with a significantly skewed one. Many researchers in the field of imbalanced learning utilize sampling as a method. It solves class imbalance by altering the dataset to generate a more evenly distributed distribution of classes. In computer linguistics, imbalance learning in a multiclass context and sampling have not been systematically investigated. This might be related to the fact that sampling in a high-dimensional environment can be computationally expensive since distances between samples are often used.

The most fundamental types are random oversampling and undersampling. To produce a more balanced distribution, most class samples are eliminated at random in random undersampling. While these two approaches appear to be functionally identical, they present their issues in possible information loss and overfitting with oversampling. Through informed undersampling, information loss might be reduced

by eliminating redundant, noisy, or borderline samples, limiting the loss of any relevant information to the classifier.

When oversampling, synthetic samples from the minority class are created to avoid overfitting and, as a result, the inability to generalize to unknown findings [3]. Although oversampling assists in balancing class distributions, it does not entirely avoid overfitting in situations when class clusters are not well defined. SMOTE + Tomek connections and SMOTE + ENN are two methods that combine well-known oversampling and data collection. These approaches do not support multiclass settings by default; thus, each algorithm was run in two distinct modes: 1 versus all and one vs. neighbor, turning the multiclass dataset into a binary class dataset.

3 Methodology

3.1 Emotions Dataset

The emotions dataset (https://www.kaggle.com/praveengovi/emotions-dataset-for-nlp) contains 40,000 tweets and is being used to evaluate RNN capsule architecture for multiclass classification of imbalanced data. In multiclass imbalanced data classification, each tweet falls into one of 13 categories, which include "anger," "boredom," "empty," "enthusiasm," "fun," "happiness," "hate," "love," "relief," "sadness," "surprise," "worry," and "neutral," The details of emotions dataset are given in Table 1.

3.2 Ensemble RNN Capsule Model to Handle Imbalanced Data

This research work presents an ensemble RNN capsule model for text classification that can handle imbalanced data. To implement this model, group the classes in a given imbalanced dataset so that each group has a balanced classes. Partition the

Table 1 Statistics of imbalanced emotions dataset

Emotion	Anger	Boredom	Enthusiasm	Empty	Hate	Relief	Fun
Number of sentences	110	179	759	827	1323	1526	1776
Percentage	0.27	0.45	1.89	2.07	3.31	3.81	4.44
Emotion	Surprise	Love	Sadness	Happiness	Worry	Neutral	
Number of sentences	2187	3842	5165	5209	8459	8638	
Percentage	5.47	9.6	12.91	13.02	21.15	21.59	

dataset for each group i denoted by D_i to train the classifier C_i. Consider m to be the average number of sentences for classes in group i. Choose m sentences at random from the remaining classes which are not considered in group i and add them to the dataset D_i as the new class O_i. The details of ensemble model are shown in algorithm 1. For new data, each classifier C_i makes a prediction. Eliminate the predictions with class O_i from the list of n_g predictions. Then, from the remaining predicted classes, select the one with the highest classifier accuracy.

Algorithm 1 Ensemble_RNN_capsule_model(D, n)

Require: Imbalanced dataset D with n classes
Ensure: Ensemble RNN capsule model, E
1. Group the classes such that each group has a balanced data. The number of groups is denoted by n_g.
2. for each group
3. n_i = number of classes in the group i
4. D_i = partitioned dataset for classes in the group i
5. n_r = number of classes not considered in the group $i = n - n_i$
6. m = average number of sentences for classes in the group i
7. select m sentences at random from the remaining classes not considered in group i and add them to the dataset D_i with the new class O_i
8. train the RNN capsule classifier C_i with the dataset D_i
9. return ensemble model E with n_g RNN capsule classifiers

4 Experimental Analysis

In this research work, the imbalanced emotion dataset is evaluated using various existing models. We uniformly shuffle the emotion dataset and partition it into three sets: train (70%), test (15%), and development (15%), using a split ratio of 14:3:3. Data were preprocessed by deleting duplicate Tweets, incomplete Tweets and Tweets shorter than 4 words, converting text to lowercase and removing punctuations and stop words from the text. The performance of RNN capsule model along with machine learning models is shown in Table 2. To improve the performance, ensemble method is used. Group the classes of emotion dataset such that each group has a balanced data to develop an ensemble model. Table 3 shows emotion dataset statistics used for ensemble model. Experimental results for ensemble model show the significant improvement in the performance measures given in Table 4. Table 5 compares the ensemble model with oversampling and undersampling methods. Oversampling produces a better F1 score than undersampling and the ensemble method. However, the drawback of oversampling is that the minority can lead to model overfitting.

Table 2 Performance of emotion classification on imbalanced emotion dataset

Model	Accuracy	Precision	Recall	F1
Decision tree	0.251	0.145	0.144	0.144
Random forest	0.307	0.176	0.164	0.169
Naïve Bayes	0.038	0.107	0.089	0.097
SVM	0.352	0.216	0.172	0.191
RNN capsule	0.325	0.16	0.194	0.175

Table 3 Emotion dataset statistics used for ensemble model

Classifier	Classes	Total number of data	Train	Test	Dev
C1	{anger, boredom, other1}	110+179+144 = 433	303	65	65
C2	{enthusiasm, empty, other2}	759+827+793 = 2379	1665	357	357
C3	{hate, relief, fun, other3}	1323+1526+1776+1541 = 6166	4316	925	925
C4	{surprise, love, other4}	2187+3842+3014 = 9043	6331	1356	1356
C5	{sadness, happiness, other5}	5165+5209+5187 = 15561	10893	2334	2334
C6	{worry, neutral, other6}	8459+8638+8548 = 25645	17951	3847	3847

5 Conclusions

In this paper, an experimental analysis is carried out on an RNN-based capsule model for the sentiment analysis of a multiclass imbalanced emotion dataset. The emotion dataset was first evaluated using existing models such as decision tree, random forest, Naive Bayes, SVM, and RNN capsule. The key idea of the RNN capsule model is to design a simple capsule structure and use each capsule to focus on one sentiment category. Each capsule outputs its active probability, reconstruction representation, and the words best reflecting the sentiment category. Learning maximizes the active probability of the capsule matching the ground truth and minimizes the reconstruction representation with the converse taking place with the other capsules. A macro-averaged F-score was used as a performance measure since it evenly weighted both precision and recall in the imbalanced class by penalizing extreme values. The experiment showed that using a single classifier for all emotions reduces accuracy, precision, recall, and F1 score due to data imbalance, yielding values of 0.325, 0.16, 0.194, and 0.175, respectively. Instead, when several classifiers were used in the ensemble RNN capsule modeling, accuracy, precision, recall, and F1

Table 4 Performance of ensemble RNN capsule model

Classifier	Model	Accuracy	Precision	Recall	F1
C1	Decision Tree	0.597	0.697	0.548	0.613
	Random Forest	0.542	0.574	0.833	0.679
	Naïve Bayes	0.583	0.65	0.619	0.634
	SVM	0.528	0.559	0.905	0.691
	RNN capsule	0.636	0.526	0.588	0.556
C2	Decision Tree	0.596	0.582	0.648	0.613
	Random Forest	0.619	0.592	0.739	0.657
	Naïve Bayes	0.508	0.505	0.260	0.343
	SVM	0.583	0.554	0.811	0.658
	RNN capsule	0.541	0.465	0.54	0.5
C3	Decision Tree	0.523	0.526	0.524	0.524
	Random Forest	0.578	0.589	0.576	0.579
	Naïve Bayes	0.364	0.394	0.392	0.348
	SVM	0.608	0.632	0.599	0.608
	RNN capsule	0.54	0.553	0.539	0.546
C4	Decision Tree	0.695	0.685	0.715	0.699
	Random Forest	0.769	0.756	0.789	0.772
	Naïve Bayes	0.664	0.632	0.773	0.696
	SVM	0.788	0.782	0.794	0.788
	RNN capsule	0.716	0.611	0.616	0.613
C5	Decision Tree	0.666	0.748	0.727	0.737
	Random Forest	0.711	0.765	0.799	0.782
	Naïve Bayes	0.447	0.705	0.249	0.368
	SVM	0.744	0.769	0.863	0.813
	RNN capsule	0.759	0.765	0.748	0.756
C6	Decision Tree	0.589	0.595	0.606	0.6
	Random Forest	0.629	0.627	0.672	0.649
	Naïve Bayes	0.604	0.592	0.718	0.649
	SVM	0.667	0.642	0.782	0.705
	RNN capsule	0.651	0.636	0.622	0.629

Table 5 Performance of models on various sampling methods

Model	Sampling method	Accuracy	Precision	Recall	F1
Decision Tree	Undersampling	0.186	0.175	0.189	0.182
	Oversampling	0.793	0.763	0.792	0.777
	Proposed ensemble method	0.611	0.639	0.628	0.631
Random Forest	Undersampling	0.177	0.17	0.177	0.173
	Oversampling	0.803	0.777	0.802	0.789
	Proposed ensemble method	0.641	0.65	0.735	0.686
Naïve Bayes	Undersampling	0.104	0.094	0.101	0.097
	Oversampling	0.306	0.328	0.305	0.316
	Proposed ensemble method	0.528	0.58	0.502	0.506
SVM	Undersampling	0.197	0.205	0.197	0.201
	Oversampling	0.444	0.414	0.443	0.428
	Proposed ensemble method	0.653	0.656	0.792	0.71
RNN Capsule	Undersampling	0.148	0.131	0.177	0.151
	Oversampling	0.72	0.72	0.76	0.739
	Proposed ensemble method	0.64	0.593	0.609	0.6

scores were significantly improved, with mean values of 0.64, 0.593, 0.609, and 0.6, respectively, representing approximately a 215% improvement in performance.

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2015
2. Sukarna Barua, Md. Monirul Islam, Xin Yao, and Kazuyuki Murase. Mwmote–majority weighted minority oversampling technique for imbalanced data set learning. *IEEE Transactions on Knowledge and Data Engineering*, 26(2):405–425, 2014
3. N. Chakrabarty, S. Biswas, Navo minority over-sampling technique (nmote): a consistent performance booster on imbalanced datasets. J. Electronics Informatics **2**, 96–136 (2020)
4. K. Cho, B.V. Merrienboer, Çaglar Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder-decoder for statistical machine translation. ArXiv, abs/1406.1078 (2014)
5. D. Díaz-Vico, A.R. Figueiras-Vidal, J.R. Dorronsoro, Deep mlps for imbalanced classification, in *2018 International Joint Conference on Neural Networks (IJCNN)* (2018), pp. 1–7

6. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 248–255
7. B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, S. Lehmann, Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm, in *EMNLP* (2017)
8. A. Fernández, V. López, M. Galar, M. José Del Jesus, F. Herrera, Analysing the classification of imbalanced data-sets with multiple classes: binarization techniques and ad-hoc approaches. Knowl.-Based Syst. **42**, 97–110 (2013)
9. M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, F. Herrera, A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. IEEE Trans. Syst. Man Cybernetics Part C (Appl. Rev.) **42**(4), 463–484 (2012)
10. S. Hochreiter, J. Schmidhuber, Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
11. B. Krawczyk, Learning from imbalanced data: open challenges and future directions. Prog. Artif. Intelligence **5**, 221–232 (2016)
12. Z. Ali Sayyed, Study of sampling methods in sentiment analysis of imbalanced data, in *CoRR*, abs/2106.06673 (2021)
13. M. Smith, T. Martinez, C. Giraud-Carrier, An instance level analysis of data complexity. Mach. Learn. **95**, 225–256 (2013)
14. P. Sobhani, H. Viktor, S. Matwin, Learning from imbalanced data using ensemble methods and cluster-based undersampling, in *Proceedings of the 3rd International Conference on New Frontiers in Mining Complex Patterns*, NFMCP'14, Gewerbestrasse 11 CH-6330 (Springer, Cham (ZG), CHE, 2014), pp. 69–83
15. Shuo Wang, Leandro L. Minku, Xin Yao, Resampling-based ensemble methods for online class imbalance learning. IEEE Trans. Knowl. Data Eng. **27**(5), 1356–1368 (2015)
16. Y. Wang, M. Huang, X. Zhu, L. Zhao, Attention-based LSTM for aspect-level sentiment classification, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (Austin, Texas, 2016), pp. 606–615. Association for Computational Linguistics
17. Y. Wang, A. Sun, J. Han, Y. Liu, X. Zhu, Sentiment analysis by capsules, in *Proceedings of the 2018 World Wide Web Conference*, WWW '18 (Republic and Canton of Geneva, CHE, 2018), pp. 1165–1174. International World Wide Web Conferences Steering Committee
18. Dandan Yan, Youlong Yang, Li. Benchong, An improved fuzzy classifier for imbalanced data. J. Intelligent Fuzzy Syst. **32**, 2315–2325 (2017)
19. Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, Hierarchical attention networks for document classification, in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (San Diego, California, June 2016), pp. 1480–1489. Association for Computational Linguistics
20. Z. Zheng, Y. Cai, Y. Li, Oversampling method for imbalanced classification. Comput. Informatics **34**, 1017–1037 (2015)

Analysis of Clustering Algorithm in VANET Through Co-Simulation

Chunduru Hemalatha and T. V. Sarath

Abstract Vehicular communications play a vital role in ITS applications. Various architectures, standards, and protocols are introduced and researched upon for emerging smart cities' vehicular networks. Clustering in VANET is introduced as a means to create smaller manageable structures in an otherwise unstable network. Various clustering algorithms are classified according to underlying application for clustering. Co-simulation of selected clustering algorithm in NetSim, MATLAB, and SUMO is analyzed in this paper.

Keywords VANET · ITS · Clustering · Vehicular communication

1 Introduction

Majority of accidents are caused due to increasing population, heavy traffic congestion, driver carelessness, lack of information about street, and violation of traffic rules [1]. Accident detection and intelligent parking system based on IoT have eased difficulties faced by drivers [2, 3]. An intelligent transport facility based on vehicular ad hoc networks (VANETs) can help reduce avoidable consequences. VANETs involve spontaneous creation of wireless network of vehicular nodes. VANET comprises vehicle-to-vehicle (V2V) communication and vehicle-to-infrastructure (V2I) communication for dissemination of messages.

VANETs are an exceptional case of mobile ad hoc networks (MANETs) and are characterized by uneven distribution of vehicles, high mobility of vehicles, and frequent changes in network topology. Unlike MANET, VANET is not constrained by lifetime of power resources and unexpected routes of nodes. Depending on the

C. Hemalatha (✉) · T. V. Sarath
Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

T. V. Sarath
e-mail: tv_sarath@cb.amrita.edu

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_32

number of vehicles in a network, large number of messages embedded with information regarding vehicle's identity and application specific data is disseminated. To handle the packet flow, communication devices like onboard unit and road side units which are stationary on road networks are placed. Messages disseminated from the source reach the destination following routing algorithms in VANETs. Novel techniques for routing algorithms in VANET using buses improve performance of high dynamic nature of a typical network [25].

Intelligent transport system (ITS) is an integral part of smart cities that provides information about the traffic, seat availability in public transport, etc. ITS services are built upon advanced communication technologies, offer applications to end users such as lane change warnings, collision ahead warning, and speed alerts. The network of vehicles created by VANET acts as a basic infrastructure for vehicles to exchange information about their surrounding and thereby reducing risk of accidental crashes.

While vehicular communications ease the exchange of information, heavy traffic on roads can burden these communication devices, thereby delaying the exchange of information. In addition, a huge network of vehicles tends to be less stable. To overcome this, an ad hoc network is partitioned into smaller groups. This process is known as clustering where each partitioned cluster functions as a whole [24]. These clusters are formed according to rules different from each other. Each cluster varies in size and is comprised of ordinary nodes and a cluster head (CH). Nodes such as cluster relays are also defined by some algorithms. CH is elected from ordinary nodes based on a set of factors such as additional network connectivity and number of neighboring nodes. Figure 1 shows the clusters with vehicular communication links and CHs. A message from ordinary node to another cluster's node is exchanged through CH of respective clusters. CHs communicate messages between clusters and road side units thereby reducing network traffic encountered in huge vehicular networks.

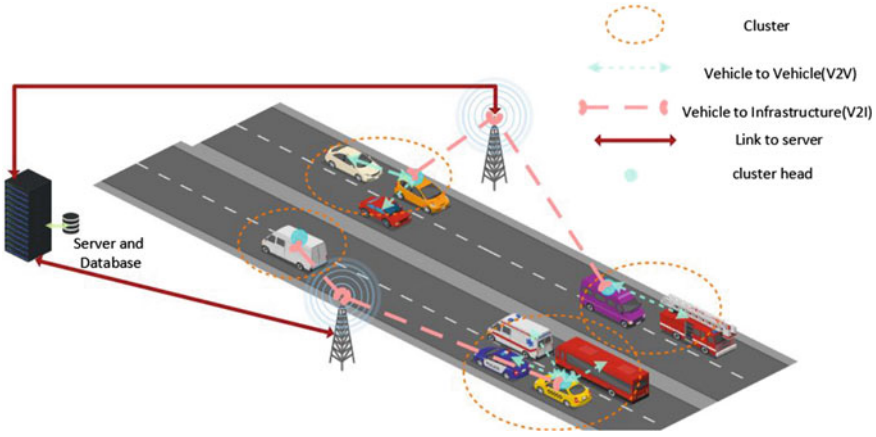


Fig. 1 Overview of vehicular network

Clustering algorithms are specifically created and modified to cater the characteristics and challenges of vehicular environments. Wide range of clustering algorithms for robust and scalable networks is categorized based on factors such as security, hop distance for messages, stability, and routing. Performance of clustering algorithms is measured with specific metrics like cluster formation time, status changes of node, and cluster lifetime. Our paper investigates the co-simulation of a clustering algorithm using MATLAB, NetSim, and SUMO, followed by the analysis on the performance metrics.

2 Related Work

The clustering technique in vehicular networks has been studied in recent years. A good number of clustering algorithms are proposed which use different parameters from the nodes to form a cluster. Clustering techniques designed for wireless networks do not cater the needs of vehicular networks. Hence, specific clustering schemes for VANETs are designed [4].

An extensive survey on clustering algorithms for VANET has given out a good number of clustering algorithms [6]. Aggregate local mobility (ALM) clustering algorithm is a beacon-based algorithm which promises to prolong cluster lifetimes in VANET [7]. Overeager reorganization is prevented using a contention-based scheme when two CHs "get in each other's range for short period of time" (Souza 2010, 1). GPS coordinates are taken for calculating ALM.

Connectivity- and mobility-aware dynamic clustering algorithm based on connectivity among neighboring vehicles and relative speed is proposed [8]. Dynamic and stable clusters are formed with multiagent system comprising of mobile and static agents with light and heavy weights, respectively.

In [9], VMasSC is introduced where CH is selected using mobility metric determined by relative speed, cluster connection with neighbors, and dissemination of cluster member details with hello packets. Main advantage of this clustering technique is the presence of minimum number of CHs.

Reduction in global overhead and prolonging the cluster lifetime is presented in AMACAD algorithm with the help of network's mobility pattern [10]. This technique mainly focuses on destination of vehicle. Size of clusters is varied with respect to traffic density, vehicle speed, and QoS.

An efficient clustering algorithm which elects the CH based on four parameters, i.e., neighborhood list of vehicles, usage of distrust value for determination of priority, entropy of vehicles, and vehicle direction is presented in [11]. In [12], density-based clustering (DBC) algorithm which is a multilevel technique has been presented. Formation of cluster is based on density connection graph, traffic conditions, and link quality which are complex clustering metrics.

In [13], direction-based clustering technique has been designed for the data communication. Communication is done only with the vehicles traveling in the same direction. GPS provides location of vehicles, and traveling directions are determined

by using digital maps. In [14], a modified distributed mobility-adaptive clustering (DMAC) has been adopted. Weight of the nodes depends on any set of node parameters including but not limited to node id, remaining power, etc. A CH is the node having biggest weight. This algorithm is suitable for mobile environment as it can adapt to the changes in the network topology caused by node mobility.

The selection of a clustering algorithm for an application needs to be carried out after analysis of the algorithm. The clustering algorithms are evaluated based on performance metrics. Depending on the nature of clustering algorithm, various performance metrics such as cluster formation time, cluster lifetime, membership duration, reaffiliation rate, status changes, cluster head selection time, and packet delivery ratio are calculated and analyzed [6–8].

Considering the message transfer in vehicular networks, a dedicated frequency range of 5.850–5.925 GHz for V2V and V2I communication is assigned by Federal Communication Commission (FCC) [20]. A dedicated short-range communication (DSRC) is developed. Wireless Access for Vehicular Environment (WAVE) has been developed by IEEE research group for VANETs. Two wireless access standards are dealt in VANET namely IEEE 802.11p for physical and MAC layer and IEEE 1609 for higher layer protocols [15].

Analysis of clustering algorithm along with calculation of performance metrics in real-world implementation is a tedious task. Instead, simulation tools are selected for implementation and testing [23]. Process of analysis of a clustering algorithm involves vehicular network simulation and road network simulation. There are wide variety of simulation tools available for vehicular and road network simulation. SIDE/SMURPH, ns-2, and ns-3 are few open-source network simulators. NetSim, OPNET, and GloMoSim are proprietary network simulators [21, 23]. For the road network simulations, SUMO, Quadstone parabic modeler, and Aimsun are available [22].

3 Co-Simulation Environment

System overview of the work is presented in Fig. 2. The process of co-simulation of clustering algorithm involves implementation of the clustering algorithm in computational engine, network simulation environment for vehicular networks, and road traffic simulation. Out of the clustering algorithms reviewed in above section, aggregate local mobility (ALM) clustering algorithm [7] is chosen. For the selected clustering algorithm, NetSim—a packet-level network simulator—simulates the network aspects of a vehicular environment. NetSim is a discrete event network simulator providing network performance metrics at each and every level [21]. MATLAB computes ALM of cluster heads and characteristics of vehicles, and road networks are configured in SUMO [17, 19].

A general concept of aggregate mobility is applied by ALM clustering algorithm for improvement of stability of clusters with a subtle set of rules. A member node (MN) converts to unidentified node (UN) instead of CH when beacons are not

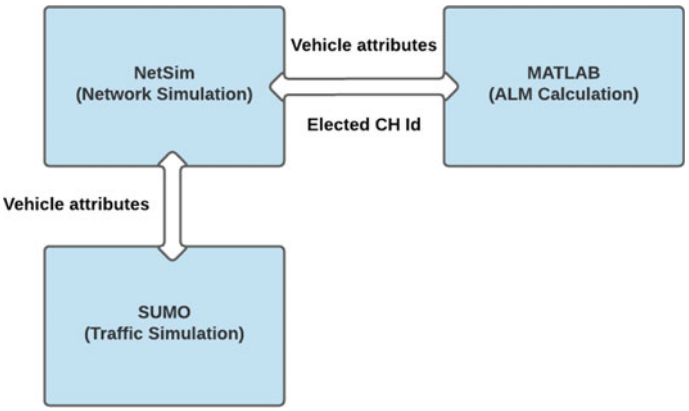


Fig. 2 System overview

received from its respective CH. Creation of new CH is postponed by this process, giving time to the MN to find another CH. Over eager reorganization of clusters is delayed when CHs come in range of each other. Instead, more than one packet is exchanged between CHs within a certain amount of time. Status change of a node is based on its perception of ALM. In this algorithm, "ratio between successive takes of distance between a node and its neighbour is the relative mobility" (Souza 2010, 2). A node's ALM is calculated as "the variance of relative mobility" (Souza 2010, 3) with all its neighbors. For lower values of ALM, CH retains its status, and the other CH changes its status to MN. A better choice for CH is "the node with less variance relative to its surroundings" (Souza 2010, 3). Table 1 lists out the parameters considered for vehicular network simulation.

To implement the above-mentioned network attributes of selected clustering algorithm, a VANET scenario in NetSim is created. Using SUMO configuration file, a scenario is opened in NetSim with predefined number of vehicles connected to an ad hoc network as shown in Fig. 3. Wireless access in vehicular environments

Table 1 Parameters for simulation

Parameters	Values
Number of vehicles	30, 45, 60, 75, 90, 105
Network simulator	NetSim
Mobility model	SUMO
Vehicle speed	50 kmph
Total length of road	7.95 km
Vehicle length	5 m
Simulation time	400 s
Vehicle arrival rate	1 vehicle/s
Network layer routing protocol	DSR

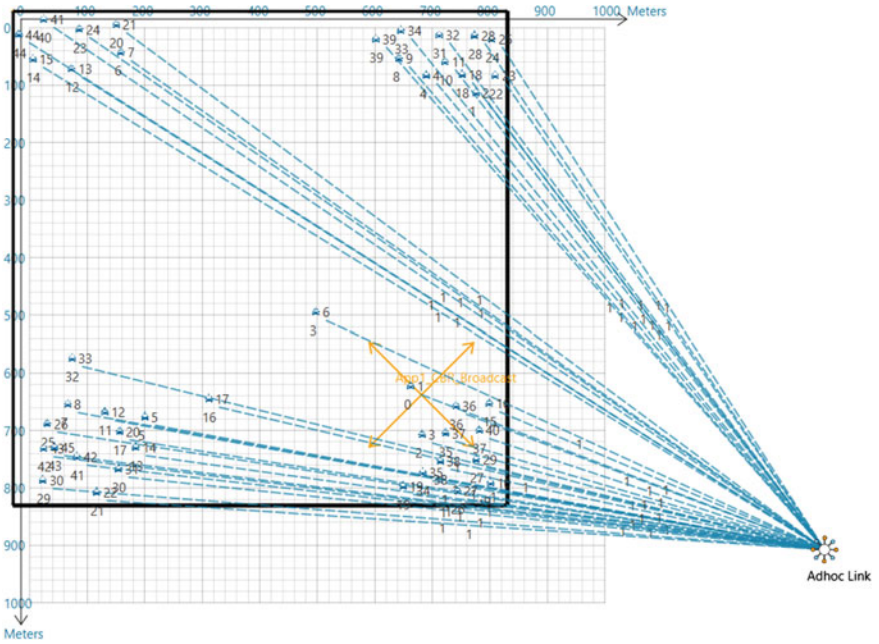


Fig. 3 NetSim scenario with 45 vehicles

(WAVE) uses IEEE 802.11p and IEEE 1609 which covers resource manager, network, and transport services, and multichannel coordination aids the packet dissemination. Among the routing protocols available for VANETs, DSR is chosen due to its better throughput, reduced packet loss, and lower delay for a huge network of vehicles [18] for our implementation. A HELLO beacon is broadcast every second to keep track of neighbors and their status. A HELLO packet comprises of vehicle's unique characteristics such as node ID, cluster ID, GPS coordinates, and node status. Based on the status of a node, a timeout event keeps track of messages received. Accordingly, the node changes its state as specified by ALM clustering algorithm.

To avoid the reorganization of cluster when two CHs exchange HELLO messages, ALM is calculated in MATLAB. Data of vehicular nodes associated with calculation of ALM is sent to MATLAB from NetSim. CH with less ALM retains its status while the other changes to a MN. This decision is passed onto NetSim. Movement and characteristics of vehicles and topology of roads are simulated in SUMO. A square network of roads is laid with each edge of 1 km length. Vehicles are characterized by vehicle length as 5 m, deceleration rate as 4.5 m/s^2, acceleration rate of 0.8 m/s^2, and maximum speed of 36 m/s. A number of vehicles in the simulation range from 30 to 105 vehicles. Input rates of vehicles are set at 1 car/s. Simulation is carried out for 400 s. Performance of the algorithm is analyzed with average CH duration, status changes, and CH changes.

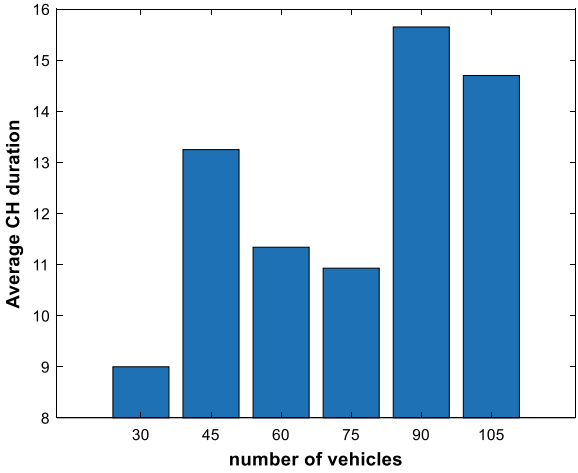


Fig. 4 Average CH duration

4 Results

The stability of clusters formed using the clustering algorithms is tested for stability of communication paths. The following performance metrics are calculated.

4.1 Average CH Duration

It is the node's duration during its role as CH. CH duration also represents cluster lifetime. Average of all the nodes' CH duration in given scenario gives average CH duration. Figure 4 shows the average CH duration for scenarios with varying number of vehicles. A stable network with lesser reorganization of clusters has higher average CH duration. From Fig. 4, we can infer that higher clustering lifetimes indicate a stable network when the number of vehicles increase.

4.2 Status Changes

It is a measure of the number of times a node changes its status during its lifetime. The average status changes in a scenario is shown in Fig. 5. Number of status changes must be lower where the clustering algorithm avoids unwanted changes. Due to lower density of vehicles in scenario with 30 and 45 vehicles, a higher status change value is recorded. This algorithm performs better in scenarios where vehicle density is higher as seen in Fig. 5.

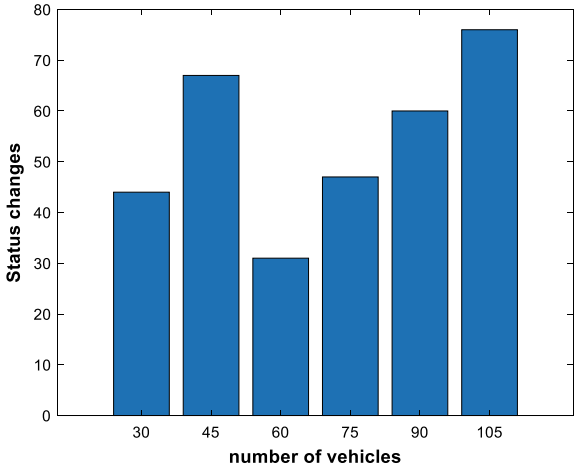


Fig. 5 Status changes

4.3 CH Changes

This metric denotes the number of times a node acting as CH changes to another status. A very small fraction of status changes as CH changes. Figure 6 shows the values of CH changes in respective scenarios. Insignificant difference of the values of CH changes when number of vehicles is varied and can be observed even when the density of vehicles is varied. The main characteristic of ALM clustering algorithm is to reduce the amount of reorganization of clusters. A CH change indicates reorganization of cluster. Thus, lesser CH changes indicate reduction in rearrangement of clusters thereby maintaining a stable network.

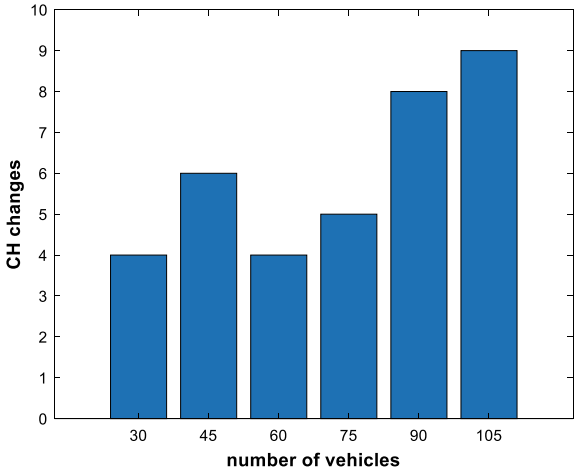


Fig. 6 CH changes

5 Conclusion

A realistic VANET simulation environment should comprise of vehicular network and traffic simulation along with a computational framework for processing the information from subsystems. Even though the use of different clustering algorithms is widely accepted for categorizing VANET, these algorithms do not possess all these three units together. We propose a co-simulation framework which includes these components using an ALM clustering algorithm. We have simulated vehicular network in NetSim and traffic simulation in SUMO. The data obtained from NetSim and SUMO is concurrently used by MATLAB for evaluating CH using ALM algorithm. We have assessed the efficacy of the proposed system with the help of performance matrices.

References

1. World Health Organization, https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
2. M. Karthikeyan, V. Manesh, L.S. Krishna, B. Vijay, R. Vishwabharan, E. Prabhu, IoT based accident detection and response time optimization, in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (2021), pp. 358–363
3. D. Annirudh, D.A. Kumar, A.T.S.R. Kumar, K.R.M.V. Chandrakala, IoT based intelligent parking management system, in *2021 IEEE Second International Conference on Control, Measurement and Instrumentation (CMI)* (2021), pp. 67–71
4. F. Li, Y. Wang, Routing in vehicular ad hoc networks: a survey. IEEE Veh. Technol. Mag. **2**(2), 12–22 (2007)
5. C. Cooper, D. Franklin, M. Ros, F. Safaei, M. Abolhasan, A comparative survey of VANET clustering techniques. IEEE Commun. Surv. Tutorials **19**(1), 657–681 (Firstquarter 2017)
6. R.S. Hande, A. Muddana, Comprehensive survey on clustering-based efficient data dissemination algorithms for VANET, in *International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)* (2016), pp. 629–632
7. E. Souza, I. Nikolaidis, P. Gburzynski, A new aggregatelocal mobility clustering algorithm for vanets, in *2010 IEEE İnternational Conference on Communications (ICC)* (2016), pp. 1–5
8. M.S. Kakkasageri, S.S. Manvi, Connectivity and mobility aware dynamic clustering in VANET, in *IJFCC* (2014)
9. S. Ucar, S.C. Ergen, O. Ozkasap, VMASC: Vehicularmulti-hop algorithm for stable clustering in vehicular ad hoc networks, in *2013 IEEE wireless communications and networking conference (WCNC)* (2013), pp. 2381–2386
10. M.M.C. Morales, C.S. Hong, Y.C. Bang, An adaptable mobility-aware clustering algorithm in vehicular networks, in *2011 13th Asia-Pacific Network Operations and Management Symposium (APNOMS)* (2011), pp. 1–6
11. A. Daeinabi, A. Ghaffar, P. Rahbar, A. Khademzadeh, VWCDPWA: an efficient clustering algorithm in vehicular ad hoc networks. J. Netw. Comput. Appl. **34**(1), 207–222 (2011)
12. S. Kuklinski, G. Wolny, Density based clustering algorithmfor vanets, in *5th İnternational Conference on Testbeds and Research İnfrastructures for the Development of Networks Communities and Workshops. TridentCom 2009* (2009), pp. 1–6
13. N. Maslekar, M. Boussedjra, J. Mouzna, L. Houda, Directionbased clustering algorithm for data dissemination in vehicular networks, in *Vehicular Networking Conference (VNC).* (IEEE, 2009), pp. 1–6

14. S. Basagni, Distributed clustering for ad hoc networks,in *Fourth İnternational Symposium on Parallel Architectures, Algorithms, and Networks, 1999. (I-SPAN '99) Proceedings* (1999), pp. 310–315
15. S. Manu, P. Sivraj, Performance comparison of communication technologies for V2X applications, in *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (2020), pp. 356–362
16. N. Maslekar, M. Boussedjra, J. Mouzna, H. Labiod,A stable clustering algorithm for efficiency applications in vanets, in *2011 7th International Wireless Communications and Mobile Computing Conference (IWCMC)* (2011), pp. 1188–1193
17. D. Krajzewicz, J. Erdmann, M. Behrisch, L. Bieker, Recentdevelopment and applications of sumo-simulation of urban mobility. Int. J. Adv. Syst. Meas. **5**(3&4) (2012)
18. S. Malik, P.K. Sahu, A comparative study on routing protocols for VANETs. Heliyon **5**(8) (2019)
19. TETCOS, https://www.tetcos.com/downloads/v12/NetSim_User_Manual.pdf
20. V. Vibin, P. Sivraj, V. Vanitha, Implementation of in-vehicle and V2V communication with basic safety message format, in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)* (2018), pp. 637–642
21. Md. Hossain, S. Hossain, S. Islam, Md.J. Hossain, Detail comparison of network simulators. Int. J. Sci. Eng. Res. 203–218 (2014)
22. G. Kotusevski, K. Hawick, A review of traffic simulation software. Technical report (2009)
23. S.A.B. Mussa, M. Manaf, K.Z. Ghafoor, Z. Doukha, Simulation tools for vehicular ad hoc networks: a comparison study and future perspectives, in *2015 International Conference on Wireless Networks and Mobile Communications (WINCOM)* (2015), pp. 1–8
24. S. Vodopivec, J. Bester, A. Kos, A survey on clustering algorithms for vehicular ad-hoc networks, in *2012 35th International Conference on Telecommunications and Signal Processing (TSP)* (2012), pp. 52–56
25. R. Dhaya, R. Kanthavel, Bus-based VANET using ACO multipath routing algorithm. J. Trends Comput. Sci. Smart Technol. (TCSST) 40–48 (2021)

Ensemble-Based Weighted Voting Approach for the Early Diagnosis of Diabetes Mellitus

S. R. Sannasi Chakravarthy and Harikumar Rajaguru

Abstract Diabetes mellitus, shortly diabetes, is a fearsome disorder that can be characterized by elevated blood glucose levels. The appropriate use of machine learning (ML) techniques aid in the earlier diagnosis of diabetes. The main goal of this research is to use an ensemble of ML algorithms for the better prediction of diabetes mellitus. For this, the work utilizes the Pima Indians Diabetes (PID) database. The ensemble-based approach of weighted voting classifier employs an ensemble of three ML algorithms for providing binary classification that includes logistic regression, random forest, and extreme gradient boosting classifiers. Here, the performance of the above three ML algorithms are individually assessed, and then the weighted voting-based ensembled approach is performed by considering standard benchmark metrics such as accuracy, precision, and F1 score. And finally, the above-said performance is validated using Matthews correlation coefficient. In this way, the proposed ensemble-based weighted voting approach of diabetic classification provides a supreme performance of 92.21% classification accuracy over other individual ML algorithms used.

Keywords Diabetes · Glucose · Ensemble classifier · Insulin · Voting · Machine learning

1 Introduction

Health professionals and doctors usually refer to diabetes as diabetes mellitus. It is a condition where the body fails in producing blood glucose, also referred to blood sugar [1]. Diabetes affects a large number of individuals globally and is categorized as Type 1 or Type 2 diabetes [2]. In this, the first one, also known as an insulin-dependent diabetic disorder, is most commonly diagnosed in children. In Type 1, the body attacks the pancreas with antibodies, causing it to damage internal

S. R. Sannasi Chakravarthy (✉) · H. Rajaguru
Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam 638401, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_33

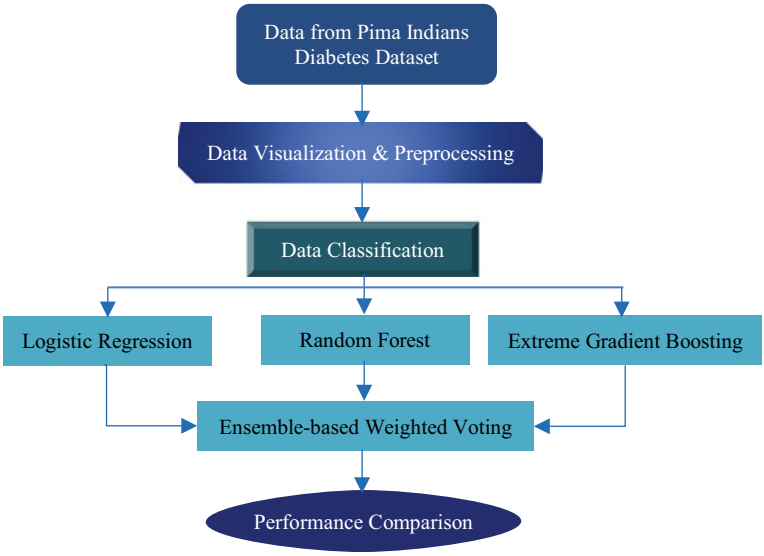


Fig. 1 Proposed workflow

body components and stop producing insulin. The next type of diabetes is sometimes termed as non-insulin-dependent or adult-onset diabetic disorder. Although it is more forgiving than type 1, it is nevertheless extremely damaging and can lead to serious complications, particularly inside. Diabetic adult patients are merely quadrupled since from the year 1980, according to the World Health Organization (WHO), growing from 4.7 to 8.5% in 2014. Diabetes claimed the lives of 1.5 million individuals in 2012 [3]. According to a WHO report, 2–17.8% of pregnant women develop gestational diabetic disorder [3]. Diabetes mellitus is one among the most important study topics in medical science because of the disease's profound societal impact, which unavoidably generates massive quantities of data. The flow of the work proposed is given in Fig. 1. As depicted in Fig. 1, the input dataset is visualized and preprocessed at the initial stage, and then the classification is performed using the four classification strategies namely logistic regression, random forest, XGBoost, and with weighted voting classifier. Finally, the results are compared for the analysis of diabetic classification.

2 Materials and Methods

The goal of this study is to improve the accuracy and outcomes of diabetes detection. The authors presented an ensemble of ML algorithms for the binary categorization of illness into positive and negative using a weighted voting methodology. Before

providing input to the model, data preprocessing is performed, followed by data cleaning as shown in Fig. 1.

2.1 Input Dataset

The Pima Indian Dataset was used for experimentation in this study. The dataset includes nine column attributes and one target output with a binary number indicating whether the victim suffers from diabetes or not [4]. The dataset includes 768 samples, 500 of which are non-diabetes and 268 of which are diabetes class output. The dataset contains the attributes, including glucose, plasma, pregnant month, pedigree function, fold thickness of triceps skin, blood pressure, BMI, insulin amount, and patient age, as well as one target class as 0 or 1 [4].

2.2 Data Visualization of Pima Indian Dataset

In every ML problem, the input data analysis is very significant for further phases of research. The univariate analysis of the input dataset is done through a distribution plot and is shown in Fig. 2.

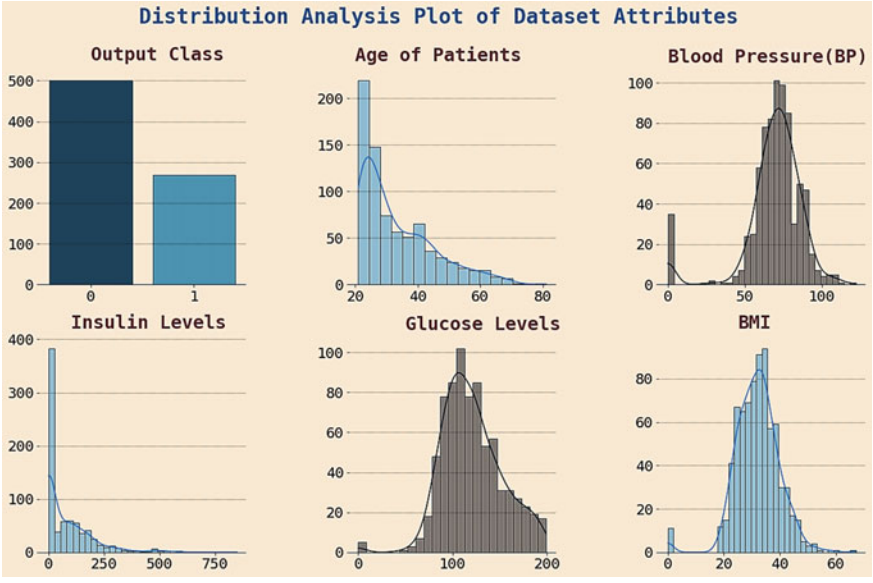


Fig. 2 Univariate distribution analysis plot of ınput dataset

As from Fig. 2, the insulin and age attributes of the input dataset are found to be much right-skewed. Also, noted that the database comprises more people between the age group of 20–40, and more people have BP in the range of 50–100 mmHg. Moreover, a larger amount of people are having insulin values of zero, in particular, these zero insulin values are found more for Type 1 diabetic people. The people with glucose levels in the range of 100–200 mg/dL are found more in the dataset. Finally, the important thing noted is that the measurement of several people having the BMI range of 20–50. From the plot of Fig. 2, it is revealed that the normalization of data is important before proceeding with classification.

2.3 Data Preprocessing

In the PID dataset, it is found that few zero values in some attributes such as Glucose, BMI, and insulin level attributes. The very important thing is that all the measurements or attributes of the input dataset considered are never being zero [5]; for example, the glucose of the living human body can never be zero at any instant. So, all these values are now imputed based on their median values. And the dataset is then normalized using the StandardScaler technique [6]. Thus, the input dataset is now completely ready for further classification as shown in Fig. 1.

3 Classification Algorithms

We employed an ensemble of ML algorithms in our proposed work that includes logistic regression (LR), random forest (RF), and XGBoost (XGB) classifiers. To improve the performance, the aforementioned algorithms were combined with a weighted soft voting approach. This section goes through these algorithms in detail.

3.1 Logistic Regression (LR) Classifier

In logistic regression (LR), statistical techniques are used to predict binary outcomes (output class = 0 or 1). A linear learning algorithm, LR model, where the odds of an event occurring are used to make logistic regression predictions. The sigmoid function is used by the LR method to map each data point in the input dataset [7]. An S-shaped curve is produced using the conventional logistic function that can be mathematically depicted as a sigmoid function as given in Eq. (1) [8],

$$\text{Sigmoid Function} = \frac{1}{1 + e^{-x}} \tag{1}$$

3.2 Random Forest (RF) Classifier

The random forest (RF) model, a sort of ensembled algorithm model which employs the decision trees (DT) model for improving prediction accuracy. It creates a large number of trees and applies the bootstrap approach to each one based on the training inputs [9]. In the prediction phase, process input is supplied to all trees in the forest, and the individual tree then votes for that target output separately. At a final point, the algorithm chooses the target output with the most voted class output. The algorithm of the RF model can be summarized as [10]:

Step 1: Begin by selecting random samples from the input dataset.
Step 2: For each sample, this algorithm will create a decision tree. The forecast result from each decision tree will then be obtained.
Step 3: Voting will be done for each anticipated outcome in this step.
Step 4: Finally, choose the prediction result with the most votes as the final forecast result.

3.3 Extreme Gradient Boosting (XGB) Classifier

Gradient boosting is a type of ensemble ML technique that may be used to solve classification or regression predictive modeling issues. Boosting refers to an ensemble approach that involves adding new models to old models to repair mistakes. Models are added in a logical order until there are no more improvements to be made [11]. XGB is a distributed gradient boosting toolkit that has been tuned for efficiency, flexibility, and portability. It uses the gradient boosting framework to construct ML algorithms. It uses parallel tree boosting for handling a wide range of data science issues quickly and accurately [11]. Because of its simultaneous and distributed processing, XGBoost is a quicker algorithm than other algorithms. XGBoost was created with careful consideration of both system optimization and machine learning techniques.

3.4 Weighted Voting Based Ensemble Classifier

This methodology includes a meta-classifier approach that uses majority vote to combine conceptually or similar different ML algorithms for classification. A voting algorithm employs both soft and hard voting algorithms. The conclusive prediction in hard voting is made via a majority voting approach where the aggregator chooses the output target which appears repeatedly among the base models [12]. The Predict proba technique [13] should be used in soft voting base models. Because it integrates the predictions of various models, the voting classification method produces better overall performance than conventional base algorithms. The employed model

combines the logistic regression (LR), random forest (RF), and the XGBoost (XGB) classifiers.

The predict proba attribute column, which specifies the probability of each target variable, was used to create a soft voting classifier. The training data and data points are then shuffled before being sent to LR, RF, and the XGB models. Each model generates an individual forecast using a vote aggregator and a soft voting approach, then computes the final prediction using majority voting.

4 Results and Discussion

The work carried out in this paper as depicted in Fig. 1 has been implemented using Google Colab online research IDE tool associated with one personal Gmail account on the Google Chrome Web browser. The processed input dataset as discussed in Sect. 2 has been split accordingly for the classification stage as the 80:20 split ratio with training samples is 80% and testing samples are 20%. As shown in the first column plot of Fig. 2, the dataset contains more 0 class labels than other classes, so there may be a class imbalance problem and so SMOTE methodology [14] to overcome this problem. The data split of inputs is given in Fig. 3. Also, before proceeding to the classification part, the processed input data is normalized using the below formula of min–max standardization methodology [6] as shown in Eq. 2,

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{2}$$

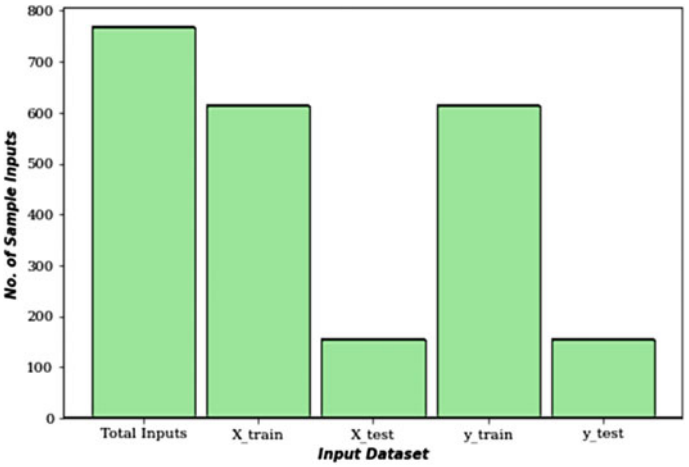


Fig. 3 Input data splitting for classification

where x is the input data point and $x_{\min}$, $x_{\max}$ refer to the minimum and maximum values of the data points, respectively.

After splitting the inputs into training and testing, the classifiers employed are fitted and tested for their effectiveness in diabetic prediction. That is, the LR, RF, and XGB algorithms along with their ensembled soft voting models are utilized for this classification. Also, all the results are fivefold cross-validated for finding the best performance. These performances in diabetic classification can be measured using the standard metrics [15] that are always a benchmark one in binary classification problems. The measures used in the work are sensitivity (Sen), specificity (Spec), accuracy (Acc), precision (Pr), and F1 score. All the metrics above discussed are taken or extracted from the confusion matrix having the elements of true and false positives and negatives. The validation of the results obtained is further done by using a metric, Matthews correlation coefficient (MCC).

The confusion matrix elements obtained for each classifier are plotted graphically in Fig. 4. In this plot, the number of true positive and negative elements is more for the random forest classifier as compared with the logistic regression algorithm. And it is more for the XGBoost classifier than the random forest algorithm. As applied with the ensemble-based weighted voting model including the above-said three algorithms, the true classification elements of the confusion matrix get increased significantly as shown in Fig. 4. Similarly, when false classifications are taken into account, the random forest classifier has lower FP and FN as compared with the logistic regression model. However, the misclassification rate of the XGBoost classifier is very lesser than the FP and FN elements of the random forest algorithm. Moreover, the performance, i.e., the number of true classification elements is increased and false classification elements get decreased while using the ensemble-based weighted voting classification approach as shown in Fig. 4. Let's discuss them in detail further.

Table 1 gives the percentage value of the performance metrics obtained for the employed four different classification models. Here, the performance of the random

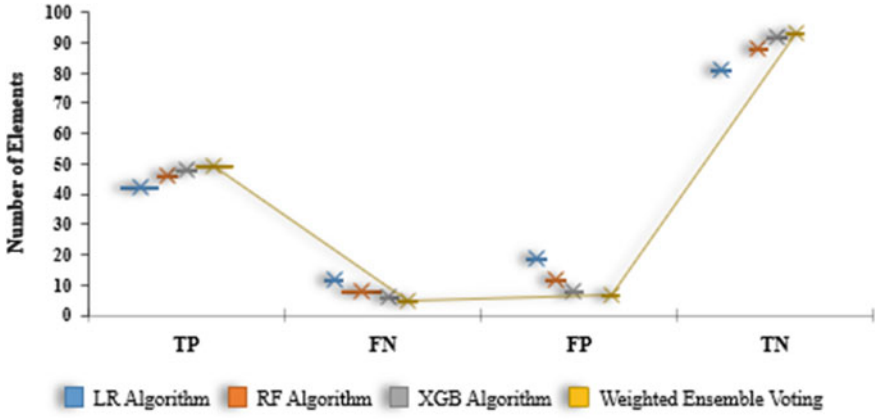


Fig. 4 Confusion matrix plot of classifiers

Table 1 Comparison of classifier performance of diabetic classification

Algorithms	Performance analysis (%)					
	Sen	Spe	Acc	Pre	F1 Score	MCC
LR algorithm	77.78	81	79.87	68.85	73.04	57.35
RF algorithm	85.19	88	87.01	79.31	82.14	72.07
XGB algorithm	88.89	92	90.91	85.71	87.27	80.24
Weighted ensemble voting algorithm	90.74	93	92.21	87.5	89.09	83.07

forest classifier is higher than the logistic regression model. But the performance of the XGBoost classifier is higher than LR and RF classifiers. That is, the XGB algorithm provides higher accuracy of 90.91%, precision of 85.71%, and F1 score of 87.27%. When assessing the individual performance of the above classifiers, the XGB classifier provides the best performance so that it provides the MCC of 80.24 which is higher than other individual algorithms. To increase further, the ensemble-based weighted voting classifier is employed. From the above discussion, it is clear that the weight of the XGB classifier is needed to be made higher than the other two algorithms. Thus, the paper adopts the unit weightage performance of the LR, RF classifiers together with the double weightage performance of the XGBoost classifiers for getting better results as discussed in Sect. 3.

The performance of this weighted voting-based classifier as compared with others is plotted in Fig. 5. From this plot, the ensemble-based weighted voting classifier provides an accuracy of 92.21% with 87.5% of precision, and 89.09% of F1 Score. This performance can be validated by the MCC value of 83.07 which is the supreme

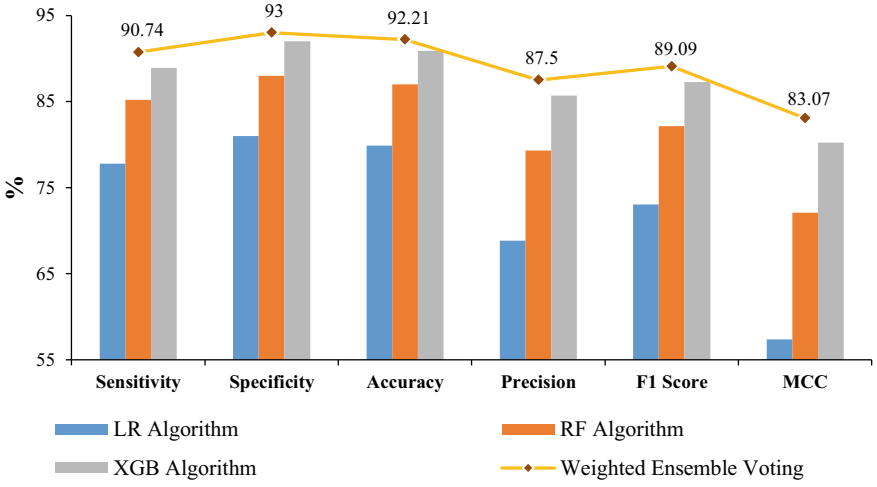


Fig. 5 Graphical plot of comparison analysis of classifiers

result over other classifiers used. Thus, as from Table 1 and Fig. 5, the ensemble-based weighted voting approach of classification provides a supreme performance than the LR, RF, and XGB classifiers.

5 Conclusion and Future Scope

Diabetes mellitus is a form of disorder which results in abundant sugar in the blood, leads to several disorders for human. The work in this paper intends to solve the binary classification in the problem of diagnosis of diabetes mellitus. In particular, the study focuses on the region of India since the dataset considered for the work is the Pima Indian Diabetes dataset taken from the UCI repository. In this dataset, several measurements were acquired from the different age groups of diabetic patients in India. The dataset is analyzed visually through the distribution analysis plot and came to know that the data needs to be normalized before classification. And the zero and missing values have been replaced with the imputed median values. Then, the inputs are split based on the standard train: test ratio through SMOTE approach. Now, three different algorithms namely logistic regression, random forest, and XGBoost algorithms are used for classification. For getting better results, an ensemble-based weighted voting approach of classification is finally implemented. The weights for this approach have been set based on the individual assessment of the above-said three classifiers. Finally, the XGB classifier provides better accuracy of 90.91% and as an ensemble weighted voting approach, provides a supreme accuracy of 92.21% with the improved MCC value of 83.07. The future work deals with the implementation of the research with different splitting ratios, different classification algorithms and through different preprocessing techniques, and with some other medical datasets.

References

1. W. Guo, M. Li, Y. Dong, H. Zhou, Z. Zhang, C. Tian, R. Qin, H. Wang, Y. Shen, K. Du, L. Zhao, Diabetes is a risk factor for the progression and prognosis of COVID-19. Diabetes/metabolism Res. Rev. **36**(7), e3319
2. C. Abirami, R. Harikumar, S.S. Chakravarthy, Performance analysis and detection of micro calcification in digital mammograms using wavelet features, in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET).* (IEEE, 2016), pp. 2327–2331
3. *Centers for Disease Control and Prevention: National diabetes statistics report, 2020.* (Centers for Disease Control and Prevention, US Department of Health and Human Services, Atlanta, GA, 2020), pp. 12–15
4. Pima Indians diabetes dataset, Available from http://archive.ics.uci.edu/ml/machine-learning-databases/pima-indians-diabetes/pima-indians-diabetesdata. Accessed: 1st of (2021, June)
5. P. Zhang, X. Zhang, J. Brown, D. Vistisen, R. Sicree, J. Shaw, G. Nichols, Global healthcare expenditure on diabetes for 2010 and 2030. Diabetes Res. Clin. Pract. **87**(3), 293–301 (2010)

6. L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, API design for machine learning software: experiences from the scikit-learn project (2013). arXiv preprint arXiv:1309.0238
7. A. De Caigny, K. Coussement, K.W. De Bock, A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. Eur. J. Oper. Res. **269**(2), 760–772 (2018)
8. S.R. Sannasi Chakravarthy, H. Rajaguru, A novel improved crow-search algorithm to classify the severity in digital mammograms. Int. J. Imaging Syst. Technol. **31**(2), 921–954 (2021)
9. H. Rajaguru, Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer. Asian Pacific J. Cancer Prev. APJCP **20**(12), 3777 (2019)
10. Z. Chen, F. Jiang, Y. Cheng, X. Gu, W. Liu, J. Peng, XGBoost classifier for DDoS attack detection and analysis in SDN-based cloud, in *2018 IEEE İnternational Conference on Big Data and Smart Computing (BigComp)* (IEEE, 2018), pp. 251–256
11. S.R. Sannasi Chakravarthy, H. Rajaguru, Lung cancer detection using probabilistic neural network with modified crow-search algorithm. Asian Pacific J. Cancer Prev. APJCP **20**(7), 2159 (2020)
12. M. Saqlain, B. Jargalsaikhan, J.Y. Lee, A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing. IEEE Trans. Semicond. Manuf. **32**(2), 171–182 (2019)
13. S. Kaur, P. Kumar, P. Kumaraguru, Automating fake news detection system using multi-level voting model. Soft. Comput. **24**(12), 9049–9069 (2020)
14. G. Douzas, F. Bacao, F. Last, Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. Inf. Sci. **465**, 1–20 (2018)
15. S.R. Sannasi Chakravarthy, H. Rajaguru, Detection and classification of microcalcification from digital mammograms with firefly algorithm, extreme learning machine and non-linear regression models: a comparison. Int. J. Imaging Syst. Technol. **30**(1), 126–146 (2020)

IoT-Based Smart Diagnosis System for HealthCare

J. Hanumanthappa, Abdullah Y. Muaad, J. V. Bibal Benifa, Channabasava Chola, Vijayalaxmi Hiremath, and M. Pramodha

Abstract COVID-19 is a pandemic situation where isolation and social distancing are enforced to surge the pandemic. Pandemic Patient Health Management Platform is presently needed to retrieve health data without visiting healthcare centres. The Pandemic Patient Health Management Platform (PPHMP) uses Internet of things (IoT) and cloud computing technology and it is a remote patient health management platform. APPHMP model is proposed, which can help patients and elderly people to receive information about their health from their premises especially in consideration of COVID-19. In the present work, an algorithm is proposed to determine the patient's current health status and send necessary information to the healthcare centre for subsequent decisions. The proposed work is implemented by utilizing a naïve Bayes machine learning algorithm for decision making, and the obtained accuracy is about 83%.

Keywords IoT · Cloud computing · PPHMS · Naïve Bayes

1 Introduction

COVID-19 is a pandemic situation across the globe, and there are so many difficulties that exist in terms of a scarcity of diagnostic tools, clinical studies on vaccination

J. Hanumanthappa · A. Y. Muaad (✉) · M. Pramodha
Department of Studies in Computer Science, University of Mysore, Mysuru, India

J. V. Bibal Benifa · C. Chola (✉)
Department of Computer Science and Engineering, Indian Institute of Information Technology, Kottayam, India
e-mail: Channabasavaphd2019@iiitkottayam.ac.in

J. V. Bibal Benifa
e-mail: benifa@iiitkottayam.ac.in

V. Hiremath
Department of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_34

and medicine availability, a scarcity of medical facilities and lockout situations. All these led to a serious problem especially with elder and COVID-infected patients. The term "Internet of Things" is an umbrella term to cover a variety of concerns linked to the physical stimulation of the Internet leveraging the Web by deploying a large number of geographically distributed devices with identifying capabilities, as well as integrated detection and actuation mechanisms [1, 2]. The Internet of things (IoT) is defined as a collection of physical items, sensors and actuators to connected to the Internet [3–6]. The IoT paradigm, that has many objects in it, will be connected to the network the domain in a single setting; IoT requires smart connectivity with existing networks as well as context-aware computation employing multiple network resources [21, 29]. Around 4 billion individuals utilize the Internet for a range of functions, including Web surfing and communication, email sending and receiving, accessing multimedia services, playing various games through social networking programmes and many more. Within a perspective limit, the term "IoT" mainly refers to the resultant interconnecting global networking of small objects by means of advanced Internet techniques and also a collection of supporting methods for identifying machine-to-machine communications among the devices. The uptake and spread of IoT and cloud computing can play a crucial role in slowing the surge in health care expenses without impacting the quality of patient care, thanks to PPHMP's reputation, efficiency and optimal results. However, by using the PPHMP, the challenge is to determine how health professionals can ensure accurate, reliable and safe follow up of the condition of their patients without a physical visit to their home. This research article supports the following contributions to further extend the research work to do qualitative advanced research.

2 Related Work

This section describes the PPHMP's related work as well as past implementation approaches. The authors in 2014 provided a theoretical model for sensor cloud designs and investigated how standard wire-free sensor network architectures might be improved [11]. Mitton et al. proposed a theoretical model for sensor cloud architectures and investigated the performance improvements that can be obtained on traditional wireless sensor network architectures [12]. The authors in 2014 presented a sensory data processing paradigm that reduces the quantity of data transferred to the cloud by processing data at the gateway layer with detectoring, riddling, prognostication, crush and recommendation and ideas [13]. The use of cloud components in a mobile healthcare system for wheel chairs is considered in [14, 29]. A sophisticated structure that spans a number of healthcare ecosystems, in which sensor data is filigreed for security and then sent to the cloud for feature extraction and classification, as detailed in [4]. The authors in [15] an architecture for data collection and safe transfer proposed. Body Cloud is a three-tier integrated software service (ttiss) that allows for the development and deployment of cloud-assisted BSN applications [16]. Zeng et al. addresses the various issues which belong to the IOT [17].

The authors in [18] introduced HealthFog, a new framework for integrating deep learning with cutting-edge computing devices and used it in a real-world application of automated heart disease analysis. Using innovative communication and model distribution techniques like enamelling, this study allows deep learning networks to be included in Edge computing paradigms, allowing for excellent accuracy with very low latencies. Jelena et al. offered a slew of policies, incentives and subsidies aimed at analysing the roadblocks to digitalizing business processes [19]. The sustainability of the FMCG supply chain is based on the implementation of IoT. Antonio et al. proposed the architecture and studied its ability to provide continuous monitoring, extended instrument integration, security and support for privacy [20]. Almaet al. suggested an ECG monitoring platform for the creation and testing of ECG monitoring applications on mobile phones, and they used the platform to record anomalous ECG signals for application development [21, 29]. Kumar et al. recognized the challenges which are faced by the healthcare system and to find good solutions in terms of popular technologies like AI and IoT [22]. Singh et al. proposed a comprehensive integrated healthcare network to address the COVID-19 pandemic by using a statistics-based method and IoT to predict an upcoming situation of this disease [23, 29]. Abawajy proposed a new system combining cloud computing and Internet of things (IoT) architecture for pervasive patient health monitoring (PPHM) [24]. The experimental results show that PPHM is a flexible, scalable and energy-efficient remote patient health monitoring system. Hariharakrishnan et al. presented a system for monitoring an athlete's thermoregulation in particular [25]. It is a new technique for identifying and training marathon athletes for race topography. In this technique, each athlete is connected with wearable sensors in their body during the training session to monitor and analyse the thermoregulation process. Wang introduced new work utilizing many techniques such as IoT, blockchain and cloud technologies for healthcare and telemedical laboratory services in the medical environment [26]. Vital signs and physiological indicators are sensed and sent in order to give meaningful, transparent and safe medical aid to patients. Suma proposed a scalable and distributed computational framework by implementing the distributed framework for medical sensors with IoT support which increases many systems; the communication dependability is inspirational and encouraging [27, 28].

3 Materials and Methods

We used a Raspberry Pi 3 model with a standard size as well as an Arduino Microcontroller board for this study. In this study, we employed a body temperature sensor and a beat rate sensor to gather data from the patient's body at the same time, and the rest was taken as a logical value from the dataset. Then we keep the data in the cloud and analyse it before sending the results to a website to allow doctor monitor there. The processor speed is predicted to be 1.6 GHz, the total RAM capacity is 4 GB, the required disc space is 1 TB, and a monitor is required. The software necessary to implement/simulate this research effort includes the Ubuntu 20 operating system,

the cloud/Arduino platform visual basic .NET frontend technologies, and MYSQL as the back end.

3.1 PPHM Platform and Its Assumptions

The proposed architecture of PPHM which consists of three-tier model in Fig. 1 assumption can be seen. The different sorts of elements which comes in this section are observation station (OS), data centre subsystem (DCS), case study congestive heart failure. As we know that the observation System (OS) could be an IoT subsystem. The main quintessence activity tracking infrastructure section of IoT is the base station of WSNs. This subsystem also consists of many n number of base stations, HJ = b_1, b_2, b_3,…, b_n. Where Each b_iHJ specifies a patient suffering from cardiac disease and is delineated as b_i = <LWSN, PS> where PS is a kind of personal server and LWSN = LWSN_i … LWSN_m as sequence of m energy reticent lightweight wireless energy nodes. Every LWSNi L made up of full capability for collecting patient's data instances and aggregate it, after performing fundamental processing forward data to a personalized servers for pre-processing. The personal server is one which acts as a main interface between IoT subsystem and a cloud infrastructure. The personalized servers are dedicated for every patients' machines like wearable health devices, tablet and smartphone.

4 Proposed Methodology

The suggested system would allow elders and patients to be monitored at home instead of needing to go to the hospital. In this work, we will develop a new platform that integrates the Internet of things with cloud computing, and we will use machine

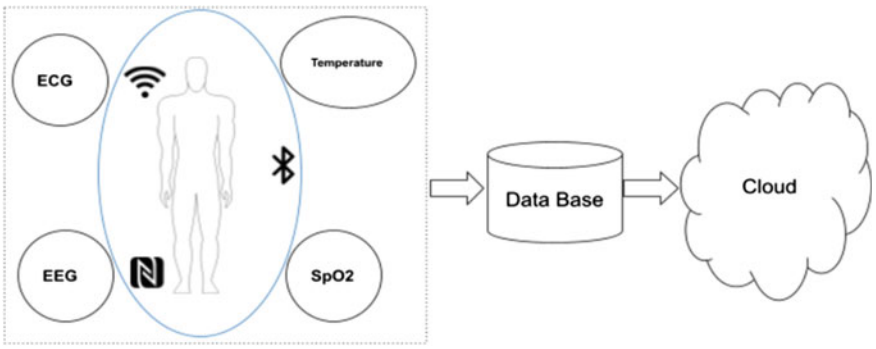


Fig. 1 PPHM architecture

learning to forecast if a patient needs to meet his doctor or not. The proposed system uses a Raspberry Pi, an Arduino and some other to gather data from patients. Finally, we construct an asp .net Web page that feeds the data into our module, which analyses patient data according on the patient ID. Model learning with predefined thresholds underpins the machine learning technique. We offer preliminary advice to the patient. The decision making mechanism's conclusion will be carried on to the next steps.

Our system provide updates to end users namely patient's as well as healthcare service providers and also helps in surveillance of on basis of continuously interval for data pushing to database (cloud) and Decision making system.

1. Storing the patients' data
2. Policy for allocating the health data
3. Cloud availability of middleware.

Patient's data storage is used to store patient's medical information (e-Health) maintaining long-term and the information from IoT subsystem (sensory data). E-Health consists of conventional clinical data like clinical observations and lab test results whereas sensory information consists of longitudinal patient data supported by base stations of WSN's. One of the pivotal aim of healthcare service provider is to collect patient clinical data and sharing with authorized healthcare professionals. In this propounded research work, the data scanning instance is highly responsive to setup the HDAP considered to regulate compliance requirements and also to provide best in class healthcare facilities for patients. The host will perform analysis and intervention of patient clinical data which is being collected by IoT.

4.1 Algorithm

Decision making system algorithm works based on input of data from personalized servers of patients cloud storage or other database. Calculation of probability for attributes for decision making support system as shown in Fig. 2 with the help of Eq. 1. We can change the attributes for making decisions; based on input data, we

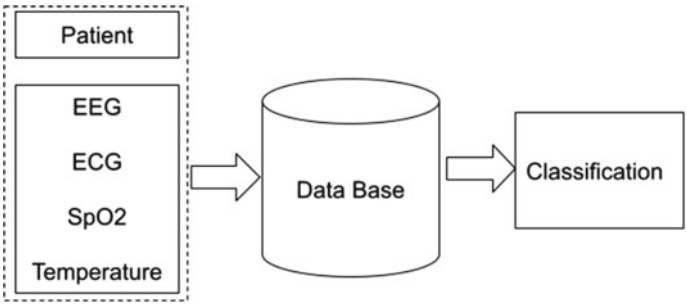


Fig. 2 Prediction system

can customize the algorithm for different diseases based on the clinical parameters selected.

$$P(C|D) = P(D|C).P(C)/P(D) \quad (1)$$

where

- P(C|D)—Probability of C being true given that D true.
- P(D|C)—Probability of D being true given that C is true.
- P(C)—Probability of C being true.
- P(D)—Probability of D being true.

Compute the probabilities by P for each classes,

In this work, we implement naïve Bayes classification for prediction. We start our work by collecting dataset from machine learning repository which need 13 parameters for 400 patients. We divide them by 70% training and 30% for testing phase; our algorithm will accept n patient with 13 parameters then predict; our model will do prediction and then it will pass alert to patient and doctor, then doctor will decide based on persons' present parameter and suggestions of prediction system.

4.2 Evaluation Metrices

Here, we are trying analyse the performance of different classification algorithms based on different feature selection and feature representation methods for different size of bench mark datasets which are available open source. Here, as performance metrics, we look at accuracy, precision, recall and F-measure [28–32].

5 Results and Discussion

We consider patient health parameters for experimentation to assess the decision-making system and understand the current health status. Here, the training process based on the input data, we use Naive Bayes classifier to evaluate the decision making system based on machine learning strategy. In this part, we will discuss about the result for our experimentation. We started our experimentation by dividing data to two part one for training with 70 out of 100 and testing with 30 out of 100. The training phase we train our model with input value for many people and giving the label for that in process called training phase which supervised machine learning. In next part model, learn the pattern then, based on the input features, then model will predict the right value for the testing part. We explore our accuracy with machine learning algorithms. Classification accuracy with naïve Bayes is 81%. For this prediction task. Finally, in this part, we will mention our experimentation in Fig. 3. Here, our

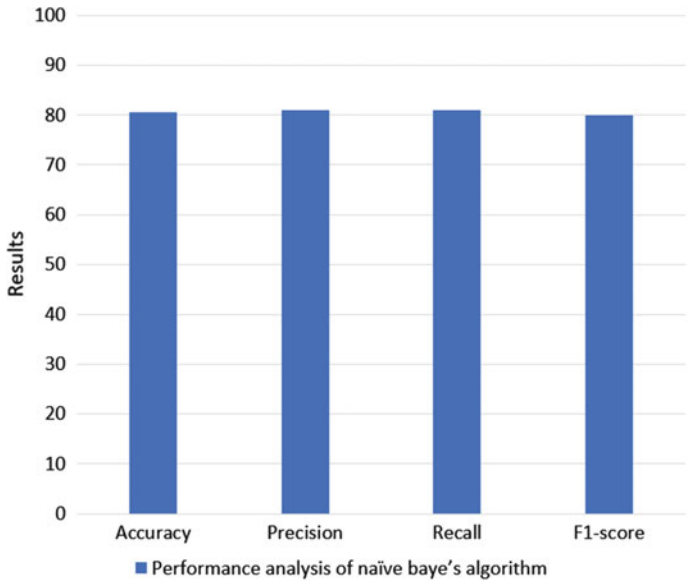


Fig. 3 Result analysis of decision making system

Table 1 Performance analysis of Naive Bayes algorithm

	Accuracy	Precision	Recall	F1-Score
Results	80.5	81	81	80

algorithm shows the evaluation metrices that were used to quantify the results (Table 1).

6 Conclusion

This research proposed that PPHMP uses both IoT and cloud computing to remotely monitor a patient's health condition. We have demonstrated and proven that the innovative study work suggested is scalable and energy-efficient with excellent accuracy using the experimental setup and performance analysis and evaluation system. We have also agreed that the patients can go without having to go to the hospital. We have decided to develop the PPHMP utilizing IoT-based cloud computing and real-world test beds in order to improve the performance of an IoT network. Finally, blockchain technology integration would be the future work.

Acknowledgements We are thankfull to Dr. Naveen maurya, Mr Govind raj, Mr Puneeth, Mr Bhairav R for the Support. We also acknowledge to HPC lab UOM and University of Mysore.

References

1. E. Borgia, The Internet of Things vision: key features, application and open issues. Comput. Commn. **54**, 1–31 (2014)
2. R. Poovendran, Cyber-physical systems: closeencounters between two parallelworlds. Proc. IEEE **98**(8),1363–1366 (2010)
3. L. Atzori, A. Iera, G. Morabito, The Internet of Things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
4. M.S. Hossain, G. Muhammad, Cloud assisted Industrial Internet of Things (IIOT) enabled-framework for health monitoring. Comput. Netw. **101**, 192–202 (2016)
5. B. Townsend, J. Abawajy, Security consideration for wireless carrier ago-nistic bio-monitoring Systems. Sec. Privacy Commun. Netw. **164**, 725–737 (2016)
6. Ghanavathi, J. Abawajy, D. Izadi, An alternative sensor Cloud architecture for remote patient health care monitoring and analysis, in *Proceedings of the IEEE International Joint Conference Neural Networks*, July 2016, Vancouver, Canada, pp. 24–29
7. L. Atzori, A. Iera, G. Morabito, The Internet of Things: a survey. Comput. Netw. **54**(15), 2787–2805
8. F. Bonomi, R. Milito, J. Zhu, Fog computing andits role in Internet of Things, in *Proceedings of the MCC*, Helsinkini, Finland (2014)
9. C. Doukas, I. Maglogiannis, Bringing IoT and cloud computing towards per-vasive healthcare, in *Proceedings of Sixth İnternational Conference on İnnovative Mobile and Internet Services in Ubiquitous Computing (IMIS)* (2012), pp. 922–926
10. L. Fei, M. Voegler et al., Towards automated IoT application deployment by acloud based approach, in *Proceedings of the EEE 6th International Service Oriented Computing and Applications (SOCA)* (2013), pp. 61–68
11. S. Misra, S. Chatterjee, M. Obaidat, On theoretical modeling of sensor cloud: a paradigm shift from wireless sensor network. IEEE Syst. J. **99**, 1–10 (2014)
12. N. Mitton, S. Papavassiliou, Combining cloud and sensors in a smart cityenvironment. EURASIP J. Wirel. Commun. Netw. **12**(1) (2012)
13. C. Zhu, H. Wang, X. Liu, L. Shu et al., A novel sensory data processing framework tointegrate sensor networks with mobile cloud. IEEE Syst. J. **99**, 1–12 (2014)
14. L. Yang et al., People centric service for mhealth of wheel chair users in smartcities, in *Internet of Things Based on Smart Objects*, April 2014, pp. 163–179
15. K. Zhang et al., Security and privacy for mobile health care networks, from a quality of protection perspective. IEEE Wirel. Commun. Mag. **22**(4), 104–112 (2015)
16. G. Fortino, D. Parisi, V. Pirrone, Body cloud: a SaaS approach for com-munity body wireless sensor networks. Fut. Gener. Comput. Syst. **35**, 62–79 (2014)
17. Y. Yin, Y. Zeng et al, The Internet of Things in healthcare: an overview. J. Ind. Inform. Integer **1**, 3–13 (2016)
18. S. Tuli, et al., Healthfog: an ensemble deep learning based smart healthcaresystem for automatic diagnosis of heart diseases in integrated iot and fog computingenvironments. Fut. Gener. Comput. Syst. **104**, 187–200 (2020)
19. J. Konˇcar, et al., Setbacks to IoT implementation in the function of FMCGsupply chain sustainability during COVID-19 pandemic. Sustainability **12**(18), 7391 (2020)
20. A.J. Jara, M.A. Zamora-Izquierdo, A.F. Skarmeta, Inter-connection framework for mHealth and remote monitoring based on the internetof things. IEEE J. Sel. Areas Commun. **31**(9), 47–65 (2013)
21. A. Secerbegovic, et al., The research mHealth platform for ECG monitoring, in *Proceedings of the 11th International Conference on Telecommunications.* (IEEE, 2011)
22. S. Kumar, R.D. Raut, B.E. Narkhede, A proposedcollaborative framework by using artificial intelligence-internet of things (AI-IoT)in COVID-19 pandemic situation for healthcare workers. Int. J. Healthcare Manag. **13**(4), 337–345 (2020)
23. R.P. Singh, et al., Internet of things (IoT) applications to fight againstCOVID-19 pandemic, in *Diabetes an Metabolic Syndrome: Clinical Research Reviews* (2020)

24. J.H. Abawajy, M.M. Hassan, Federated internet of thingsand cloud computing pervasive patient health monitoring system. IEEE Commun. Mag. **55**(1), 48–53 (2017)
25. J. Hariharakrishnan, N. Bhalaji, Adaptability analysis of 6LoWPAN andRPL for Healthcare applications of Internet-of-Things. J. ISMAC **3**(02), 69–81 (2021)
26. H. Wang, IoT based clinical sensor data management and transfer usingblockchain technology. J. ISMAC **2**(03), 154–159 (2020)
27. V. Suma, Wearable IoT based distributed framework for ubiquitous computing. J. Ubiquitous Comput. Commun. Technol. (UCCT) **3**(1), 23–32 (2021)
28. J.V. Bibal Benifa, J. Philip, C.B. Chola, Detection and classification of brain tumour from MRI images by different classifiers, Chapter: 7, in *High-Performance Medical Image Processing*. (Apple Academic Press, 2021). ISBN 9781774637227
29. C. Chola et al., IoT based intelligent computer-aided diagnosis and decision making system for health care, in *2021 International Conference on Information Technology (ICIT)* (2021), pp. 184—189.https://doi.org/10.1109/ICIT52682.2021.9491707
30. A.Y. Muaad, H. Jayappa, M.A. Al-antari, S. Lee, ArCAR: a novel deep learning computeraided recognition for character level arabic text representation and recognition. Algorithms **14**, 216 (2021). https://doi.org/10.3390/a14070216
31. M. Pramodha, A.Y. Muaad, J.V.B. Bibal, J. Hanumanthappa, C. Chola, M. Al-antari, A hybrid deep learning approach for COVID-19 diagnosis via CT and X-ray medical images, in *Proceedings of the 1st Online Conference on Algorithms*, (2021), MDPI: Basel, Switzerland. https://doi.org/10.3390/IOCA2021-10909
32. A.Y. Muaad, M.A. Al-antari, C. Chola, J.V.B. Benifa, J. Hanumanthappa, Detection of misogyny from Arabic Levantine Twitter tweets using machine learning techniques, in *Proceedings of the 1st Online Conference on Algorithms*, (2021), MDPI: Basel, Switzerland. https://doi.org/10.3390/IOCA2021-10880

Adaptive Load Balancing Scheme for Software-Defined Networks Using Fuzzy Logic Based Dynamic Clustering

Ashish Sharma, Sanjiv Tokekar, and Sunita Varma

Abstract This paper presents an intelligent load balancing strategy among the controllers at the control plane of software-defined networking (SDN) paradigm. Due to the separation of user plane and control plane in SDN framework, the assignment of switches and the respective controller is a tremendously complex task. The use of a central allocation scheme for dynamic controller assignment in SDN results into a huge amount of data at the central super controller. The processing of this large amount of data presents a trade-off situation between the quality of service (QoS) parameters. This issue has been resolved in this work by proposing a cluster-based approach for a dynamic controller allocation scheme. A non-uniform intelligent clustering scheme is proposed in this work to provide a robust load distribution technique in the presence of dynamic and uncertain network and traffic environments. The formation of clusters and the cluster head selection is done using fuzzy logic control to incorporate a greater number of parameters for decision making and to enhance the performance in terms of QoS requirements. The theoretical analysis of the optimal load balancing technique for SDN is supported by the experimental analysis. The performance of fuzzy logic-based clustering has shown a significant improvement in the average latency and packet loss as compared to the cluster-less load balancing and conventional cluster-based method for the Open Network Operating System (ONOS) controllers.

Keywords Software defined network · Fuzzy logic · Load balancing · Clustering

A. Sharma (✉)
Government Women's Polytechnic College Indore, Indore, MP, India

S. Tokekar
IET DAVV Indore, Indore, MP, India

S. Varma
SGSITS Indore, Indore, MP, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_35

1 Introduction

The ever-increasing number of internet users over the last decade has posed a serious challenge to the available IP network structure. The diversified services like audio, video, live streaming, webcast, online gaming, etc. with the required level of QoS, has completely changed the paradigm of the field of computer networking. The huge amount of data required to be communicated to fulfill the requirement of users cause heavy traffic in the network and huge load on the servers. The problem becomes more severe with the availability issues of the limited resources. A very efficient resource management strategy is the desperate need for the networking infrastructure to deal with this situation. The problem of resource allocation becomes more complex in the case of providing the services to the heterogeneous network [1–3].

This problem has attracted many researchers to come up with some efficient and flexible resource allocation schemes that could provide the required QoS also. However, the architecture of the conventional IP network, as shown in Fig. 1, makes it very difficult to use the available resources among all the users to its full extent. In traditional networks, the complete traffic flow is based on the source to destination routing framework only, which does not guarantee the best and most efficient path between the source and destination [4]. This may result in a compromise with the QoS parameters like throughput, packet delivery rate, latency, bandwidth, etc. and leads to congestion.

The flow control and data forwarding in a conventional network are closely coupled to each other thereby reducing the flexibility. Due to this, the distributed network devices are used to decouple them, but this technique is complex and difficult to implement. Another problem with conventional architecture was the implementation of high-level network policies. These policies cannot be directly implemented

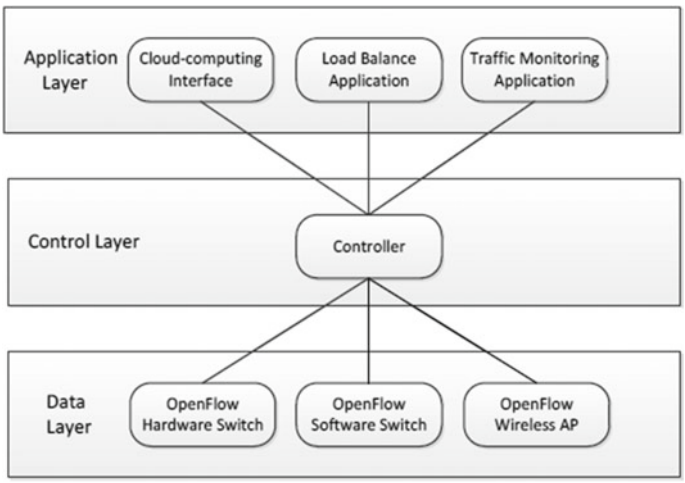


Fig. 1 SDN Architecture

here and need to be converted first to vendor level and then could be applied to each network device. This rigidness of the conventional networks has evolved a need for a structural modification in the network architecture, which can be flexible enough to balance the Load without compromising much with the QoS requirements [5].

Software-defined Networking (SDN) has emerged as a new networking paradigm where the architectural modifications have been done by separating the data plane and control plane. This logical decoupling between these two planes has provided optimal network traffic management and flexible resource management. The control plane is dedicated to the coordination between all the nodes which in turn allows the network operators to apply the high-end policies for flexible network management and network traffic control [6]. These high-end control logics are implemented to satisfy the end user's requirements. Although for a small network, one controller can be implemented to control the flow of data within the network, it is not an efficient scheme but for large networks because of the limited processing capabilities. The corresponding parameters governing the network performance like latency, throughput, power consumption, packet delivery ratio, etc. are also affected in this scheme [7].

One possible way to deal with this situation is to implement multiple controllers in the control plane of the network with limited computation and processing capabilities. These controllers are dedicated to the task of controlling several groups of switches which are known as clusters. These controllers are also connected to each other to take the communication further as per the requirements. This is a cluster-level multi-tier architecture of the control plane in SDN. The most suitable node as per the optimal parameters' conditions will be the cluster head and be allotted as controller. The optimal controller placement and the matching of control among switches is a big concern in the case of static network deployment. Therefore, the allocation has to be dynamic in nature and adaptive in operation to match the changing environment. The scheme for optimal controller allocation or resource management to deal with the load balancing among different controllers cannot be based on some statistical method as the communication parameters are very random and uncertain [8–10].

Machine learning tools can be used to deal with the situation of random and uncertain communication environments in resolving the problem of cluster head selection and load balancing. Over the last decade, fuzzy logic has proved its potential in intelligent decision making in various applications. The capability of fuzzy logic control to map the complex mathematical framework into simpler linguistic architecture has really changed the paradigm of the intelligent decision-making process. The computational complexity and the accuracy of the fuzzy logic-based control architecture are also superior to the conventional statistical techniques [11–13].

This paper presents an intelligent adaptive clustering-based strategy for the multi-tier architecture of the control plane for SDN. The artificial intelligence in this work has been incorporated through fuzzy logic, which is utilized for cluster formation and the respective cluster head selection. The decision on allocation and selection is based on the multiple network parameters like latency, distance between controller and switch, energy consumption, etc. The controller allocation scheme is done in such a way that the load balancing could be attained in the most optimal manner.

The major contribution of this work is the novel adaptive intelligent multi-tier framework for the control plane of SDN using fuzzy logic. The proposed model is network environment-sensitive and adaptive in nature which adds novelty to the research. It has combined the advantages of clustering-based distribution of the complete network to optimize the Load among the controllers and the intelligence of fuzzy logic to make a decision in the presence of randomness, noise and uncertainties. This combined approach has added robustness to the communication model without compromising with the QoS parameters and computation complexity. It also has presented a non-uniform clustering algorithm where the number of controllers in a cluster is not equal and it will be allotted in such a way that the best optimal load distribution is ensured. The algorithm is dynamic in the sense that the performance of the clustering and cluster head selection is evaluated periodically to cope up with the changing network parameters and QoS requirements.

The remaining paper is organized as follows: Sect. 2 discusses the related work in the field of load balancing in SDN. The mathematical model of the clustering-based resource management is presented in Sect. 3. Section 4 proposes the fuzzy logic-based clustering framework implemented for the control plane of SDN. The performance of the proposed technique is evaluated in Sect. 5 through the simulation study, while Sect. 6 concludes the paper.

2 Related Work

The potential of SDN in high-end communication has been recognized and explored by many researchers over the last decade [14]. Various aspects of SDN, like resource management and load distribution, have also attracted the attention of many researchers [15–17]. Some remarkable works are discussed in this section to identify the challenges and the opportunities in the field of SDN to enhance the network performance further.

Various challenges and issues of the high-end computing technologies like cloud, big data, SDN etc. have been addressed in [18]. They have also discussed the importance of the strategies of data streaming and processing for further analysis. They presented that there is a huge amount of data communicated in the high bandwidth network needs to be processed using a programmed framework in a distributed environment known as Hadoop, which supports a new programming paradigm known as MapReduce. Cloud computing resources are also to be allocated optimally. Authors in [19] presented the concept of green networking to address the enhancement in networking, caching, and computing in the SDN. This concept helped to reduce power consumption and increase the lifetime of the wireless node. They proposed a solution using a Software-Defined network to improve network planning and development. A bandwidth-based joint algorithm was proposed in [20] by emphasizing the 4Cs. Optimal load balancing was tried to be achieved by them establishing a trade-off between the parameters governing the communication, computation, caching, and control parameters.

A detailed survey of load balancing algorithms for SDN has been carried out by [21]. They have encompassed the various techniques included into the three types of load balancing, static, dynamic, and hybrid. A flow table-based load balancing scheme was proposed in [22] by designing a combination of a single flow table and group flow table algorithm. The traffic monitoring of a single node is monitored by a single flow table, while the group flow table was responsible for the load balancing the server by categorizing the client nodes. This algorithm not only avoided the excessive number of flow tables effectively but solve the matching range defect in the wide flow tables too. The concept of cloud load balancing was proposed by [23] in the form of an SDN-enhanced Inter-cloud Manager (S-ICM). It comprises two important aspects of communication, monitoring and decision making for efficient load balancing. For monitoring, S-ICM uses an SDN control message that observes and collects data, and decision-making is based on the measured network delay of packets. Measurements are used to compare S-ICM with a round-robin (RR) allocation of jobs between clouds, which spreads the workload equitably, and with a honeybee foraging algorithm (HFA). S-ICM is better at avoiding system saturation than HFA and RR under heavy load formula using RR job scheduler. However, this study did not consider a multi-controller for load balancing using SDN.

The idea of efficient configuration and management of SDN infrastructure was presented by [24] in the context of data centers and Wide Area Networks (WAN). They have addressed the issue of Load balancing across multiple host network interfaces. They have explored and implemented different SDN-based Load balancing approaches based on OpenFlow software switches and have demonstrated the feasibility and potential of their approach. They reviewed some parameters for the SDN-based data center environment also.

Authors in [25] introduced the concept of clustering and control plane. They proposed that clustering can be used inside the Software-Defined network approach to explore relevant node groups and plan the network accordingly. Authors also insist on the importance of authentication inside network planning and the involvement of massive data processing with data analysis. Some resource management techniques on the basis of QoS parameters using some intelligent techniques like game theory, reinforcement learning, neural network, etc., have been proposed [26]. However, the performance of these techniques varies for the varying scale of the network.

3 Clustering-Based Network Framework

The network architecture presented in this work is a two-level network. One level comprises the Master controller (MC) or primary controller (PC) and the second level includes the other controllers known as “team members”. In the architecture of the control plane of SDN, the two levels are interfaced with each other in such a way that all the controllers are always connected to the MC. The load information and the other communication parameters are constantly transferred from all controllers to the MC. Unlike the conventional architecture where all the controllers were connected to the

MC, this paper presents cluster-based network architecture. The MC will gather the information received from all the controllers and divide team members into groups, known as clusters. The clustering is performed to balance the load among all the controllers and the MC as well. Once the clustering configuration is completed, the MC performs the configuration and monitoring of the controllers and their switches.

The time complexity in the clustering approach has also been found to be reduced to $O(M^2)$ [27]. The cluster allocation is adaptive in nature and the dimension and the structure of the cluster keep on changing as per the change in the communication environment and the required QoS. This results in the reduction of overhead and the response time too. The controllers having the best and optimal performance scenario are selected as cluster head (CH). The selection of clusters has to be done intelligently so as to maintain the network performance and QoS parameters also. It also has to be taken care that the cluster head selection process doesn't disturb the load balancing. Therefore, a great level of synchronization among various controllers and between CH and MC is of the highest importance. Also, if MC goes down, the next selected controller becomes the new MC. When the failed MC recovers, it resumes operation only as a team member in the cluster.

The network configuration considered in the proposed work is shown in Fig. 2. The mathematical framework of the presented network architecture is discussed as follows: Assuming the controllers set as $C = \{C_1, C_2, \ldots, C_N\}$ where N represents the number of controllers and the respective distances between the controllers, computational power and latencies are represented as $D \in R^{N \times N}$, $E \in R^N \wedge L \in R^{N \times N}$ respectively. Here d_{ij} is the distance between ith and jth controller. The clusters formed in the network are shown as $X = \{X_1, X_2, \ldots, X_K\}$ where K is the number of clusters in the network. The value of K will be varying and will be evaluated in each time cycle.

Unlike the previous works, where the number of controllers per cluster was constant, here it will depend upon the instantaneous values of the network parameters and the desired QoS parameters. This aspect of the SDN for the non-uniform clustering of controllers is addressed in this paper to deal with the random traffic scenario. The number of controllers in each cluster is represented by $M(m)$, where m varies from 1 to K. This type of traffic environment may be created in real-time where the switches may not always be requiring the same rate of data flow because of varying user demands.

Advance communication technologies resembling the fourth and fifth-generation are providing the bandwidth as per the service in demand concept. The selection of the number of controllers in each cluster is also variable and will be selected on the basis of the communication and QoS parameters. MC at the control plane of the SDN will be responsible for taking the multi-level decisions on the number of controllers in a cluster and about the cluster head. Also, the assumption that all the controllers in the network have the same processing capability may not be true in a real-time network situation. Considering all these non-uniform and uncertain traffic and network environments, an intelligent MC is proposed in this work and thereby making it adaptive in true sense.

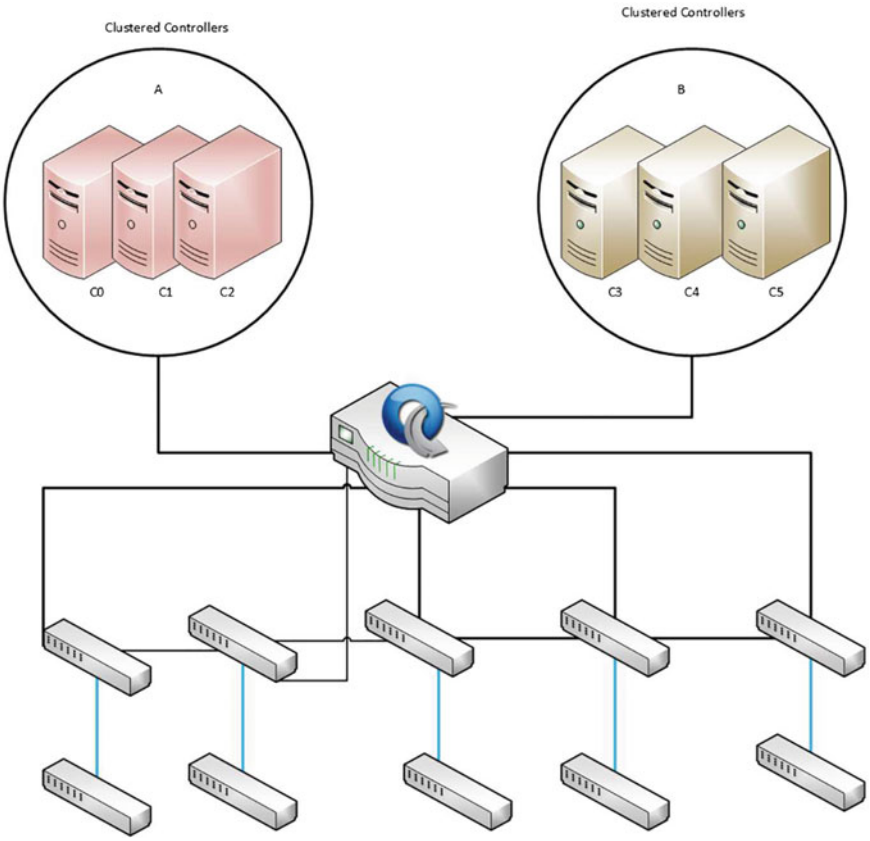


Fig. 2 Clustering-based SDN configuration

4 Proposed Fuzzy Logic-Based Clustering for Load Balancing

4.1 Fuzzy Logic Control

Fuzzy logic is a decision-making framework that is very close to the nature of the human thought process. Unlike the crisp set decision-making structure, fuzzy logic converts the linguistic strategy into the overlapping decision sets. The selection of these sets is made on the basis of human experiences, which makes the decision realistic in practice. The implementation of fuzzy logic control is done through a fuzzy inference system (FIS) which comprises of three stages: fuzzification, fuzzy inference logic and defuzzification. The complete process of fuzzy logic control is depicted in Fig. 3. The first stage of fuzzification deals with the measurement of the input variable and its mapping with the linguistics scale. The Fuzzy inference logic

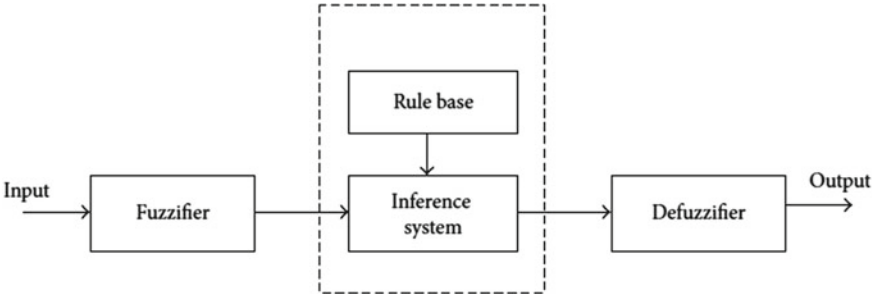


Fig. 3 Fuzzy logic system

stage is the real brain of the fuzzy logic control system. It includes the rule base, which is responsible for making the decision. It defines the membership function for each parameter. The value of the membership function decides the outcome of the system. The whole intelligence of the fuzzy logic lies under the smartness of deriving these rule bases and respective membership functions. The Defuzzification stage covers the linguistic variables again to the numerical values so as to generate the machine interfaced parameters.

4.2 Fuzzy Logic-Based Clustering for Load Balancing

The proposed load balancing framework for the software-defined network incorporates the potential of artificial intelligence tools with the clustering-based architecture. Here Fuzzy logic has been used to derive the complete dimension of the clusters along with the cluster head. The decision taken by fuzzy logic is based upon various network parameters like the distance of a controller from other switches, the load-carrying capacity of the controllers and the respective latency. Cluster Vector Load (L) will be shared and communicated by the MC. It is the vector representing the controller loads (l), i.e., the average flow request per second, in the cluster. The overall cluster load (Θ) will be evaluated as

$$\ominus(k) = \sum_{m=0}^{K} l(m)Y(m) \tag{1}$$

The performance of the clustering is evaluated in this work by the cluster cost (Ψ)

$$\Psi(k) = \ominus(k) + l(m, k). \tag{2}$$

where $l(m, k)$ represents the load of a newly added controller in the kth cluster with the mth controller of the respective cluster. The algorithm for clustering-based load balancing is divided into two phases: initial clustering and rearranging. Initial

clustering phase deals with the distribution of controllers in a cluster on the basis of the load capacity, respective distances between the controllers and the instantaneous cluster load. The controllers are sorted in decreasing order on the basis of load capacity and other controllers are evaluated on the basis of distance and latency with respect to the respective controller. The decision-making about the inclusion of a controller in a specific cluster is performed through fuzzy logic.

A fuzzy inference system (FIS) is designed for the decision-making about accepting or rejecting a controller in a respective cluster. The inputs to the FIS are the distance vector, Latency vector and cluster load vector n and the output of the system is Y. The value of Y decides that whether the respective controller be included in the cluster or not. $Y = 1$ resembles inclusion and $Y = 0$ resembles rejection. Fuzzy logic control is a decision-making framework on the basis of overlapping membership functions, which resembles the rules of making the decision. This overlapping intermediate approach of generating output is similar to the linguistic control, which is based on expert knowledge. The rule base is derived in such a way that the optimal load balance is achieved for every cluster. The evaluation of the inclusion eligibility of the nodes in a respective cluster will continue till the cluster cost reaches the threshold value of $\Psi(k)$, i.e., Ψ_{th}. After the cluster reaches its threshold cost, the list of remaining nodes is updated and the same evaluation using FIS is done for the next cluster head with higher load capacity and the remaining controllers. The process of cluster allocation continues till all the controllers are allotted with any one cluster. By the end of phase 1, all the controllers are distributed in clusters. However, the cluster allocation obtained from phase 1 cannot be claimed to be optimal in terms of load because of the dynamic environment and traffic uncertainties in high-end communication networks which offer a bandwidth on demand services. So, a continuous assessment has to be done and the cluster needs to be restructured to attain the optimal load balance.

$$\Psi_{\text{av}} = \ominus(k)/K \tag{3}$$

A strategic utility function (χ) is derived from the average and threshold values of Ψ as

$$\chi = \frac{\Psi_{\text{av}} - \Psi_{\text{th}}}{\Psi_{\text{av}}} \tag{4}$$

The Strategic utility function will assess the quality of the clustering of the whole network and evaluate that whether the threshold load is not much larger or smaller than the average value. The decision of restructuring is taken on the value of the magnitude of χ. Phase 1 of clustering will be repeated if $|\chi| > 0.2$. The complete algorithm of fuzzy logic-based clustering to achieve the optimal load balancing is shown in Algorithm 1.

Algorithm 1 Fuzzy logic-based clustering Algorithm

input:

$\boldsymbol{n_t}$ Network contain $C = \{C_1, C_2, \ldots, C_N\}$ Controller list, N, number of controllers, respective distances between the controllers, computational power and latencies as $D \in \Re^{N \times N}$, $E \in \Re^{N}$ and $L \in \Re^{N \times N}$ respectively, Cluster Vector Load (L).

output:

$X = \{X_1, X_2, \ldots, X_K\}$ clusters formed in the network where K is the number of clusters in the network, overall cluster load (Θ), cluster cost (Ψ) and Y.

Procedure:

Phase 1:

1: *SortedListD* descending order of controllers list according to their loads.

2. *SortedListA* ascending order of controllers list according to their loads.

3. Candidates from SortedListD $\leftarrow X$

4. **for** all c belongs to SortedListC **do**

$Y =$ ***fuzzy*** (distance vector, Latency, cluster load).

5. end for

6. **if** $Y = 1$, **then**

include c in x, evaluate Θ and Ψ, increment m.

end if

7. **If** $\Psi < \Psi$th**, then**

repeat step 4 to 6.

8. **Else**

Remove the selected c from *SortedListA* and *SortedListD,* go to step 3.

9. **End if**

Phase 2:

10. Calculate Ψ_{av} and χ.

11. **if** $|\chi| > 0.2$ **then**

Repeat the steps 1 to 7.

12: **end if**

13. **Return X**

The time complexity of the proposed fuzzy logic-based adaptive clustering is compared with some conventional load distribution techniques, as shown in Table 1. Considering the number of switches that are associated with N number of controllers,

Table 1 Time complexity comparison [27]

Technique	Time complexity
Switches Matching Transfer	$O(N.S.\log S)$
Balance Flow	$O(\max((S^2).\log(S^2), S^2.N))$
Dynamic Controller Placement (DCP)—Assignment Phase	$O(N.S.\log S)$
Hybrid Flow	$O(S.N^2)$
Proposed Fuzzy logic-based clustering	$O(N^2)$

it is shown in Table 1 that the complexity reduction in the proposed technique is considerable and encouraging.

5 Experimental Setup and Analysis

The performance of the proposed load balancing strategy for SDN is evaluated using the experimental analysis. The experimental setup is established by installing ONOS controller over Amazon EC2 Ubuntu Server 64-bit 14.0 LTS edition over three separate servers. Rest API has been used here to implement the clustering of multiple controllers and to manage the Open-flow switches which are connected to the master controller with the IP addresses indicating the standby controllers for each connected switch. The distribution of controllers among the clusters is governed by fuzzy logic-based decision-making. A standard network topology from the Internet topology zoo (ITZ) has been used here, which stores the data of network topologies in the graphical description in extensible markup language (XML) format. We have used Agis Network topology as a testbed network from ITZ, which contains 25 switches, 25 hosts with 30 connected links. The topology map of the Agis network is shown in Fig. 4, which provides sufficient information to build a testbed network as close to a real-world topology.

The average latency and packet drop of the network are evaluated through an application known as Controller bench marker (Cbench), which generates new packet-in events to test the SDN OpenFlow controllers. PACKET_IN message is sent to the controller by Cbench and it waits for the response before sending another packet. The controller IP address is passed into the network topology and transformed into the python script. Then, Mininet is used as the network emulator to conduct the experiment testing.

A fuzzy logic-based decision-making scheme is derived for the clustering and the respective cluster head election. The number of controllers considered in this analysis is 100. Network topologies from the Internet Topology Zoo are assumed here for the distance matrix, load and latency vector. The input parameters for the fuzzy inference system considered in this work are distance matrix, Latency vector and cluster vector load. These parameters are converted into the linguistic version

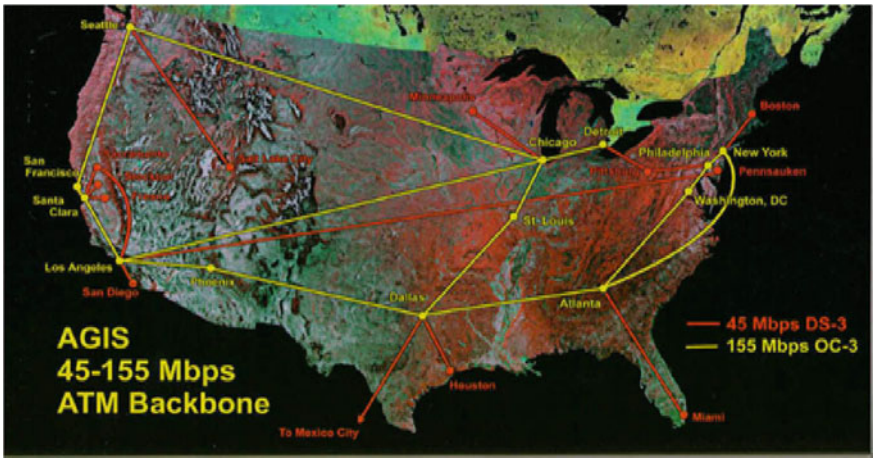


Fig. 4 Agis network topology map

through the fuzzification process. Distance is fuzzified in 4 linguistic sets, Very Low, Low, Medium and High. Latency and load capacity are fuzzified as low, medium and high. Similarly, the output is derived in terms of Y fuzzified as Yes and No. Yes, it resembles the selection as a cluster element and no resembles exclusion from the cluster. The rule base is derived for the FIS system to take an efficient and accurate decision which is as shown below:

- if distance is very Low and the latency is Low and Load is high THEN Y is Yes.
- if distance is Low and the latency is Low and Load is high THEN Y is Yes.
- if distance is Low and the latency is Low and Load is medium THEN Y is Yes.
- if distance is Low and the latency is Low and Load is low THEN Y is No.
- if distance is Very Low and the latency is Low and Load is medium THEN Y is Yes.
- if distance is medium and the latency is Low and Load is low THEN Y is No.
- if distance is high and the latency is Low and Load is low THEN Y is Yes.
- if distance is Low and the latency is high and Load is low THEN Y is No.
- if distance is High and the latency is Low and Load is high THEN Y is No.
- if distance is Very Low and the latency is medium and Load is medium THEN Y is No.
- if distance is medium and the latency is medium and Load is medium THEN Y is No.
- if distance is medium and the latency is Low and Load is low THEN Y is No.
- if distance is very Low and the latency is high and Load is low THEN Y is No.

The membership function type used in this fuzzy inference system is Gaussian, considering the uncertainties in the traffic and communication environment. The respective FIS is as shown in Fig. 5. The clustering performance of the proposed technique is shown in Fig. 6. The number of nodes allotted in each cluster is unequal

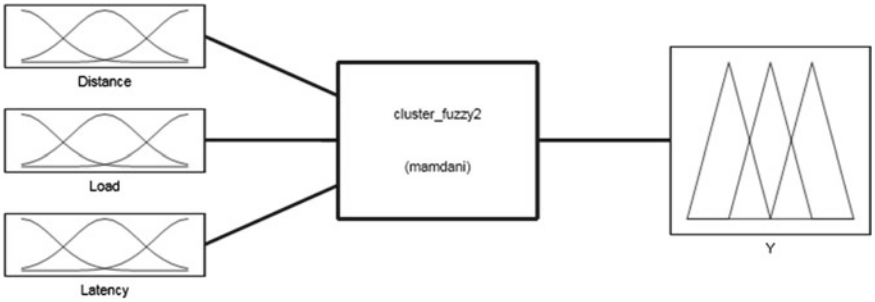


Fig. 5 Proposed fuzzy inference system

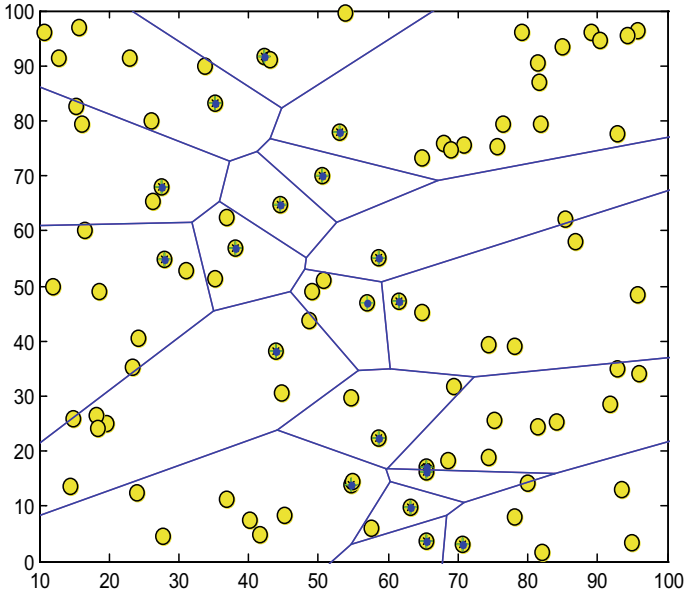


Fig. 6 Fuzzy based clustering

and dynamic depending on the load condition and performance cost in terms of the strategic utility function.

The efficacy of the proposed fuzzy logic-based load balancing strategy has also been verified through analysis of other network parameters like the number of clusters, average energy consumption and the number of iterations with respect to the number of nodes in the network. Figure 7 shows the variation of number of clusters with respect to number of nodes. It depicts that variation is uniform and realistic even when the number of nodes is close to 200. Similarly, Fig. 8 shows the average energy consumption with the increase in number of nodes. The most important point to note from this variation is that the average energy consumption is kept within

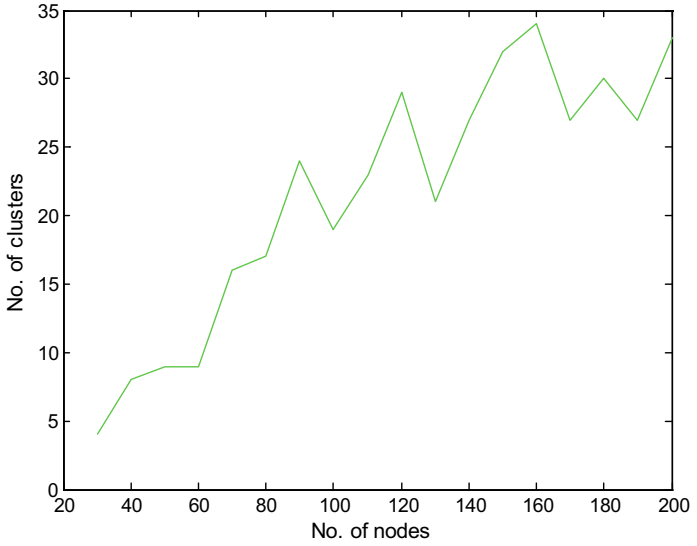


Fig. 7 No. of clusters versus no. of nodes

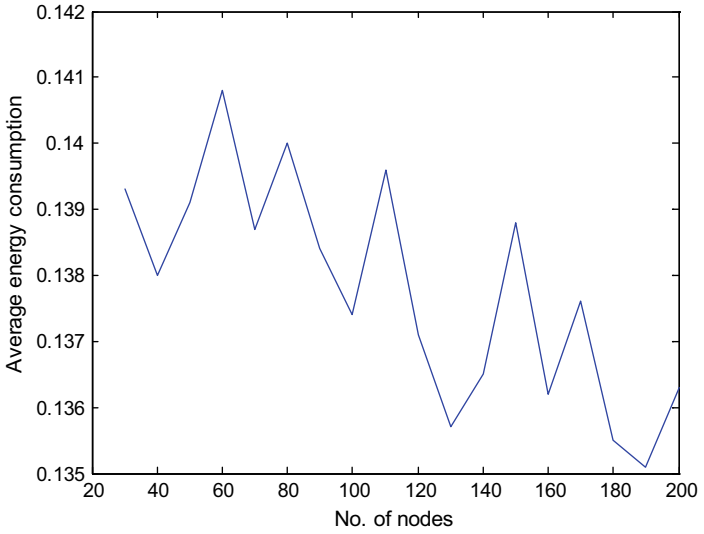


Fig. 8 Average energy Consumption versus no. of nodes

the limits only even when the nodes are larger in numbers. The clusters are derived in such a manner in this algorithm that the average energy consumption is kept within the practical limits. Figure 9 reflects the complexity and dynamic nature of the proposed clustering algorithm through the number of iterations required to reach

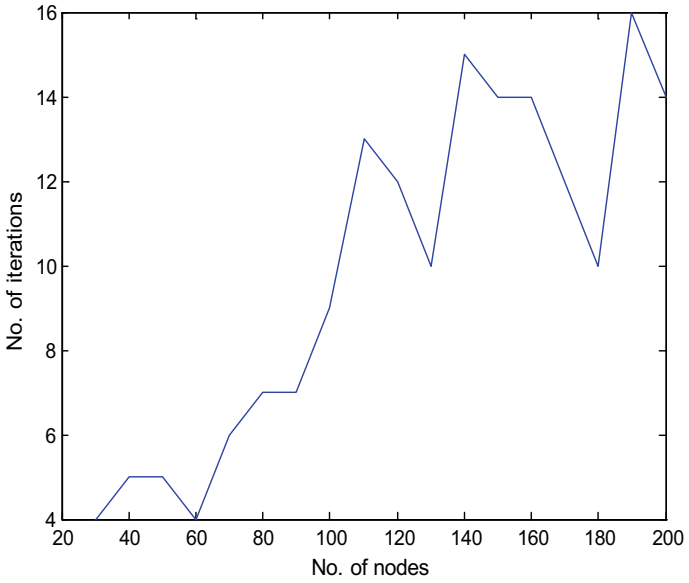


Fig. 9 No. of iteration versus no. of nodes

to the optimal clustering. It shows that number of iterations required attaining the best possible distribution of nodes in the clusters and the selection of respective cluster head is reaching to the maximum value of 16 only.

The variation of average latency for ONOS controller clustering with respect to the increase in the number of switches is presented in Fig. 10. The respective variation of average latency is compared with the case of no clustering and conventional clustering [28]. Similarly, Fig. 11 shows the comparative analysis of the proposed technique with no clustering-based Load balancing and conventional clustering-based load balancing in terms of packet drop ratio with respect to the number of sending packets. UDP packets with the size of 64,000 KB are sent between the end devices to evaluate the packet loss ratio.

The improvement in the latency and the packet loss ratio for the proposed fuzzy logic-based clustering technique for Load balancing in SDN is due to the efficiently coordinated operation among controllers within a cluster. Also, each controller is well aware of the network state, which is shared across the other clustered nodes. The adaptive clustering presented an efficient and optimal load balancing by redistributing the load among different nodes in the cluster and thereby improves the scalability too in comparison with the distributed cluster model without clustering and conventional clustering.

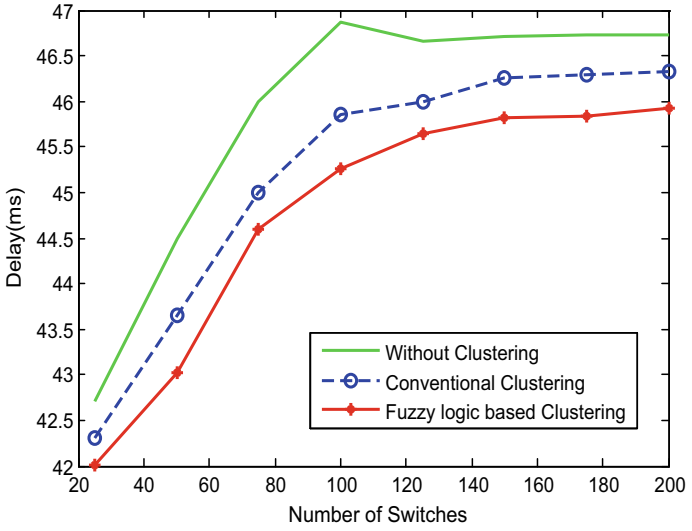


Fig. 10 Latency wrt no. of switches

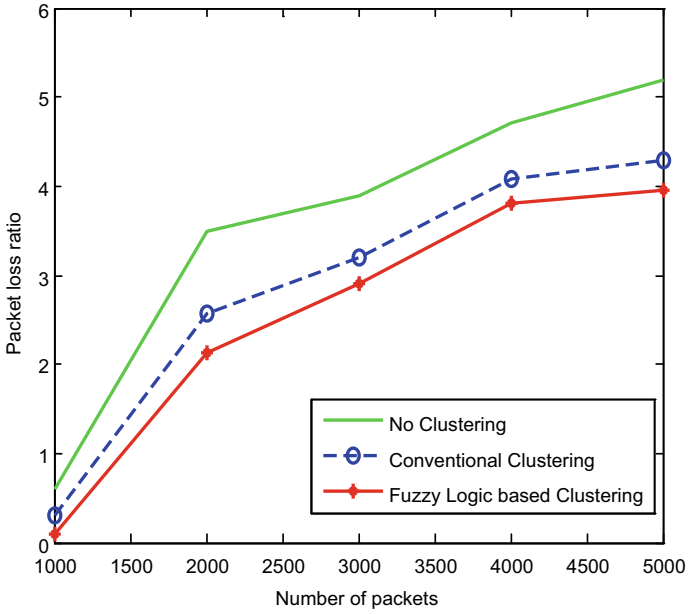


Fig. 11 Packet Loss ratio with respect to no. of packets

6 Conclusion

An adaptive intelligent load balancing framework is proposed in this paper, which uses fuzzy logic for the load-sensitive clustering. The proposed model is used for the load balancing of controllers in the control plane of software-defined networks. The parameters considered for the cluster head selection is the load capacity of the controllers and the clustering is performed on the basis of distance, latency and the cluster load. The performance of the proposed algorithm is ensured through the strategic utility function which resembles the best possible load distribution by minimizing the standard deviation in the iteration of clustering. The scenario of 200 nodes is evaluated in this work and the randomness and uncertainties in the load and latencies of the controllers is considered here to present a robust clustering framework. The simulation analysis showed the effectiveness of the proposed technique.

References

1. R. Schatz, T. Hoßfeld, L. Janowski, S. Egger, *From Packets to People: Quality of Experience as a New Measurement Challenge*. (Springer, Berlin Heidelberg, 2013), pp. 219–263
2. G. Fortino, C. Savaglio, C.E. Palau, J.S. de Puga, M. Ganzha, M. Paprzycki, M. Montesinos, A. Montesinos, M. Llop, Towards multi-layer ınteroperability of heterogeneous IoT platforms: the INTER-IoT approach, in *Integration, Interconnection, and Interoperability of IoT Systems*. (Springer, Cham, Switzerland, 2018), pp. 199–232
3. L. Cui, F.R. Yu, Q. Yan, When big data meets Software Defined networking: SDN for big data and big data for SDN. IEEE Netw. **30**(1), 58–65 (2016)
4. A. Osseiran, V. Braun, T. Hidekazu, P. Marsch, H. Schotten, H. Tullberg, M.A. Uusitalo, M. Schellman, The foundation of the mobile and wireless communications system for 2020 and beyond: challenges, enablers and technology solutions, in *Proceedings of the 2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, Dresden, Germany, 2–5 June 2013, pp. 1–5
5. K. Kirkpatrick, Software-defined networking. Commun. ACM **56**(9), 16–19 (2013)
6. T. Hu, Z. Guo, P. Yi, T. Baker, J. Lan, Multi-controller based software-defined networking: a survey. IEEE Access **6**, 15980–15996 (2018)
7. Y. Hu, W. Wang, X. Gong, X. Que, S. Cheng, Reliability-aware controller placement for software-defined networks, in *Proceedings of the 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*, Ghent, Belgium, 27–31 May 2013, pp. 672–675
8. S.H. Yeganeh, A. Tootoonchian, Y. Ganjali, On scalability of software-defined networking. IEEE Commun. Mag. **51**, 136–141 (2013)
9. A. Krishnamurthy, S.P. Chandrabose, A. Gember-Jacobson, Pratyaastha: an efficient elastic distributed SDN control plane, in *Proceedings of the ThirdWorkshop on Hot Topics in Software Defined Networking*, Chicago, IL, USA, 22 August 2014, pp. 133–138
10. Y. Liu, A. Hecker, R. Guerzoni, Z. Despotovic, S. Beker, On optimal hierarchical SDN, in *Proceedings of the 2015 IEEE International Conference on Communications (ICC)*, London, UK, 8–12 June 2015, pp. 5374–5379
11. I. Tal, G.-M. Muntean, Towards reasoning vehicles: a survey of fuzzy logic based solutions in vehicular networks. ACM Comput. Surv. **50**, 80 (2018)
12. P. Tillapart, T. Thumthawatworn, P. Viriyaphol, P. Santiprabhob, Intelligent handover decision based on fuzzy logic for heterogeneous wireless networks, in *Proceedings of the 2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Hua Hin, Thailand, 24–27 June 2015, pp. 1–6

13. A.B. Zineb, M. Ayadi, S. Tabbane, An enhanced vertical handover based on fuzzy inference MADM approach for heterogeneous networks. Arab. J. Sci. Eng. **42**, 3263–3274 (2017)
14. Y. Fu, J. Bi, Z. Chen, K. Gao, B. Zhang, G. Chen, J. Wu, A hybrid hierarchical control plane for flow-based large-scale software-defined networks. IEEE Trans. Netw. Serv. Manag. **12**, 117–131 (2015)
15. P.D. Bhole, D.D. Puri, Distributed hierarchical control plane of software defined networking, in *Proceedings of the 2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, India, 12–14 December 2015, pp. 516–522
16. D. Kreutz, F.M. Ramos, P.E. Verissimo, C.E. Rothenberg, S. Azodolmolky, S. Uhlig, Software-defined networking: a comprehensive survey. Proc. IEEE **103**, 14–76 (2015)
17. M.F. Bari, A.R. Roy, S.R. Chowdhury, Q. Zhang, M.F. Zhani, R. Ahmed, R. Boutaba, Dynamic controller provisioning in software defined networks, in *Proceedings of the 2013 9th International Conference on Network and Service Management (CNSM)*, Zurich, Switzerland, 14–18 October 2013, pp. 18–25
18. B. Görkemli, A.M. Parlakı¸sık, S. Civanlar, A. Ula¸s, A.M. Tekalp, Dynamic management of control plane performance in software-defined networks, in *Proceedings of the 2016 IEEE NetSoft Conference and Workshops (NetSoft)*, Seoul, Korea, 6–10 June 2016, pp. 68–72
19. Y. Hu, W. Wang, X. Gong, X. Que, S. Cheng, Balanceflow: Controller load balancing for openflow networks, in *Proceedings of the 2012 IEEE 2nd International Conference on Cloud Computing and Intelligent Systems (CCIS)*, Hangzhou, China, 30 October–1 November 2012, vol. 2, pp. 780–785
20. A. Ndikumana, N.H. Tran, T.M. Ho, Z. Han, Joint communication, computation, caching, and control in big data multirnell-access edge computing, in *IEEE Transaction Conference Location Cornell University*, 30 Mar 2018
21. M.H. Al Bowarab, N.A. Zakaria, Z.Z. Abidin, Load balancing algorithms in software defined network. Int. J. Rec. Technol. Eng. (IJRTE) **7**(6S5), 686–693 (2019) ISSN 2277–3878
22. M. Qilin, S. WeiKang, A load balancing method based on SDN, in *Seventh International Conference on Measuring Technology and Mechatronics Automation* (2015), pp. 18–21
23. B. Kang, H. Choo, An SDN-enhanced load-balancing technique in the cloud system. J. Supercomput. 1–25 (2016)
24. R. Kanagavelu, K.M.M. Aung, Software defined load balancer in cloud data centers, in *ICCIP'16 Proceedings of the 2nd International Conference on Communication and Information Processing* (2016), pp. 139–143
25. J. Wu, K. Ota, J. Li, Z. Guan, Big data analysis-based secure cluster management for optimized control plane in software defined networks. IEEE Trans. Netw. Serv. Manag. **15**(1) (2018)
26. H. Sufiev, Y. Haddad, DCF: dynamic cluster flow architecture for SDN control plane, in *Proceedings of the 2017 IEEE International Conference on IEEE Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 8–10 January 2017, pp. 172–173
27. H. Sufiev, Y. Haddad, A dynamic load balancing architecture for SDN, in *Proceedings of the IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, Eilat, Israel, 16–18 November 2016, pp. 1–3
28. A. Abdelaziz, A.T. Fong, A. Gani, U. Garba, S. Khan, A. Akhunzada, et al., Distributed controller clustering in software defined networks. PLoS ONE **12**(4), e0174715 (2017)

Optical Character Recognition for Test Automation Using LabVIEW

Srinivas Perala, Ajay Roy, and Sandeep Ranjan

Abstract Every embedded product is getting updated with advanced features every year. Because of this reason, the embedded software verification became significant and critical. Testing all functionalities together is a big challenge, which takes a considerable amount of time. Test automation helps to fast up all large test scenarios. However, due to technical challenges, some of the advanced features cannot automate, for example, LCD image verification. LCD image verification is about verifying the expected output image on the LCD screen. Automotive infotainment is a common feature in all automotive-embedded software and has improved functionality to support all mobile OS like iOS and Android. These features need to monitor on the LCD screen as per the functionality. Image processing algorithms are practical to compare reference images and test images on the LCD screen. Nevertheless, the accuracy level of these algorithms is not as expected. This paper gives a solution to this problem statement. This paper presented a novel solution to compare test images and reference images using OCR techniques to get more accurate embedded software testing.

Keywords OCR · Test automation · LabVIEW

1 Introduction

Embedded software impacts the quality of a product and, in turn, most of the aspects of our society [1]. Its effectiveness is critical for many commercial and societal actions. As necessary, assuring the precision and quality of software systems and components becomes dominant. Embedded products (e.g., automotive, appliances,

S. Perala (✉) · A. Roy
Department of Electronics, LPU, Phagwara, Punjab 144411, India

A. Roy
e-mail: ajoy.22652@lpu.co.in

S. Ranjan
Apeejay Institute of Management and Engineering Technical Campus, Jalandhar, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_36

and medical devices) come with user interfaces where LCD, LED and other display devices are used to visualize the output for using them effectively. Users can select options from graphical displays based on the product features. All these aspects increase the product functionality along with the probability of bugs in software. The quality assurance (QA) team tests each feature in detail to improve the software quality. However, as the number of test cases increases, the QA team needs more time, which leads to delay in the product release. Manual testing involves observing the expected results by human testers and reporting the results. Test automation is converting manual observation into machine-attainable by writing a programming script [2]. Once the automation script is ready, we can execute the procedure "*n*" number of times, which helps speed up the regression test. An image processing algorithm makes it helpful to automate manual processes like validating the expected image on the display units.

Digital image processing is about applying mathematical analysis to each pixel to find the required information about the image. A digital image is a two-dimensional array considered with x horizontal coordinates and y vertical coordinates. At these coordinates, each combination has a value that is called pixel value. The digital images taken at different time frames have different intensity values. For this reason, identifying the similarity between the two images with an absolute difference is not accurate.

The optical character recognition OCR is a typical image processing algorithm practical to extract text from digital images. This algorithm combines the mechanical and electrical transfer of scanned images of handwritten, typewritten text into software text. An OCR algorithm is the extraction of features based on pattern recognition.

2 Literature Review

Image comparison transactions the resemblance between two images. It processes the identical image lists from a reference image gathering to calculate the similarity of each feature between the test image and each reference image and allocates weight to each feature to fuse these features via the "query-adaptive weighting method." It targets the test image and its neighborhood set selected from the repossession dataset as the test class. It utilizes the image-to-class similarity to re-rank the repossession results [3]. Inserting soft-tissue patterns and interventional instruments into the image can considerably change the performance of some resemblance rates formerly used on 2D and 3D image registration [4].

Support vector machine (SVM) used a supervised learning method to compare and identify biomedical image processing [5]. A similarity index for images represents both pixel intensity changes and geometric alterations. Classification of a large number of images can use these two parameters [6]. Magnetic resonance imaging (MRI) is widely used in medical investigations. Label and clustering unsupervised learning methods improve the accuracy of MRI.

Image reconstruction and finely tuned parameters also improve the accuracy [7, 8]. The adjustment interaction for the mathematical mutilation created by a wide-point lens introduces outspread twisting antiquities brought about by non-direct resampling [9]. The approval technique introduced is a significant advancement toward more nonexclusive recreations of biomechanically potential tissue distortions and evaluation of tissue movement recuperation utilizing non-rigid picture enrollment. It is a reason for improving and analyzing the changed non-rigid enlistment method for various clinical applications [10].

A novel method has been described to mark student attendance using image processing methods by comparing students' faces [11]. Machine learning algorithms have been used to generate test scripts from test cases using LabVIEW and Dspace test management software [12]. Natural language processing techniques are helpful to convert software requirements to test cases which is the most significant component of the agile verification and validation process [13].

Physical user interface simulation is helpful to test automation in case of changing physical hardware [14, 15]. The standard deviation-based algorithms give the similarity score between the test image and the reference image, which helps to calculate the difference [17]. Artificial neural networks have been employed to reduce the power consumption of LED street lights [16]. ANN helps to use the hardware in an energy-efficient way.

3 Methodology

A novel algorithm has been proposed to recognize optical characters for test automation. The algorithm extracts text from a selected region of interest of the test image. The OCR algorithm was enhanced with training fonts used in embedded vision software. This program has been tested on Automotive Infotainment software connected to Apple Care. The flow chart in Fig. 1 is based on the algorithm implementation. Image acquisition is the process of grabbing an image from a test setup using the industrial camera. The output of this step is a digital RGB image, as shown in Fig. 2. Once the image is acquired, the vision algorithm processes the text extraction. Initially, the image is converted to grayscale; then, the smoothing filter is applied with the local average to remove the noise. Here, the kernel size is 3*3.

This filter reduces the measure of the forced variety between a pixel and its next following pixel. Mean filtering replaces each pixel esteem in the picture with its neighbors' mean ("average") esteem, including itself. Figure 3 shows the output of the smoothing filter.

After the smoothing process, the algorithm considers the region of interest (ROI) gas per the requirement—the threshold configuration represented in Fig. 4. The auto linear mode is selected by applying the character's light–dark background—the spacing and size configuration represented in Fig. 5. Flow chart steps are implemented and shown in Fig. 6.

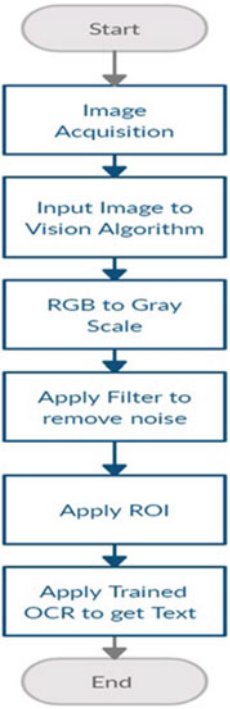


Fig. 1 Flow chart of the proposed algorithm

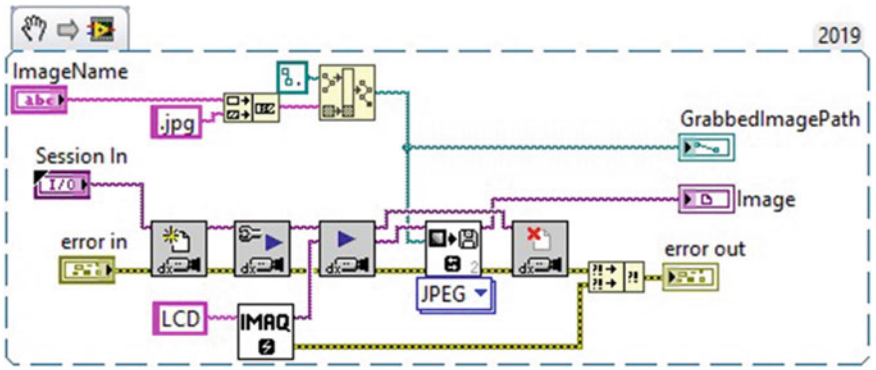


Fig. 2 Grabbing an image from the camera

4 Results

The results and the front panel are displayed in Fig. 7. The region of interest (ROI) has been highlighted with red color, which can easily track and check the results. Each

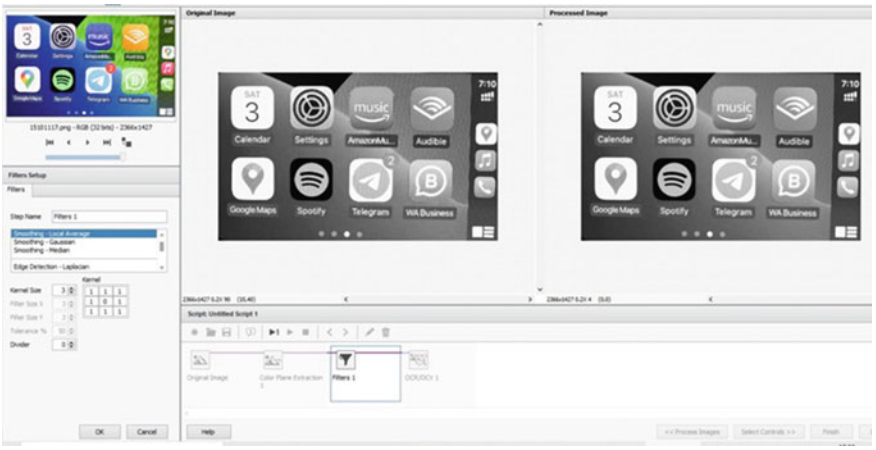


Fig. 3 Image processing with smoothing filter

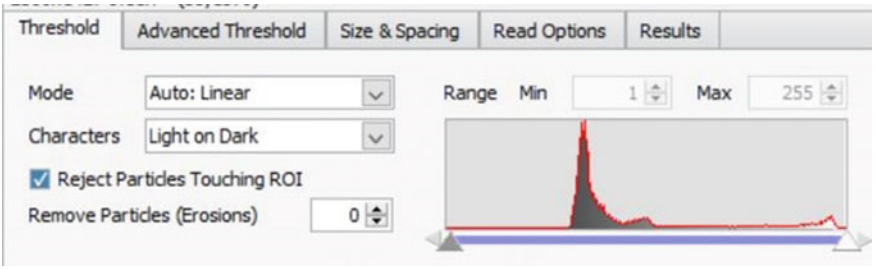


Fig. 4 OCR algorithm threshold configuration

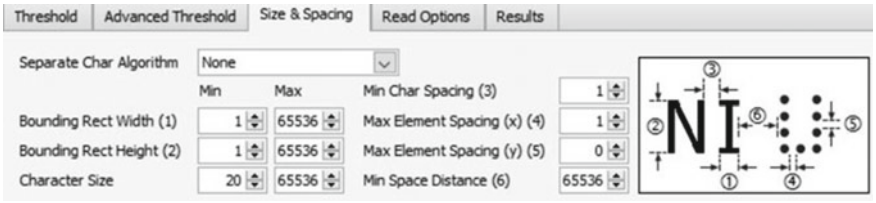


Fig. 5 Spacing and size configuration

character has a classification score and identification score. The algorithm extracted the text as "Settings" from the selected ROI and shown complete result analysis in Fig. 8. This OCR method showed more accuracy compared with the traditional pixel-to-pixel comparison method.

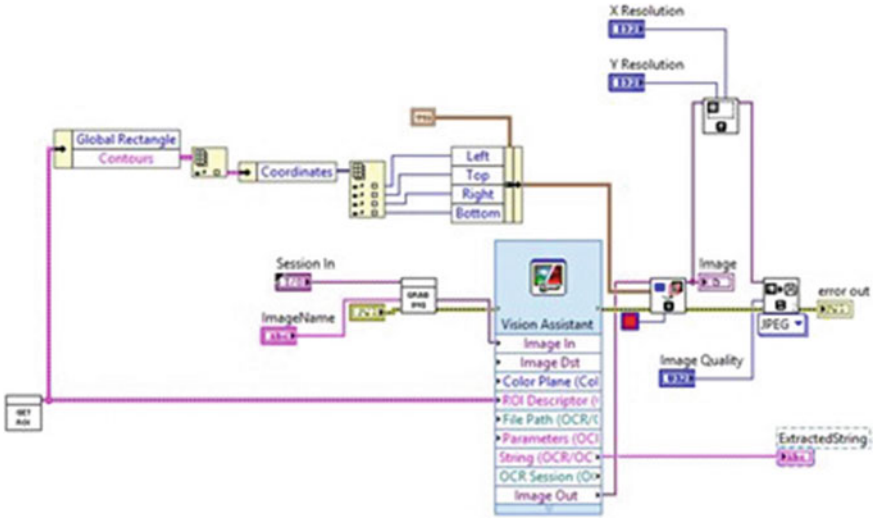


Fig. 6 Block diagram of code

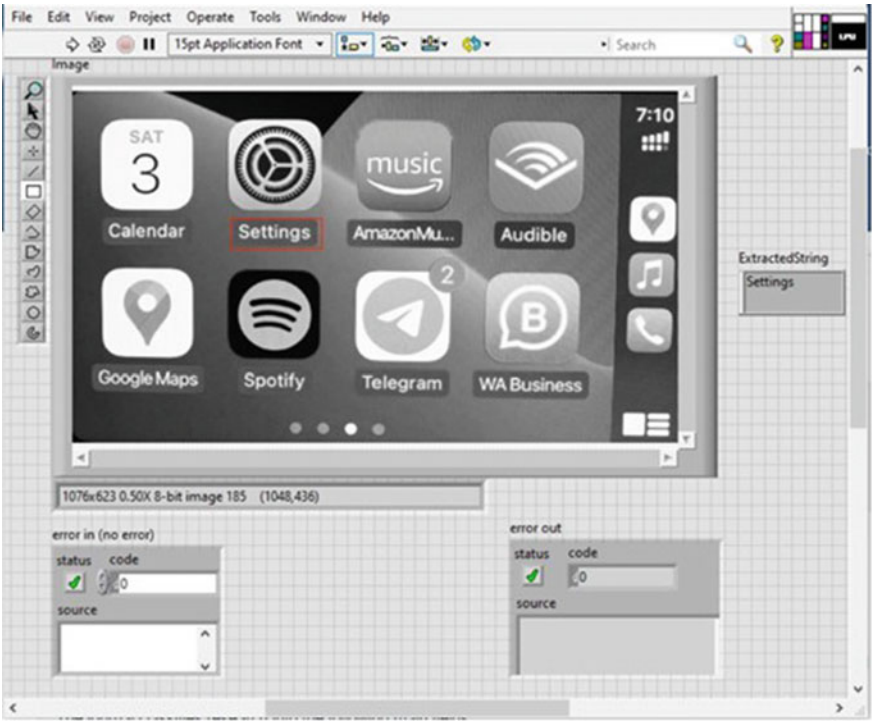


Fig. 7 Front panel with result1

Results …	S	e	t	t	i	n	g	s
Classification Score	803	951	997	998	794	997	995	804
Identification Score								
Left	652	695	737	761	791	810	850	896
Top	542	555	546	547	541	557	557	558
Width	38	36	22	23	9	33	37	31
Height	55	41	50	50	56	41	55	41

Fig. 8 Result1 details

5 Conclusion

This work improved the accuracy of test automation results for image comparison in embedded vision software. Instead of comparing pixels, extract the expected string or icon from the region of interest can give more accuracy. As increasing the features of infotainment in automotive and other industries, this algorithm helps to fast up the software testing. This traditional pixel comparing method and the OCR method combination gives better results based on the requirement. The novelty of this paper is adapting the OCR techniques in validating the embedded vision software.

References

1. D. Akdur, B. Say, O. Demirörs, Modeling cultures of the embedded software industry: feedback from the field. Softw. Syst. Model. **20**(2), 447–467 (2021)
2. K.K.D. Ramesh, G.K. Kumar, K. Swapna, D. Datta, S.S. Rajesh, A review of medical image segmentation algorithms, in *EAI Endorsed Transactions on Pervasive Health and Technology* (2021)
3. D. Hou, S. Wang, H. Xing, Query-adaptive remote sensing image retrieval based on image rank similarity and image-to-query class similarity. IEEE Access **8**, 116824–116839 (2020)
4. G.P. Penney, J. Weese, J.A. Little, P. Desmedt, D.L. Hill, A comparison of similarity measures for 2-D-3-D medical image registration. IEEE Trans. Med. Imaging **17**(4), 586–595 (1998)
5. M.M. Rahman, S.K. Antani, G.R. Thoma, A learning-based similarity fusion and filtering approach for biomedical image retrieval using SVM classification and relevance feedback. IEEE Trans. Inf Technol. Biomed. **15**(4), 640–646 (2011)
6. M.P. Sampat, Z. Wang, S. Gupta, A.C. Bovik, M.K. Markey, Complex wavelet structural similarity: a new image similarity index. IEEE Trans. Image Process. **18**(11), 2385–2401 (2009)
7. X. Bai, Y. Zhang, H. Liu, Z. Chen, Similarity measure-based possibilistic FCM with label information for brain MRI segmentation. IEEE Trans. Cybern. **49**(7), 2618–2630 (2019)
8. A. Bustin, D. Voilliot, A. Menini, J. Felblinger, C. de Chillou, D. Burschka, F. Odille, et al., Isotropic reconstruction of MR images using 3D patch-based self-similarity learning. IEEE Trans. Med. Imag. **37**(8), 1932–1942 (2018)
9. D. Kim, J. Park, J. Jung, T. Kim, J. Paik, Lens distortion correction and enhancement based on local self-similarity for high-quality consumer imaging systems. IEEE Trans. Consum. Electron. **60**(1), 18–22 (2014)
10. J.A. Schnabel, C. Tanner, A.D. Castellano-Smith, A. Degenhard, M.O. Leach, D.R. Hose, D.J. Hawkes, Validation of nonrigid image registration using finite-element methods: application to breast MR images. IEEE Trans. Med. Imaging **22**(2), 238–247 (2003)
11. A. Kumar, P.K. Verma, S. Perala, P.R. Chadha, Automatic attendance visual by programming language lab view, in *IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems* (2016), pp. 1–5

12. S. Perala, A. Roy, A novel framework design for test case authoring and auto test scripts generation. Turkish J. Comput. Math. Educ. **12**(6), 1479–1487 (2021)
13. S. Perala, A. Roy, A review on test automation for test cases generation using NLP techniques. Turkish J. Comput. Math. Educ. **12**(6), 1488–1491 (2021)
14. S. Perala, A. Roy, Koushik, A novel method of test automation for testing embedded software. Think India J. **22**(37) (2019)
15. J.I.Z. Chen, L.T. Yeh, Analysis of the impact of mechanical deformation on strawberries harvested from the farm. J. ISMAC **3**, 166–172 (2020)
16. S. Perala, A. Roy, S. Ranjan, Image processing algorithm to compare test image with reference image to validate embedded software of display application, in *2 Days International Conference on Recent Innovations in Science, Engineering, Humanities and Management (IEI Chandigarh)* (2021), pp. 63–68
17. S. Smys, A. Basar, H. Wang, Artificial neural network based power management for smart street lighting systems. J. Artif. Intell. **2**(01), 42–52 (2020)

Analyzing and Detecting Advanced Persistent Threat Using Machine Learning Methodology

Vijaya Chandra Jadala, Sai Kiran Pasupuleti, CH. M. H. Sai Baba, S. Hrushikesava Raju, and N. Ravinder

Abstract Advanced persistent threat is a well-organized targeted attack, which turns from passive attack to active based on spear phishing. Due to fast technology growth and vast internet usage, network exploitation and spear phishing emails became common gateway to risks, threats and vulnerabilities in areas of data privacy and security. Generally, security breaches lead to vulnerabilities which transforms to threats and zero-day exploits. Most sophisticated threats, risks and vulnerabilities resides for a long period of time in an intangible stage, even creates a backdoor in order that it can be hide and seek as per its requirement. We did a strong literature review, regarding history and attack procedure of advanced persistent threat. Among different attack procedures adopted by the APTs, for experimental analysis, we observed regarding the spear phishing emails. Data science is a correctional that communicates to huge quantity of data, which is processed, mined and understandable by machines based on statistical methods and data analysis. We discussed the machine learning classification methodologies on datasets. We used K-fold cross validation to evaluate the model, as the dataset consists of attributes and class where the binary-based classification can be implemented to overcome the hyperparameters. The proposed algorithmic representation is different from the existing algorithms where we used the ensemble learning methods, which is a blend of multiple algorithmic approaches. In experimental setup, we used the Anaconda navigator which includes the open-source Python packages such as NumPy and Pandas. Where Pandas is used to extract the dataset from the repository called UCI that is spambase.csv from machine learning

V. C. Jadala (✉) · S. K. Pasupuleti · CH. M. H. Sai Baba · S. Hrushikesava Raju · N. Ravinder
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur 522502, India
e-mail: drvijayachandra@kluniversity.in

S. K. Pasupuleti
e-mail: psaikiran@kluniversity.in

CH. M. H. Sai Baba
e-mail: saibaba.ch77@kluniversity.in

N. Ravinder
e-mail: ravindernellutla@kluniversity.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_37

datasets. We use Jupyter notebook for coding and used scikit-learn. The scikit-learn is used to interact with supervised and unsupervised machine learning algorithms. The result analysis given in graphical representation using the libraries such as Matplotlib and Seaborn.

Keywords Advanced persistent threat · Machine learning · Data science · Risks · Threats · Zero-day Exploits · Python · Pandas · Algorithms

1 Introduction

Security analysis of large quantity of confidential data is the major research area in modern Internet-based cloud storage environment. All over the world, cyber security played a major role to address solutions to different risks, threats, vulnerabilities, and attacks. The researchers explored defense mechanisms related to different threats majorly polymorphic threats, blended threats, and APTs. The furthermost familiar type of threats are related to the worms, trojans, bots, viruses and keyloggers. The Cyber Threat Taxonomy consists of different exploitation methods such as information gathering, intrusion attempts, frauds, malwares and attacks. Information gathering is the primary step involved in any cyberattack, generally known as passive attack done through scanning, sniffing and social engineering. The different advanced persistent threat samples are grouped and established supported by characteristics where APT01 to APT41. The frequently occurred advanced threat is plugX which is an attachment categorical APT [1].

The significant research areas are related to data security, data mining, cloud security, machine learning and data science. A dataset can be viewed as a collection of data objects, which are often also called as records, points, vectors, patterns, events, cases, samples, observations or entities. The steps used for data preprocessing usually fall into different categories such as selecting data objects and attributes for the analysis, creating or changing the attributes. The machine learning algorithms are furnished to dataset, and the accuracy is measured. The precision of the methodology calculates how far the machine learning algorithm correctly predicts that identifies and provides the prediction in percentage based on the samples. It even shows how correctly prediction is identified over existing samples [2].

Advanced persistent threat is one of the fast-growing information and data security threat challenging big data and cloud storage today. APT uses unique attack processes, practices and procedures followed by attackers; the major attack technologies that support advanced threat are spear phishing emails. It is an attack technique adopted by intruder to reach goal by sending an email with malicious link or malware attachment [3]. Artificial intelligence-based machine learning techniques are implemented on datasets to classify the spam and ham emails. Intelligent data analysis plays a major role in detecting risks, threats, and attacks. The classification is a mechanism differentiated based on attributes, abilities and classes of feedbacks

and clarification which specifies the class that the data object belong to two classes or more classes.

2 Related Work

The researchers started exploring different mechanisms to detect advanced persistent threats. In 2012, Wei Wang et al. proposed a framework based on signature, policy and profile for detecting APTs, and later in the same year, Johannes de Vries et al. proposed a roadmap based on artificial intelligence and machine learning based on data analysis for detecting APT based on signatures and behaviors. A step-by-step procedure is adopted to detect the APT based on the detection rules such as context-based detection established on key issues associated to APT activities. In next year, Tarique Mustafa et al. focused on data leak prevention solutions for APT management, where intelligence-based approach is proposed to classify different risks, threats and attacks [4].

In 2014, APT detection procedure was proposed by Parth Bhatt et al. toward a framework to detect multi-stage APT attacks; the defense mechanism adopted is constructed on creating a blacklist and implementation of blocking procedure based on blacklist. The malware will be identified based on the signature identification techniques. In the same year, Andrew Vance et al. proposed a mechanism based on flow analysis of APTs in cloud computing for network-based anomaly detection. The flow-based collection and detection anomaly is constructed using signature. Patterns, procedures, policies, profiles, protocols, signatures and behaviors perform a major role in identification of threats [5].

In 2015, J. Vukalovie et al. focused on prevention procedures such as user education that is general awareness regarding risks, threats, vulnerabilities and attacks. Even a user should be educated regarding precaution measures to be adopted related to APTs as precaution is better than cure, implementing access based on authentication, and authorization security mechanism is major step in defending APT. They concentrated on usage policies, controlling external media, protecting vulnerable data, managing end-point security, implementing network access control, blocking high-risk application, blocking known malware servers, and analyzing security breaches. In the same year Pengfei Hu et al. proposed a dynamic defense strategy based on artificial intelligence mechanism such as MAX and MIN two player game. A general framework of two-layer differential game between defender and attacker [6].

In 2018, W. Matsuda et al. proposed machine learning mechanism for detecting advanced persistent threat attacks against active directory. Mainly concentrated on the threat detection methodologies based on the prediction mechanism. Data security analysis predicts different detection technologies that are used for detecting attack activities from event logs with outlier detection that means odd man out procedure using unsupervised learning mechanism based on event logs related to process creation [7].

In 2020, Liu Yan et al. proposed a Web-based defense system for the advanced persistent threat model using the concept of self-translation machine mechanism with attention. The model proposed is based on unsupervised anomaly detection algorithm which is majorly used to identify unusual data points. The identification of security vulnerabilities is the major task discussed based on the artificial intelligence-based game technology [8]. They proposed unsupervised learning algorithms for identification of Web-based attacks into self-translation mechanism.

In 2021, different game-based APT detection methodologies are identified, whereas Talal Halabi et al. proposed Stackelberg game, whereas in the same year, Wen Tian et al. proposed theoretic game. The Stackelberg game is meant for the standard analysis of data based on the sequence of steps. Where the first player is called as leader and remaining followers, the other technology is of theoretic game where the concept is in between attacker and defender. The game concept is based on set of rules and mechanisms based on the outcomes [9].

3 Advanced Persistent Attack Procedure

The steps involved in advanced persistent attack procedure are based on goal, that is, target will be identified as the primary step, where the target may be an individual or an organization, implementation of targeted attack using spear phishing emails, organization mapping is to explore the organization and to identify vulnerabilities, privilege escalation for high-end privileges, back door creation so that the attacker can be in and out as per the requirement and to hide from anti-virus software and intrusion detection system. Stealth fighters that use the self-destruction mechanism to hide from security monitoring mechanisms and re-engineering mechanism for reconstruction of attack procedures. Finally, the attacker uses command and control mechanism to operate the target remotely and to stay for a long period without any identity [10] (Fig. 1).

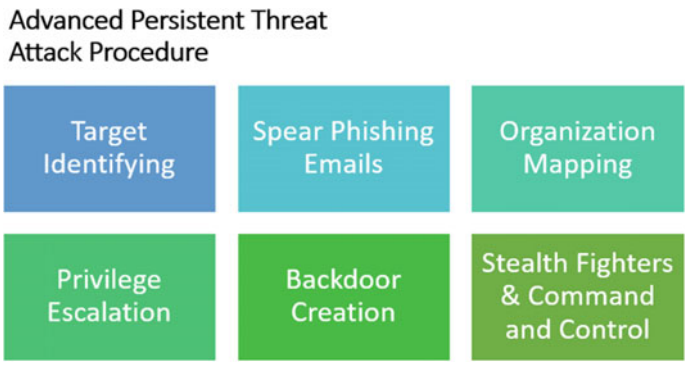


Fig. 1 The steps in advanced persistent threat attack procedure

The intruder selects the target; after identifying the target, he collects data based on social engineering techniques, where the most confidential data will be collected such as victims name, mobile number, email ID, residential address, workplace details, banking, pan card, Aadhar and credit card details. Based on the gathered information, intruder sends spear phishing email, that is, a spam email which seems to be a legitimate email. The email consists of malware attachment or malicious uniform resource locator which relinks to phishing websites. On receiving such emails, the victim unknowingly by downloading the attachment gives a way to intruder to enter the organization [11]. Once the intruder enters the organization, he starts exploring himself and maps the total organization for identification of vulnerabilities. Then, he tries to get the high-end privileges, even creates a backdoor to escape from general monitoring systems and intrusion detection and prevention system. By using backdoor, the intruder can be in and out as per the requirement. Stealth fighters are used to be in undetectable stage, when an intruder reaches the final stage, that is, command and control the total organization will be under the control of intruder [12].

4 Datasets

A dataset is a collection of data objects, also known as records, points, vectors, patterns, occurrences, instances, samples, observations or entities. Data preprocessing is a data mining technique used to turn raw data into a usable and efficient format [13]. Data preprocessing refers to the steps used to prepare data for data mining. The steps required for data preprocessing are often divided into two categories:

1. Choosing the data items and properties for analysis.
2. Adding/changing characteristics.

Data preprocessing approaches include data cleansing, data transformation and data reduction. There could be a lot of useless and missing information in the data. Data cleansing is performed in order to handle this section.

It involves handling of missing data, noisy data etc. Data transformation is the procedure that is part of mining process that transfers data into the appropriate forms using normalization, attribute selection, discretization and concept hierarchy generation. The important concept in the data preprocessing is data reduction, where the redundant data will be reduced to increase the storage efficiency and reduce data storage and analysis costs. Sampling is a commonly used approach for selecting a subset of the data objects to be analyzed. The most important component of sampling is to use a representative sample. A sample is considered representative if it has the same property (of interest) as the original set of data. If the mean (average) of the data items is the property of interest, then a sample is considered representative, if its mean is close with that of the original data [14].

After selecting the data, it must be preprocessed using the following steps:

1. Preparing the data for ML by formatting it (structured format)

2. Cleaning up the data to remove any missing variables
3. Increasing the sampling rate of the data to reduce algorithm running times and memory requirements.

5 Experimental Analysis

In experimental setup, we used the Anaconda navigator which includes the open-source Python packages such as NumPy and Pandas. A spam-base dataset [6] consists of 1813 spam emails means junk emails and 2788 ham emails means legitimate emails, and the total instances are 4601. The machine learning methodology is based on the training and testing, where in general, the training will be 70% and the testing will be of 30% [15] (Fig. 2).

Experimental and graphical analysis is given by using the spam email database that is spambase.csv, where the ham are 2788 instances, and the spam are 1813 instances identified through the graph as above.

Day by day, there is a large amount of data transactions through emails; emails are the cheap and latest source of data transmission in very less time, which increases usage of emails drastically, if huge quantity mails of spam have produced a demand for trustworthy anti-spam filters. Nowadays, it is usual a user receives many emails daily where almost 92% of emails are promotions, social and spam. These emails are unsolicited malware attachments or embedded with malware URLs. Email classification of the spam and ham which is of filtering can be done with a binary classification. Data science produces the statistical methods for huge quantity of data, the data will be gathered based on attributes, the class term is the last column which classifies based on true or false, which can be solved by machine learning algorithms and implementation is done using different software's and tools related to data science and machine learning. The most common used tool "Waikato Environment for Knowledge Analysis (Weka)" is used for the data analysis and predictive analysis based on the machine learning algorithms and used for implementation of

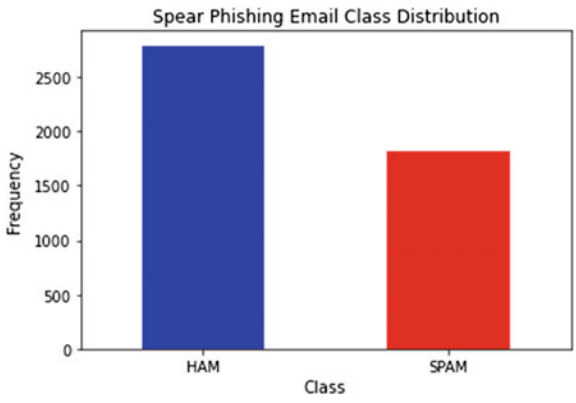


Fig. 2 The dataset class distribution of spam and ham emails

data mining techniques. The extension Meka is used for multi-label classification, whereas generally Weka for binary classification or trinary classification [16].

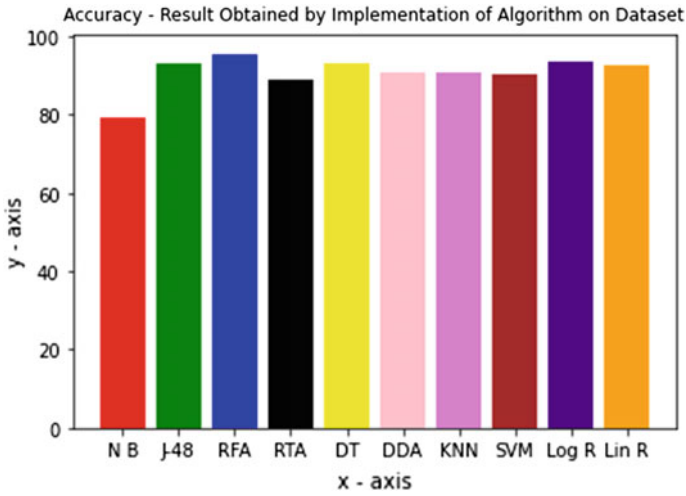
Weka provides many machine learning algorithms related to supervised and unsupervised learning algorithms. It even supports the data preprocessing procedures. For the experimental purpose, the total number of instances loaded are 4601 through spambase dataset. The different machine learning algorithms that are implemented on the dataset are naive Bayes algorithm, J-48 algorithm, random tree algorithm, random forest algorithm, decision tree algorithm, support vector machine algorithm, k nearest neighbor algorithm and related to regression are linear regression and logistic regression [17].

6 Result Analysis

For the experimental purpose, we took above ten machine learning algorithms and implemented on the spambase dataset. The dataset consists of the different attributes and a class, where the class consists of 0 or 1, the total number of instances is 4601, and the time taken for the execution is calculated. Different evaluation metrices are measured such as correctly classified instances, incorrectly classified instances, kappa statistics, mean absolute error, root mean squared error, relative absolute error and root relative squared error [18].

The different machine learning algorithms are implemented on the dataset called spambase. The experimental results obtained are given in the tabular form where the metrices are time taken to implement the algorithm. The correctly classified instances, incorrectly classified instances and finally the accuracy are measured based on correctly classified instances and the total number of instances. The other metrices are kappa statistics, which is used to calculate the interior reliability of the correlation statistics. Mean absolute error is used to calculate the absolute difference between the actual values and predicted values. Root mean squared error, relative absolute error and root relative squared error are the other factors that influence in measuring the error rate and accuracy [19].

The result analysis is given based on the result obtained by implementation of different algorithms on the dataset. The obtained parameters of result are tabulated, and graphical analysis is given using bar chart. The first graph is based on the time taken to implement the algorithm on the dataset, where decision tree algorithm consumed least time in implementation of algorithm that is 0.04 s. As the decision tree takes only one iteration of branch, that is, binary classification verifying true or false [20].



Next bar chart explains the correctly classified instances, where random forest algorithm has given the maximum correctly classification, with highest correctly prediction results when compared to other algorithms with 4394 instances among 4601 instances. Next graph is for incorrect classification of instances naïve Bayes produced highest number of incorrect classification instances, so it stood least at the comparison of all algorithms. Finally, graphical analysis for accuracy is given by calculating result obtained by implementation of algorithm on dataset, where we found random forest algorithm with highest percentage of accuracy.

7 Future Work

Researchers all over the world are working on the major research concepts such as cyber security, machine learning and data science. The research issues are based on the risks, threats, vulnerabilities and attacks. The future of security and safety of online and cloud computing depends on the defense procedures adopted by the software industries. Researchers and defenders should concentrate on the defending procedures to deploy machine learning to acquire and prevent malicious activities on the Internet-based online and cloud computing. We discussed and implemented different machine learning methodologies that are used to secure and calculate the rate of data privacy and security. It is also used to detect the anomalies for data protection.

References

1. F. Yang, An implementation of naive bayes classifier, in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, 2018, pp. 301–306. https://doi.org/10.1109/CSCI46756.2018.00065
2. J.V. Chandra, N. Challa, M.A. Hussain, Data and information storage security from advanced persistent attack in cloud computing. Int. J. Appl. Eng. Res. **9**(20), 7755–7768 (2014)
3. S.H. Raju, M.N. Rao, N. Sudheer, P. Kavitharani, Visual safe road travel app over google maps about the traffic and external conditions. Int. J. Eng. Technol. (UAE). ISSN: 2227–524X. https://doi.org/10.14419/ijet.v7i2.32.15697
4. J. Vijaya, C. Narasimham, C. Sai, K. Pasupuleti, Intelligence based defense system to protect from advanced persistent threat by means of social engineering on social cloud platform. Indian J. Sci. Technol. **8**(28), (2015)
5. R. Krishnasrija, K.P. Rani, P.S. Kiran, P. Dharani, B.S.C. Goud, V. Sushmitha, Smart secure student bag pack, in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2021, pp. 1885–1890. https://doi.org/10.1109/ICICCS51141.2021.9432256
6. Weblink: UCI machine learning repository: Spambase data set: https://archive.ics.uci.edu/ml/datasets/Spambase
7. D. Sivaganesan, A data driven trust mechanism based on blockchain in IoT sensor networks for detection and mitigation of attacks. J Trends Comput Sci Smart Technol (TCSST) **3**(01), 59–69 (2021)
8. M. Adithya, P.G. Scholar, B. Shanthini, Security analysis and preserving block-level data DE-duplication in cloud storage services. J Trends Comput. Sci. Smart Technol. (TCSST) **2**(02), 120–126 (2020)
9. A.K. Manekar, G. Pradeepini,Cloud based big data analytics a review, in *2015 International Conference on Computational Intelligence and Communication Networks (CICN),* 2015, pp. 785–788https://doi.org/10.1109/CICN.2015.160
10. CH.M.H. Saibaba, S.H. Raju, M.V.B.T. Santhi, S. Dorababu, S.F. Waris, International currency translator using IoT for shopping, in *IOP Conference Series:Materials Science and Engineering*, vol. 981, no. 4, Dec (2020)
11. J.V. Chandra, N. Challa, S.K. Pasupuleti, A practical approach to Email spam filters to protect data from advanced persistent threat, in *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT),* Nagercoil, 2016, pp. 1–5
12. CH.M.H. Saibaba, V. Uday Kumar, K. Praveenkumar, D. Bhattacharyya, A new privacy approach using graph-EMD. Int. J. Softw. Eng. Appl. **10**(7), 11–16 (2016)
13. S.L. Bangare, G. Pradeepini, S.T. Patil,Brain tumor classification using mixed method approach, in *2017 International Conference on Information Communication and Embedded Systems (ICICES)*, 2017, pp. 1–4https://doi.org/10.1109/ICICES.2017.8070748
14. G. Pradeepini, S. Jyothi, Tree-based incremental association rule mining without candidate itemset generation. Trendz Inf. Sci. Comput. (TISC2010). 78–81 (2010). https://doi.org/10.1109/TISC.2010.5714603
15. S. Sharma, J. Agrawal, S. Agarwal, S. Sharma, Machine learning techniques for data mining: a survey, in *2013 IEEE International Conference on Computational Intelligence & Computing Research, Enathi*, 2013, pp. 1–6
16. J.V. Chandra, N. Challa, S.K. Pasupuleti, Advanced persistent threat defense system using self-destructive mechanism for cloud security, in *2016 IEEE International Conference on Engineering and Technology (ICETECH), Coimbatore*, 2016, pp. 7–11
17. L. Greeshma, G. Pradeepini, Unique constraint frequent item set mining, in *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 2016, pp. 68–72. https://doi.org/10.1109/IACC.2016.23

18. J.V. Chandra, N. Challa, S.K. Pasupuletti, Machine learning framework to analyze against spear phishing, Int. J. Innov. Technol. Exploring Eng. (IJITEE). **8**(12), (2019). ISSN: 2278–3075
19. G. Ranganathan, Real time anomaly detection techniques using PySpark framework, J. Artif. Intell. Capsule Netw. **02**(01), 20–30
20. J.S. Manoharan, P. Muralidharan, G. Jayaseelan, A Cloud FCM based intrusion detection model in a MANET environment. Int J Res Dev Technol **7**(8), 812–817 (2017)

Software-Defined Networking Security System Using Machine Learning Algorithms and Entropy-Based Features

Shankaraiah and S. Shashank

Abstract Software-defined networking (SDN) is an exceptional technology among the recent trends in IT. The static quality of the classical network architecture design deteriorates to the demands of modernized corporation IT. SDN technology does not replace existing networks but adds benefits and enhanced features like traffic programmability, pliability, granular security, operational control, reduced hardware footprint, unique network policy, lower operating costs, and improved uptime. SDN-based networks can have enhanced security systems that would detect and prevent cyberattacks. These applications will be capable of blocking malicious intruders before they invade the critical regions of the network. Even though the SDN networks try to provide a more secure network there will be some security issues in the configurations of the system. One of the crucial security concerns is that the SDN system has distributed denial of service (DDoS). The implementation of machine learning applications to automate the network can be doable through programmable and ascendable networking solutions proposed by SDN. Using the SDN solution, it is possible to analyze the network traffic and allocate the resources, which is a main concern in the classical method. This paper proposes the implementation of machine learning algorithms like decision tree (DT), random forest (RF), support vector machine (SVM), and multilayer perceptron (MLP) with the desire of detecting DDoS attacks. Another approach to recognize DDoS attacks is done by using entropy-based features. Also a firewall system is created in an SDN simulated environment to provide a more secure SDN network.

Keywords Software-defined network · Distributed denial of service · Machine learning · Decision tree · Random forest · Support vector machine · Multilayer perceptron

Shankaraiah · S. Shashank (✉)
Department of Electronics and Communication Engineering, JSS Science and Technology University, Mysuru, India

Shankaraiah
e-mail: Shankaraiah@sjce.ac.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_38

1 Introduction

Modern IT needs networking technology that makes the network accessible and frisky. SDN is a network architecture that provides networks newer programmability and flaccidity by bifurcating the control layer from the data layer. The use of SDN delegates network structure more adequately and enhances network performance and monitoring. It employs a centralized SDN controller to provide software-based network services. The SDN architecture shown in Fig. 1 comprises three layers. They are the infrastructure layer, the control layer, and the application layer.

The application layer also called as application plane consists of the applications that run on the network. It comprises broader network composition, monitoring, and administration actions across all layers of the networking heap. The control layer, also known as the control plane, characterizes the traffic routing, network topology and shows the centralized SDN controller program that operates just like the intellect of the SDN. The controller lies on a server and administers programs and also the flow of traffic throughout the network. The infrastructure layer is otherwise called the data plane physically controls the traffic supported on the configurations delivered from the control plane. It comprises routers, switches, and the supported physical hardware. For connection between the layers, SDN adopts northbound application program interfaces (APIs) and southbound application program interfaces. The northbound application program interface connects between the application and the control layers. The southbound application program interface connects between the infrastructure and control layers. A controller in SDN is a program function that implements a centralized aspect of and management over the integrated network. Network

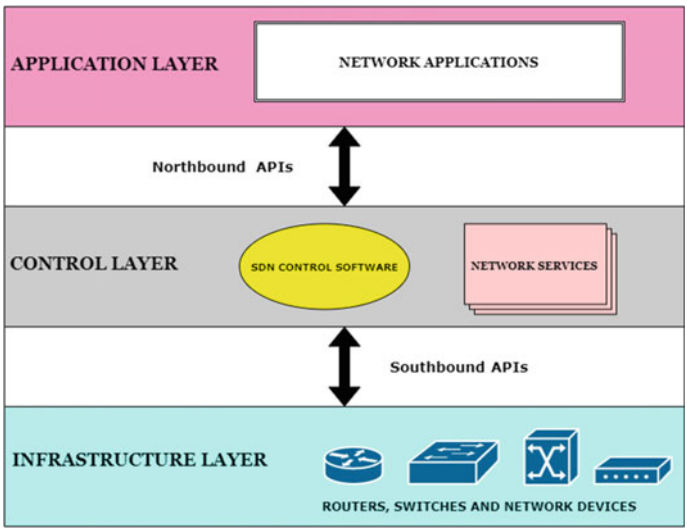


Fig. 1 SDN architecture

managers operate the controller to administer and the operation of elemental infrastructure's forwarding plane controlling the traffic. Also the controller implements protocols that guide network management. Network administrators create protocols that are enforced constantly on multiple nodes in the network. The network protocols are regulations that are employed to traffic, which decides the degree of access it has to the network, the resources it is permitted, or what priority it is selected.

1.1 Problem Statement

Despite the phenomenal advance of SDN in many companies, the control layer partitions from the elemental infrastructure layer will broaden the possibility of jeopardizing the SDN environment to many risks and threads. So [1] the SDN technologies are susceptible to distinctive network attacks, including volumetric attacks like DDoS attacks and service-specific attacks. Several techniques are developed to recognize and alleviate DDoS attacks in the SDN network. These defense methods are based on the gathering of the flow components like frequency of IP addresses, the average number of packets, and average number of bytes from packets received. In an instance of malevolent traffic, the amount of the observed packets will be greater than the preceding delineate brink. However, the existing approaches do not lack large processing capabilities to supervise and control the flow of data but they cannot defend the network for prolonged life. The breakneck advance in the achieved data and the enormous amount of packets that require to be examined and labeled in problem-solving time is the primary deficiencies in the prevailing scheme. Incurring the affinity of benevolent traffic and malicious packets, the result turns into an arduous exercise to identify malevolent traffic. The assailant can effortlessly customize the packet header to appear as a benign traffic environment and delude the security structure. DDoS attacks can easily breach the working of the victim server. The high efficiency of DDoS attack and their ability to disrupt normal network functions was the main motivation for taking this project.

1.2 Research Objective

The primary ambition of this research is to establish a software-defined networking security system to identify and prevents DDoS attacks using machine learning (ML) methods and entropy-based features and also create a firewall to prevent malicious attacks. Machine learning can bring better efficiency, more dynamic, and newer agile keys for SDN security, management, and development. For safeguarding network security, [2] the detection of DDoS attacks is extremely crucial that helps to take the obligatory measures promptly. The operation developed from flow data [3] of SDN with the ML-based DDoS attack detection techniques integrated into SDN

architecture results drives to an autonomy network that can learn and act. This detection method can be employed in social multimedia [4] security systems in SDN environment.

2 Literature Review

Dehkordi et al. [5] proposed an advanced approach that comprises the three collectors, entropy-based and analysis categories. The empirical outcomes achieved by employing the datasets such as CTU-13, UNB-ISCX, and ISOT imply that the approach surpasses its doppelganger in terms of accuracy for the disclosure of DDoS invasions in SDN. Xu and Liu [6] examined employing SDN for recognizing DDoS interventions by catching the flow quantity factor and flow rate asymmetry factor. Yan et al. [7] applied distinctive characteristics to recognize if an intervention has appeared. As there exist many main parameters in detecting DDoS invasions, the momentous matter is concerned with how these specifications are determined. Jisa David et al. [8] put forward a DDoS intervention disclosure structure that is based on rapid entropy adopting flow-based scrutiny. Their proposed method shows better accuracy of detection. They analyze network freight and compute the rapid entropy of requests per outflow. The method submitted by Seyed et al. [9] are based on the entropy juxtaposition of ensuing packet patterns to determine differences in their arbitrariness. The proposed method extends the recent work done by Kia [10], which is dependent on an entropy alteration of several parameters. The proposed method is a lightweight DDoS attack detection at its early stage. Sivatha Sindhu et al. [11] proposed a neural decision tree for element collection and analysis. The features used were subtlety and particularity for the conduct appraisal. Keisuke Kato et al. [12] proposed an inventive DDoS invasion disclosure scheme employing packet scrutiny and an SVM. The detection system used SVM with a radio basis function (RBF) neural networks. Experiments were performed using CAIDA DDoS attack 2007 dataset. SPHINX [13] is a framework proposed to detect interventions in SDN in actual time with low-performance overheads. It can detect both acknowledged and probably anonymous interventions on network topology. It is mainly based on an approximation of a real network into a flow graph. It uses these flowcharts to expose surveillance intimidations in the network topology. DDoS attack detection and prevention based on ensemble classifier RF proposed by Alpna et al. [14] use a combination of classifiers to improve the performance of their model. Experimental results were conducted on the UCLA dataset. The results show high accuracy with minimum error. From the survey, few methods gave a lightweight detection system at an early period but our proposed method gives a strong detection system as the output. Most of the methods in the survey have a single machine learning algorithm in implementation. In our approach, we have four ML algorithms that are suitable for considering according to the environment of the system that gives better precision of detection.

3 Proposed Modeling

The network emulator we use is Mininet [15] which builds a network that consists of switches, controllers, virtual hosts, and links. Scapy is used to generate packets. It is a significant interactive packet manipulation program and is applied for the generation of packets, detecting, examining, producing packets, and assaulting. It is adopted for the generation of UDP packets and deluding the origin IP address of the packets. We use POX as an SDN controller using the Python programming language which provides an infrastructure for communicating with SDN switches using OpenFlow protocol. OpenFlow manages the switches in the network and allows the controller to handle the outflow of packets through the network. The function of SDN controllers like pox is to authorize users to compose their applications that run the controller as an abstraction layer between network applications and the network equipment. Figure 2 shows a typical example of network topology using Mininet.

3.1 DDoS Attacks Detection Using Machine Learning Algorithm

Nowadays, many software tools can be used to identify DDoS attacks. But malignant traffic can be in any condition. The assailants alter their invasion methods frequently. So there is a demand to gain from the incident. This can be achieved by machine learning. The ML method can be used to analyze traffic to recognize an attack pattern. The advanced network features like a global picture of the network make it easier for us to mitigate and detect DDoS attacks. The methodology that we use is machine

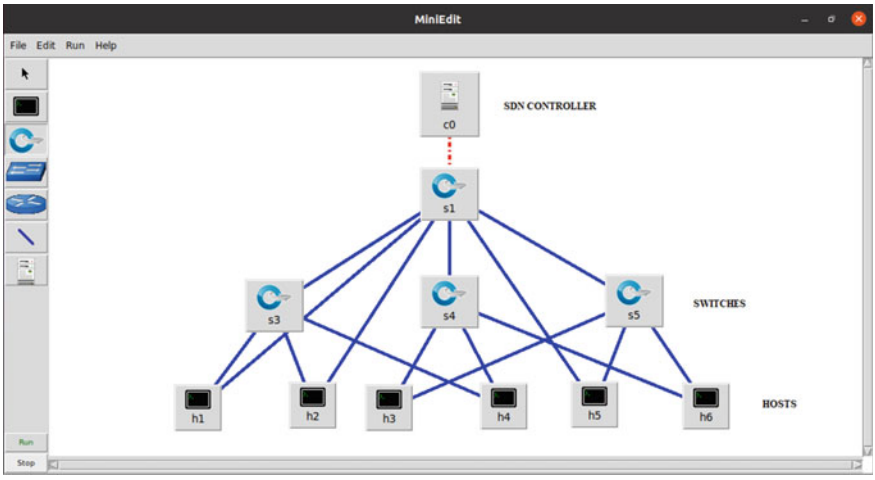


Fig. 2 A typical topology in mininet

learning. The purpose of ML is to extricate between benign traffic and malicious traffic.

Figure 3 shows the flowchart of the proposed model. The flow from start to an initialize traffic flow represents the traffic that is initialized in the simulation. Two types of traffic are initialized, the first one is the normal one which is the random number of packets and the second one is malicious (DDoS attack) with random numbers of IPs. After traffic initialization, the system will extract the features from both of the traffic. If the features are not relevant, they are deleted from the dataset, and if the features are relevant, then they will be updated in the dataset which is the main part of this system used for the ML model. It also separates the data by labeling them according to the features that have been extracted in the previous step. The labels will be 0 (normal) and 1 (malicious). The dataset will be fed to an already trained ML model which just classifies the coming dataset with its tested data and based on that comparison it will generate the result as its DDoS attack or not and end.

Supervised learning is a method in ML where sample labeled data are given to the ML system to train it, and from that basis, the output is predicted. DT, RF, SVM, and MLP are the types of supervised ML that are applied in the approach. DT is an ML where data are consistently classified according to parameters. Decision nodes

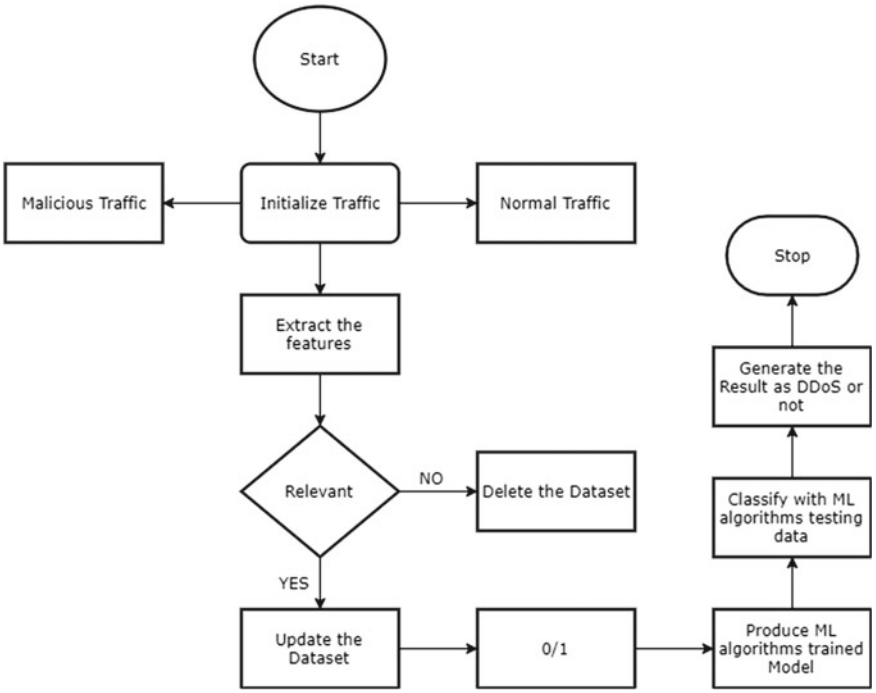


Fig. 3 Methodology for DDoS attacks detection using machine learning algorithm

and leaves are the two entities in the tree. RF is derived from DT. SVM is used for classification and regression challenges. MLP uses backpropagation for training.

3.2 DDoS Attack Detection Using Entropy-Based Features

The main motive for dealing with entropy is its capacity for measuring haphazardness in a network. The bigger the haphazardness the more noteworthy is the entropy and the other way around. So a DDoS attack will be transpired whenever the entropy merit is less than a threshold merit. Figure 4 shows the proposed methodology. First, we make a count of 50 packets in the window and then calculate the entropy and compare it with a threshold that we set and make a count of consecutive entropy values lower than a threshold. If this count reaches more than five, then we can say that DDoS had occurred otherwise not. Here, the entropy is detected with the help of two factors which are destination IP and the number of times it frequented.

The entropy is calculated by the following:

- Window size = 50
- Pi = Probability of a destination IP occurred in the window
- $Pi = Xi/n$
- X = the number of event in the set
- n = window size.
- Entropy $H = -$ sum of all (Pi) $\log(Pi)$ where, i is from 0 to n

After calculating entropy value, it is stored in entropy dictionary which is used in the L3 learning module of pox to compare and say whether a DDoS attack had occurred or not. A host first launches the traffic and then other hosts starts attacking in the network. Entropy is calculated after traffic is launched and also after other hosts attacks. A threshold value is fixed and is compared with the entropy values calculated. If the entropy value is less than a threshold value, a DDoS intrusion is identified by the proposed security system.

3.3 Firewall

Firewalls are the techniques that govern the arriving and departing packets to and from the inner network. An SDN controller can filter the traffic based on flow controls, this feature can be employed to delineate a firewall within the switch independently. The function of a firewall can be set up using feedback from the switches and controller. The packets are filtered according to user-defined policies. Figure 5 shows the methodology employed for a firewall system where a pox controller is used and programmed. The network traffic is listened to by the switch. If the event has occurred, then the user-defined firewall policies are applied to it. Accordingly,

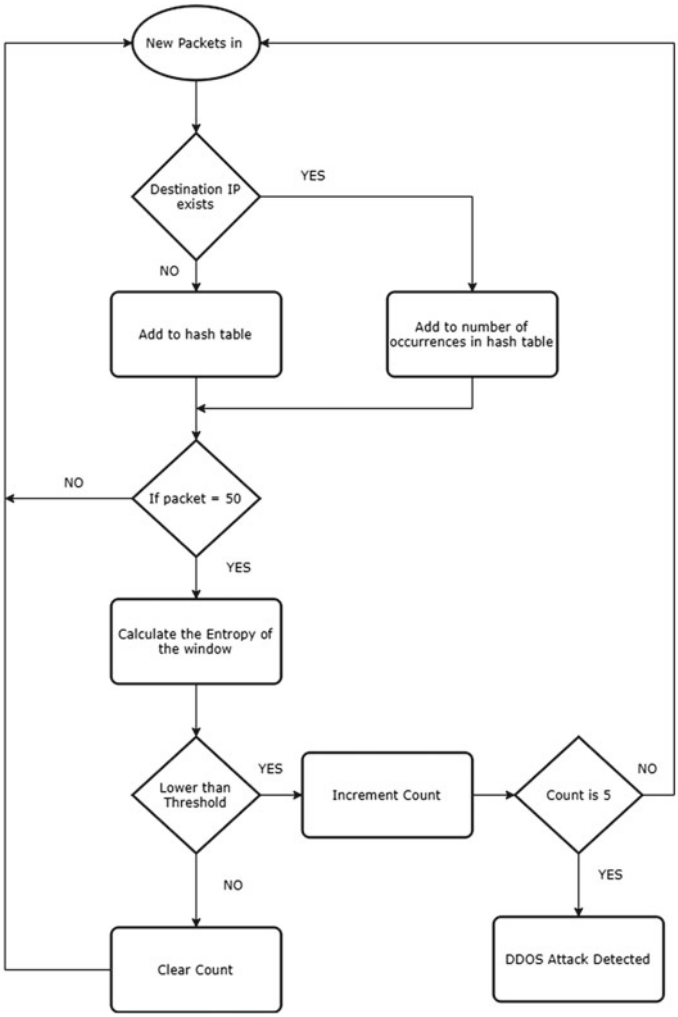


Fig. 4 Methodology for DDoS attack detection using entropy-based features

the packets are dropped if the rule applies and if the rule does not apply, the learning switches come into action.

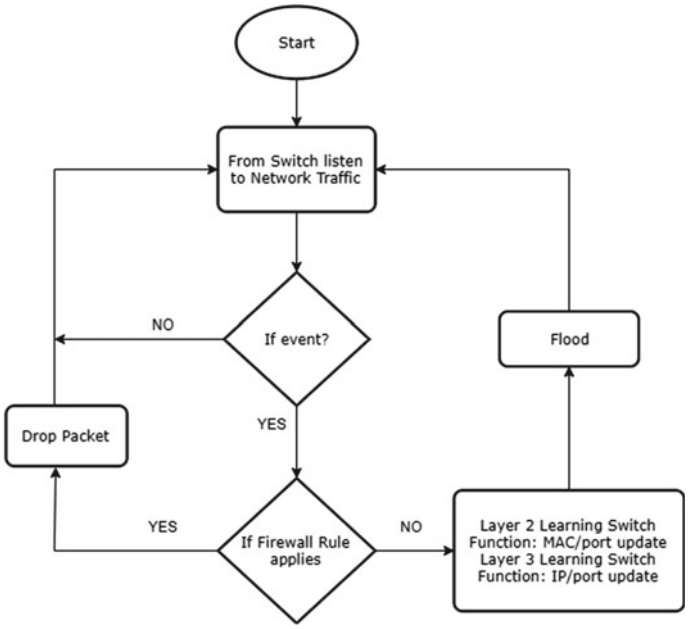


Fig. 5 Methodology for firewall

4 Experimental Results

4.1 DDoS Attacks Detection Using Machine Learning Algorithm

Network simulation was performed for benign UDP, TCP, and ICMP traffic, and malicious traffic that is the assemblage of UDP flood attack, TCP SYN attack, and ICMP attack using Scapy tool. The dataset was created and filtered, and features were extracted. Machine learning deals with data and parameters which were used from dataset for each ML algorithm. Figure 6 represents the receiver operating characteristic (ROC) curve that presents the performance of classification models at all classification thresholds. This curve has two parameters which are true positive rate and false positive rate. It can be seen from the graph that the decision tree and random forest are performing far better than MLP and SVN.

Feature importance allocates a record to input features based on how profitable they are at anticipating an objective variable. Figure 7 shows the feature importance of parameters used from a dataset that plots parameters with corresponding values. It can be seen that byte per-flow and byte count have the highest values.

Accuracy is the analysis applied to determine which model is best at determining patterns and relationships between variables in a dataset formulated on the input or training data. The data are divided into trained and tested set and the values

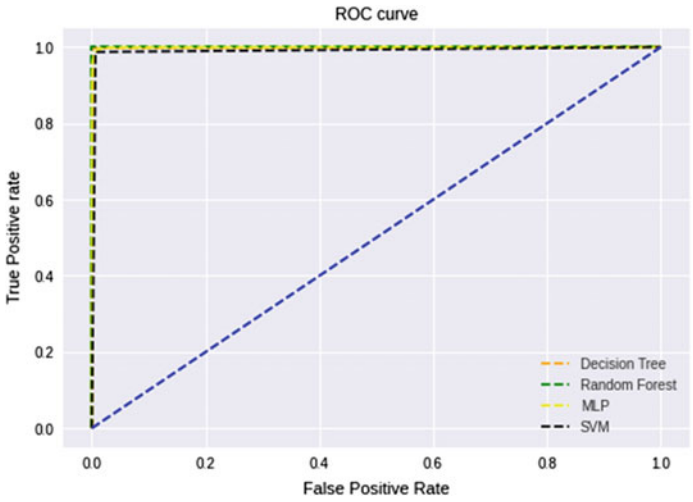


Fig. 6 ROC curve

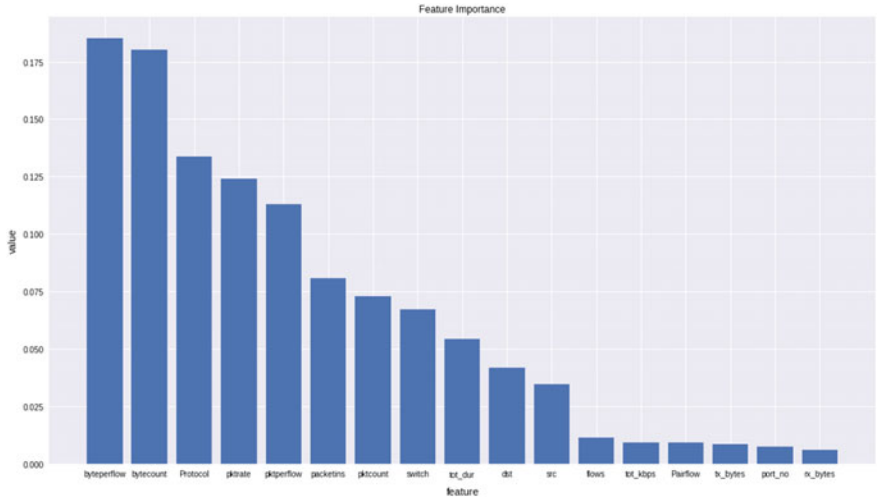


Fig. 7 Feature ımportance

are predicted for each model. K-fold cross-validation is done to get a compute of accuracy. Figure 8 displays the accuracy of each algorithm, and Table 1 shows the corresponding values, where random forest has more accuracy than other algorithms.

Figure 9 shows the time taken by each algorithm to process the task, and Table 2 shows the corresponding values. It can be observed that the decision tree takes less time than other algorithms.

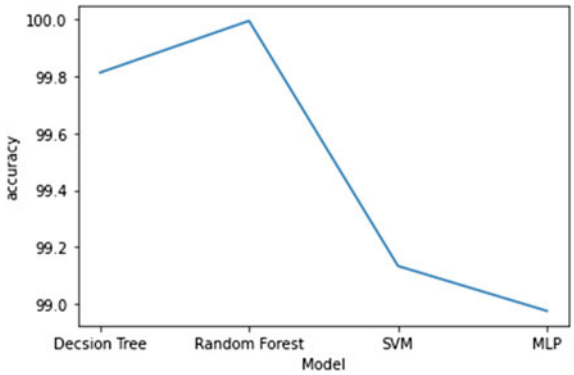


Fig. 8 Accuracy

Table 1 Accuracies of ML algorithms

ML algorithm	Accuracy (%)
Decision tree	99.81
Random forest	99.99
SVM	99.13
MLP	98.97

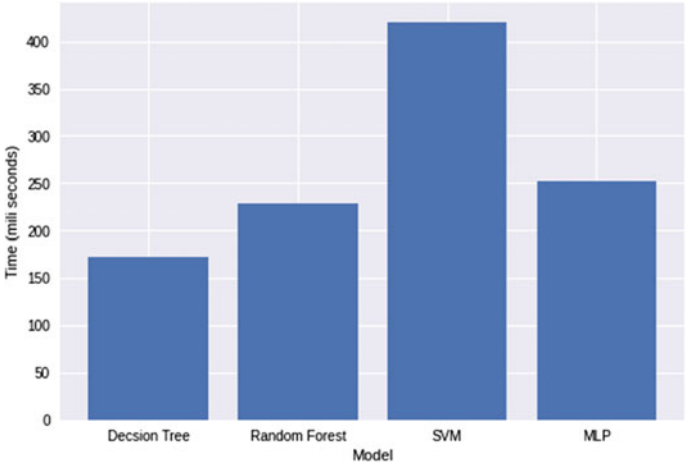


Fig. 9 Time taken by ML algorithms to detect DDoS attack

Table 2 Time taken by ML algorithms to detect DDoS attack

ML algorithm	Time taken (ms)
Decision tree	172
Random forest	228
SVM	420
MLP	252

Table 3 Entropy values

Entropy values after traffic launched	Entropy values after the attack
1.464	0.606
1.466	0.607
1.463	0.605

4.2 DDoS Attack Detection Using Entropy-Based Features

The Mininet topology will start running by configuring the hosts and switches adding the links between them as we have made them and connecting a pox controller which is the main part of SDN. DDoS attacks can be recognized early by observing some hundred packets and considering changes in entropy, which blocks the controller from going down. The phrase "early" is described as the resistance level and traffic being controlled by the controller. From this, the jolt of malicious packet flooding can be controlled. Such a system needs to be featherweight and high feedback time. The high feedback time conserves the controller during the attack stage for recovering the regulation by terminating the DDoS attack. The threshold value calculated is 0.65.

Using xterm in Mininet, a host launches traffic. The corresponding pox terminal gives entropy values after traffic launched, where the entropy value is more than 1 shown in Table 3. At the same time, other hosts in the network launch attacks. The corresponding pox terminal during these attacks give the entropy value which is less than the threshold shown in Table 3. When the entropy value is less than threshold value, DDoS attack is detected. Figure 10 illustrates when DDoS attacks are detected, the port number and switch ID are shown in the pox terminal and blocked.

4.3 Firewall

A topology is created in Mininet, and the pox controller is programmed with some user-defined firewall rules. When the pingall command is used, ping reachability for all hosts is tested which is shown in Fig. 11. It can be observed that the user-defined firewall rules are applied, so a 53 percent drop is the result. This is a simple security approach which monitors and filters packets based on policies.

5 Conclusion

A secure software-defined network has been implemented in this paper, using machine learning algorithms and entropy-based features, which helps in building a secure SDN system. The clustered network data by the SDN controller use the granted data analytics approach to scrutinize and employ machine learning designs

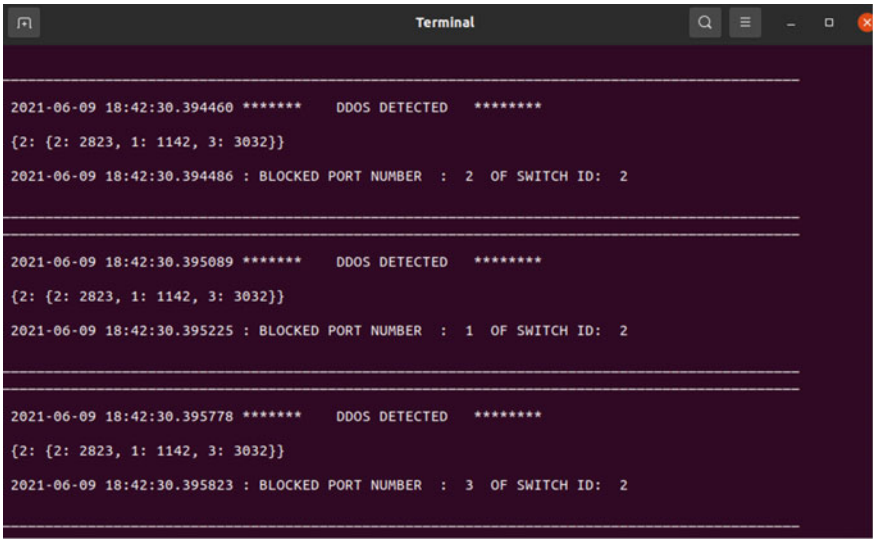


Fig. 10 DDoS attack detected

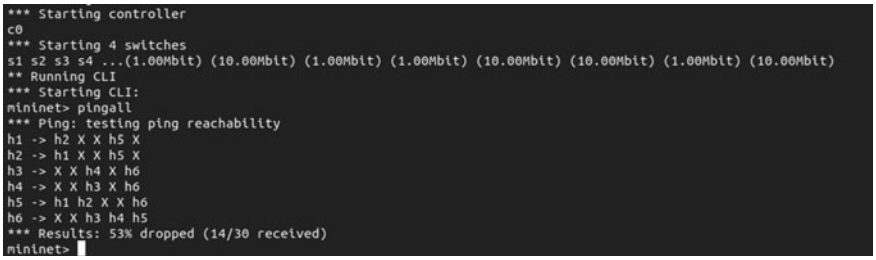


Fig. 11 Firewall

for customizing the network operation process. The ML algorithms like decision tree, random forest, MLP, and SVM were used to expose DDoS invasions in the paper. Where "random forest" showed the best accuracy in identifying DDoS attacks, whereas in the "decision tree" algorithm, the time taken to detect the DDoS attack is less. Since in the "random forest" algorithm, the time taken is almost equivalent to the "decision tree" algorithm and also it has the best accuracy, the "random forest" algorithm is considered to be the best algorithm among the four in our experiment. Using entropy-based features, DDoS attacks were detected by setting a threshold value in the environment. A firewall application was created using OpenFlow controller pox. The major benefit of an SDN-enabled security system is that it provides a smarter and faster response on granular support by selectively blocking malicious traffic while granting normal traffic outflow in the network. The SDN security applications are self-capable of developing anomalies by deflecting the specific network flows to the

appropriate applications or security services (such as firewalls and intrusion detection/prevention). Once the SDN security system is enforced, the system will itself have the capability for accomplishing a more secure network and also provides a way for advancing the pace of implementing new security services.

References

1. M.S. Elsayed, N.-A. Le-Khac, S. Dev, A.D. Jurcut, Machine-learning techniques for detecting attacks in SDN, in *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, 2019
2. H. Polat, O. Polat, A. Cetin, Detecting DDoS attacks in software-defined
3. L. Yang, H. Zhoa, DDoS attack ıdentification and defense using SDN based on machine learning method, in *2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN)*, 2018
4. J.I.Z. Chen, S. Smys, Social multimedia security and suspicious activity detection in SDN using hybrid deep learning technique. J. Inf. Technol. **2**(02), 108–115 (2020)
5. A.B. Dehkordi, M.R. Soltanaghaei, F.Z. Boroujeni, The DDoS attacks detection through machine learning and statistical methods in SDN. (2020) https://doi.org/10.1007/s11227-020-03323-w
6. Q. Yan, Q. Gong, F.-A. Deng, Detection of DDoS attacks against wireless SDN controllers based on the fuzzy synthetic evaluation decision-making model. Adhoc. Sens. Wirel. Netw. **33** (2016)
7. Y. Xu, Y. Liu, DDoS attack detection under SDN context [C]// INFOCOM 2016—the, in *IEEE International Conference on Computer Communications*, pp. 1–9, (IEEE, 2016)
8. J. David, C. Thomas, DDoS attack detection using fast entropy approach on flow-based network traffic. Procedia Comput. Sci. pp. 30–36, (2015)
9. S.M. Mousavi, M.S. Hilaire, Early detection of DDoS attacks against SDN controllers, in *International Conference on Computing, Networking and Communications, Communications and Information Security Symposium*, 2015
10. M. Kia, *Early Detection and Mitigation of DDoS Attacks in Software Defined Networks*, Master's thesis, Ryerson University, Canada 2015
11. S. Sindhu, G. Subbiah, K. Arputharaj, Decision tree based lightweight intrusion detection using a wrapper approach, Expert Syst. Appl. 129–141, (2012)
12. K. Kato, V. Klyuev, An ıntelligent DDoS attack detection system using packet analysis and support vector machine. IJICR **5**(3), (2014)
13. M. Dhawan, R. Poddar, K. Mahajan, V. Mann, SPHINX: detecting security attacks in software-defined networks. IBM Res. (2009)
14. Alpna, S. Malhotra, DDoS attack detection and prevention using ensemble classifier (RF). IJARCSSE (2016)
15. http://mininet.org/

Intrusion Detection System Using Big Data Based Hybrid Hierarchical Model (BDHHM) of Deep Learning

U. B. Mahadevaswamy and Meghana Nagaraju

Abstract The world interconnectivity for the Internet infrastructure has been increasing incredibly because of the enormous increase of data day by day. As most of the communications is dependent on data, the generation and use of the data have become very high so the maintenance of these data is the challenging part. For the security of data, intrusion detection system (IDS) has become a vital component. In machine learning, many IDS models proposed the previous single shallow technique used was not effective in identifying intrusion for the unique pattern. This paper introduces a machine learning model big data-based hybrid hierarchical model (BDHHM) with an Apache spark framework. BDHHM is a hybrid of two hierarchical models such as k-means and random forest tree, hence will increase the detection rate of the intrusion attack. Improved k-means and random forest tree are implemented in this work to identify the unique patterns. In deep learning, one of the artificial intelligence (AI) functions is used to process the data and comparison is made with deep learning models such as FC, CNN, and RNN. This model is effective compared to other models [14] with a true positive rate (TPR) of 96.16%, an accuracy of 95.3%, and a false positive rate of 9.1%.

Keywords Intrusion detection system (IDS) · Big data-based hybrid hierarchical model (BDHHM) · K-means · Random forest tree

1 Introduction

With the large amounts of data creating every day, the world interconnectivity for the Internet infrastructure has been increasing incredibly many security vulnerabilities are creating every often. The security vulnerability creates an opportunity for the

U. B. Mahadevaswamy · M. Nagaraju (✉)
Department of Electronics and Communication Engineering, JSS Science and Technology University, Mysuru, India

U. B. Mahadevaswamy
e-mail: mahadevaswamy@sjce.ac.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_39

Table 1 Comparison between shallow learning and deep learning

Shallow learning	Deep learning
The shallow learning model contains fewer computational layers	The deep learning model contains more computational layers
Limited ability to combine features	Can combine more features
Works only for small number of samples	Works for both large and small number of samples

cyber attackers to perform the malicious activity to destroy the network and the valuable data, which will affect national security. Hence, IDS is necessary to protect the computers and network.

The traditional IDS was using particular rules to determine the intrusion attacks, as it was a labor intensive technique rule-based identification got failed when the number of intrusion attacks and the type of intrusion attacks got increased.

Apart from the traditional detection system, a machine learning method was introduced in the later stage. Compared to the previous intrusion detection system, the machine learning technique was more effective, automatic and it was able to fit into model parameters. The machine learning technique was able to identify the features and unique patterns by examining the network traffic.

For intrusion detection in machine learning, there are two major techniques. 1. Shallow learning model and 2. Deep learning model. The shallow learning model contains fewer computational layers and this include a decision tree, regression, and a support vector machine. When the number of samples increased, this model got failed to identify the intrusion. In recent years, the deep learning model used most as it works for more computational layers and for the large number of parameters. The result of the deep learning model was better when compared to the shallow learning method. Further, single deep learning was effective only when the number of samples is low (Table 1).

Significantly, to enhance the performance, proposed a big data-based hybrid hierarchical model (BDHHM), which contains a hybrid of hierarchical deep learning models. This model proposed to divide the samples into multiple level clusters and each level cluster is trained using a trained dataset.

The major steps followed in the model are

- Extraction of the behavioral and a content feature
- Train the system using trained dataset and apply the improved k-means
- Divide the dataset and make a cluster from each iteration until it reaches the threshold
- Apply the deep learning models to each model cluster such as decision fusion algorithm and random forest tree
- Decision values of the models are merged to check the sample is intrusive or not.

Extraction of the feature plays a vital role because its impacts the performance of the algorithm. The proposed model combines two features, behavioral and content

features. Combined features help to analyze the network traffic patterns. Behavioral features are the network traffic characteristics, and the content feature is the payload information. The combination of these two features will improve the efficiency of the algorithm.

To evaluate the performance of the comparison of an algorithm is made with the different computational model such as

- Decision tree (DT)
- Support vector machine (SVM)
- Convolutional neural network (CNN)
- Combination of recurrent neural network (RNN) and feature extraction. The contributions of this work include utilization of both behavioral and content feature
- Implementation of the model on Apache spark framework for clustering and selecting features
- Adaption of different deep learning models such as k-means, random forest tree for identification the network traffic pattern, and to detect the sample is intrusive or not.

2 Literature Review

S. No.	Paper name	Methodology	Conclusion
1	Big data analytics in cybersecurity: network data and intrusion prediction	Wang and Jones [1] proposed network issues in cybersecurity by taking two characteristics such as "variety"and "veracity" of big data using R function, also analyzed "spam base" form e-mail database by using naive Bayesian classification on 'VAST 2013' dataset. Along with the naive Bayesian also used decision tree (DT), SVM, and very fast decision tree (VFDT). The prediction is done by using k-means and histogram	R helps to perform the statistical analysis and to analyze the correlation of variables in dataset. The spambase survey shows that the R employed datasets are effective as R has the potential in solving the challenges faced by big data such as variety and veracity

(continued)

(continued)

S. No.	Paper name	Methodology	Conclusion
2	Intrusion detection system with recursive feature elimination by using random forest and deep learning classifier	Ustebay et al. [2] proposed to detect the malicious attack in a computer network using the CICIDS2017 dataset. The method used was recursive feature elimination performed via the random forest tree, and hence, there was an increase in the accuracy by deep multilayer perceptron (DMLP) structure	Based on the machine learning method used, dataset can reduced without affecting the accuracy and speed and hence recursive feature elimination is used. DMLP model proposed archives accuracy of 89%
3	Improving detection accuracy for imbalanced network intrusion classification using cluster-based under-sampling with random forests	Miah et al. [3] have proposed a system to detect an intrusion system by taking the KDD99 dataset. The system has developed by using machine learning (ML) and data mining (DM) algorithms such as artificial neural network (ANN), naive Bayes (NB) classifier, random forest, and bagging techniques	The proposed method is the multilayer classification and cluster-based sampling technique is used with random forest tree to solve over fitting problem, hence this method increased the detection rates of low frequency attack in network and decreases the false positive rate
4	Performance evaluation ofintrusion dtection streaming transactions using apache kafka and spark streaming	Tun et al. [4] have proposed a time-consuming rate to detect an intrusion in a system. Multiple brokers of Kafka and spark streaming were used by taking the UNSWNB-15 dataset. The machine learning models like conventional neural networks proved the time required to detect the system is reduced in the proposed framework	The result of the proposed system multiple partitions of Apache Kafka and the batch intervals between 10 and 50 s of spark streaming get better performance in the integration of Kafka and spark streaming

(continued)

(continued)

S. No.	Paper name	Methodology	Conclusion
5	An improved İntrusion detection algorithm based on GA and SVM	Tao et al. [5] used the characteristics of the genetic algorithm(GA) and the support vector machine (SVM) algorithm and proposed an algorithm FWP-SVM-GA (feature selection, weight, and parameter optimization of support vector machine based on the genetic algorithm) this reduce the error rate and time consumption of classification and increased the true positive rate	This model uses an alarm intrusion detection algorithm based on support vector machine and genetic algorithm. Hence, the intrusion detection rate is high with high accuracy and true positive rate
6	Deep learning approach for intelligent intrusion detection system	Ishaque et al. [6] have proposed the deep neural network (DNN) model to identify the continuously changing attack of intrusion at both network-level and host-level, by using the KDDCup 99 dataset. The scale-hybrid-IDS-AlertNet (SHIA) algorithm is used in real-time to identify at both network-level and host-level to alert the cyberattacks	The proposed hybrid model for intrusion detection using deep learning model with DNN for analyzing the large scale data. The proposed model can perform for both HIDS and NIDS
7	Clustering approach based on mini batch k-means for intrusion detection system over big data	Peng et al. [7] have proposed the system by using k-means+ + to initialize the centroids on mini batch k-means hybrid, with principal component to reduce the local optimum of the algorithm. Calsski Harabasz indicator used for clustering results	The proposed model has a clustering method on k-means with PCA for Intrusion detection system. The PCA method used to reduce the dimension to improve the clustering efficiency

(continued)

(continued)

S. No.	Paper name	Methodology	Conclusion
8	Developing a hybrid intrusion detection system using data mining for power systems	Pan et al. [8] have proposed the synchrophasor systems for accessing large quantity of data for monitoring the system demand of reliable energy, an automated system is built to identify the pattern by fusion the path mining with synchrophasor systems	The proposed system uses the common path mining algorithm with time-domain data analysis. The common paths used are hybrid signatures and patterns of system behavior. The classification is done by taking twenty-five scenarios consisting of stock ticker SLG faults, control actions, and cyberattacks. This model shows the average accuracy of 73.43% for zero day attack
9	Investigation of network intrusion detection using data visualization method	Bulavas [9] has proposed a system that is a hybrid of two methods, decision tree, and PCA. Is designed by setting a certain threshold to the PCA and has focused on linear projection for the parameter analysis	The proposed model uses the PCA analysis with decision tree, for classification of Smurf, Satan, Neptune, Portsweep, Ipsweep with probabily of above 95%
10	Building an intrusion detection system using a filter-based feature selection algorithm	Ambusaidi et al. [10] have proposed an intrusion system by using KDD Cup 99 dataset. The attack was detected by identifying feature selection on various parameters such as mutual information, and liner correlation co-efficient. The framework is based on the least square support vector machine to normalize the data, classifying the data, and recognizing the attack	The proposed method used supervised filter-based feature selection algorithm such as flexible mutual information feature selection for identifying the intrusion. The implemented algorithm result in identifying the intrusion with more accuracy

(continued)

(continued)

S. No.	Paper name	Methodology	Conclusion
11	Enhancing trust management for wireless intrusion detection via traffic sampling in the era of big data	Meng et al. [11] have proposed the clustered wireless sensor network to detect the attack by using the machine learning models like the Bayesian model and observes trust management via traffic sampling. Random n out N sampling (randomly generating n samples from 1-N) is used for evaluating trust value in a simulated WSN by setting the threshold in different models	This work uses wireless sensor network for sense and control of interconnect object along with Bayesian based on the result with the threshold enhanced the high traffic environment for identification of malicious nodes
12	Knowledge discovery from big data for intrusion detection using LDA	Huang et al. [12] have proposed the intrusion attack with the latent Dirichlet allocation method of big data. The methods used are the signature-based method and an anolamy method map-reduce method for analyzing the timely manner of the latent Dirichlet allocation method	The proposed model has an hybrid method using LDA to capture the patterns using the "bag of words" method along with the LDA an MapReduce method is used for large volume logs
13	Research on hadoop-based intrusion detection by IP address	Huang[13] proposed the research on the attack based on the IP address by using the machine learning models like FP growth and a k-means algorithm, Hadoop framework is used for distributed computing ability for the dataset. This method is implemented and tested on the node hence reduces the false report of the attack detection	The proposed system used an Hadoop method to identify the intrusion, the IP is used to identify the intrusion using association rule clustering analysis
14	Applying big data based deep learning system to intrusion detection	Zhong et al. [14] have proposed a big data-based Hierarchical deep learning system that is used to find the intrusion type with k-means and decision fusion with three different types of the dataset	In this paper, they have proposed a BDHDLS model to identify the intrusion and uses big data techniques and parallel strategies for feature selection. This model achieved more performance gain compared to the previous model

(continued)

(continued)

S. No.	Paper name	Methodology	Conclusion
15	Network intrusion detection using improved genetic K-means algorithm	Sukumar et al. [15] have proposed an improved genetic algorithm implemented to show the performance of the genetic k-means algorithm	The proposed work uses IGKM algorithm to identify the intrusion and concludes that the IGKM as the less accuracy when compared to the k-means

3 Proposed Modeling

In the proposed work, a BDHHM algorithm is used to identify the intrusive type.

BDHHM algorithm

Step 1: Extraction of behavioral and a content feature and combining both the feature using a big data technique
Step 2: creating a one level clustering by using the improved k-means algorithm, an initial point k is chosen from each iteration
While (k < cluster threshold) {
From the iteration select the k initial point
In an Apache, spark environment run the k-means to a certain iteration until it meets the threshold point
Evaluate the cluster quality from the initial point by calculating the maximum distance of the cluster increase the quality of the cluster until it reaches the maximum threshold

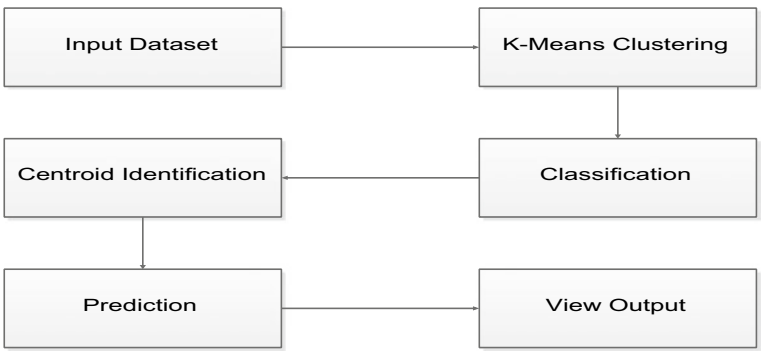


Fig. 1 Block diagram of proposed system

} End while
Step 3: Proceeding multilevel from each single-level cluster
{ Perform the hybrid hierarchical method of clustering in the Apache spark framework
Perform decision fusion algorithm and metric method in parallel,
Perform random forest tree, a "bagging method" in each one level cluster until a maximum threshold meet }
The root of each one level clusters is merged until it reaches a maximum threshold and generates a high-level cluster
Step 4: In an Apache spark framework train, each cluster with the machine learning model
Step 5: From the model classification of cluster, identify the status of the intrusive attack (Fig. 1).

System will be given dataset as input and it will perform clustering on the given dataset. After the clustering, it will classify it and then will identify the centroid of the dataset. Based on the centroid value, the system will predict what type of anomalies has been identified in the network and the result will be displayed to user in the end.

3.1 Dataset and Experimental Setup

The dataset used in this work is a combined dataset of ISCX2012, CICIDS2017, and CICIDS2018. The dataset used is the public benchmark dataset. The dataset used for building an intrusion detection system needs a complete and considerable dataset with real-time background network traffic. As the experiment carried out for the content feature, dataset with payload information is chosen. Payload information plays a vital role in getting complete data on how the intrusion affects the network. The combined public datasets like ISCX2012, CICIDS2017, and CICIDS2018 contain the raw data with the labels. Each dataset network traffic flow communication between source and destination.

Multiple packets from the traffic flow are combined into a single sample. The dataset contains the intrusive such as dos, probe, U2R, and R2L.

In this work, machine learning models tested using the combined dataset. Firstly, samples are randomly selected for the big data-based hybrid hierarchical model (BDHHM) for the trained data to train the environment, then the remaining dataset is used as the independent testing set.

3.2 BDHHM for Intrusion Detect

Module 1: Login to the application and train the model

3.2.1 Application Login Page

To login into the intrusion detection application flask is used, it is a Web framework used to design a Website. By triggering the flask URL will generate, using the URL user can log in to the application after entering certain credential.

The login page which is the entry point to enter into an application. This is done by using flask, a Python lightweight WSGI Web framework that uses a Jinja template engine to build a dynamical HTML page. By using this, the required dataset can be selected without changing hardcoded data to make the task easier. Here, this have created two buttons for browsing required file, one for loading and the other for processing the dataset.

3.2.2 Training the Model

In this step, the trained and processing dataset is added to the environment to train the system. Any machine learning models need to train with the trained dataset, and this will be the base for detecting the intrusive. Hence, in this step, the system is trained with the trained dataset by adding the trained dataset to the environment (Fig. 2).

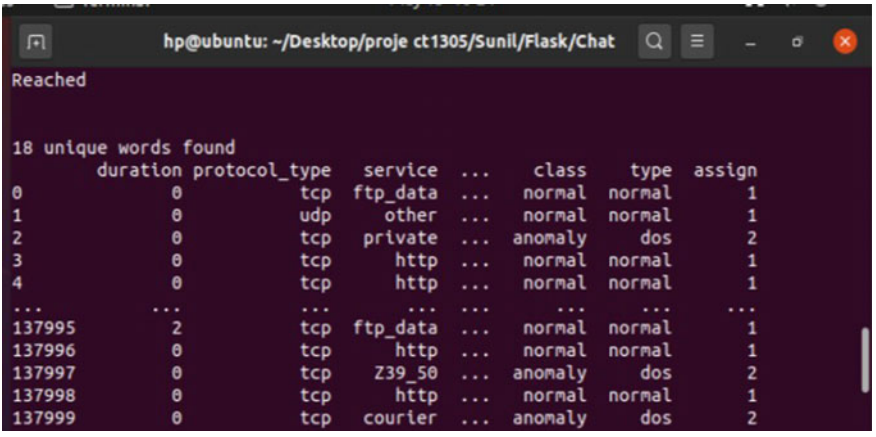


Fig. 2 Processing of trained dataset

Module 2: Loading the test data and processing the dataset using a clustering algorithm

After training a machine learning model by trained dataset, evaluation is done by testing the system with the test data.

The processing of the test data is carried out by using a clustering algorithm. A clustering algorithm chosen is the k-means clustering algorithm which is an unsupervised learning method in machine learning.

K-means iteration takes place in two steps,

- Mapping
- Reducing

This technique is called MapReduce. Hadoop MapReduce is a framework used for writing applications for a large amount of data and parallel processing.

In mapping, it will split the dataset into chunks and processes in an independent and parallel manner. The framework will sort the output of the maps and then it is used as an input for the reduce phase. Here, centroid for each phase is calculated by using the divide and conquer technique.

$$\text{Dist}(y, c) = \sum_{i=0}^{n} |(ay(\text{i}) - ac(i)| \tag{1}$$

where n is the number of attributes in each sample, ay is the value of attribute at position i for a sample y, and ac is the value of attribute at position i for the centroid. Based on the number of types of intrusive attacks present in the dataset that many iterations are carried out and for each type of intrusive attack, a cluster is created by identifying the centroid.

Figure 3 represents the cluster formation for the non-intrusive attack. The type of intrusive can be determined from the class column of the cluster table. Along with

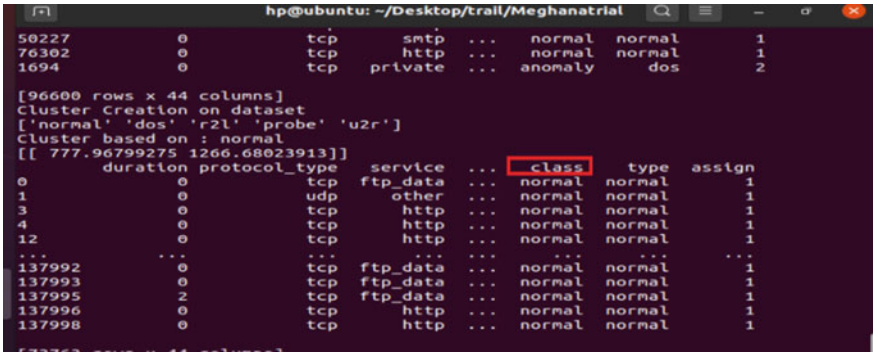


Fig. 3 Represents the cluster formation for the non-intrusive

both behavioral and content feature, duration of the connection, source bits protocol are used.

Figure 4 represents the cluster formation for the r2l intrusive attack, and r2l attack is an unauthorized type of attack, raises when user tries to enter the application without root access. Intrusive type can be determined from the type column of the cluster table. Based on the number of types of attacks cluster formation will form, cluster formation directly depends on the type of intrusive attacks in the dataset.

In the similar way, cluster formation is done for dos, probe, and u2r.dos attack raises when the inappropriate shutdown of network or machine done. Probe raises when an unauthorized user tries to get access of file from machine. U2r raises when the unauthorized user tries to get access for the password of the regular user. All intrusive type can be determined from the type column of the cluster table. Based on the number of types of attacks cluster formation will form, cluster formation directly depends on the type of intrusive attacks in the dataset.

Figure 5 shows a result of k-means clustering, where it identifies the number of types of attacks found in the test dataset and classifies them by using a centroid. Here, for this particular dataset, a graph identifies the four types of attacks and also classifies them for the non-intrusive data that are for the normal type. Types of attacks identified are r2l, u2l, probe, and dos.

Module 3: Training the cluster with deep learning model:

From the single-level cluster, high-quality subclusters can be formed using a hierarchical clustering method. In this model, on combine features of both behavioral and content features is used to train each cluster in the multilevel, this is done to enhance the performance of intrusion detection rate. To select the best one, the model is evaluated with the deep learning models like FC, RNN, and RNN in the multilevel clustering.

A forward connected (FC) network will work in a forwarding connected manner and will not recognize the unique characteristics as it does not focus on the structure. CNN concentrates on spatial patterns of the network traffic from the network payload,

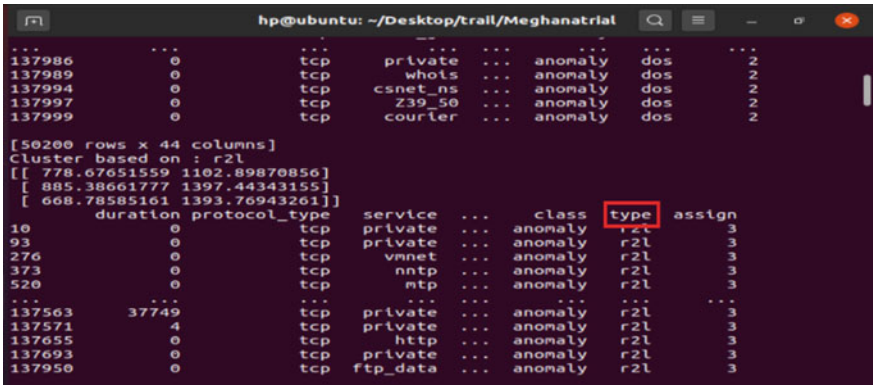


Fig. 4 K-means cluster for r2l intrusive attacks

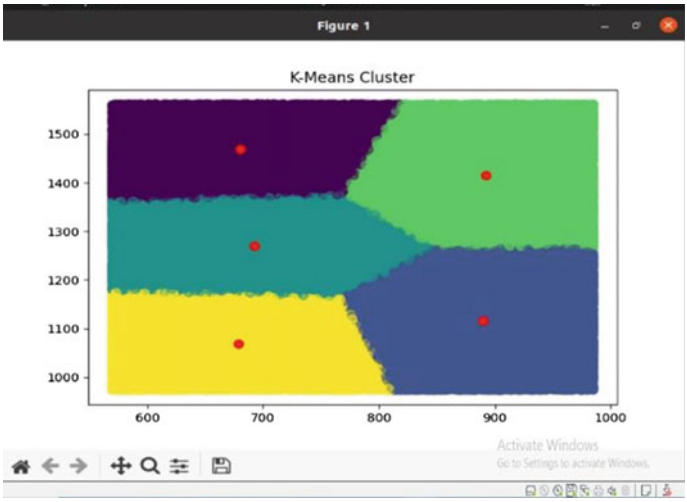


Fig. 5 Result of k-means clustering

and the representation is done in the matrix. CNN works in two major steps, one with the sparsity of connection and the weight sharing to extract the feature and CNN can identify the lower to higher-order features. RNN used to examine the sequence of the network payload. Each of these model focus on the highly similar samples in each cluster. FC, CNN, and RNN is a mapping function from input vector to output vector at every layer.

Module 4: Decision fusion algorithm and random forest tree:

3.2.3 Decision Fusion Algorithm

Decision fusion algorithm in deep learning model is used to identify an intrusive attack and it is calculated by high dimensional feature space for each cluster, and representation is done using matrix. The result of this decision value helps in predicting the sample as intrusive.

$$fk(x) = \text{Sigmoid}(Wk * yk + bk) \tag{2}$$

where yk is the "activation" of the last cluster of the neural network for the cluster k. Wk is the weight matrix between the last layer and the output layer of neural networks for cluster k.

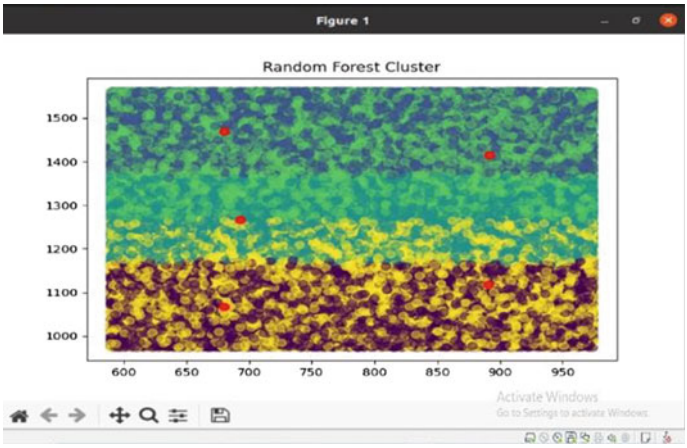


Fig. 6 Random forest tree result

3.2.4 Random Forest Tree

Random forest tree (RFT) includes an n number of decision tree with different subdataset and parameters, and it is trained using the bagging method.

After building, the multiple decision tree will merge them to get an accurate value. The main reason to choose this algorithm is to get an accurate result to an assigned input feature.

Figure 6 shows the random forest result, where an RFT is implemented at each cluster level as shown, and each cluster is represented by a centroid.

Module 5: calculation of the TPR, FPR, and accuracy:

Performance evaluation of the system is done by calculating the true positive value (TPR), false positive rate (FPR), and accuracy. This is called a confusion matrix in machine learning.

True positive rate (TPR):

True positive rate is calculated using,

$$\mathrm{TPR} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{3}$$

where true positive (TP) represents the correctly recognized intrusive sample, and false negative (FN) represents the incorrectly recognized samples. Figure 7 shows the result of TPR.

False positive rate (FPR):

False positive rate calculated using,

$$\mathrm{FPR} = \mathrm{FP}/(\mathrm{FP} + \mathrm{TN}) \tag{4}$$

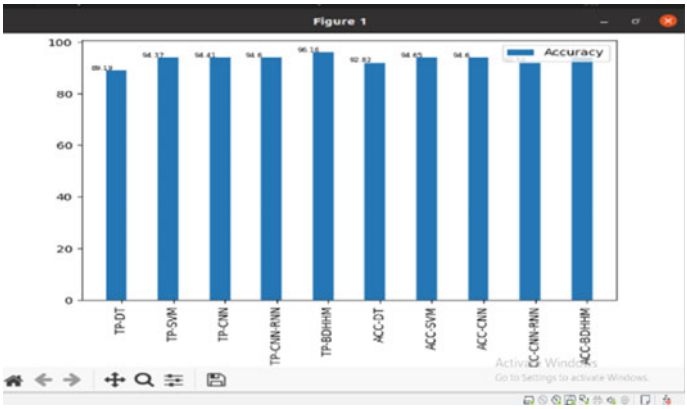


Fig. 7 TPR and accuracy

where false positive (FP) represents the number of non-intrusive samples that are incorrectly recognized as intrusive samples, and true negative (TN) represents the number of non-intrusive samples that are correctly recognized. Figure 8 shows the graph of FPR.

Accuracy:

Accuracy is calculated by

$$\text{Accuracy} = \text{TP} + \text{TN}/(\text{TP} + \text{FP} + \text{TN} + \text{FN}) \tag{5}$$

Loss function:

The loss function is used to calculate the difference between the predicted output and the real output.

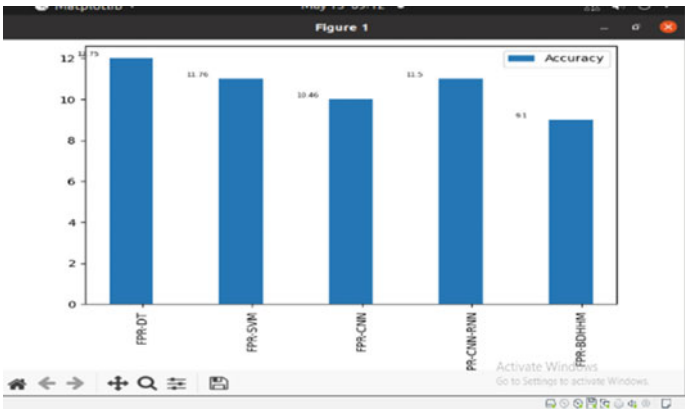


Fig. 8 Result of FPR

Table 2 Performance comparison table

Machine learning model	TPR (%)	Accuracy (%)	FPR (%)
DT	89.19	92.82	12.75
SVM	94.37	94.65	11.56
CNN	94.41	94.6	10.46
CNN-RNN	94.6	92.58	11.5
BDHHM	96.16	95.3	9.1

$$\text{Loss}(x, y) = \sum_{i=1}^{n} (xi \log yi + (1 - xi)lo) \tag{6}$$

where xi represents the true label for sample i, ∈ 0, and 1. 0 represents the non-intrusive sample and represents the intrusive sample, and n is the total size of the dataset.

Cost function:

To get the optimized parameters for all weights(w) and bias(b), the model cost function is calculated (Table 2).

$$K(w1, b1) = \frac{1}{n} \sum_{i=1}^{n} (\text{Loss}(x, y)) \tag{7}$$

4 Conclusion

In this work, a big data-based hybrid hierarchical model (BDHHM) is proposed. This includes a hybrid of hierarchical models such as k-means, decision fusion algorithm, and random forest tree. This method is effective to capture the data patterns at each layer to identify the intrusive attack and this strategy also reduces the detection rate.

BDHHM adopts both behavioral and content features, this helps to detect the intrusive at both in network traffic and payload content. This strategy helped in improving the IDS. This model also shows the big data techniques and parallel strategies for feature selection, training datasets, and clustering the data into multilevels.

When compared to the single deep learning approach, BDHHM uses more computational resources to achieve performance efficiency. Performance efficiency is proved by calculating the true positive rate (TPR), false positive rate (FPR), and accuracy. The performance gain of this model is high when compared to the previous model [14] with a TPR rate and accuracy of 96.16% and 95.3%, respectively, and the false positive rate of 9.1% as shown in Table 1. Random forest tree is the fast bagging technique to generate multilevel clusters hence reduced the false positive rate and

implementation of this increased the efficiency and proves as a better computational technique.

Further, in future can be implement by increasing the number of machines and reducing the computational resource to achieve the same performance.

References

1. L. Wang, R. Jones, Big data analytics in cybersecurity: network data and intrusion prediction, 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 0105–0111 (2019). https://doi.org/10.1109/UEMCON47517.2019.8993037
2. S. Ustebay, Z. Turgut, M.A. Aydin, Intrusion detection system with recursive feature elimination by using random forest and deep learning classifier, in 2018 International congress on big data, deep learning and fighting cyber terrorism (IBIGDELFT), pp. 71–76 (2018). https://doi.org/10.1109/IBIGDELFT.2018.8625318
3. M.O. Miah, S. Shahriar Khan, S. Shatabda, D.M. Farid, Improving detection accuracy for imbalanced network intrusion classification using cluster-based under-sampling with random forests, in 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), pp. 1–5 (2019). https://doi.org/10.1109/ICASERT.2019.8934495
4. M.T. Tun, D.E. Nyaung, M.P. Phyu, Performance evaluation of intrusion detection streaming transactions using Apache Kafka and spark streaming, in 2019 International Conference on Advanced Information Technologies (ICAIT), pp. 25–30 (2019). https://doi.org/10.1109/AITC.2019.8920960
5. P. Tao, Z. Sun, Z. Sun, An improved intrusion detection algorithm based on GA and SVM, in IEEE Access **6**, 13624–13631 (2018). https://doi.org/10.1109/ACCESS.2018.2810198
6. M. Ishaque, L. Hudec, Feature extraction using deep learning for intrusion detection system, in 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), pp. 1–5 (2019). https://doi.org/10.1109/CAIS.2019.8769473
7. K. Peng, V.C.M. Leung, Q. Huang, Clustering approach based on mini batch kmeans for intrusion detection system over big data, in IEEE Access 6, 11897–11906 (2018). https://doi.org/10.1109/ACCESS.2018.2810267
8. S. Pan, T. Morris, U. Adhikari, Developing a hybrid intrusion detection system using data mining for power systems. IEEE Trans. Smart Grid **6**(6), 3104–3113 (Nov 2015). https://doi.org/10.1109/TSG.2015.2409775
9. V. Bulavas, Investigation of network intrusion detection using data visualization methods, in 2018 59th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), pp. 1–6 (2018). https://doi.org/10.1109/ITMS.2018.8552977
10. M.A. Ambusaidi, X. He, P. Nanda, Z. Tan, Building an intrusion detection system using a filter-based feature selection algorithm. IEEE Trans. Comp. **65**(10), pp. 2986–2998 (1 Oct 2016). https://doi.org/10.1109/TC.2016.2519914
11. W. Meng, W. Li, C. Su, J. Zhou, R. Lu, Enhancing trust management for wireless intrusion detection via traffic sampling in the Era of big data. IEEE Access 6, 7234–7243 (2018). https://doi.org/10.1109/ACCESS.2017.2772294
12. J. Huang, Z. Kalbarczyk, D.M. Nicol, Knowledge discovery from big data for intrusion detection using LDA, in 2014 IEEE International Congress on Big Data, pp. 760–761 (2014). https://doi.org/10.1109/BigData.Congress.2014.111
13. Table of Contents, in 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), pp. i–ix (2020). https://doi.org/10.1109/ICBAIE49996.2020.00004

14. W. Zhong, N. Yu, C. Ai, Applying big data based deep learning system to intrusion detection. Big Data Mining Anal. **3**(3), 181–195 (Sept 2020). https://doi.org/10.26599/BDMA.2020.9020003
15. J.V. Anand Sukumar, I. Pranav, M. Neetish, J. Narayanan, Network intrusion detection using improved genetic k-means algorithm, in 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 2441–2446 (2018). https://doi.org/10.1109/ICACCI.2018.8554710

A Model for Managing the Procedure of Continuous Mutual Financial Investment in Cybersecurity for the Case with Fuzzy Information

Berik Akhmetov, Valeriy Lakhno, Volodimir Malyukov, Bakhytzhan Akhmetov, Bagdat Yagaliyeva, Miroslav Lakhno, and Yakiyayeva Gulmira

Abstract The paper looks at the development of a mathematical model of the process of continuous mutual investment of projects in the sphere of information security and information protection within the framework of a scheme with fuzzy information. The proposed model is the core of an intelligent support system of accepting decisions in the analysis of different investment plans of information security systems for information objects of informatization, particularly national centers for responding to cyber threats from investors from different countries. The model makes it possible to use the toolkit of a quality game surfaces in the case when the information support of investors is specified by means of fuzzy sets. Particularly, this information support may relate to fuzzy data on the size of investors' financial resources or technologies used to protect information and the corresponding risks of their implementation. The paper presents the outcomes of experiments accomplished in the MATLAB modeling simulation environment. The online platform of the support system of making decisions of investors is also described when choosing an investment plan of information protection systems of an informatization object. The outcomes of simulation have confirmed the performance and capability of the model for the analysis of different strategy investment in information protection systems of an informatization object, taking into account fuzzy information.

B. Akhmetov · B. Yagaliyeva (✉)
Yessenov University, Aktau, Kazakhstan
e-mail: bagdat.yagaliyeva@yu.edu.kz

B. Akhmetov
e-mail: berik.akhmetov@yu.edu.kz

V. Lakhno · V. Malyukov · M. Lakhno
National University of Life and Environmental Sciences of Ukraine, Kiev, Ukraine

B. Akhmetov
Kazakh National Pedagogical University, Almaty, Kazakhstan

Y. Gulmira
Khoja Akhmet Yassawi Kazakh-Turkish International University, Turkistan, Kazakhstan
e-mail: gulmira.yakiyaeva@ayu.edu.kz

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_40

Keywords Information security · Information security systems · Investment · Optimal strategies · Differential game · Decision support system · Fuzzy sets

1 Introduction

The quick development of information technology (IT) made it necessary to pay due attention to ensuring information security (IS) of various objects of informatization (OIN). This solves the problem of ensuring compliance of the state of a specific IT with rapid changes in the landscape of information threats for OIN. This, in turn, reduces the likelihood of risks associated with information threats. In the formation of information security systems (ISS) in many companies, enterprises, and institutions, the greatest attention is paid, as a rule, to the fulfillment of the requirements of the regulatory and methodological framework in the field of information security (IS), defining these requirements as the fundamental basis for the formation of ISS. However, without a suitable degree of investment in the ISS of OIN, these activities, by themselves, do not yet create guarantees of a sufficient level of IP. Nevertheless, the question of the ratio of the costs of building an information security system (ISS) and possible losses from the implementation of information threats in the absence or insufficient reliability is still poorly studied. Taking this into account, the question of determining the amount of investment, that is advisable to invest in the information protection of OIN, should be recognized as still relevant.

An organization can allocate significant resources to ensure the stability and stability of the functioning of corporate information systems, but this does not guarantee the achievement even of the minimum level of security of information resources. The essence of the problem is that in the design and construction of information security systems, the main attention should be paid not to minimizing the impact of a certain list of typical information threats (which is often compiled for a certain imaginary environment for the OIN functioning) but to the search for optimal investment management strategies, including mutual strategies for different ratios of criteria of the investment procedure in the OIN in the context of fuzzy information about investors.

The provision of IS of OIN is possible only through the comprehensive and continuous application of organizational, legal, and technical protection methods at different levels of implementation. In order to develop common approaches in countering cyber threats, to consolidate efforts in the investigation and prevention of cybercrimes, to prevent the use of cyberspace for illegal and military purposes, many leading states [1, 2] have stepped up their participation in organizing joint international projects to build cyber potential, which in fact are examples mutual investment in the OIN.

In this context, Ukraine and Kazakhstan continue to apply European and international standards in the field of cybersecurity, develop the work of relevant bodies that are able to effectively interact with the relevant bodies of the EU and NATO. The experience of Ukraine and Kazakhstan allows them to be not only recipients of

assistance from the EU and NATO states but also sources of new knowledge, skills and ways to counter modern cyber threats.

Many experts in the field of cybersecurity have noted that the use of intelligent information systems, which certainly include decision support systems or expert systems for finding the optimal strategy for investing in cybersecurity circuits, can be useful. This primarily concerns multilateral interstate projects. That is, such projects in which the optimal solution should take into account the balance of interests of many players in the investment market of information security systems and cybersecurity. An expert person, no matter how qualified he is, is unable to cover dozens of interrelated factors that can affect the success of investing in such a complex area as information security and cybersecurity [3, 4].

In its turn, the increasing complexity of the architecture of an intelligent system entails the need to apply more complex algorithms and corresponding mathematical models. In such problems, it is impossible to do with simple linear dependencies, which were used 10–15 years ago. Today, it is not enough just to calculate the payback period of investments in information security and cybersecurity projects. It is crucial to take into account dozens of external factors, and most importantly, to realize clearly that the procedure for investing in information protection and cybersecurity circuits takes place in conditions of constant confrontation with the attacker. Moreover, the attacking party of the defense is not bound by any ethical or legislative norms and is aimed at achieving his goals at any cost.

2 Literature Review and Problem

The economic efficiency of the information security system is an important and often a determining indicator of the effectiveness of such systems [3, 4]. A description of the investment model in information security systems and probabilistic models of losses from attacks are proposed in the work [5]. These models allow describing the mathematical expectation and variance of losses for the information security system in an analytical form. On this basis, a methodology has been developed for assessing the effectiveness of investments and economic risk for ISS [3, 4]. As a generalized indicator of investment efficiency, it is proposed to use the degree of risk for a random variable (R.V) of the net present value (NPV) of total costs for the ISS. This measure of risk is equal to the sum of the mathematical expectation of R.V. cost and its standard deviation multiplied by the coefficient k, but it is noted that a necessary condition for the adequate application of stochastic models is the mandatory availability of reliable statistical and expert data on attacks and security measures [4, 5]. And this is not always possible. Therefore, naturally, in relation to the models considered in [3, 5], the question arises about the influence of input data errors on the resulting indicators.

The model proposed in [6] uses optimization methods to analyze the investment levels in cybersecurity measures and insurance for owners of critical infrastructure

facilities. This model can be used to develop strategies to minimize cybersecurity risks. However, the authors do not provide a software solution.

In recent years, many researchers have increased interest in the scientific substantiation of the solution to the problem of defining optimal methods for investing in information security systems. Particularly, such fundamental research can be mentioned in [4, 7].

It is shown in [8] that it is possible to achieve a given level of IS of the OIN only by comprehensively solving financial, design, production, organizational, research and other interrelated tasks. Consistency has undoubtedly become an advantage of this approach. However, in the work, the authors did not provide an assessment of the potential for using the DSS in such complex tasks to assure the information security of the OIN.

The model proposed in [3, 4, 9] (hereinafter referred to as the GL model) has become one of the most popular for practical assessment of investment strategies in the IS of the OIN. However, this model, and its numerous modifications [10, 11], do not take into account the real mechanisms of return on investment to investors. This led to the limitations of the practical aspects of the application of this model.

The development of intelligent computing [12] gave a powerful impetus to such an independent direction of applied research as the development of intelligent DSS in the process of choosing optimal strategies for investing in ISS. It should be noted that the results of this research, and particular works [12–14], showed that often the proposed ones do not allow generating real recommendations for investors in the information security system. This is especially manifested in situations where there is no clear information about the aspects of investment, for example, the maximum amount of resources allocated for investment projects to create information security information objects of informatization. As the authors admit in [14, 15], the proposed models lack the properties of adaptability. That is, it is necessary to make adjustments to them even with a slight change in the initial parameters and boundary conditions in the process of analyzing investment strategies in projects related to IS and ISS.

In [16, 17], the authors showed that investing in information security should be considered comprehensively from the point of view of various tasks arising in the course of providing information security for the OIN. Investment areas include: anti-virus software (software), firewalls, cryptographic systems, intrusion detection systems; automated backup systems, etc.

The aforementioned necessitated the development of new adaptive models for the DSS [18] in terms of determining optimal strategies for mutual financial investment in ISS projects.

As shown in [19], hackers are often more motivated to achieve their goals, while the defense side is often satisfied only with the return on investment in the information security system. While the defense side can spend huge sums of money on the cybersecurity of OIN, hackers may have to invest only a small portion of their financial resources in the attack, for example, by bribing an unscrupulous employee who is willing to “help” to overcome OIN security perimeters.

Taking into consideration the results presented in [20, 21], it can be stated that the use of intelligent information systems can give a new impetus to seek for solutions in

the problems of optimizing investment strategies in information security and cybersecurity systems of complex multi-circuit distributed information systems. Moreover, nowadays such distributed systems often form the basis of the business processes of many companies and organizations around the world. Consequently, the search for mathematical models of the computing core for such intelligent systems is still relevant. And the game theory acts, namely its subsection concerning the description of the quality games procedure as a variant of the solution that has confirmed its functionality.

3 Purpose and Objectives of the Research

The purpose of the work is to develop a model for the module of a computer support system of decision making in the course of discrete mutual investment in ISS on condition of fuzzy information about investors. To achieve the goal, the following tasks are solved:

- the optimal strategies of investors have been determined for a situation when there is unclear information from the defense party;
- simulation modeling is performed in the MATLAB environment using the developed online DSS platform for various strategies for investing in information security systems in a fuzzy formulation.

4 Models and Methods

The landscape and scale of cyberattacks force the OIN defense side to prioritize defensive methods and techniques. This means that an organization or an enterprise must take into account the full range of information security threats to which they are exposed. The risks of losing information resources as a result of an attack must also be considered, and actions must be taken to minimize the vulnerabilities that are identified. All of the above tasks are quite difficult. In doing so, bear in mind that: (1) often information security administrators and ISU management do not always have a clear budget for ISS; (2) do not have clear information on the ratio of cost of attack/size of losses; (3) the pans for financing the information security system in the short term is not always defined.

Despite the fact that investments in information security are constantly being given great attention by practitioners and the academic community, the number of cyber incidents, violations of the information security perimeter, and unauthorized intrusions into information systems is steadily increasing.

In most cases, this is due to a lack of understanding of investment strategies in the information security of OIN. And this, in turn, leads to the adoption of erroneous decisions. Such solutions will not be viable in terms of cost/benefit ratio. This is due to the fact that attempts to correct potential vulnerabilities of OIN information

systems in "manual mode" in order to avoid information security violations often leads to excessive investments in information security systems.

Let us consider this situation. An investor (player 1 or *RG1*) in the field of information security (IS) from a state where a stronger currency (*VL1*) is used in monetary circulation, having free financial resources (hereinafter *FRE*), strives to accept the most desirable options for its placement in information protection technology, for example, for a national cyber monitoring center in another country. To do this, he must choose a counterparty (player 2 or *RG2*). The counterparty uses a weaker currency by default $-VL2$. This situation is typical, for example, when investing in projects to create national information security centers in developing countries.

Investors need to assess the priority of investing their financial resources in such areas of development and relevant technologies that provide IS of OIN (for example, an information security situation center or a monitoring center) as: (1) ensuring the cybernetic stability of OIN; (2) innovative technologies in the tasks of monitoring the risk indicators of the implementation of information threats and ensuring the required level of information security; (3) culture of information security at OIN; (4) information security of the network infrastructure; (5) security of software (software); (6) security of data processing technologies; (7) and others.

The problem of studying strategies for investing in the ISS of OIN can have many different, nonequivalent mathematical formulations. Depending on the setting of the task and the mathematical apparatus used for their analysis, various approaches can be used. This proves the importance of a flexible approach to the mathematical formulation of the problem.

Based on the analysis of the attractiveness of investment strategies for different investors representing different states, as indicated above, the mathematical apparatus of game theory was used.

Conceptually, the interaction of players (hereinafter denoted as $RG1$ and $RG2$) will be described this way: $RG1$, having some free financial resources (*FRE*), increases them at time *g1* (g1 is the rate of growth of resources *RG1).* Further, for example, using the DSS, it is decided what part of these resources will be directed to active operations to create a national center for monitoring information security and cyber threats. These operations involve the allocation of resources $RG1$ in investment projects as a part of building an information security system for OIN. The part of the resource is used to pay off the debt that exists at *RG1* in this period of time. We believe that it does the same with respect to *RG2*. In the proposed model, the following assumptions are made:

(a) $RG1$ FiR h valued at $VL1$ (currency 1);
(b) $RG2$ controls the FIR of q valued at $VL2$ (currency 2);
(c) throughout the interaction, the ratio of $VL1$ to $VL2$ (the exchange rate) k_d remains constant; player $RG1$ has no idea of the financial resources of $RG2$. He has only information that they belong to the fuzzy set $\{X, m(.)\}$. Here X is a subset $R_+ m(.)-$ function of the second investor's FiR $q(0)$ value belonging to the set $X, m(q(0)) \in [0, 1]$ for $q(0) \in X$. In addition, at each moment t $(t \in [0, T])$ his states are known $h(\tau)$ for $\tau \leq t$. The following conditions

are satisfied: $h(\tau) > 0$ when condition $h(\tau) \succ 0$ is satisfied with reliability $\geq p_0 (0 \leq p_0 \leq 1)$ and $h(\tau) \prec 0$ when condition $h(\tau) \prec 0$ is satisfied with reliability $\prec p_0$, and the values of realizations of strategy $u(\tau)(\tau \leq t)$, allocated for interaction with $RG2$ are also known.

4.1 A Model Describing Player Interactions

Further in the text of the paper, we will assume, respectively, that the players are designated as RG1 and RG2. The players have their own resources, which they are ready to invest either permanently or for a certain period of time in the information security or cybersecurity systems of the company. Players can have active operations. For example, at certain moments of investment, resources must be directed to active operations. As an example of such an active operation from the point of view of the classical approach of the interacting parties of the investment process, one can point to the mutual repayment of debts that accumulate among the parties during the implementation of information protection projects. The interactions between players and their resources are described by the following system of equations:

$$\begin{aligned} dh(t)/dt = & -h^+(t) + g_1 \cdot h^+(t) \\ & + [(1 - f_1(t)) \cdot (m_1(t) + p_1(t)) - 1] \cdot u(t) \cdot g_1(t) \cdot h^+(t) \\ & + [1 - (m_2(t) + p_2(t)) \cdot (1 - f_2(t))] \cdot v(t) \cdot g_2(t) \cdot \frac{q^+(t)}{k_d}; \end{aligned} \quad (1)$$

$$\begin{aligned} dq(t)/dt = & -q^+(t) + g_2 \cdot q^+(t) \\ & + [(1 - f_2(t)) \cdot (m_2(t) + p_2(t)) - 1] \cdot v(t)(t) \cdot q^+(t) \\ & + [1 - (m_1(t) + p_1(t)) \cdot (1 - f_1(t))] \cdot u(t) \cdot g_1(t) \cdot h^+(t) \cdot k_d. \end{aligned} \quad (2)$$

and

$$h^+ = \left\{ \begin{matrix} h, h \geq 0 \\ 0, h \prec 0 \end{matrix} \right\}., q^+ = \left\{ \begin{matrix} q, q \geq 0 \\ 0, q \prec 0 \end{matrix} \right\}.$$

Thus, at time t, the value of $dh(t)/dtRG1$ (in $VL1$) is equal to:

$g_1(t) \cdot h^+(t)$, the amount of interest $m_1(t) \cdot (1 - f_1(t)) \cdot u(t) \cdot g_1(t) \cdot h^+(t)$ for the invested FiR $RG1$;

$(1 - f_1(t)) \cdot u(t) \cdot g_1(t) \cdot h^+(t)$—the size of the invested FiR of $RG1$;

$p_1(t) \cdot (1 - f_1(t)) \cdot u(t) \cdot g_1(t) \cdot h^+(t)$—the value, which characterizes the share of the "returned" investment resource (hereinafter InR) $RG1$;

$(1 - f_1(t)) \cdot u(t) \cdot g_1(t) \cdot h^+(t)$—R $RG1$ for ISS;

$\left[\{(1 - p_2(t)) \cdot (1 - (f_2(t)/k_d))\} \cdot v(t) \cdot g_2(t) \cdot q^+(t)\right]$—the value of the "unrecovered" asset (investment) $RG2$ (in $VL1$);

$\left[\{f_2(t)/k_d\} \cdot v(t) \cdot g_2(t) \cdot q^+(t)\right]$—resources to repay the debt $RG2$ to $RG1$.;

$u(t) \cdot f_1(t) \cdot g_1(t) \cdot h^+(t)$—the resource allocated to pay off the debts incurred by $RG1$ at time t to $RG2$;

$u(t) \cdot (1 - f_1(t)) \cdot g_1(t) \cdot h^+(t)$—the resource allocated to carry out the investment in ISS OIN at time t;

$\{g_2(t) \cdot (1 - (f_2(t)/k_d))\} \cdot v(t) \cdot g_2(t) \cdot q^+(t)$—interest charge for the InR $of\, RG2$;

$\{(1 - (f_2(t)/k_d))\} \cdot v(t) \cdot g_2(t) \cdot q^+(t)$—InR $of\, RG2$.

$h^+(t)$ is the value subtracted from this sum.

Similar terms will be for expression (2). Thus, the value of $dq(t)/dt$ (in $VL2$) at time t is equal to the sum of such terms:

$g_2(t) \cdot q^+(t)$, values of interest $m_2(t) \cdot (1 - f_2(t)) \cdot v(t) \cdot g_2(t) \cdot q^+(t)$ for invested FiR $RG2$;

$(1 - f_2(t)) \cdot v(t) \cdot g_2(t) \cdot q^+(t)$—the size of InR $of\, RG2$;

$p_2(t) \cdot (1 - f_2(t)) \cdot v(t) \cdot g_2(t) \cdot q^+(t)$—the value characterizing the share of the "returned" InR $RG1$ to $RG2$;

$(1 - f_2(t)) \cdot v(t) \cdot g_2(t) \cdot q^+(t)$—InR $RG2$ on ISS;

$(1 - p_1(t)) \cdot (1 - f_1(t)) \cdot u(t) \cdot k_d(t) \cdot g_1(t) \cdot h^+(t)$—is the value of the "unreturned" asset (investment) in $RG1$ by player $RG2$;

$u(t) \cdot f_1(t) \cdot k_d(t) \cdot g_1(t) \cdot h^+(t)$—is the value characterizing the repayment of $RG1$ debt to $RG2$;

$v(t) \cdot f_2(t) \cdot g_2(t) \cdot q^+(t)$—is the amount allocated to $RG2$ to repay the debt it has owed to $RG1$ at time t;

$(1 - f_2(t)) \cdot v(t) \cdot g_2(t) \cdot q^+(t)$—the value allocated by $RG2$ to make investments in ISS at time t;

$m_1(t) \cdot (1 - f_1(t)) \cdot k_d(t) \cdot u(t) \cdot g_1(t) \cdot h^+(t)$—is the percentage charge for the InR $RG1$;

$(1 - f_1(t)) \cdot u(t) \cdot g_1(t) \cdot h^+(t)$—InR $of\, RG1$.

The value of $q^+(t)$ is subtracted from this amount;

The interaction ends when the conditions are met:

$$(h(t), q(t)) \in S_0 = \big\{\big((h(t),\ q(t) \in S_0^*, \text{with reliability} \geq p_0, (h(t), q(t)) \in S_1^*, \text{with reliability} \geq p_0\big\} \tag{3}$$

$$(h(t), q(t)) \in F_0 = \big\{\big(((h(t),\ q(t)) \in S_0^*, \text{with reliability} \prec p_0, (h(t), q(t)) \in S_1^*, \text{with reliability} \prec p_0\big\} \tag{4}$$

where

$$S_0^* = \big\{(h, q) : (h, q) \in R_+^2, h > 0\big\},$$
$$S_1^* = \big\{(h, q) : (h, q) \in R_+^2, q = 0\big\},$$

If it turns out that condition (Eq 3) is fulfilled, then we will say that in the process of investing in the ISS of OIN has achieved the desired result with confidence $p \geq p_0$ and the procedure is completed.

If it turns out that the condition (Eq 4) is fulfilled, then we will say that in the procedure of investing in the SPI, the SPI has achieved the desired result with confidence $p \succ 1 - p_0$ and the process is completed.

If both condition (Eq 3) and condition (Eq 4) are not carried out, then the process of investing in ISS of OIN continues further.

Define the function $F(.) : X \to R_+$, $F(x) = \{\sup m(y) \text{ for } y \leq x\}$.

Denote by Φ the set of such functions, by $T^* = [0, T]$,—the time segment.

Strategy of $RG1$ is the rule that allows him to determine the amount of FiR based on the available information, that $RG1$ allocates to invest in ISS of OIN.

The second player $RG2$ chooses his plan $v(.)$ on the base of any information that is available.

The first player *RG1* tries to figure out the set of his initial states. The set of such states and the preference set of the first player W_1 are presented [21]. Then, the plans of the first player will be called as his optimal plans. The goal of the first player is $RG1$ to find the preference set and to find his strategies. Applying them he will obtain the fulfillment of condition (3).

The formulated game model corresponds to the classification of the decision making theory and the problem of decision making in terms of fuzzy information. In order to describe the preference sets *of* $RG1$ it is crucial to include the value:

$$\phi(0) = \inf\{\phi'\},$$
$$F(\phi') \geq p_0.$$

Further, the solutions are made, i.e., "preference" sets Z_1 and optimal strategies $u_*(.)$ with all game parameter ratios. It is the set of such initial states $(h(0), \phi(0))$.I If the game starts from them, there exists a plan of $RG1$, which, for any realizations of plan $RG2$, "leads" at time t of the system state $(h(0), \phi(0))$ in which condition (3) will be accomplished. In that case, $RG2$ is lack of the strategy that can "lead" to the fulfillment of condition (4), at one of the previous times.

The paper touches on the following issue. How to determine the time of possible loss of capitals (i.e., INR) with a given degree of confidence using information about the initial FiR(capitals), the exchange rate, the growth rate of resources of *RG1* and *RG2*, percentage rates on allocated capitals, levels of payable and receivable debts, fuzzy information on FiR of the second player related to the use of new information security technologies, and cybersecurity?

We have used the apparatus of the theory of multistep quality games as a toolkit to find out the problem [21, 22]. This method allows determining the areas of possible initial states of resources (capitals) of parties. Therefore, we assume that the objects have the following property: if the interaction begins from these states, then the loss of capital is possible at time t either by one side of the party or by the other, and it gives the answer to the given question. A multistep quality game with two quality surfaces

has been defined in order to find such areas. The solution involves the determination of the preferences of the parties. Furthermore, the optimal strategies of the parties have been revealed while investing in the OIN information security system (using the example of an international situational center on information security).

Within the framework of the research, we have made an attempt to consider a plain option for interaction allowing us to draw qualitative conclusions about the financial condition of the subjects. And it can also be quite easily applied algorithmically in any high-level programming language.

4.2 The Solution of the Problem

The solution of the problem consists of finding the preference set and its optimal strategies (the problem from the first ally player's point of view [21, 23]). Similarly, the problem is set from the point of view of the second ally player. Due to the symmetry of the problem statement, it is sufficient to solve the problem from the viewpoint of the first allied player. Solving the problem from the second allied player's point of view is similar.

The solution to **Problem 1** is found using the toolkit of the theory of multistage games with complete information [24], which allows finding the solution to the game for various ratios of the game parameters. Let us give the solution to the game, i.e., sets of preferences and optimal strategies $RG1$.

Suppose that the conditions are carried out at any time t:

$$g_1(t) = g_1; \quad g_2(t) = g_2;$$
$$f_1(t) = f_1; \quad f_2(t) = f_2; \quad p_1(t) = p_1; \quad p_2(t) = p_2.$$

Denote through z_1 & z_2 the following quantities:

$$z_1 = (1 - f_1) \cdot (m_1 + p_1) - 1, z_2 = (1 - f_2) \cdot (m_2 + p_2) - 1.$$

There are four possible cases:

$$a) \quad z_1 \geq 0; \quad z_2 \geq 0; \quad b) \quad z_1 \prec 0; \quad z_2 \prec 0;$$
$$c) \quad z_1 \succ 0; \quad z_2 \leq 0; \quad d) \quad z_1 \leq 0; \quad z_2 \succ 0.$$

Let us give the solution to the game, i.e., a set of preferences W_1 and optimal strategies of the first player.

For the case a) we have:

$$W_1 = \left\{(h(0), \phi(0)) : (h, \phi) \in \text{int } R_+^2, \ \phi(0) \prec w^* \cdot h(0)\right\} \tag{5}$$

for

$$w^* = \left\{ \frac{-[z_2 \cdot g_2 + g_2 - z_1 \cdot g_1 - g_1]/[2z_2 \cdot g_2] +}{\sqrt[2]{\{[z_2 \cdot g_2 + g_2 - z_1 \cdot g_1 - g_1]/2z_2 \cdot g_2\}^2 + (z_1 \cdot g_1)/(z_2 \cdot g_2)}} \right\}; \quad (6)$$

$u_*(h, \phi) = \{1,\ \phi < w \cdot h,\ (h, \phi) \in \text{int}R_+^2\}$ is not defined, else}.
In case (b) and (c) the quantity of W_1 is empty.
In case (c) and $g_2 > g_1 + z_1 \cdot g_1$ we get

$$W_1 = \{(h(0), \phi(0)) : (h(0), \phi(0)) \in \text{int}R_+^2,\ \phi(0) < \delta \cdot h(0)\}. \quad (7)$$

For $\delta = (z_1 \cdot g_1)/(g_2 - z_1 \cdot g_1 - g_1)$; $u_*(h, \phi) = \{1, \phi < \delta \cdot h,\ (h, \phi) \in \text{int}R_+^2\}$, is not defined, else
In case c) and $g_2 \le g_1 + z_1 \cdot g_1$ we have:

$$W_1 = \text{int}R_+^2,\ u_*(h, \phi) = \{1, (h, \phi) \in \text{int}R_+^2\}, \quad (8)$$

is not defined, else.
Problem 2 is defined symmetrically (from the second ally player's standpoint).

5 Imitation (Simulation) Experiment

In order to illustrate the results of the calculation, they are carried out for the data that are adopted for the information protection and cybersecurity systems of the situation center of the Ministry of Transport of Kazakhstan. This center is a vivid example of investment interaction in the field of information protection and cybersecurity of many states, including Kazakhstan, China, countries of the European Union, etc. Simulation modeling is performed in the MATLAB package. Some of the outcomes obtained during the simulation are illustrated in Figs. 1 and 2.

In the graphs Figs. 1 and 2, h-axis means "million" $ (in our case *VL1*). In Fig. 1, the tangent of the angle is equal to "2." In Fig. 2, the tangent of the angle is "3." Axis ϕ means million in local currency (e.g., Kazakhstan tenge or Ukraine hryvnia). In Figs. 1 and 2, the trajectories of investors are illustrated. In Fig. 1, the trajectory is in the preference area of the second investor and shown by a green dashed line with round green markers (line number 1). In Fig. 2, the trajectory of investors follows the ray of balance, which is the boundary of the preference area of the first investor, shown by a blue dotted line with markers in the form of rhombuses (line number 1). The balance beams are shown in Figs. 1 and 2 in red solid line with red round markers.

Figure 3 describes an example of the implementation of the proposed model on the online DSS platform.

Figure 3a shows an example of a solution describing the ratio of players' resources for a situation in which the trajectory (shown by the yellow line) of the first investor's movement is located in his area of preference.

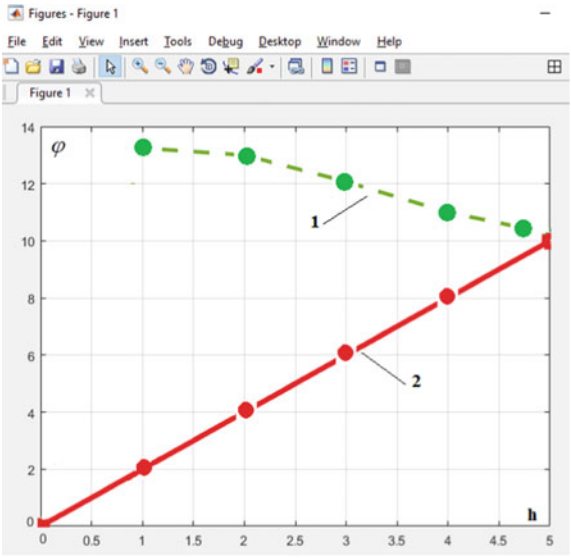


Fig. 1 Computational experiment No.1

1- Player trajectory; 2- Beam of balance

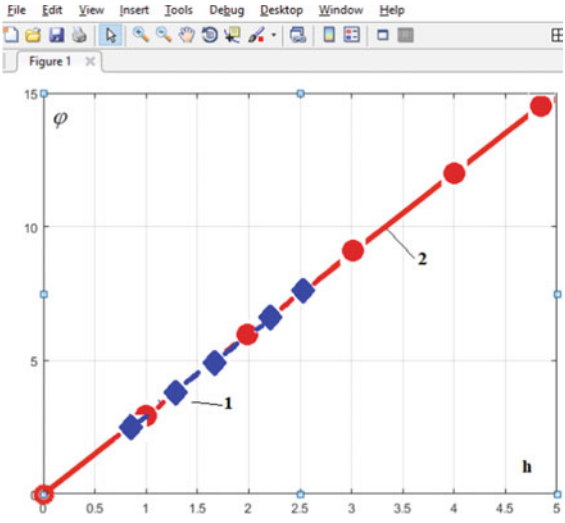


Fig. 2 Computational experiment No.2

1- Trajectory of the player's movement; 2- Beam of balance

Figure 3b shows histograms of step-by-step changes in the size of the FiR of the players for the first investor in red for the second in green.

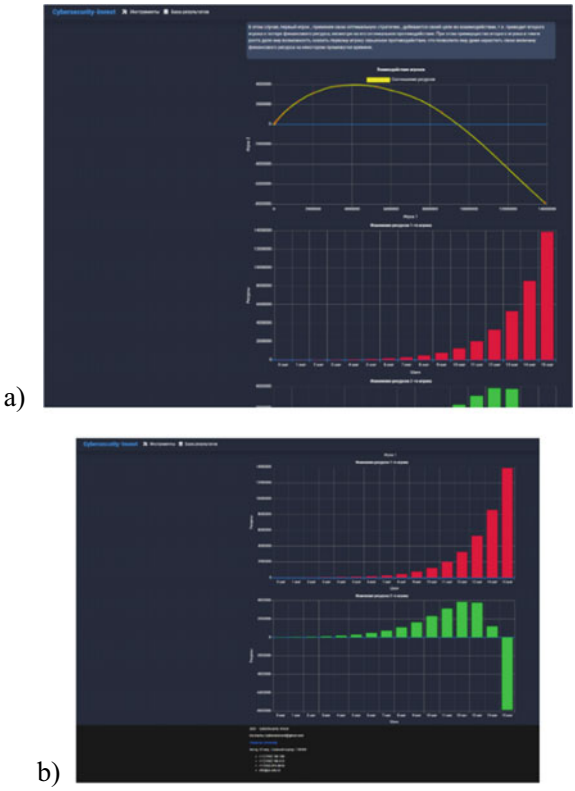


Fig. 3 General view of the DSS

6 Discussion

Thus, a new model is presented that describes the investment process in information security and cybersecurity systems of an informatization object. The model is based on the apparatus of game theory. The graphs in Figs. 1 and 2 are the results demonstrating the effectiveness and functionality of the game model. The graphs in Fig. 1 correspond to simulation experiment number 1. For this experiment, a result was obtained that would be typical for situations when RG2 player used the non-optimal behavior of RG1 at the initial time. If the trajectory moves under the balance beam (red line), then on the contrary, RG1 used the non-optimal behavior of RG2. Such a graph is not shown in the paper. But a situation is possible when the actions of the players and their investment strategies satisfy both parties. It will be a balanced investment strategy for both parties. In the case of a balanced strategy, both players and their investment trajectories will coincide with the balance beam.

To confirm the functionality of the proposed model, the obtained results were compared with other approaches that various authors proposed [6, 8–10, 25]. The outcomes achieved by us were close enough. However, the complexity of calculations

is less in accordance with our model. It has taken 11–14% of less time to obtain data than, for example, the models described in the works [8–10].

7 Conclusions

As part of the study, the following tasks have been solved:

A model of the process of continuous mutual investment of projects in the sphere of information security and information protection within the framework of a scheme with fuzzy information has been developed. The model has served as the core of the computing module of the intelligent support system in the analysis of various investment strategies in information protection systems of information objects. The model is based on the application of the tools of a quality game surfaces in case when the information support of investors is given by means of fuzzy sets;

Simulation experiments have been carried out in the MATLAB simulation environment;

The online platform of the decision making support system for investors is described while choosing a strategy of investing in ISS of OIN.

Acknowledgements This research was funded by Committee of Science MES RK, grant number AP08855887—«Development of intelligent decision support system in the process of investing in cybernetic security systems».

References

1. H.S. Lallie, L.A. Shepherd, J.R. Nurse, A. Erola, G. Epiphaniou, C. Maple, X. Bellekens, Cyber security in the age of covid-19: a timeline and analysis of cyber-crime and cyber-attacks during the pandemic. Comput. Secur. **105**, 102248 (2021)
2. M.K. Kagita, N. Thilakarathne, T.R. Gadekallu, P.K.R. Maddikunta, S. Singh, A Review on cyber crimes on the Internet of Things, (2020). arXiv preprint arXiv:2009.05708
3. L. Gordon, M. Loeb, W. Lucyshyn, Information security expenditures and real options: a wait-and-see approach, Comput. Secur. J. **19**(2), 1–7 (2003)
4. L.A. Gordon, M.P. Loeb, L. Zhou, Information Segmentation and Investing in cybersecurity. J. Inf. Secur. **12**(1), 115–136 (2020)
5. D. Kosutic, F. Pigni, Cybersecurity: investing for competitive outcomes. J. Bus. Strategy. ahead-of-print(ahead-of-print), (2020). https://doi.org/10.1108/JBS-06-2020-0116
6. D. Young, J. Lopez Jr., M. Rice, B. Ramsey, R. McTasney, A framework for incorporating insurance in critical infrastructure cyber risk strategies. Int. J. Crit. Infrastruct. Prot. **14**, 43–57 (2016)
7. A. Yang, Y.J. Kwon, S.-Y.T. Lee, The impact of information sharing legislation on cybersecurity industry. Ind. Manag. Data Syst. **120**(9), 1777–1794 (2020). https://doi.org/10.1108/IMDS-10-2019-0536
8. L.A. Filimonova, N.K. Skvortsova, On issue of algorithm forming for assessing investment attractiveness of region through its technospheric security, in *IOP Conference Series: Materials Science and Engineering—IOP Publishing*, vol. 262, no. 1, p. 012196, (2017). https://doi.org/10.1088/1757-899X/262/1/012196

9. L.A. Gordon et al., The impact of the Sarbanes-Oxley Act on the corporate disclosures of information security activities. J. Account. Public Policy **25**(5), 503–530 (2006). https://doi.org/10.1016/j.jaccpubpol.2006.07.005
10. W. Qin, Z.H.U. Jianming, Research on the game of information security investment based on the Gordon-Loeb model. J. Commun. **39**(2), 174 (2018). https://doi.org/10.11959/j.issn.1000-436x.2018027
11. X. Li, Decision making of optimal investment in information security for complementary enterprises based on game theory. Technol. Anal. Strategic Manage.1–15 (2020). https://doi.org/10.1080/09537325.2020.1841158
12. Y. Li, L. Xu, Cybersecurity investments in a two-echelon supply chain with third-party risk propagation. Int. J. Prod. Res. **59**(4), 1216–1238 (2021)
13. M. Benaroch, Real options models for proactive uncertainty-reducing mitigations and applications in cybersecurity investment decision making. Inf. Syst. Res. **29**(2), 315–340 (2018)
14. K.K.F. Yuen, Towards a cybersecurity investment assessment method using primitive cognitive network process, in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)* . IEEE, 2019), pp. 068–071
15. A. Fielder, E. Panaousis, P. Malacaria, C. Hankin, F. Smeraldi, Decision support approaches for cyber security investment. Decis. Support Syst. **86**, 13–23 (2016)
16. Y. Lee, R. Kauffman, R. Sougstad, Profit-maximizing firm investments in customer information security. Decis. Support Syst. **51**(4), 904–920 (2011)
17. B. Srinidhi, J. Yan, G.K. Tayi, Allocation of resources to cybersecurity: the effect of misalignment of interest between managers and investors. Decis. Support Syst. **75**, 49–62 (2015)
18. B. Akhmetov, V. Lakhno, B. Akhmetov, Z. Alimseitova, Development of sectoral intellectualized expert systems and decision making support systems in cybersecurity, in *Proceedings of the Computational Methods in Systems and Software* (Springer, Cham, 2018), pp. 162–171
19. L. Xu, Y. Li, J. Fu, Cybersecurity investment allocation for a multi-branch firm: modeling and optimization. Mathematics **7**(7), 587 (2019)
20. B.B. Akhmetov, V.A. Lakhno, B.S. Akhmetov, V.P. Malyukov, The choice of protection strategies during the bilinear quality game on cyber security financing. Bull. Nat. Acad. Sci. Repub. Kaz **3**, 6–14 (2018)
21. V. Lakhno, V. Malyukov, N. Gerasymchuk et al., Development of the decision making support system to control a procedure of financial investment. Eastern-Eur. J. Enterp. Technol. **6**(3), 24–41 (2017)
22. R. Casado-Vara, F. Prieto-Castrillo, J.M. Corchado, A game theory approach for cooperative control to improve data quality and false data detection in WSN. Int. J. Robust Nonlinear Control **28**(16), 5087–5102 (2018)
23. A. Agah, S.K. Das, K. Basu, A game theory based approach for security in wireless sensor networks, in *IEEE International Conference on Performance, Computing, and Communications*, (IEEE, 2004), pp. 259–263
24. B. Yang, C. Lai, X. Chen, X. Wu, Y. He, Surface water quality evaluation based on a game theory-based cloud model. Water **10**(4), 510 (2018)
25. L. Gordon, M. Loeb, W. Lucyshyn, L. Zhou, The impact of information sharing on cybersecurity underinvestment: a real options perspective. J. Account. Public Policy **34**(5), 509–519 (2015)

QoS Provisioning in MANET Using Fuzzy-Based Multifactor Multipath Routing Metric

S. Venkatasubramanian, A. Suhasini, and C. Vennila

Abstract Volatile topology is the basic nature of MANET. Haphazardly, the nodes may move which makes and breaks connection between nodes. Because of this nature, providing quality of service (QoS) is a challenging task. For the applications that use real-time data packets delivery, quality of service (QoS) has to be provisioned. This paper discusses a technique to provide QoS in a heterogeneous network with multipath routing is given. The multipath routes are to be discovered primarily using the parameters such as delay, channel occupation, link quality and residual energy. To choose the best routes between the originating node and the target node, a fuzzy logic mechanism is used. This Fuzzy-based multi-parameter and multipath QoS routing (FMMQR) find out the optimal path for data transmission between the two nodes. This technique gives better performance than some of the existing method. The simulation experiments prove our Fuzzy-based multi-parameter and multipath QoS routing (FMMQR) performs better than other methods.

Keywords MANET · Routing · Fuzzy based · Multifactor · Multipath · QoS

S. Venkatasubramanian (✉)
Department of Computer Science and Engineering, Saranathan College of Engineering, Trichy 620012, Tamil Nadu, India
e-mail: veeyes@saranathan.ac.in

A. Suhasini
Department of Computer Science and Engineering, Annamalai University, Annamalainagar 608002, Tamil Nadu, India

C. Vennila
Department of Electronics and Communication Engineering, Saranathan College of Engineering, Trichy 620012, Tamil Nadu, India
e-mail: vennila-ece@saranathan.ac.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_41

1 Introduction

MANET consists of a network of wireless nodes without any access point. Here, the nodes form a self-managed network to forward data packets. Each node communicates through multihop with them. There will be frequent changes in the topology as nodes move in different direction. Thus, the provision of quality of service (QoS) routing protocol becomes a challenging task compared to the conventional networks [1]. MANETs are used in emergency situations such as natural catastrophes, military conflicts and medical facilities among others. With the demand for supporting multimedia services, it is the need of the hour to make MANETs to support QoS. Providing real-time media traffics such as audio and video become problematic in the presence of changing network architecture due to the demand of high data rate and delay limits [2].

The two major solutions for providing QoS in MANETs can be either [3] based on resource reservation called stateful approach. Eg: INSIGNIA [4] which do not reserve resource and uses service differentiation called stateless approach. Eg: SWAN [5]. These techniques are applicable when the node mobility is low and for homogeneous nodes. But, with a reasonable mobility and with the heterogeneous nodes, an alternative light weight approach should be used. One such a technique is the theme of this paper.

2 Related Works

Manoharan et al. [2] proposed an adaptive gateway management (AGM) method, in which an excellent gateway from the gateway candidate nodes has to be selected. It uses metrics like balance energy, signal strength and movement speed to select these. When an elected gateway loses its energy and due to the dynamic unstable topology, it makes it unavailable to other nodes. Adaptive gateway migration method was used to sustain the connectivity between networks. A predefined threshold value is used in the gateway selection method by selecting optimal gateway.

Load balancing and congestion avoidance are the core in the adaptive multipath routing protocol with congestion controlled proposed by Soundararajan et al. [6].The algorithm for determining multipath routes computes fail-safe multiple paths, which supply the principal path with multiple routes to all intermediate nodes on the path. The fail-proof multipath includes nodes for the minimum load, necessary battery power, and appropriate energy. A threshold value is used to see whether the load on a node along the route exceeds a certain threshold. If this is the case, traffic is divided into multiple multipath routes to alleviate the strain on a crowded network.

A routing strategy considering the stability of the link and energy-aware routing protocol with energy monitoring was suggested by De Rango et al. [7], which has tried to account for stability of the link and for minimum energy consumption rate.

To validate the exactness of their proposed strategy, a bi-objective optimization was formulated.

Chandrakant et al. [8] have projected heterogeneity in packet sizes, protection mechanisms, mobility among nodes in MANETs. The security management and communication management are the focus of their work. Diverse encryption techniques such as packet sizes and protocols are used by every node in such a way that every node did not understand security methods of every other node.

Power-aware QoS multipath routing (PAQMR) protocol was proposed by Santhi et al. [9] to eliminate the loop formation in heterogeneous MANET; therefore to reduce congestion in the channel. An enhancement version of ad hoc on demand multipath distance vector (AOMDV) protocol was suggested.

In power-aware rate adjustment with the cross-layer-based routing MAC protocol [10], fuzzy logic system 1 is used for best path selection and rate adjustment were done by fuzzy logic system 2. By fuzzifyng the inputs such as delay and packet loss ratio, state of transmission rate was estimated. By comparing the output of fuzzy logic system 2 with the initial transmission rate of path, the current transmission rate of the path is tuned.

Santhi et al. [11] proposed a multi-constrained QoS routing with mobility prediction based on fuzzy costs (FCMQR). Their goal was to create a cost metric that took into account available bandwidth, end-to-end delay and hop count. Our paper compares the performance with this work.

Suraki et al. [12] proposed a cross-layer approach at transport, network and MAC layers in which fuzzy logic system is used in intermediate and destination nodes as a dynamic tool for controlling the congestion problem. The fuzzy rule was drafted based on the packet length, buffer size and congestion level.

Chen et al. [13] in their paper proposed how an AODV can adopt to the topological change and devised TA-AODV routing protocol. To provide QoS, this can adapt to high-speed node mobility. A stable path selection technique is used in this protocol to give connection stability probability between nodes by using path selection characteristics such as residual energy, available bandwidth and queue length.

Yas et al. [14] proposed "A trusted MANET routing algorithm based on fuzzy logic", in which the routing algorithm relies on two main principles as trust level and shortest path (hops count).

Mukesh KumarGarg et al. [15] have taken fuzzy of security Level of trust and a value of trust as main parameters to decide the routing.

Ran et al. [16] have taken the states of the nodes and form a blockchain and the nodes that match the QoS restrictions are filtered out via a smart contract on the blockchain. S. Venkatasubramanian et al. [17] have considered the multi-factors in MANET to provide QOS. But the fuzziness of the values are considered in that work.

3 System Design and Protocol

For MANET, a routing metric with support for QoS is designed in a way each node along with a congested link should use optimal available bandwidth and least end-to-end latency. So our algorithm should rely on the individual links weightage which is based on parameters including channel busyness, end-to-end delay, link quality and node energy.

3.1 Link Quality Estimation

The nodes in the network compute its quality of links. During a slot T_{win}, if N_h is the number of HELLO message packets received and P_h is the HELLO message packets percentage received during the previous time duration, then the quality of the link L_{qual} is

$$L_{\text{qual}} = \beta.P_h + (1-\beta).N_h \tag{1}$$

where β represents the degree of weighting factor between 0 and 1. Then, the route quality is the summation of all link quality along the path.

3.2 End-to-End Delay Estimation

RREQ and RREP report a node with hop count $H_{\text{count.}}$ The number of packets comprised in each up and down stream is $2H_{\text{count}}$. The node uses the latency between RREQ and RREP and finds the average delay D_{avg} of the packets received. The selection of route also depends in the delay and transmits data packets only when the average is within the threshold level.

3.3 Channel Occupancy Estimation

The channel occupancy is the time and a node utilizes the channel for its transmission. For this estimation, the packets such as request-to-send (RTS), clear-to-send (CTS), data and acknowledgement (ACK) of the IEEE 802.11 MAC with the distributed coordination function (DCF) are used. Then, the channel occupation will be

$$C_{\text{occ}} = t_{\text{RTS}} + t_{\text{CTS}} + 3t_{\text{SIFS}} + t_{\text{acc}} \tag{2}$$

where *CTS* and *RTS* times are t_{CTS} and t_{RTS}, respectively, and s the *SIFS* period is t_{SIFS}. The time taken during contention is t_{acc}. If congestion occurs but is not handled, it can diminish a link's capacity and cause it to become congested.

3.4 Residual Energy

Whenever a node communicates, certain amount of battery power will be drained. Owing to the limited energy resources for communication will have an impact on the lifetime of the network and nodes [18].

The energy consumed for transmitting RTS and CTS are given by:

$$\mathrm{Energy}^{\mathrm{RTS}} = r * t_N^L * p_{\mathrm{RTS}} \tag{3}$$

$$\mathrm{Energy}^{\mathrm{CTS}} = r * t_N^L * p \tag{4}$$

where r denotes the receiver's threshold, t_{N} is the maximum transmission range, L denotes the path loss exponent and p_{RTS} and p_{CTS} denote the RTS and CTS transmission times, respectively. The source uses Eq. 5 to compute the overall energy consumption of each path.

$$\text{Total Energy } E = \mathrm{Energy}^{\mathrm{RTS}} j + \mathrm{Energy}^{\mathrm{CTS}} j \tag{5}$$

The residual battery power across each route is determined using the battery power consumed at each node along the path.

$$R_i^b = \sum_{\in M_L} R_i \tag{6}$$

where

$$R_{\mathrm{i}} = \text{Initial Energy}_i^E - (\mathrm{Energy}_i^{\mathrm{RTS}} + \mathrm{Energy}_i^{\mathrm{CTS}}) \tag{7}$$

The number of nodes in the chosen path and the path with the most leftover energy are denoted by M_L. The W for the link from node i to a specific neighbouring node for an intermediate node i having established transmission with several of its neighbours is given by

$$W = \left(L_{\mathrm{qual}} + C_{\mathrm{occ}} + R_i^b\right) / D_{\mathrm{avg}} \tag{8}$$

Each node calculates the weight of its entire links and when an incoming RREQ packet comes with the previous nodes' weight W1, the current node adds its weight to

Table 1 Link weight table maintained at each node

Neighbour node	Destination	C_{occ}	Link weight
1	*D*	24%	0.7
7	*D*	32%	0.72

it. When the RREP packets are sent to the source from the target node, the originating node, which initiates the route discovery process, will receive response from multiple paths. It stores the response in the local table a shown (Table 1)

The same table is maintained in all the nodes. The node will use the link with best link weight. If that link breaks due to mobility of the node, then the node selects the alternative link by selecting the next best metric from table. Channel occupation due to MAC contention ($\boldsymbol{C_{occ}}$) is an important parameter used in this paper for deciding about the link congestion. A threshold value is chosen for this parameter depends upon the nature of the data packet. If the $\boldsymbol{C_{occ}}$ value is below the threshold, then it is a normal traffic. But if near to the threshold value, then the node changes the Bluetooth interface from "sniff mode" to "connected mode (active)" and starts sending the data packets via Bluetooth interface to any of the neighbouring node from the table. This switching occurs only for the packets marked with real-time data. The numbers indicated in the Fig. 1 are the paths to reach the destination node.

A. *Route Request*

The route finding process starts with the source node sending the RREQ packet as shown in Fig. 2.

Few new fields are incorporated in the format of AOMDV, which includes link quality, average path delay, channel occupancy and residual energy of a node. When

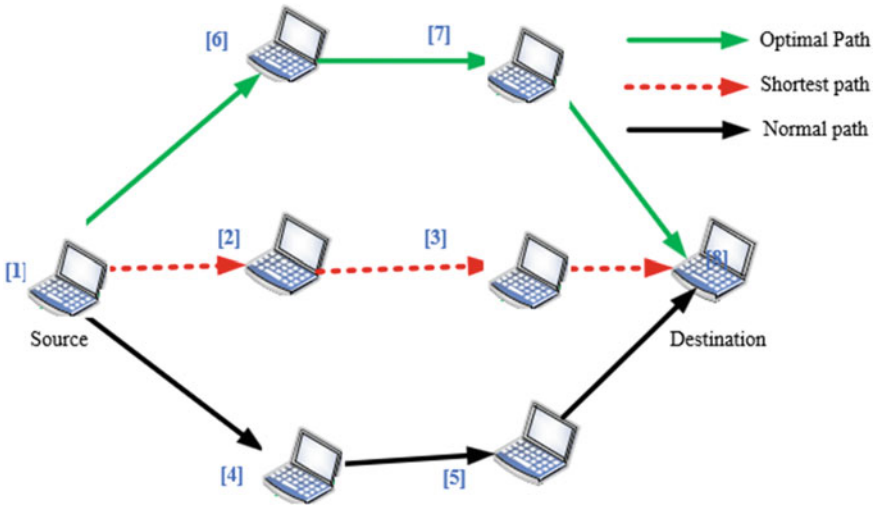


Fig. 1 Optimal path selection

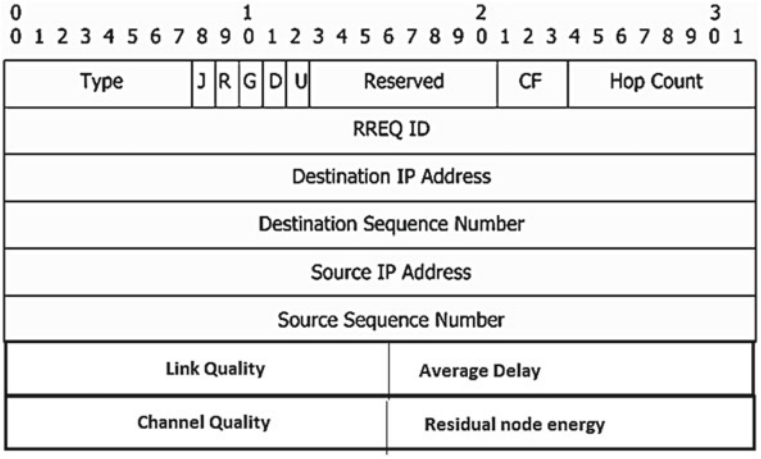


Fig. 2 Format of modified RREQ packet

an intervening node gets a RREQ, FMMQR updates the quality of the link, average path delay, channel occupancy and residual energy values in the RREQ field and then sends a new RREQto its one-hop neighbour.

The QoS-aware route discovery starts with the node S broadcasting RREQ. When a neighbour node receives an RREQ packet, it calculates the required metrics using the estimations stated in the preceding section.

The weight W_{RI} of the node $H1$ is computed using (8).

$$RREQ_{HI} \xrightarrow{W_{R1}} H2$$

In a similar manner, $H2$ computes its weight and adds it to the $H1$ weight. The packet is subsequently forwarded with the additional weight.

$$RREQ_{H2} \xrightarrow{W_{R1}+W_{R2}} H3$$

Finally, the aggregate of node weights along the path reaches the target node.

$$RREQ_{H3} \xrightarrow{W_{R1}+W_{R2}+W_{R3}} D$$

B . ***Route Reply***

The target node D unicast the route reply packet to the upstream neighbour node$H3$, together with the overall node weight.

$$RREP \xrightarrow{W_{R1}+W_{R2}+W_{R3}} H3$$

Now, $H3$ calculates the cost depending on the data provided by $RREP$ as

$$C_{R3} = (W_{R1} + W_{R2} + W_{R3}) - (W_{R1} + W_{R2}) \tag{9}$$

The intervening hosts calculate its cost by the same method. On reception of $RREP$, the source chooses the route with the best cost value out of all the options.

Based on simulation outcomes in the next section, it is seen that these calculations will never raises the overhead and thereby will not be reducing the network performance.

4 Fuzzy-Based QoS Provisioning

This paper proposes a protocol which is designed to select a route with an objective of (i) selecting a route to reduce the end-to-end delay (ii) increase delivery of packets and (iii) maximize the path lifetime. To achieve these objectives, various metrics have been used to come up with a single cost metric (C) is end point latency (D), count of hops (N), channel occupancy (Cocc) and node residual energy (REi). This is the novel work proposed in this paper.

The cost function C_f and the other metrics have the following relationship:

The three key methods such as fuzzification, fuzzy inference and defuzzification as depicted in Fig. 3 are used in the fuzzy logic system. The data that our fuzzy logic system accepts are end-to-end delay, channel occupancy, link quality and node's energy.

The IF–THEN structure forms the fuzzy rules. The AND operator is used to combined the inputs.

The rule mapping example is as follows:

If (quality of link is "High") AND (delay is "High").

AND (channel occupancy is "Low") AND (residual energy is "High") Then weight is "Medium".

There are four input variable and the number of feasible fuzzy inference rules for each of the three linguistic states is 4* 3 * 3 = 36. Table 2 lists a few of the fuzzy rules utilized in the fuzzy controller.

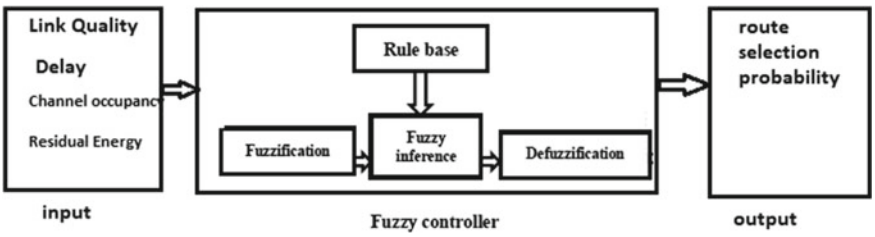


Fig. 3 Fuzzy logic system key processes

Table 2 Fuzzy set rules

Rule#1	Link quality	Delay	Channel occupancy	Residual energy	Output
1	Low	Low	Low	Low	Bad
2	Low	Medium	Low	Medium	Bad
3	Medium	Medium	Medium	Medium	Medium
4	High	Low	Low	Medium	High
5	High	High	Low	High	Bad
6	Medium	Medium	Low	High	High
7	High	Medium	High	Medium	Medium
8	Low	Low	Low	Medium	Medium

Defuzzification is a method for obtaining a plump value as a representation value from a fuzzy collection. The type of defuzzification technique taken in this paper is the centroid of area.

$$x^* = \frac{\int x\mu_A(x)\mathrm{d}x}{\int \mu_A(x)\mathrm{d}x}$$

where $\mu_A(x) =$ the membership function's aggregated output.

The algorithm for route request and reply is as follows:

1. Each node calculates its energy, delay to neighbour, link quality
2. And initial channel occupancy.
3. If a node's buffer got a route to destination then goto step 7
4. Else, broadcast RREQ (with packet format as shown Fig. 2)
5. If not destination node, stores the required parameters in the RREQ packet, computes the fuzzy value and send the broadcast RREQ.
6. If destination node, send RREP packet and computes the fuzzy value.
7. From the fuzzy values received, the node selects one path and keeps the next choice in reservation.
8. Emit the data packet
9. If link breaks, choose the next path stored.

5 Simulation Experiment and Results

5.1 Setup and Settings for Simulation

NS2 tool was used in the simulation of the proposed scheme. The channel capacity of node is set to 2 mbps. IEEE 802.11's MAC layer protocol distributed coordination function (DCF) is used. The simulation's settings are summarized in Table 3.

Table 3 Simulation experiment parameters

Nodes	From 25 to 100
Area dimension	1200 × 1100 m
Layer 2	802.11
Coverage range	250 m
Simulation run time	100 s
Data type	CBR
Packet size	512
Mobility model	Random way point
Velocity	5–20 m/s
Pause time	5 s

5.2 Metrics of Performance

The FMMQR protocol is evaluated with the AOMDV [6] and FCMQR [11] protocol**.** The following performance metrics are considered for the evaluation of our FMMQR protocol.

5.2.1 Average End-To-End Delay

The average latency in sending data packets from source node and receiving acknowledgement at the same node.

5.2.2 Average Packet Delivery Ratio

It is the proportion of successfully received valid data packets to total packets delivered (Fig. 4).

From Fig. 5, in our FMMQR, we have noticed a reduction in latency because FMMQR uses multipath routing. Even if one path breaks, it takes alternate path to send the data. Also our FMMQR provides a minimum latency, as it preciously finds the optimal path using fuzzification.

From Fig. 5, it is proved that our FMMQR nodes forward the data via Bluetooth interface, even the link congestion occurs due to burst data traffic. The packet loss in our proposed method is less as the optimal path with good conditions for the packet travel is chosen.

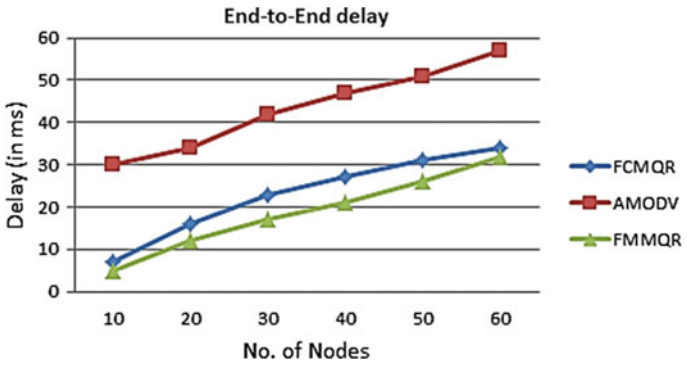


Fig. 4 Nodes versus delay

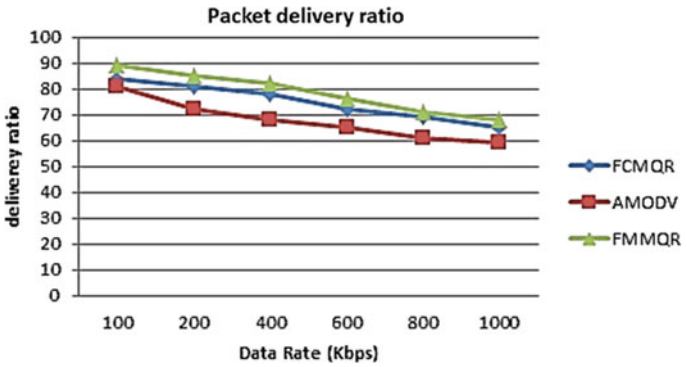


Fig. 5 Data rate versus packet delivery ratio

6 Conclusion

This work proposed a Fuzzy-based multi-parameter and multipath QoS routing (FMMQR) to facilitate quality of service in mobile ad hoc networks by allotting individual links weightage that depends on the metrics such as end-to-end delay, link quality, channel occupation and residual energy. The fuzzy weight cost value helps the routing protocol to select a best path and sending data packets via the congested area is avoided. Thus, provides a balanced traffic and an improved network capacity. The performance of the proposed scheme achieves increased delivery of packet with less latency is also demonstrated by simulation results. This paper can be further extended by formation of clustering and security in transmission of data.

References

1. .P. Zhao, X. Yang, W. Yu, X. Fu, A loose virtual clustering based routing for power heterogeneous MANETs, in *IEEE Transaction Accepted Paper For Future Publication*, (2011)
2. R. Manoharan, E. Ilavarasan, S. Rajarajan, Adaptive gateway management in heterogeneous wireless networks, Int. J. Res. Rev. Ad-Hoc Netw. **1**(2), 2011
3. P. Magula, QMMAC—New QoS model with admission control for mobile ad hoc networks, (2011)
4. S. Venkatasubramanian, V. Gopalan, A quality of service architecture for resource provisioning and rate control in mobile ad-hoc networks. Int. J. Ad-hoc, Sens. Ubiquitous Comput. **1**(3), 2010
5. S. Suri, P. Tomar, Cluster based QoS routing protocol for MANET. Int. J. Comput. Theory Eng. **2**(5), (2010)
6. B. Soundararajan, Adaptive multi-path routing for load balancing in mobile ad-hoc networks. J Comput Sci **8**, 648–655 (2012)
7. F. De Rango, P. Fazio, Link-stability and energy aware routing protocol in distributed wireless networks. IEEE Trans Parallel Distrib Syst **23**(4), 713–726 (2012)
8. N. Chandrakant, D.P. Shenoy, k.R. Venugopal, Middleware services for security in scalable and non-scalable heterogeneous nodes of MANETs. Int. J. Future Gener. Commun. Netw. **4**, (2011)
9. S. Santhi, G.S. Sadasivam, Power aware QoS multipath routing protocol for disaster recovery networks. Int. J. Wirel. Mobile Netw. (IJWMN) **3**(6), (2011)
10. S. Narayanan, R. Thottungal, Cross-layer based routing and power aware rate adjustment (CBR-PARA) MAC protocol for MANET. Res. J. Appl. Sci. Eng. Technol. **11**(5), 460–467, 2015. https://doi.org/10.19026/rjaset.11.1848
11. G. Santhi, A. Nachiappan, Fuzzy-cost based multiconstrained QoS routing with mobility prediction in MANETs. Egyptian Inf J. **13**(1), 19–25 2012. ISSN 1110- 8665, https://doi.org/10.1016/j.eij.2011.12.001
12. M.Y. Suraki, FCLCC: fuzzy cross-layer congestion control in mobile ad hoc networks. IJCSNS Int. J. Comput. Sci. Netw. Secur.**18**(1), 155–165 (2018)
13. Z. Chen, W. Zhou, S. Wu, L. Cheng, An Adaptive on-demand multipath routing protocol with QoS support for high-speed MANET. IEEE Access **8**, 44760–44773 (2020). https://doi.org/10.1109/ACCESS.2020.2978582
14. Q.M. Yas, M. Khalaf, A trusted MANET routing algorithm based on fuzzy logic, in *Applied Computing to Support Industry: Innovation and Technology. ACRIT 2019. Communications in Computer and Information Science*, vol 1174, ed. by M. Khalaf, D. Al-Jumeily, A. Lisitsa (Springer, Cham, 2020). https://doi.org/10.1007/978-3-030-38752-5_15
15. M.K. Garg, N. Singh, P. Verma, Fuzzy rule-based approach for design and analysis of a Trust-based Secure Routing Protocol for MANETs, Procedia Comput. Sci. **132**, 653–658 (2018). ISSN 1877–0509. https://doi.org/10.1016/j.procs.2018.05.064
16. C. Ran, S. Yan, L. Huang et al., An improved AODV routing security algorithm based on blockchain technology in ad hoc network. J Wireless Com Network **2021**, 52 (2021). https://doi.org/10.1186/s13638-021-01938-y
17. S. Venkatasubramanian, V. Gopalan, A QoS-Based robust multipath routing protocol for mobile ad-hoc networks. IACSIT Int. J. Eng. Technol. **1**(5), 2009
18. S. Uma, S.P. Shantharajah, Optimization of power consumption in mobile adhoc network communication systems. **2**(4), (2013). ISSN 2319–9725

IoT-Based Embedded System for Monitorization of Healthcare

M. Thilagaraj, R. Krishna Kumar, I. S. Mary Jency, U. Gokul Krishna, R. S. G. Santhana Prabhu, A. Reshma, and U. Ramani

Abstract Health is consistently a significant worry in each development humanity is progressing regarding innovation. Like the new COVID assault that has destroyed the economy of China to a degree is a model how medical services have happened to significant significance. In such territories where the plague is spread, it is consistently a superior thought to screen these patients utilizing distant well-being observing innovation. This scheme is answerable for gathering beat, internal heat level and heart beat from the patient's body and send the information into IoT Cloud stage by utilizing Wi-Fi-module and ailment of patient put away in the cloud. It empowers the clinically trained professional or approved individual to screen patient's well-being, where the clinical subject matter expert or approved individual can constantly screen the patient's condition on the cloud worker. In our new scheme, the conclusion is to produce effective and health services to patients.

Keywords Internet of Things · Healthcare · Wi-Fi module · Embedded system

1 Introduction

Distant patient monitoring game plan enables perception of patients outside of standard clinical settings (for example, at home), which grows admittance to human administrations workplaces at cut down costs [1, 2]. The centre target of this venture is the plan and usage of a shrewd patient well-being global positioning framework that utilizes sensors to follow persistent well-being and utilization Web to illuminate their friends and family if there should be an occurrence of any issues. The target of creating checking frameworks is to diminish medical services costs [3].

M. Thilagaraj (✉) · R. Krishna Kumar · I. S. Mary Jency · U. Gokul Krishna · R. S. G. Santhana Prabhu · A. Reshma
Karpagam College of Engineering, Coimbatore, India

U. Ramani
K.Ramakrishnan College of Engineering, Trichy, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_42

Specialist's offices ceaselessly require remarkable organization. The information base of every single patients should be useful and adequate. Be that as likewise, there should be an opportunity to be data shirking. Similarly, the lenient data should additionally reinforcing be kept hidden in the occasion. Social protection might be the lion's share basic worry from guaranteeing various countries in the universe. Improving the existing system of patients especially in the more vulnerable pieces of the specific social request which join those old, genuinely. Also, sanely crippled and moreover the persistently wiped out patients might be the fundamental thought will make advanced [4]. On existing framework, those data is recorded in the sign from asserting desk work or investigating general amassing worker. Anyway, all around that data will be congenial on every single staff besides specialists. In this way, we need help proposing another course the spot lenient what is more specialists fit to relate through adaptable demand furthermore Web demand [5].

To specialist's offices, there need help acquisitions to constant screening from guaranteeing patients. Their pulses need help interminably checked. There might be no acquisition on checking those boundaries the moment they trade home [6]. What is all the more. Consequently, there is a chance that the illness may return once more. Tolerant health's data (high-temperature, cardiovascular recurrence, position) will be occasionally estimated and communicated through networker. Time about sending (say every 3 min) could an opportunity be arranged. Checking singular takes in lenient specific edge [7].

Roughly, the standard internal heat level of open minded is 37 °C, while solitary persnickety faculties are hot in as much body temperature is 37 °C. By using an averaging technobabble in a reasonably lengthy timespan, observer could take these limits for patients [8]. Using same arrangement previously, specialist's high-level cell phone, expert may point of view as much the patient's prosperity status. At any of the boundary plunges past the edge regard, he will get an alert notification. Using Android arrangement secured close by patient, then again as much overseer's sharp phone, those open minded could see as much prosperity status [9]. Exactly on schedule, recognizable proof of what's more finding of possibly dangerous accident physiological states for instance, with the end goal that heart strike requires relentless after about patients prosperity imitating trade from centre on home [10]. Examinations demonstrated that 30% of patients for a delivery finding from guaranteeing heart disillusionment need help readmitted no not as much as When inside multiple times for degrees reaching out from 24 to 55% inside 4–6 months [11]. In view of the resistance to such necessities, prosperity checking systems are ceaselessly proposed as a low baby result. Such a game plan includes physiological data that stores, change and also relate through a close by route for model, with the end goal that sharp telephones, singular PCs [12]. Such systems should additionally reinforcing satisfy severe well-being, security, dependability. Also, long stretch constant operation necessities [13]. In the suggested structure, we show a prosperity noticing system that usage of the detecting hubs for social occasion data from enduring individuals, scholarly tip top gauges individuals' prosperity position furthermore gives notion ought to specialists through their adaptable devices' hosting arrangement. The patients will partake in the human administrations strategy Eventually Tom's examining their adaptable

devices and subsequently may right their prosperity information from wherever any event [14].

2 IoT Health Sensors

2.1 Blood Pressure Sensor

The blood pressure sensor is made ward on the Plethysmography speculation. In this, condition of the thumps is additional fundamental in structures everywhere; the heart beat rate is to be followed.

2.2 Internal Heat Level LM35 Sensor

The LM35 plan is definitely smoothed out hotness circuits through yield power, which is straightforwardly similar with the hotness in centigrade.

2.3 Room Temperature Sensor (DHT11)

DHT11 is a sensor for temperature and soddenness which is by and large used. The sensor goes with a dedicated temperature assessment NTC and an eight-cycle microcontroller for getting ready of temperature and tenacity regards in plan (Fig. 1).

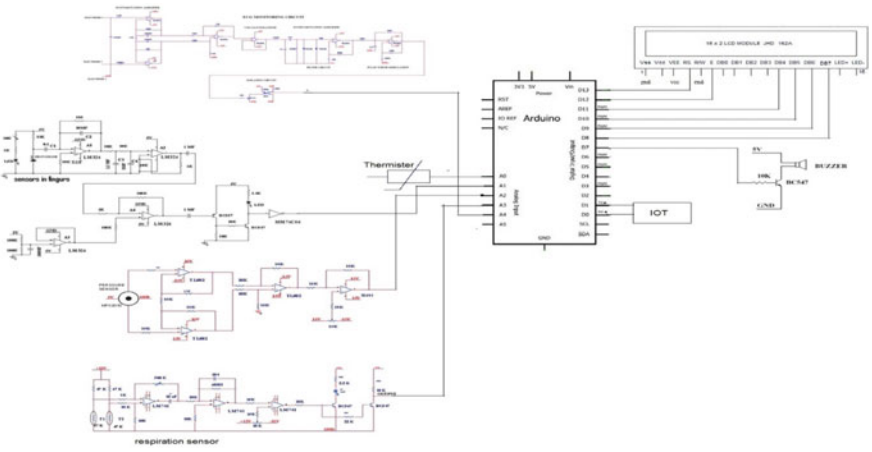


Fig. 1 Proposed health monitoring architecture

2.4 CO_2 Sensor (MQ-135)

The value arranges structures, the MQ-135 gas sensors are used for NH_3, nicotine, benzene, smoke and CO_2 revelation similarly as assessment. The MQ-135 sensor unit goes by means of a modernized join to facilitate engages this feeler to work even without a microcontroller and is profitable for perceiving express gases.

3 Result and Investigation

The overall meaning of clinical consideration programming plans is difficult to overestimate as development pledges to make clinical consideration benefits seriously convincing and ease up the weight put on clinical consideration providers. This is fundamental concerning the developing people and the extension in the quantity of progressing contaminations (Figs. 2, 3 and 4).

The principal points of interest of IoT execution in medical care:

Distant checking: Constant distant seeing through related IoT devices and insightful cautions can dissect afflictions, treat disorders and save lives if there ought to emerge an event of a well-being-related emergency.

Expectation: Smart sensors research clinical issue, lifestyle choices and the environment and propose hindrance measures, which will decrease the occasion of sicknesses and serious states. Reduction of clinical consideration costs: IoT lessens costly visits to subject matter experts and facility affirmations and makes testing more moderate.

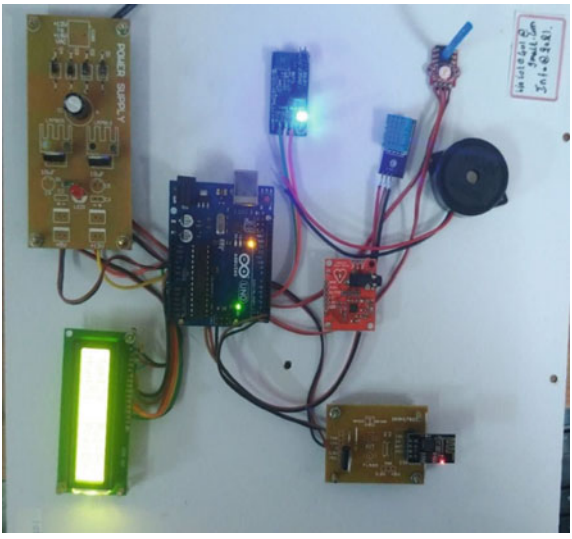


Fig. 2 Model of the working kit

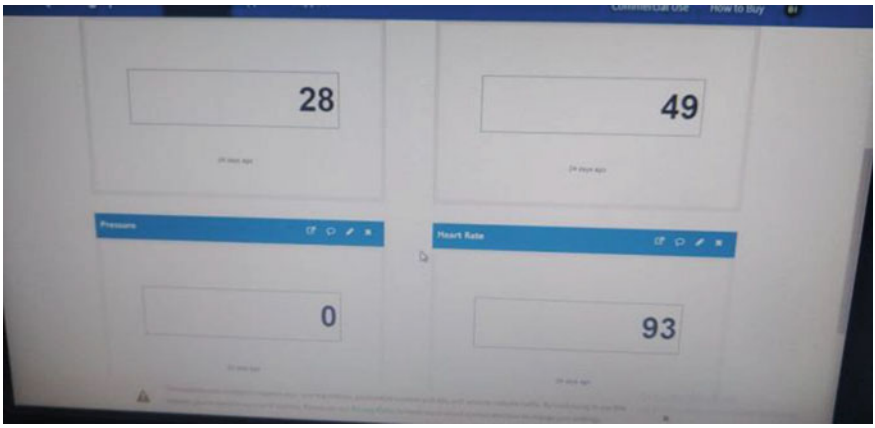


Fig. 3 Health parameters limit

Fig. 4 Simulation of the model

Clinical data accessibility: Accessibility of electronic clinical records license patients to get quality thought and help clinical consideration providers make the right clinical decisions and prevent complexities.

Further developed treatment the leaders: IoT contraptions assist with following the association of drugs and the response to the treatment and diminishing clinical bungle. Further developed clinical benefits the chiefs: Using IoT contraptions, clinical consideration experts can get significant information about stuff and staff reasonability and use it to propose improvements.

Investigation: Since IoT devices can assemble and analyse a colossal proportion of data, they have a high potential for clinical assessment purposes (Fig. 5).

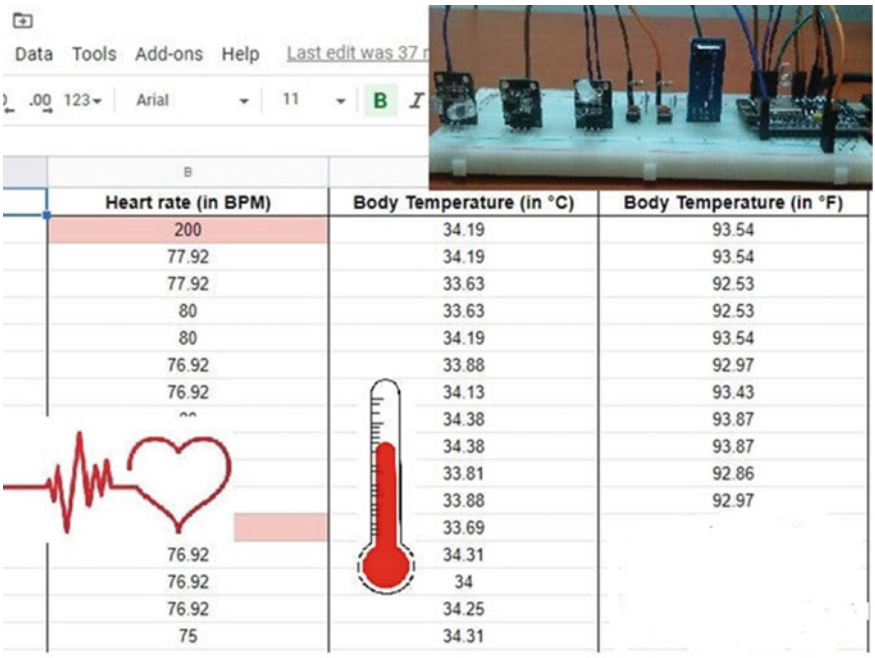


Fig. 5 Real-time monitoring of health parameter

4 Conclusion

Internet of things can be of incredible advantage to medical care, and there are as yet significant difficulties to address before full-scale usage. The dangers and hindrances of utilizing associated gadgets in medical care are as per the following: Security and protection: Security stays a significant concern preventing clients from utilizing IoT innovation for clinical purposes, as medical care checking arrangements can possibly be penetrated or hacked. The break of touchy data about the patient's well-being and area and interfering with sensor information can have grave results, which would counter the advantages of IoT. Risk of disappointment: Failure or bugs in the equipment or even force disappointment can affect the exhibition of sensors and associated hardware putting medical care activities in danger. Also, skirting a booked programming update might be significantly more unsafe than avoiding a specialist check-up. Integration: There is no agreement with respect to IoT conventions and norms, so gadgets delivered by various makers may not function admirably together. The absence of consistency forestalls full-scale mix of IoT, in this manner, restricting its expected adequacy. Cost: While IoT vows to lessen the expense of medical services in the long haul, the expense of its execution in emergency clinics and staff preparing is very high.

In our proposed system, we have implemented and demonstrated the representation for a programmed framework that ensures a consistent observing of different well-being boundaries and expectation of any sort infection to paying continuous visits to the emergency clinics. The new scheme can group in the clinics, and a gigantic measure of information can be gotten and put away in the online data set. Indeed, even the outcomes can be made to be gotten to from portable devices through an application. The framework can be additionally improved further by adding man-made consciousness framework parts to encourage the specialists and the patients. The data, including clinical history of various patients' limits and contrasting outcomes, can be researched using data mining, searching for dependable models and proficient associations in the affliction. For instance, if a patient's prosperity limits are changing in a comparable model as those of a past open minded in the informational collection, the outcomes can similarly be surveyed. If similar models are found reliably, it would be less difficult for the trained professionals and clinical investigators to find an answer for the issue.

References

1. B.G. Ahn, Y.H. Noh, D.U. Jeong, Smart chair based on multi heart rate detection system, in *2015 IEEE Sensors*, pp. 1–4, (2015)
2. S.H. Almotiri, M.A. Khan, M.A. Alghamdi, Mobile health (m-health) system in the context of IoT, in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pp. 39–42 (2016)
3. T.S. Barger, D.E. Brown, M. Alwan, Healthstatus monitoring through analysis of behavioral patterns. IEEE Trans. Syst. Man, Cyben Part A Syst. Humans **5**(1), 22–27 (2005). ISSN 1083–4427
4. I. Chiuchisan, H.N. Costin, O. Geman, Adopting the internet of things technologies in health care systems, in *2014 International Conference and Exposition on Electrical and Power Engineering (EPE)*, pp. 532–535 (2014)
5. A. Dwivedi, R.K. Bali, M.A. Belsis, R.N.G. Naguib, P. Every, N.S. Nassar. Towards a practical healthcare information security model for healthcare institutions, in *4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, pp. 114–117 (2003)
6. M. S. D. Gupta, V. Patchava, and V. Menezes. Healthcare based on iot using raspberry pi. In 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), pages 796–799, Oct 2015
7. P. Gupta, D. Agrawal, J. Chhabra, P.K. Dhir, IoT based smart healthcare kit, in *2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT)*, pp. 237–242 (2016)
8. N.V. Lopes, F. Pinto, P. Furtado, J. Silva, IoT architecture proposal for disabled people, in *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 152–158 (2014)
9. R. Nagavelli, C.V. Guru Rao, Degree of disease possibility (ddp): a mining based statistical measuring approach for disease prediction in health care data mining, in *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*, pp. 1–6 (2014)
10. P.K. Sahoo, S.K. Mohapatra, S.L. Wu, Analyzing healthcare big data with prediction for future health condition. IEEE Access, **4**, 9786–9799 (2016). ISSN 2169–3536

11. S. Tyagi, A. Agarwal, P. Maheshwari, A conceptual framework for IoT-based healthcare system using cloud computing, in *2016 6th International Conference—Cloud System and Big Data Engineering (Confluence)*, pp. 503–507 (2016)
12. B. Xu, L. D. Xu, H. Cai, C. Xie, J. Hu, F. Bu. Ubiquitous data accessing method in IoT-based information system for emergency medical services. IEEE Trans. Ind. Inf. **10**(2), 1578–1586 (2014). ISSN 1551–3203
13. E.E.B. Adma, Sathesh, Survey on medical imaging of electrical impedance tomography (EIT) by variable current pattern methods, J. ISMAC 3(2), 82–95, (2021)
14. J.I.Z. Chen, L.-T. Yeh, Analysis of the impact of mechanical deformation on strawberries harvested from the Farm. J. ISMAC **10**(3), 166–172

A Comprehensive Evaluation of Traditional MPPTS and Fuzzy Rule-Based Algorithms at Varying Solar Irradiance Levels

Aarti S. Pawar and Mahesh T. Kolte

Abstract One of the most significant techniques in photovoltaic systems is maximum power point tracking (MPPT). The objective of this technique is to get the biggest power out of a solar array in any situation by taking partial shadowing, temperature changes, and solar irradiation levels into consideration. The conventional MPPT techniques of perturb and observe (P&O) and incremental conductance are addressed in this article. These techniques have some technical difficulties such as a long tracking time, an unstable operating point, and a low tracking efficiency. To address these issues, a fuzzy rule-based technique has been adopted. Using the MATLAB Simulink tool, the effectiveness of these strategies is evaluated on a 720 W photovoltaic system at different solar irradiance levels. According to simulation results, the fuzzy rule-based MPPT technique surpasses traditional techniques in terms of stabilization. Other strategies discussed in this article follow the fuzzy rule-based technique in terms of decreasing oscillation in power output.

Keywords Solar irradiance level · Photovoltaic cell · Maximum power point tracking · Perturb and observe · Fuzzy logic controller

1 Introduction

Academics are focusing on nonconventional energy sources like renewable energy as a result of rising power demand and a scarcity of traditional energy sources. Electricity generation from traditional sources is quite costly. Solar, wind, and other renewable energy sources are examples. Photovoltaic (PV) power generation is a fast promising renewable energy source due to its environmental benefits, cheap maintenance, and lack of noise, thanks to the abundance of solar energy. A crucial

A. S. Pawar (✉) · M. T. Kolte
Department of Electronics and Telecommunication,, Pimpri Chinchwad College of Engineering, Pune, Savitribai Phule Pune University, Pune, India
e-mail: aarti.pawar@pccoepune.org

M. T. Kolte
e-mail: mahesh.kolte@pccoepune.org

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_43

element of all PV systems is the efficiency of its MPPT. Because of the PV panel's nonlinear current–voltage characteristics, tracking maximum output is neither simple nor straightforward. Tracking an accurate MPP from a PV module is difficult. Various factors, including changes in solar irradiation, temperature variations, and partial shadowing conditions, have an impact on the performance of PV modules. The tracking of the operating point from a PV module is simple under uniform solar irradiation. However, in nonuniform conditions, numerous peaks can be created using a PV array. It is a difficult task to trace down a single worldwide peak from it. As a result, keeping track of the precise MPP has always been a tricky issue. To increase efficiency, a variety of MPPT algorithms is available. A large number of researchers are working on it. Fractional open-circuit voltage (FOCV) [4, 22], hill climbing [3], perturb and observe [4–7, 9, 18], incremental conductance [4–6, 9, 13, 17], fractional short circuit current [4, 6, 22], constant voltage, constant current, fuzzy logic-based MPPT [2, 7, 17–21], and hybrid MPPT [10] are among the most commonly used MPPT algorithms. Mohammed A. Elgendy et al. developed the P&O algorithm for farming applications [11]. A lot of authors collaborated on the P&O algorithm. When the Sun is shining evenly, they are commonly utilized to extract the most electricity from a photovoltaic panel. Incremental conductance is the most often utilized alternative algorithm. This method is evaluated using the power output of a PV module [6, 12, 13, 17]. Jubaer Ahmed and Zainal Salam have proposed a technique for enhancing perturb and observe MPPT efficiency by taking into account the dynamic perturbation step size [14]. The P&O MPPT in PV systems was explored in depth by Rozana Alik, Awang Jusoh, and Nur Ameda Shukri. They spoke about the traditional P&O and modified P&O with variations in step size, as well as difficulties with partial shading [16]. A comprehensive comparative analysis of MPPT methods was described by Bidyadhar Subudhi et al. [8]. They classified MPPT approaches based on factors including software complexity, hardware complexity, and performance metrics like tracking time, MPPT efficiency, the number of control variables, and accuracy.

Researchers have lately created a slew of sophisticated MPPT algorithms. The paper looked at both traditional MPPT and fuzzy logic techniques. These algorithms are implemented, and their effectiveness is expressed in terms of tracking time, tracking efficiency, and power consumption. This paper is divided into four sections that are structured in a logical order. The first portion is an introduction, while the second half has a section on PV modeling. The final portion delves into the specifics of MPPT methods. The fourth section explains the simulation findings and discussions. Finally, in the concluding part, the conclusion is given.

2 PV Modeling and System Description

2.1 PV Modeling

The single diode photovoltaic model is used to demonstrate the link between photovoltaic cells current, voltage, and power, as well as how solar irradiation and temperature impact the characteristics of PV arrays [10]. Figure 1 represents the relevant circuit.

In a single diode model, the output current "I" is calculated as

$$I = Ipv - I_s\left[\exp\left(\frac{V + R_s I}{aV_t}\right) - 1\right] - \frac{V + R_s I}{R_P} \tag{1}$$

The solar irradiance S determines the photovoltaic current, which may be expressed as

$$I_{pv} = \left(\frac{S}{S_{\text{ref}}}\right) I_{sc} + k_i (T - T_{\text{ref}}) \tag{2}$$

The temperature dependent diode saturation current may be represented as

$$I_s = I_{rs}\left(\frac{T}{T_{\text{ref}}}\right)^3 \exp\left[\frac{qE_g}{ak}\left(\frac{1}{T_{\text{ref}}} - \frac{1}{T}\right)\right] \tag{3}$$

The mathematical equation that reflects a real PV module's characteristic is

$$P = V \cdot \left[Ipv - I_s\left[\exp\left(\frac{V + R_s I}{aV_t}\right) - 1\right] - \frac{V + R_s I}{R_P}\right] \tag{4}$$

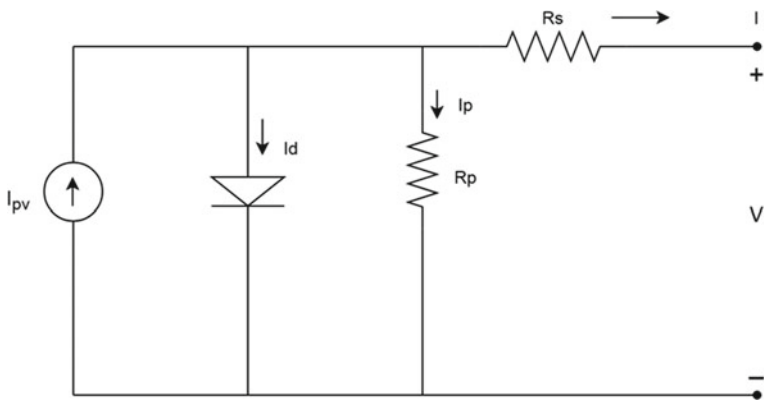


Fig. 1 Equivalent model of solar cell

Nomenclature,

R_s, R_p	Array's equivalent series and parallel resistance
I_{pv}, I	Photovoltaic and output current.
Is, I_{rs}	Currents of saturation and reverse saturation
V_t	The diode's thermal voltage
a	Diode ideality factor
S, S_{ref}	Irradiance and irradiance at the standard test condition
k_i	Current coefficient
T, T_{ref}	Actual and reference temperatures of cell
q	Charge of an electron
E_g	Bandgap energy
k	Constant of Boltzmann

2.2 Detailed About the System

The system is made up of six PV modules connected in parallel. On the right side of the I-V curve, PV panels have low internal impedance, whereas, on the left, they have high impedance. The performance of a PV module is determined by its operating point upon this PV curve. Impedance matching between both the PV module and the load connected to it is the most important component. By altering the converter's duty cycle supplied by the MPPT controller, a dc-dc converter of the PV system functions as an intermediary to accomplish at MPP. Figure 2 illustrates the system architecture of a proposed system. To figure out how to control the duty cycle, the maximum power transfer theorem is employed. This case makes use of the boost converter topology. Take the boost converter's average output

$$V_o = \frac{V_i}{1 - D} \tag{5}$$

Assuming 100% converter efficiency

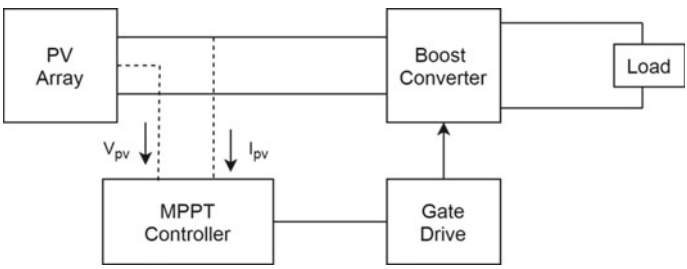


Fig. 2 Detailed architecture of PV system

$$\frac{I_o}{I_i} = \frac{V_i}{V_o} = 1 - D$$
$$I_o = (1 - D) I_i \tag{6}$$

The input resistance was seen by the PV module

$$R_{eq} = \frac{V_i}{I_i} \tag{7}$$

From Eqs. (5) and (6)

$$R_{eq} = (1 - D)^2 R_L \tag{8}$$

Modifying the duty cycle will align the PV module's measured input resistance with the optimal resistance at which the PV system will produce the most electricity. As a result, by adjusting the load impedance value, the most power is transferred to the load. The following formula is used to determine the duty cycle:

$$D = 1 - \sqrt{\frac{R_{eq}}{R_L}} \tag{9}$$

2.3 Effects of Environmental Parameters on PV module's Capability

At a single position known as the maximum power point, the PV module generates the most electricity. At MPP, the power at this stage is referred to as cell power (P_m). Cell voltage at MPP (V_m) is the voltage corresponding to this, and cell current at MPP (V_m) is the current at this moment (I_m). PV panels have a nonlinear P–V curve. The typical characteristics of the photovoltaic cell are depicted in Fig. 3. The operational point of a PV panel fluctuates as temperature and solar irradiance change. When the solar irradiance is changed from 250 to 450, 650, 850, and 1000 W/m^2 under stable cell temperature (25 °C), the PV current rises and the voltage changes minimally with the relevant solar irradiance levels, and the PV array output power grows as well. Figure 4 depicts the impact of solar irradiance at a constant temperature. The current of the PV array rises slightly, the PV array's voltage decreases considerably, and the PV array's power output is reduced when the cell temperature changes between 25, 35, 55, 75, and 85 °C under constant solar irradiation (1000 W/m^2). Under a constant solar irradiation situation, Fig. 5 demonstrates the effect of solar cell temperature.

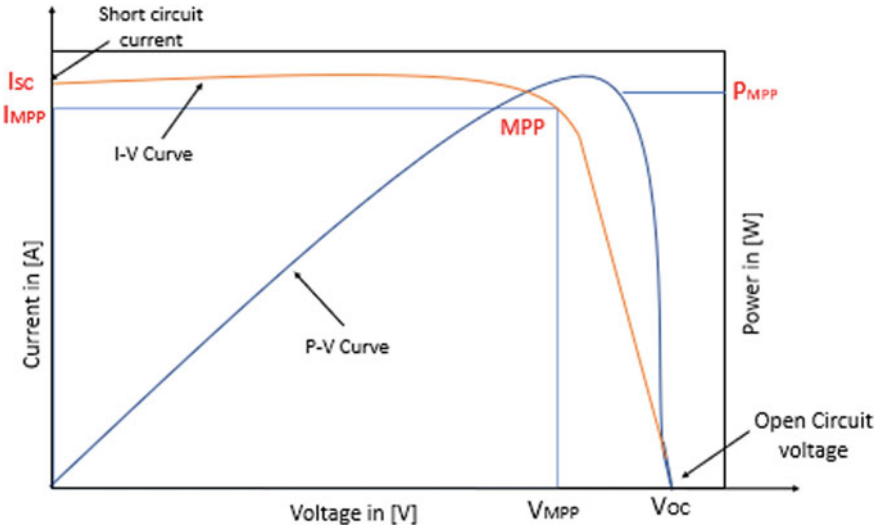


Fig. 3 Typical characteristics of photovoltaic cell

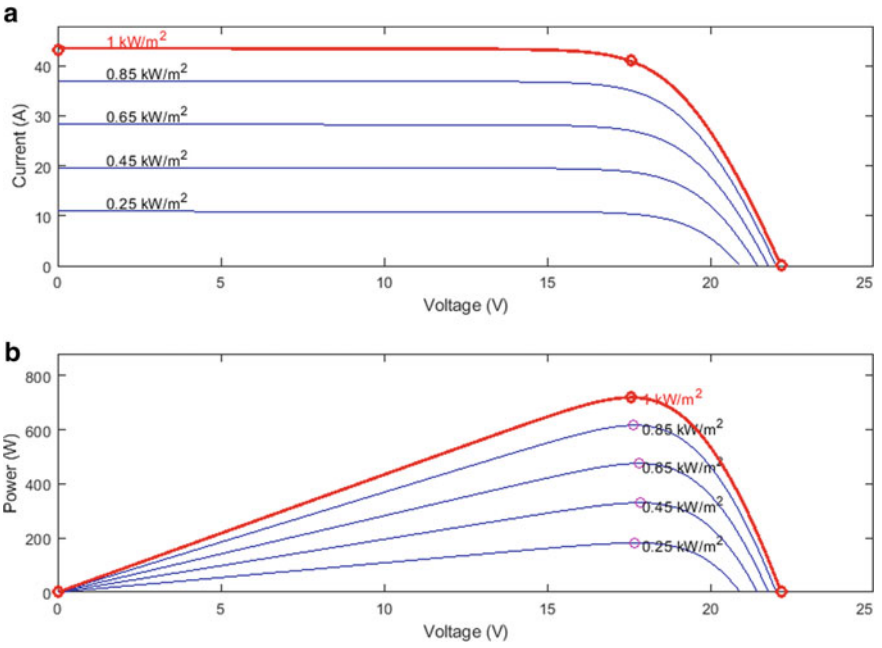


Fig. 4 **a** PV module's current–voltage graph at cell temperature (25 °C), **b** PV module's power-voltage graph at cell temperature (25 °C)

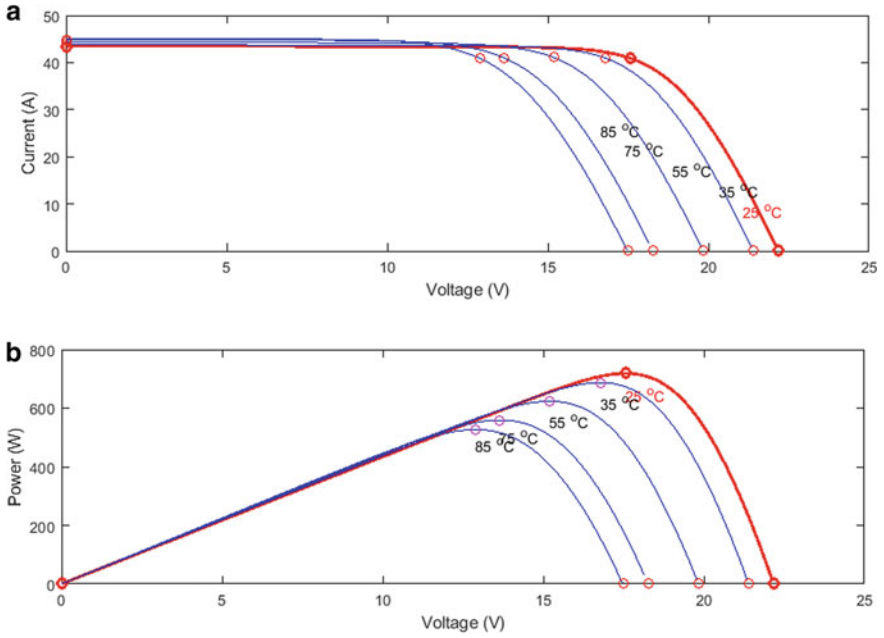


Fig. 5 **a** PV module's current–voltage graph at solar irradiance level (1000 W/m^2), **b** PV module's power–voltage graph at solar irradiance level (1000 W/m^2)

3 Working of MPPT Techniques

3.1 P&O Technique

It is a popular MPPT algorithm. It is easy to set up and utilizes fewer parameters. The research of this algorithm is focused on the link between PV module output current and voltage [10]. Figure 6 depicts the PV module's behavior by showing the condition of the module's operating point in response to changes in PV power. When the slope is positive, the MPP is on the P–V curve's left-hand side. The voltage perturbation of a PV module should rise toward the MPP as the generated power of the PV module increases. The voltage perturbation should be minimized near the MPP if the module's operating point is on the P–V curve's right side (the slope is negative). It monitors MPP by comparing the power before and after disturbance. It starts by estimating the PV array's power using voltage and current measurements. After then, the PV power change (dP_{pv}) will be analyzed to determine if it is positive or negative. If dP_{pv} is larger than zero, the perturbations will persist in almost the same direction. If this is not the case, the following perturbation is reversed. Using this technique, the method can accomplish MPP and fluctuate about it in steady-state conditions. Figure 7 depicts the flowchart for the traditional P&O technique. The MPP is where the system truly fluctuates. The perturbation step size is reduced to

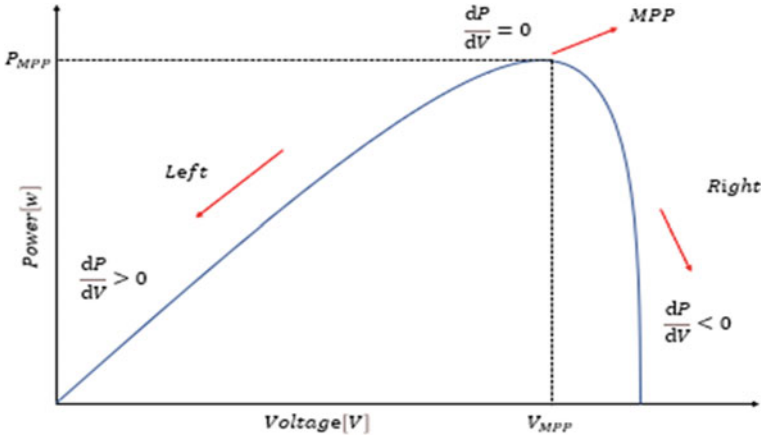


Fig. 6 Operation of PV module at the operating point

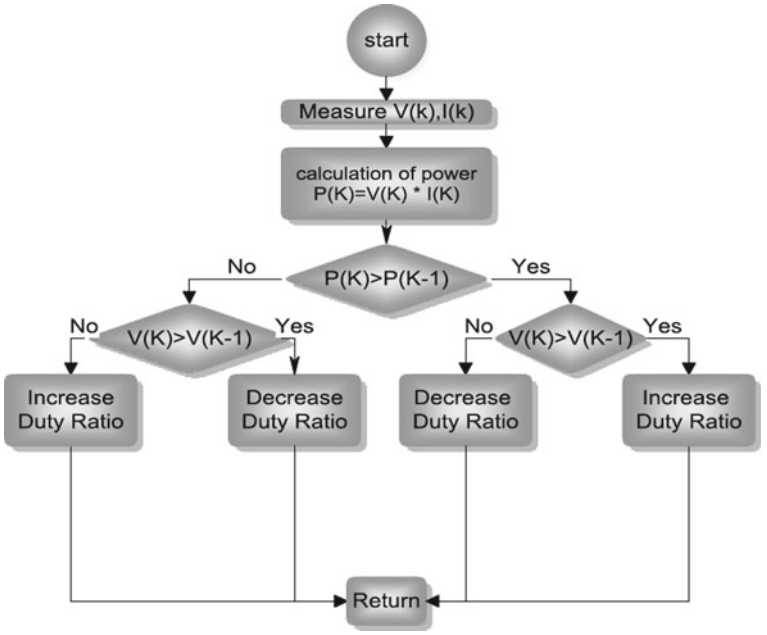


Fig. 7 Working flow of traditional P&O technique

reduce oscillation. Small step sizes, on the other hand, slow down MPPT's reaction. For different irradiance and panel temperatures, the PV array would have different characteristic curves. The method's biggest limitation is that it generates MPP oscillations under various Sun irradiation circumstances. Obtaining MPP in this situation is difficult.

3.2 Incremental Conductance Technique

It is one of the most often utilized MPPT techniques. By making some modifications in the P&O technique, the MPP is tracked here. In this method, the change in the current parameter is also considered along with a change in voltage and power to track the operating point. It tracks MPP using instantaneous conductance as well as incremental conductance. The slope of the P–V curve detects the presence of MPP. It decides whether the PV system is proceeding to the MPP's right or left side. Based on this, this technique tracks the MPP. It uses the following equation to detect the presence of MPP in the correct direction.

$$P = VI \tag{10}$$

$$\frac{dP}{dV} = I + V \cdot \frac{dI}{dV} \tag{11}$$

The *PV* module is functioning at MPP when the fraction of incremental conductance equals the negative of instantaneous conductance, according to the preceding equation. However, it is assumed that the MPP is obtained when the operation is performed with some marginal error (ε)

$$\frac{dP}{dV} = \pm\varepsilon \tag{12}$$

At the MPP, the marginal error (ε) is approaching zero. Adjusting the step size prevents tracking beyond the margin. The PV module is functioning on the MPP's left side, if $\frac{dI}{dV} > \frac{-I}{V}$, i.e., the slope is positive $\left(\frac{dP}{dV} > 0\right)$. Depending on the scenario, the duty cycle is then lowered or increased. If $\frac{dI}{dV} < \frac{-I}{V}$, the PV module is operating on the MPP's right side. It solves the problem with the P&O technique. When the irradiance is abruptly increased, however, it fails to make a smart judgment. Figure 8 depicts the incremental conductance flowchart.

3.3 Fuzzy Rule-Based MPPT Technique

It is one of the most effective methods for getting the most electricity out of a solar system. It is a strategy for multivariable evaluation and resolution based on several rules. It offers several advantages including the ability to deal with imprecise inputs, the lack of the requirement for an exact mathematical model, a resilient structure, simplicity in comparison with other design techniques, and the ability to manage nonlinearity [12]. The designer's knowledge and experience in acquiring acceptable inputs and modifying the rule-based table evaluate the strength of a fuzzy rule-based logic controller. The membership functions of the fuzzy system determine how the

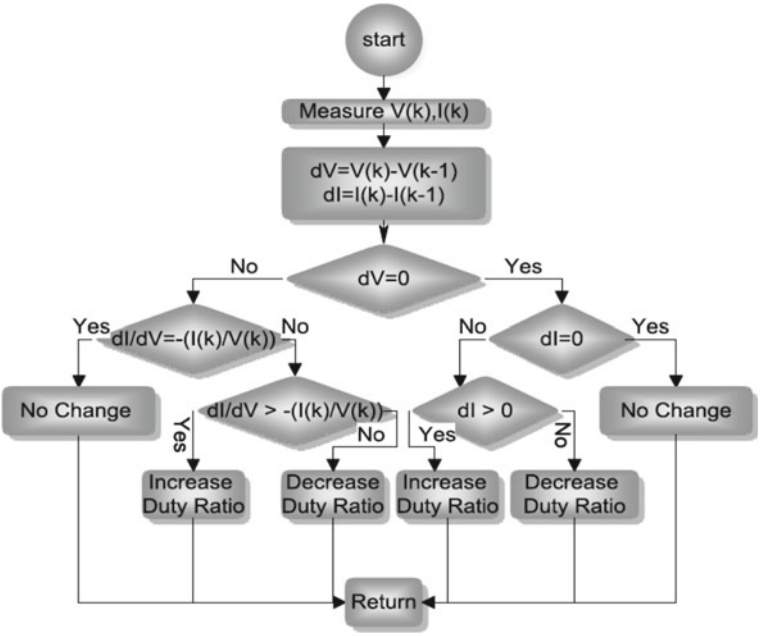


Fig. 8 Flowchart of incremental conductance technique

fuzzy logic controller works. The type of membership function employed determines the controller's complexity. Figure 10 depicts a typical fuzzy logic controller. The initial stage in the operation of a fuzzy logic controller is fuzzification. In the functioning of the fuzzy logic controller, the inference engine is followed by the defuzzification process.

1. Fuzzification: The recommended fuzzy logic controller's inputs are errors and change in errors. These inputs are denoted by the letters (C) and (CE). They are calculated using Eqs. (13) and (14) for a sample interval of x.

$$E(x) = \frac{P_{PV}(x) - P_{PV}(x-1)}{V_{PV}(x) - V_{PV}(x-1)} \tag{13}$$

$$CE(x) = E(x) - E(x-1) \tag{14}$$

where the PV system's power and voltage are represented by $P_{PV}(x)$ and $V_{PV}(x)$, respectively. Based on the inputs, the fuzzy controller utilizes membership functions and a rule table to identify the next operational point. The $E(x)$ parameter specifies whether the operational point is on the MPP's left or right side. The CE(x) parameter indicates how quickly the operating point approaches or retreats from MPP. The output of the suggested controller changes the duty cycle of the converter. As a consequence, when the slope is zero, the slope $E(x)$ will determine whether the

operating point is on the appropriate point or not by converting the converter's duty ratio. Using the fundamental fuzzy collection, five exact fuzzy levels are employed to offer membership function values to linguistic variables like ZS, NS, NB, PB, and PS. Figure 9 depicts input and output variables of the triangle membership function of five fundamental fuzzy levels.

2. Inference Engine:

The values of the input variables are fuzzified into fuzzy numbers once they have been calculated. The fuzzy controller employs linguistics processes to predict which regulatory action should be taken in response to a collection of input values. The rule structure is a series of if–then rules that keep track of the data associated with the parameters under consideration. Table 1 lists the 25 rules that will be used in the fuzzy logic technique that will be developed. This regulation intends to change the

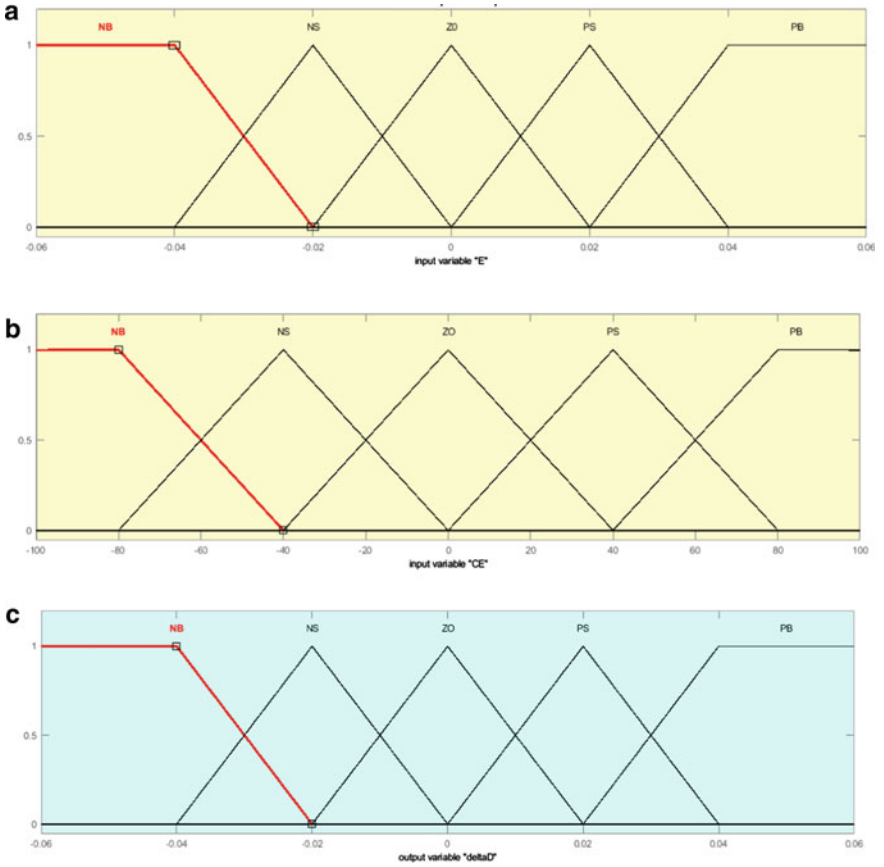


Fig. 9 **a** Error (E) function, **b** Change in error (CE) function, **c** Duty cycle function

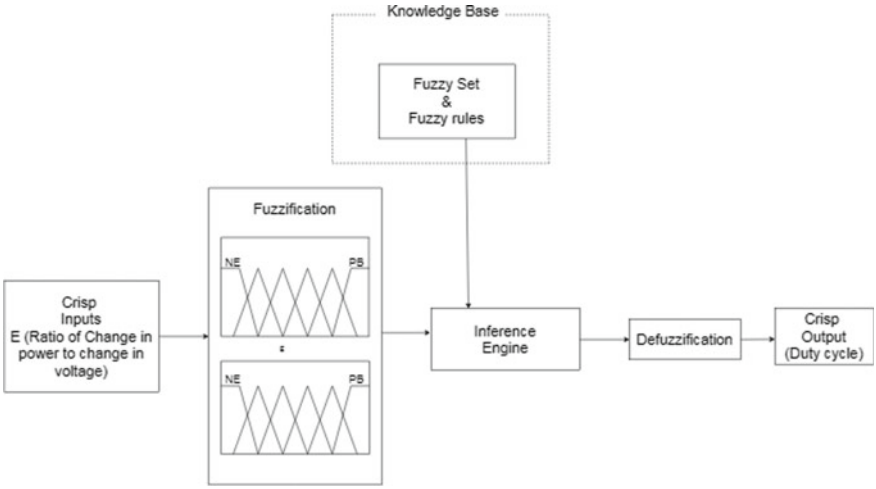


Fig. 10 Architecture of fuzzy rule-based controller

Table 1 Regulation of fuzzy

E	CE				
	NB	NS	ZS	PS	PB
NB	ZS	ZS	NB	NB	NB
NS	ZS	ZS	NS	NS	NS
ZS	NS	ZS	ZS	ZS	PS
PS	PS	PS	PS	ZS	ZS
PB	PB	PB	PB	ZS	ZS

duty cycle based on the position of the MPP, bringing the operational point closer to the MPP. The following is the rule's structure.

If (E is PB) and (CE is NB) then (dD is PB).

The duty cycle should be altered to raise the output voltage if the operating point differs from MPP and the slope (E) is larger than zero. Depending on the fuzzy rule architecture, the inference engine reaches a logical conclusion and converts the decision rules to a fuzzy language output. There are many fuzzy inference methods from which to pick. The fuzzy inference technique utilized in this article is Mamdani's fuzzy inference method, which is a max–min fuzzy combination [2]. This step generates a fuzzy output value that is tailored to each type of event.

3. Defuzzification:

The inference engine, it was discovered, provides ambiguous data. However, numerical data are required by the dc-dc converter. The system typically allows non-fuzzy values. Therefore, the fuzzy linguistic variables are converted to crisp numerical variables at this point. Defuzzification can be accomplished using a variety

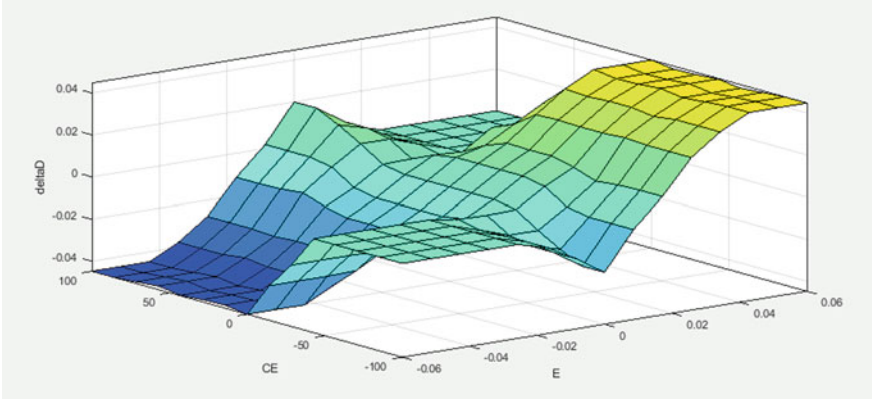


Fig. 11 Surface function

of techniques including the center of area and the max of criterion approaches. The center of gravity approach is the most often used defuzzification method. In this context, this method is utilized to create non-fuzzy output. The gravity center (dD) is determined as [1].

$$\mathrm{d}D = \frac{\sum_{j=1}^{n} \mu\left(\mathrm{d}D_j\right) \cdot \mathrm{d}D_j}{\sum_{j=1}^{n} \mu\left(\mathrm{d}D_j\right)} \tag{15}$$

The value must be measured by the following duty cycle after computing the duty cycle dD, and the final duty cycle should be applied to the MOSFET gate drive.

Surface viewer:

The surface function depicts the connection between fuzzy inputs and outputs as a three-dimensional curve (design) [2]. Figure 11 illustrates a simulation of the surface function. The output duty cycle is supplied to the converter, and the conversion is performed via centroid defuzzification.

4 Discussions on Simulation Results

The three MPPTs' functionality is evaluated using the MATLAB Simulink tool for tracking time, efficiency, and power loss (oscillations). Table 2 lists the exact specifications of the six-parallel string of 120 W, 12 V PV modules that were employed. In the PV system stated, the boost converter is used and the details of the components used in the converter are provided in Table 3. The functionality of P&O, incremental conductance, and also fuzzy rule-based MPPTs is compared using two different solar irradiation levels. The simulation results are depicted in Figs. 12, 13 and 14.

Table 4 shows the simulation results to look at the various MPPT techniques under

Table 2 Technical details of 120 W PV module

Technical parameter	Values
Open-circuit voltage (V_{oc})	22.17 V
Short -circuit current (I_{sc})	7.21A
The voltage at P_m (V_m)	17.56 V
The current at P_m (I_m)	6.84A
Maximum power (P_m)	120.11 W
Voltage temperature coefficient (K_v)	−0.35(%/°C)
Current temperature coefficient (K_i)	0.06
No. of cells in series per module	36

Table 3 Technical details of boost converter

Parameter	Values
Inductor (L)	5 mH
Capacitor (C1, C2)	12 mF, 0.1 mF
Switching frequency (f)	5 kHz
Switching device (IGBT)	6.84A

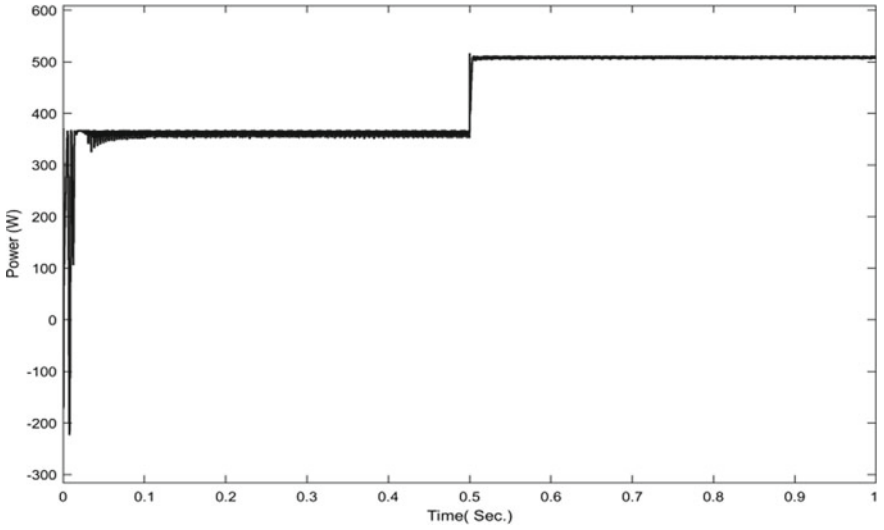


Fig. 12 The output of the P&O controller

various solar irradiation conditions. The simulation results in Figs. 12, 13, and 14 indicate that the time required to monitor the MPP in the P&O algorithm is rather considerable, and this technique clearly shows the ongoing fluctuation of the operating point. It happens because the operational voltage is continually disrupted in an attempt to avail the MPP. While this fluctuation is less noticeable in the incremental

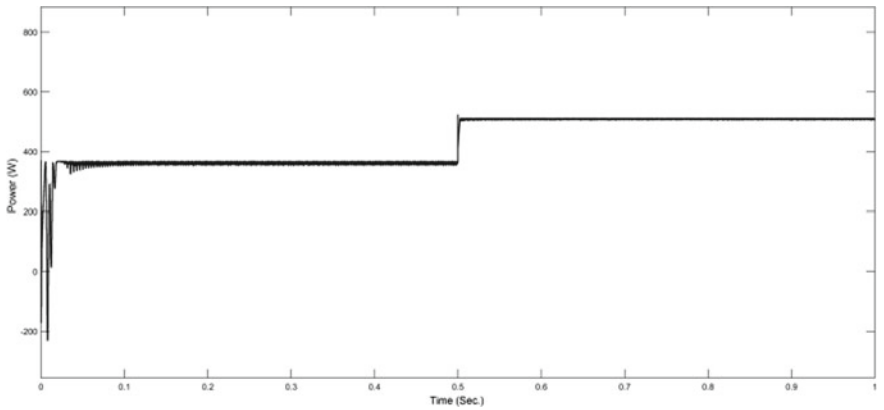


Fig. 13 The output of the incremental conductance controller

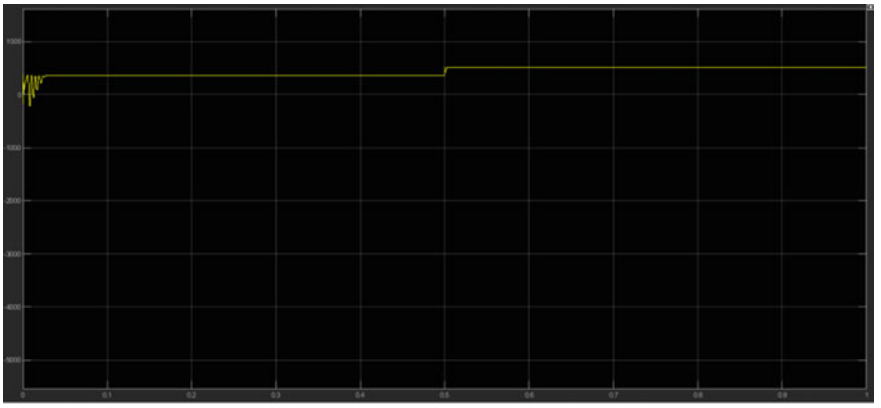


Fig. 14 The output of the fuzzy rule-based controller

Table 4 Performance table of three MPPT Techniques

Irradiance (W/m^2)	MPPT techniques	Tracking time (s)	Tracking efficiency (%)
500,700	P&O	0.509	99.30
	Incremental Conductance	0.504	99.80
	Fuzzy rule-based MPPT	0.50	99.85

conductance method, it is not seen in the fuzzy rule-based MPPT method, where the power signal is nearly constant. As a result, there have been fewer power losses. The response time of the fuzzy rule-based MPPT technique is faster than the other two traditional MPPTs, namely P&O and incremental conductance. Table 5 shows the results of a comparative examination of MPPTS in terms of features. Figures 15, 16 and 17 depict a simulation model of the system. A rapid rise in the solar irradiance from 500 to 700 W/m^2 was replicated in 0.5 s. The temperature of cells is kept at 25 °C. The fuzzy rule-based MPPT technique is more significant under these

Table 5 Comparison of three MPPT Techniques in terms of functionality

Parameters	P&O	Incremental conductance	Fuzzy rule-based MPPT
Response time	Large	More	less
Tracking speed	Slow at varying solar irradiation levels	Medium	Fast
Power loss	Maximum	Medium	Minimum
Implementation complexity	Simple	Moderate	Moderate
Sense parameter	V, I	V, I	P, V

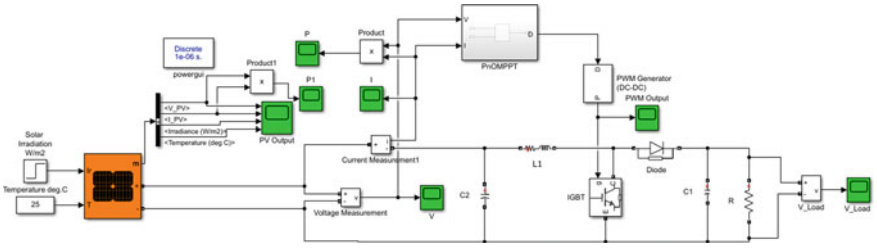


Fig. 15 Simulation model of P&O controller system

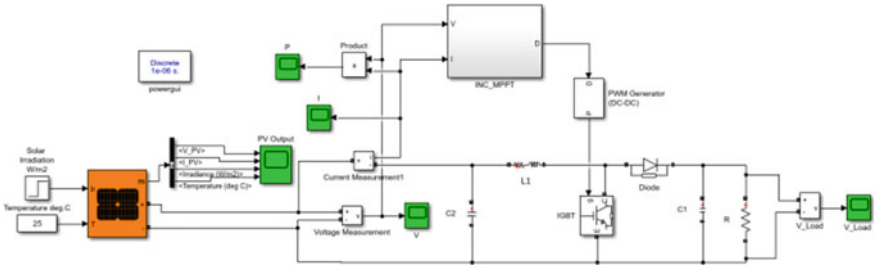


Fig. 16 Simulation model of incremental conductance controller system

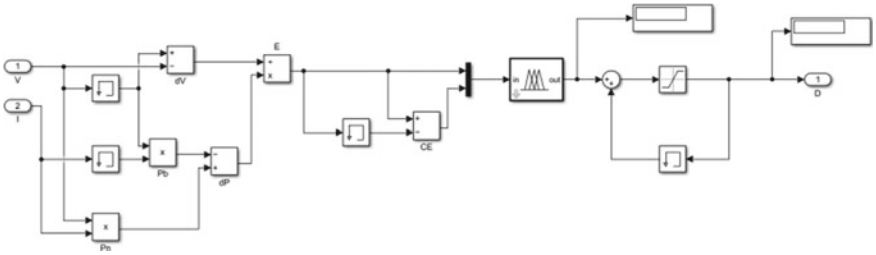


Fig. 17 Simulation model of fuzzy rule-based controller

operating conditions. The generated power of the fuzzy rule-based MPPT rises gradually, whereas the traditional P&O MPPT approach, as shown in Fig. 12, deviates significantly from the MPP.

5 Conclusion

This research focuses on three MPPT algorithms based on varying amounts of solar irradiation, notably traditional P&O, incremental conductance, and fuzzy rule-based method. The performance of these algorithms is carried out in this research using the MATLAB Simulink tool. The PV panel can produce the greatest amount of electricity with the three controllers, according to the simulation. The performance of fuzzy logic-based MPPT, on the other hand, is superior to that of traditional MPPT controllers. When compared to other MPPTs, it has a lower power loss. In comparison with traditional MPPTs, it delivers a substantial increase in tracking efficiency and a faster response time to achieve MPP.

References

1. B. Bendib, F. Krim, H. Belmilia, M.F. Almi, S. Boulouma Advanced fuzzy MPPT controller for a stand-alone PV system, (Energy Procedia, Science Direct, Elsevier, 2014)
2. E.H. Mamdani, S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller. Int. J. Man-Mach. Stud. (1975)
3. S.B. Kjaer, Evaluation of the hill climbing and the incremental conductance maximum power point tracker for photovoltaic power systems. IEEE Trans. Energy Convers. (2012)
4. A. Mohapatra, B. Nayak, P. Das, K.B. Mohanty, A review on MPPT techniques of PV system under partial shading condition. Renew. Sustain. Energy Rev. (2017)
5. S. Saravanan, N.R. Babu, Maximum power point tracking algorithms for the photovoltaic system—a review. Renew. Sustain. Energy Rev. (2016)
6. R.I. Putri, S. Wibowo, M. Rifa'i, Maximum power point tracking for photovoltaic using incremental conductance method. Energy Procedia **68**, 22–30 (2015)
7. U. Yilmaz, A. Kircay, S. Borekci, PV system fuzzy logic MPPT method and PI control as a charge controller. Renew. Sustain. Energy Rev. (2018)

8. B. Subudhi, R. Pradhan, A comparative study on maximum power point tracking techniques for photovoltaic power systems. IEEE Trans. Sustain. Energy **4**(1), 89–98 (2013)
9. T.K. Soon, S. Mekhilef, A fast converging MPPT technique for photovoltaic system under fast varying solar irradiation and load resistance. IEEE Trans. Ind. Informat. **11**(1), 176–186 (2015)
10. S. Mohanty, B. Subudhi, P.K. Ray, A grey wolf-assisted perturb & observe MPPT algorithm for a PV system. IEEE Trans. Energy Convers. **32**(1), 340–347 (2017)
11. M.A. Elgendy, Assessment of perturb and observe MPPT algorithm implementation techniques for PV pumping applications. IEEE Trans. Sustain. Energy **3**(1), 21–33 (2012)
12. M. Kumar, S.R. Kapoor, R. Nagar, A. Verma, Comparison between IC and fuzzy logic MPPT algorithm based solar PV system using boost converter. Int. J. Adv. Res. Electr. Electron. Instrum. Eng. **4**(6), 4927–4939 (2015)
13. M.A. Elgendy, B. Zahawi, D.J. Atkinson, Assessment of the incremental conductance maximum power point tracking algorithm. IEEE Trans. Sustain Energy **4**(1), 108–117 (2013)
14. J. Ahmed, Z. Salam, An improved perturb and observe (P&O) maximum power point tracking (MPPT) algorithm for higher efficiency. Appl Energy (2015)
15. M.A. Enany, M.A. Farahat, A. Nasr, Modeling and evaluation of maximum power point tracking algorithms for photovoltaic systems. Renew. Sustain. Energy Rev. **58**, 1578–1586 (2016)
16. R. Alik, A. Jusoh, N.A. Shukri, An improved perturb and observe checking algorithm MPPT for photovoltaic system under partial shading condition, (IEEE, 2015)
17. N.K. CH, V.S.V.K. Pravallika, Fuzzy based improved incremental conductance MPPT algorithm in PV system, in *IEEE Conference*, 2020
18. D. Remoaldo, I. Jesus, Analysis of a traditional and a fuzzy logic enhanced perturb and observe algorithm for the MPPT of a photovoltaic system, in *Algorithms, MDPI*, 2021
19. K. Bataineh, An intelligent maximum power point using fuzzy log controller under severe weather conditions. Int. J. Photoenergy (2018)
20. W.-S. Kim, H.-S. Kim, A new maximum power point tracker of photovoltaic arrays using fuzzy controller, (IEEE, 1994)
21. G. Balasubramanian, S. Singaravelu, Fuzzy logic controller for the maximum power point tracking in a photovoltaic system. Int. J. Comput. Appl. (2012)
22. T. Esram, P.L. Chapman, Comparison of photovoltaic array maximum power point tracking techniques. IEEE Trans. Energy Convers. **22**(2), 439–449 (2007)

Object Tracking Using Moderate Derivative Gain Kalman Filter

Arjun Nelikanti, G. Venkata Rami Reddy, and G. Karuna

Abstract Object tracking in a video sequence is becoming challenging and important research area, and it becomes more complex when there are multiple objects. Many researchers have proposed several algorithms for detection and tracking in various videos including sports. In this paper, Kalman gain is calculated using moderate derivative of Grunwald–Letnikov function, and the modified Kalman filter is used to track the object in video sequence. The performance of the proposed method is analyzed by root mean square error (RMSE) metric and also compared with other methods. The results demonstrate that the proposed algorithm showing improvement by up to 12%.

Keywords Object tracking · Kalman filter · Grunwald–Letnikov · Fractional derivative and RMSE

1 Introduction

In Kalman filter, series of calculations are made at time intervals. It is an algorithm where state variables are estimated with noise and other inaccuracies for given parameters. Since 1968, constant gain has been analyzed in Kalman filter; an algorithm is proposed which determined piecewise-constant feedback gains for a linear system over a finite interval [1]. Kalman filter proposed with constant gain [2–5] for target tracking in both standalone and data fusion mode. Kalman filter with adaptive gain [6–8] was proposed based on minimization of variance to help in improving performance with constant gain. For evaluating the filter performance, the difference of actual and estimated state variables are calculated. There are few cases where the

A. Nelikanti (✉)
CSE, JNTUH, Hyderabad, India

G. Venkata Rami Reddy
JNTUH (SIT), Hyderabad, India

G. Karuna
CSE, GRIET, Hyderabad, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_44

filter does not converge on innovation process. There are many adaptive versions of filter proposed by researches to improve the performance of Kalman filter.

Kalman filter is a state estimator with two stages, i.e., predict and update. In tracking an object, the filter can give more weight for the measurements to predict close values by having high gain. The filter can follow measurement closely with model having optimal gain and also smoothening noisy signals resulting low responsiveness. The Kalman filter gain can perform well in few cases when the gain can be tuned for desired results. The extended Kalman filter (EKF) will have high convergence rate if the difference between estimated state and actual state is low [9]. The EKF diverges if the error is large; the solution for controlling this divergence is by inserting a feedback loop at the calculation of gain. Fractional calculus is gaining more importance in many research areas for different applications related to material science, electrical, computer networking and mathematical formulations of mechanics [10]. This is also applicable in image processing which showed good accuracy results for some systems [11]. The objectives for this work are:

- Track the object accurately even under different occlusion conditions
- Normalize the Kalman gain to handle abrupt changes of the object in video sequence.

To address these above objectives, in this paper, we have tried to integrate Kalman filter and fractional calculus concepts for tracking an object in a video sequence under different foreground situations.

1.1 Motivation

In numerous sports, exceptionally competitive expert associations have developed, where technically flawlessness has gotten fundamental for competitors to stay aware of their opponents. Cricket has consistently been quite possibly the most mainstream sports in numerous nations in Asia. Because of its wide range of components, it is in fact exceptionally testing, and the need of analyzing these events became important to do in efficient way.

The rest of the paper is organized as follows: Sect. 2 provides overview of related work done in same area. Section 3 discussed proposed work, Sect. 4 provides details of experiment and results obtained, and Sect. 4 presents conclusions of the work done.

2 Literature Review

Kalman proposed Kalman filter in 1965 [12]. It has attracted many researches due to its capabilities and scope of applications. Various versions of Kalman filter were proposed based on an important parameter, i.e., Kalman gain.

In [2], Cook and Dawson proposed a version of Kalman gain computed using integral equation which showed good results. There were also different equations used to compute the Kalman gain matrix, for example, Chandrasekhar equation proposed in [13].

Kalman filter also handles occlusion and the adjustments made to occlusion rate designed adaptive Kalman filter discussed in [14]. Chung et al. [7] proposed a concept of adaptive control for calculating Kalman gain without taking initial state estimation and process noise. In [15], adaptive high gain EKF is proposed for navigation system where the gain is increased by changing the innovation. Xu et al. [16] proposed an adaptive iterated extended Kalman filter to increase the performance by statistics estimator for noise in iteration of Kalman filter.

In [17], geo-location of mobile terminals was predicted using digital fractional integration (DFI). The trajectories of these mobile terminals are predicted in two indoor scenarios with noisy environment, namely spiral and sinusoidal trajectories. Autocorrelation of paths obtained showed good results for proposed method.

In [18] an algorithm for tracking bearings only passive maneuvering target is proposed. A modified function of Galkowski-Islam was used to propose the algorithm. The proposed algorithm predicts the target with chi-square in sliding window format; while tracking the maneuvering, the noise is increased. The process noise is decreased after tracking is completed. Using adaptive fating, EKF and adaptive UKF was proposed in [19] to decrease the noisy effect.

In [20], an adaptation process was proposed where the desired Kalman gain was achieved by choosing appropriate window length of sample and latency time. The gain was modified to normalize the prediction error autocorrelation. P. Li et al. proposed a new tracker based on UKF, and it employs a nonlinear model for measurement. The results showed better estimation of the state of system having same order as EKF [21]. Van Der Merwe [22] had extended EKF different algorithms based on derivative-less sampling of Gaussian statistics to the category of algorithms called as Sigma-Point Kalman Filters. Using this, the results obtained were better when compared with the results of EKF.

In [23], using UKF, an aerodynamic model is designed which uses six auxiliary states for computation with less intense, and the results are compared with EKF. Adelina Ioana Ilieş et al. [24] proposed a system to find the state of charge in a battery and the performance is compared using UKF and EKF.

In tracking of an object in different situations, the state of object can show abrupt changes. For these scenarios, Kalman can diverge while tracking, so optimization of high gain could provide better results. By introducing derivate gain feedback loop at Kalman gain, the divergence can be reduced where the loop can normalize the sudden variations and increases the gain. This loop does not diverge as it uses moderate derivation. This process is explained in the following section III.

3 Proposed Method

This section provides the proposed modifications in Kalman and also some concepts described which are useful for the discussions here. The below Eq. 1 is the *n*th order derivative of function 'f' in terms of x:

$$f^{(n)}(x) = \frac{d^n f}{dx^n} = \lim_{h \to 0} \frac{1}{h^n} \sum_{r=0}^{n} (-1)^r \binom{n}{r} f(x - rh) \tag{1}$$

Accordingly, fractional derivative of the Grunwald–Letnikov[10] for single variable function 'f' can be defined as in below Eq. 2:

$$D^{\alpha}_{G-L} f(x) = \lim_{h \to 0} \frac{1}{h^{\alpha}} \sum_{r=0}^{\left[\frac{x-a}{h}\right]} (-1)^r \binom{\alpha}{r} f(x - rh) \tag{2}$$

where
$\binom{\alpha}{r} = \frac{\Gamma(\alpha+1)}{\Gamma(r+1)\Gamma(\alpha-r+1)}$,
and Γ is gamma function.

Fig. 1 shows the block diagram of Kalman filter with a change at Kalman gain 'K,' where a loop f(KG) is fed back to the Kalman gain block.

The color intensity of image at point (x, y) can be defined in terms of a 2D function using two spatial coordinates x and y as $f(x, y)$. In the field of image processing, the Grunwald–Letnikov derivative [25] in 2D in the x-direction can be defined as follows:

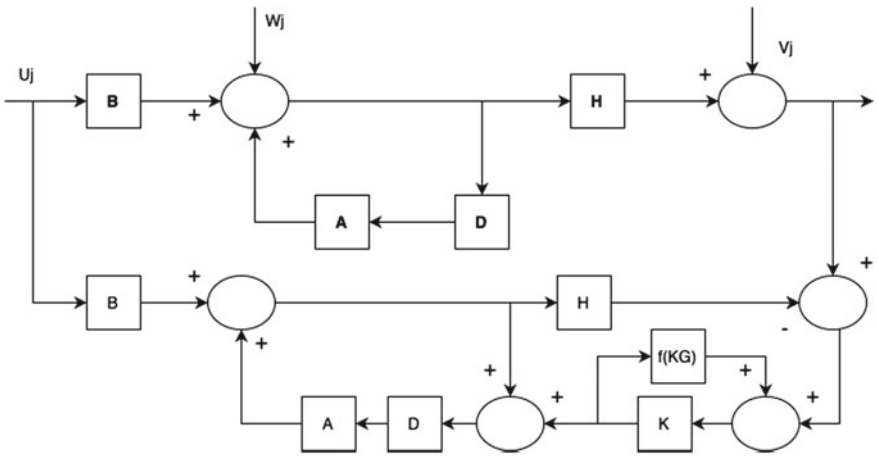


Fig. 1 Kalman filter block diagram

$$D^{\alpha}_{G-L} f_x(x, y) = f(x, y) - \alpha f(x-1, y) + \frac{\alpha(\alpha-1)}{2} f(x-2, y) \tag{3}$$

The derivation of function f in y direction can be expressed using *G–L* derivative. Hence, fractional derivative of *G–L* can be defined by Eq. 4:

$${}_m D^{\alpha}_{G-L} f(x, y) = \sqrt{({}_m D^{\alpha}_{G-L} f_x(x, y))^2 + ({}_m D^{\alpha}_{G-L} f_y(x, y))^2} \tag{4}$$

We express a modification of *G–L* derivative in both the directions as follows, the function M is the minimized of function f can be defined as Eq. 5:

$$M(x, y) = \frac{1}{s^n} \min\{f(x, y), f(x-1, y), f(x-2, y)\} \tag{5}$$

We obtain the below Y value shown in Eq. 6 and X value shown in Eq. 7 both from Eq. 5

$$Y(x, y) = M(x, y)\left(\frac{s - X(x, y)}{s}\right) \tag{6}$$

$$X(x, y) = \left| f(x, y) - \alpha f(x-1, y) + \frac{\alpha(\alpha-1)}{2} f(x-2, y) \right| \tag{7}$$

Using Eq. 7, we define the modified *G–L* in X direction given in Eq. 8

$${}_m D^{\alpha}_{G-L} f_x(x, y) = \frac{f(x, y) - \alpha f(x-1, y) + \frac{\alpha(\alpha-1)}{2} f(x-2, y)}{Y(x, y) + 1} \tag{8}$$

Kalman gain is modified using the Eq. (8)

$$KG_{\text{new}} = KG_g + \Delta^{\alpha} KG_{GL}, \tag{9}$$

where α is moderate fractional order Eq. (8), and $\Delta^{\alpha} KG_{GL}$ is modified fractional derivative of previous Kalman gain.

Equation 9 is the new moderate derivative Kalman gain. The modified Kalman gain KG_{new} having the terms, KG_g is previous Kalman gain, and $\Delta^{\alpha} KG_{GL}$ is calculation of mean fractional difference of previous Kalman gains.

Uj is input, *Wj* is noise and *Vj* is output noise at instant *K*, *D* is delay element, *A*, *H* and *B* is transition, measurement and control input matrix, respectively.

4 Results and Discussions

The proposed work performance validation is evaluated with both 1D and 2D data. 1D dataset includes square and sinusoidal waveforms. 2D dataset includes videos taken from online benchmark datasets [26] as well as data captured indoor and outdoor using two cameras, Canon Power Shot A4000 IS and second camera is SonyDSC-T70 with a standard tripod. We have implemented the proposed method and executed on the PC that has Intel(R) CoreTM i5 CPU and 6144 MB memory running MATLAB R2018a.

While applying Kalman filter for object tracking, various types of noise may corrupt the data. Measurement and process noises are the main noise which can be seen in Kalman filter. For evaluation of proposed method, measurement noise which is added at acquisition like sensor noise is added to signal with varying SNR. This signal to noise ratio ranging [−10, 10 dB], the noise added is to simulate the Gaussian noise with various values of variance and 0 mean. In Fig. 2a, b is shown the results of RMSE for square and sinusoidal noisy inputs, respectively. Equation 10 is used to calculate the RMSE values, and the results shows the proposed work has the lower values than Kalman filter.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{10}$$

Detection of the object is done using background subtraction. The masked object is tracked using the proposed method, a rectangular blue box in Figs. 3 and 4 shows the object being tracked in the frame. The proposed method is evaluated on the TB50 David3 and walking benchmark datasets shown in Fig. 3 and also on the secondary dataset captured indoor and outdoor where the object is a ball as shown in Fig. 4.

In Fig. 3a, b is the dataset of David3 and c, d is walking dataset from TB-50 sequences available online. The object is person in the sequences, and the person is tracked using the proposed method and the results are compared.

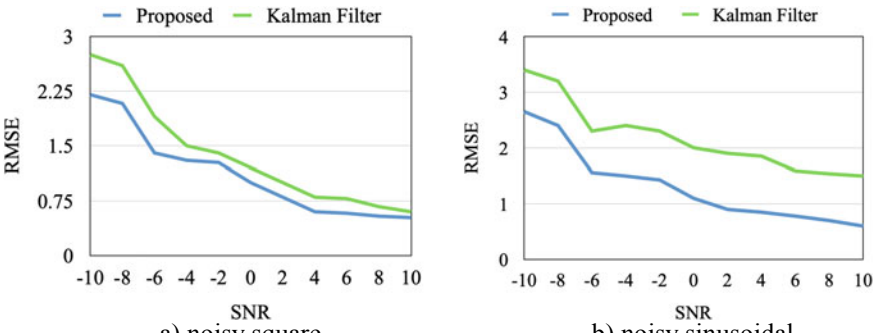


Fig. 2 Depict RMSE values for 1D inputs

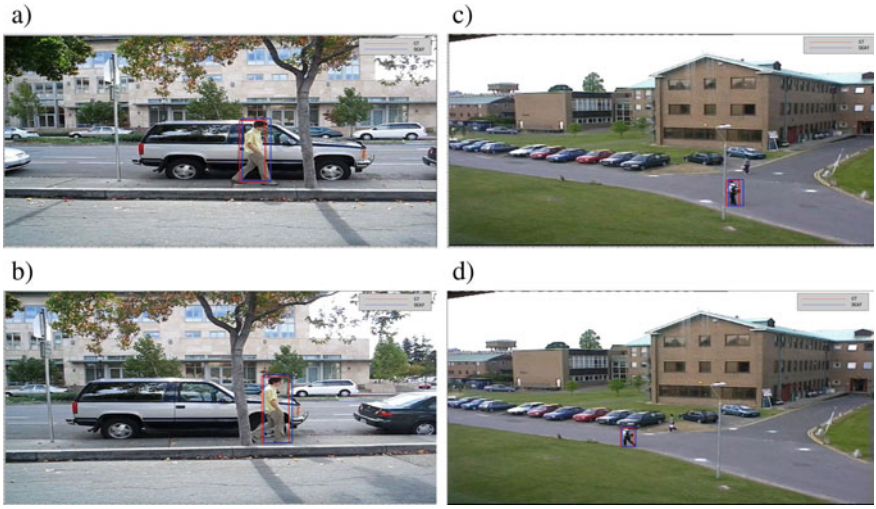


Fig. 3 **a**, **b** David3 and **c**, **d** walking from TB-50 dataset

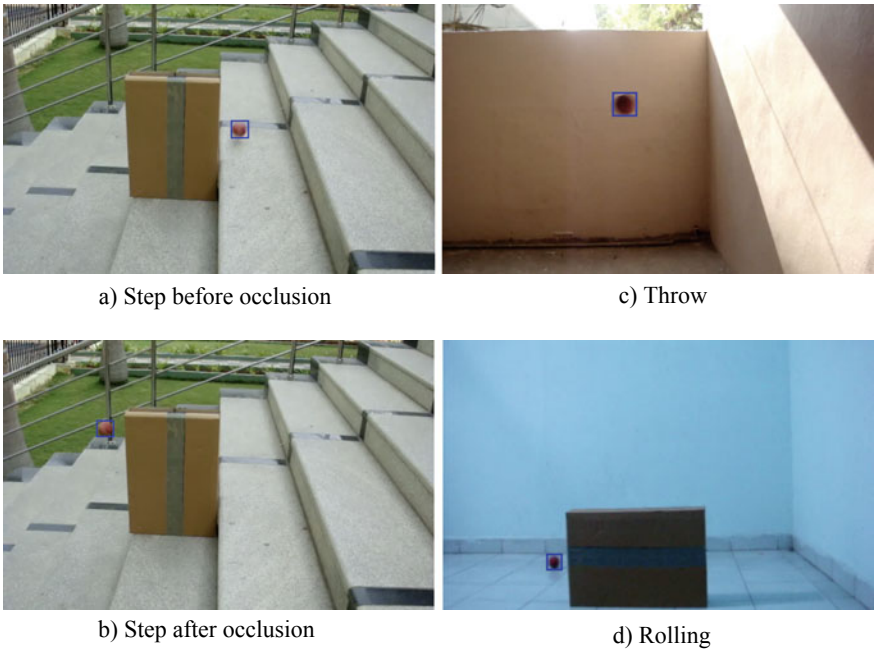


Fig. 4 Outdoor ball thrown dataset

In Fig. 4a, b and d, the ball undergoes occlusion in different situations, and the ball is successfully detected and tracked before and after the complete occlusion. The performance of the proposed method is evaluated using RMSE for both the datasets.

4.1 Limitations

The proposed work is evaluated on few benchmark datasets of the object tracking where the size of object is large. To track the smaller object in a video sequence is the actual challenge.

5 Conclusion and Future Work

For tracking object, a novel method is proposed in this paper which uses moderate derivative gain Kalman filter. This modified gain is used as feedback to Kalman which is obtained by fractional derivation using modified G–L function of previous Kalman gain. For evaluation of proposed method derivative gain Kalman filter, RMSE metric is used which has shown better capability and improved by up to 12%. The results shown in this paper on 1D data on different waveforms, on benchmark data and also on secondary captured data indicate the proposed approach exhibits good improvements on comparison with Kalman filter. In the future, the proposed method can be evaluated on different dataset containing smaller objects which undergoes occlusion.

References

1. M.D. Kleinman, T. Fortmann, M. Athans, On the design of linear systems with piecewise-constant feedback gains. IEEE Trans. Automat. Control **13**(4), 354–361 (1968)
2. A. Anilkumar, M. Ananthasayanam, P.S. Rao, A constant gain Kalman filter approach for the prediction of re-entry of risk objects. Acta Astronaut. **61**(10), 831–839 (2007)
3. G. Cook, D.E. Dawson, Optimum constant-gain filters, IEEE Trans. Ind. Electron. Control Instrum. **IECI-21**(3), 159–163 (1974)
4. K. Wilson, An optimal control approach to designing constant gain filters. IEEE Trans. Aerosp. Electron. Syst. **AES-8**(6), 836–842 (1972)
5. A. Yadav, P. Awasthi, N. Naik, M. Ananthasayanam, A constant gain kalman filter approach to track maneuvering targets, in *Proceeding of IEEE International Conference on Control Applications (CCA)* (2003), pp. 562–567
6. N. Chernoguz, Adaptive-gain tracking filters based on minimization of the innovation variance, in *Proceeding of IEEE International Conference on Acoustics Speech and Signal Processing,* vol. 3 (2006)

7. H.Y. Chung, T.W. Kim, S.H. Chang, B. Lee, Adaptive kalman gain approach to on-line instrument failure detection with improved GLR method and suboptimal control on loft pressuriser. IEEE Trans. Nucl. Sci. **33**(4), 1103–1114 (1986)
8. A. Almagbile, J. Wang, W. Ding, Evaluating the performances of adaptive kalman filter methods in GPS/INS integration. J. Global Positioning Syst. **9**(1), 33–40 (2010)
9. N. Boizot, E. Busvelle, J.P. Gauthier, An adaptive high- gain observer for nonlinear systems. Automatica **46**(9), 1483–1488 (2010). [Online] http://www.sciencedirect.com/science/article/pii/S0005109810002542
10. F. Riewe, Mechanics with fractional derivatives. Phys. Rev. E. **55**, (1997). Art. no. 3581
11. D. Sierocuik, A. Dzielinski, Fractional kalman filter algorithm for the states, parameter and order of fractional system estimation. Int. J. Appl. Math. Comput. Sci **16**(1), 129–140 (2006)
12. R.E. Kalman, A new approach to linear filtering and prediction problems. ASME J. Basic Eng. **82**, 35–45 (1960)
13. R.F. Brammer, A note on the use of chandrasekhar equations for the calculation of the kalman gain matrix (corresp.). IEEE Trans. Inf. Theory **21**(3), 334–336 (1975)
14. S.K. Weng, C.M. Kuo, S.K. Tu, Video object tracking using adaptive kalman filter. J. Vis. Commun. Image Represent. **17**(6), 1190–1208 (2006)
15. K. Sebesta, N. Boizot, A real-time adaptive high-gain EKF, applied to a quadcopter inertial navigation system. IEEE Trans. Ind. Electron. **61**(1), 495–503 (2014)
16. Y. Xu, X. Chen, Q. Li, Adaptive iterated extended kalman filter and its application to autonomous integrated navigation for indoor robot. Sci. World J. **2014**, (2014). Art. no. 138548
17. A. Nakib, B. Daachi, M. Dakkak, P. Siarry, Mobile tracking based on fractional integration. IEEE Trans. Mobile Comput. **13**(10), 2306–2319 (2014)
18. S. Rao, Modified gain extended kalman filter with application to bearings only passive manoeuvring target tracking, in *IEE Proceedings-Radar, Sonar and Navigation*, vol. 152, no. 4, pp. 239–244 (2005)
19. K.S.L.J.V. Fathabadi, Mehdi Shahbazian, Comparison of adaptive kalman filter methods in state estimation of a nonlinear system using asynchronous measurements, in *Proceedings of the World Congress on Engineering and Computer Science*, 2009
20. B. Han, X. Lin, Adapt the steady state kalman gain using the normalized autocorrelation of innovations. IEEE Signal Process. Lett. **12**(11), 780–783 (2005)
21. P. Li, T. Zhang, B. Ma, Unscented kalman filter for visual curve tracking. Image Vis. Comput. **22**(2), 157–164 (2004). [Online]. http://www.sciencedirect.com/science/article/pii/S026288560300146X
22. R.V an Der Merwe, Sigma-point kalman filters for probabilistic inference in dynamic state-space models, Ph.D. dissertation, Oregon Health Sci. Univ., Portland, OR, USA, 2004
23. W. Muhammad, A. Ahsan, Airship aerodynamic model estimation using unscented Kalman filter. J. Syst. Eng. Electron. **31**(6), 1318–1329 (2020). https://doi.org/10.23919/JSEE.2020.000102
24. A.I. Ilieş, G. Chindriş, D. Pitică, A comparison between state of charge estimation methods: extended Kalman filter and unscented Kalman filter, in *2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, pp. 376–381. https://doi.org/10.1109/SIITME50350.2020.9292232
25. H. Jalalinejad, A. Tavakoli, F. Zarmehi, A simple and flexible modification of Grünwald-Letnikov fractional derivative in image processing. Math Sci **12**, 205–210 (2018). https://doi.org/10.1007/s40096-018-0260-6
26. TB-50 Visual Tracker Benchmark [Online]. Available: http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html

Reconfigurable 1:4 Wilkinson Power Divider Used in ISM Band Applications

Aparna B. Barbadekar and Pradeep M. Patil

Abstract Many researchers have contributed to the technical developments of a variety of multistage power dividers based on the Wilkinson topology by replacing the quarter wave transmission line section with other means. It was possible to reduce the size of the power divider to a great extent but the S-parameters were largely affected. This was due to the effect of termination, discontinuties, mismatching losses and manufacturing tolerance. Amplitude and phase imbalance also contributed to destabilised performance of power dividers. The current paper presents a reconfigurable 1:4 Wilkinson power divider (WPD) with 3 sections of 2-way WPDs operating in ISM Band applications in order to work around such limitations. The simulation of the power divider is carried out with CST software which is capable of analysis of microwave circuits, and calculation of amplitude and phases of the scattering parameters can be done very accurately. The average measured values of the S-parameters are: insertion loss −6.28 dB, the bandwidth at −10 dB and the return loss is 39.69%. The output ports have no reflection, and hence, they are perfectly isolated. The surface current of the power divider has no coupling of current between output ports when port 1 is excited. It indicates that the output ports are perfectly isolated. The dimensions of the power divider are 112 × 157 mm.

Keywords Power divider · Insertion loss · Microstrip · Return loss · Reflection coefficient

1 Introduction

The history of a conventional WPD is summarized in [1, 2] by different authors. The main drawback of the WPD is its narrow bandwidth and larger size because of the use

A. B. Barbadekar (✉)
Department of Electronics and Telecommunication, AISSMS IOIT, Pune, India

Department of Electronics and Telecommunication, VIIT, Pune, India

P. M. Patil
SNDCOERC, Yeola, Nashik, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_45

of λ /4 transmission line sections and its single frequency operation. In [3], the use of lumped passive components instead of a transmission line section was suggested to reduce the circuit size. However, a limited bandwidth because of inductors with a high Q factor and a higher insertion loss were the drawbacks. A novel idea was presented by Chiu et al. in [4] which included the use of parallel strip line structure as shown in Fig. 1a. However, the structure introduced a discontinuity and reduction in performance of the circuit. In 2008, Jiang et al. introduced the use of a bent microstrip line in place of a straight microstrip line to reduce the circuit size [5]. In [6], a novel compact CNS power divider was designed which used an equivalent low pass filter circuit instead of λ/4 transmission lines. The observation here was that the reflection coefficient of the center frequency was offset slightly to the low frequency. Using structures like a two branch directional coupler was suggested by Ding et al. [7]. Even/odd modes analysis was used to design and analyze the power dividers useful in an antenna fed network. A structure based on a single layer microstrip line with capabilities to design different types of power dividers was proposed in [8]. It could reduce the occupational area and suppress the harmonic, multiband, and arbitrary dividing ratios as shown in Fig. 1b.

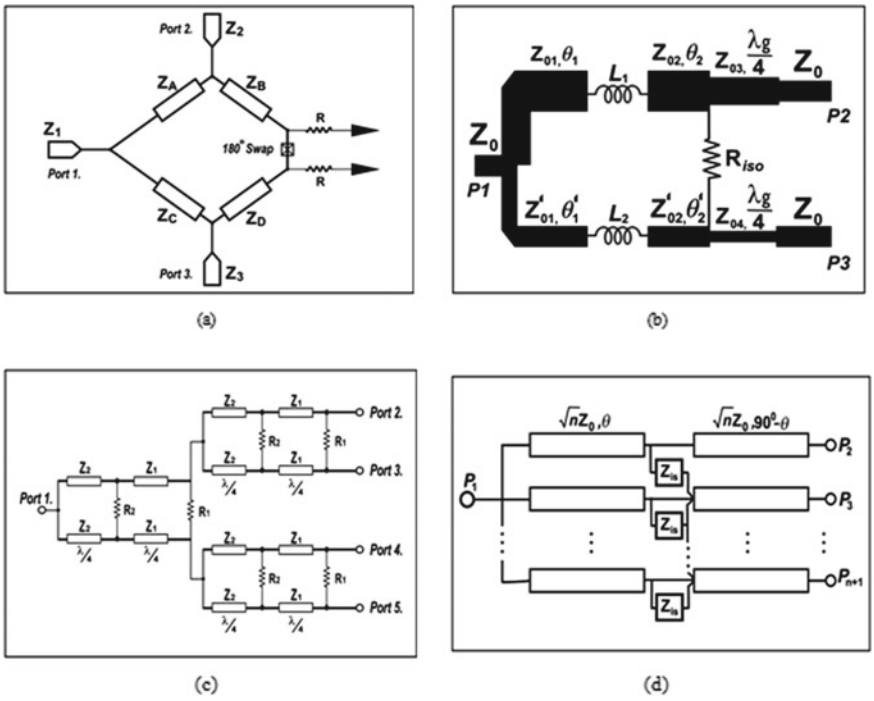


Fig. 1 **a** Modified WPD with a swap and shunt resistors, **b** Layout of unequal power divider, **c** Circuit model of 4-way power divider, **d** Schematic of n-way power divider

A compact 1:1 and 1:9 divider at 1 GHz frequency totally based on Wilkinson topology was designed with the use of two microstrip circuits [9]. In the present topology, the first stage was a 1:2 power divider, whereas the second stage used two power dividers but converted to three output ports. A signal coming in could be separated by the resultant 1:6 power divider with equivalent amplitude of outputs into 6 output ports. Replacement of all λ /4 lines [10] with the use of step impedance resonator (SIR), implementation of glass-based TFIPD technology and the use of Bridge T-coils [11], as shown in Fig. 1c were made to reduce the size without affecting the operational bandwidth of the proposed power divider. N-type divider with an n quarter wave transmission line was presented in [12]. Each line had a characteristic impedance that was $\sqrt{n}$ Z0 and isolation impedance Zis as shown in Fig. 1d. Because the symmetry of the structure isolation impedance acted as an open circuit, perfect input match and equal power division were possible irrespective of the electrical length θ.

To address the demands of the standards of telecommunication, the primary focus of researchers is now the design of the dual band, the multiband and the broadband power dividers.

In 2013, a planar six-way power divider [13] was exhibited as per the Nobuo Nagai theory to reduce the size area with a good harmonic suppression. The structure shown in Fig. 2a was two dimensional, easy to design and symmetrical. A UBW power divider was designed in [14]. The structure was based on three open stubs on each branch and defected ground structure (DGS) on the back of inter digital coupled lines. Multistage, quarter wave transformers were used to increase the bandwidth and reduced the size of proposed UWB power divider in [15]. The physical model shown in Fig. 2b was simulated with HFSS software to get a desirable result and operate between the frequency band (2–18 GHz) [16]. The dielectric is enclosed between thick layer so that the width of the strip lines can be increased and the effects of the fabrication tolerances can be decreased. The performance of an unequal dual frequency power divider was analyzed based on the modeling approach line theory [17]. The structure shown in Fig. 3a minimizes the size of the dual band power divider [18] using isolation resistors and line couplers. Based on theoretical analysis it could

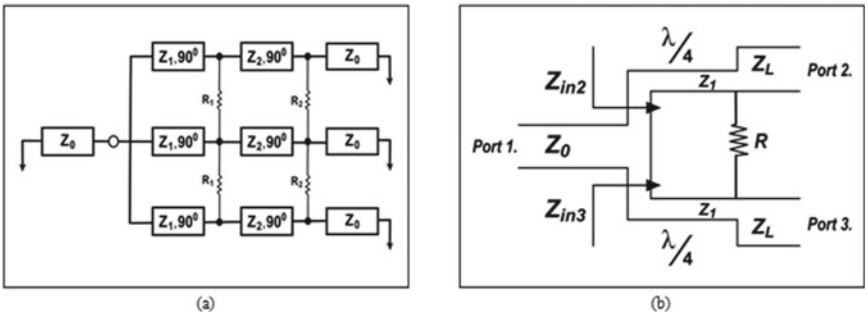


Fig. 2 **a** Three-way power divider based on Nabuo Nagai, **b** WPD in microstrip form

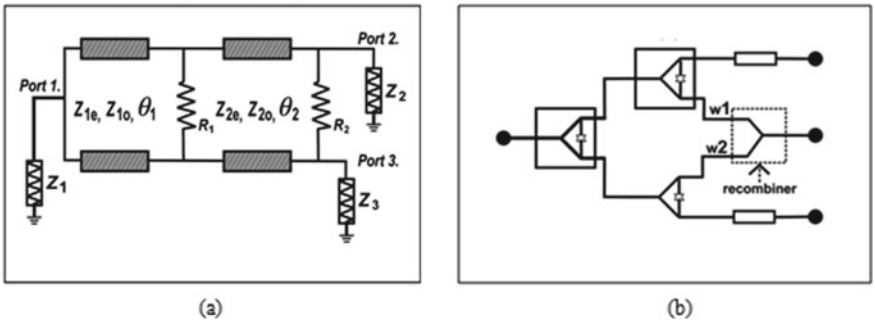


Fig. 3 **a** Dual band coupled line WPD, **c** Standard model cell for dual band power divider

operate over a wide frequency range (1 to 3 GHz) without additional requirements on the microstrip fabrication and could cover all other dividers. The planar structure proposed in [19] and based on a recombined structure concept was useful to design an arbitrary planner power divider (it could be generalized to any number). The large separation between W1 and W2 shown in Fig. 3b) reduced the parasitic effect without reducing the efficiencies of the power divider. Multisection impedance transformers were proposed to increase the operation band [20]. The offset double side parallel strip provided high impedance in unequal power dividers. A presentation in [21] brought forth a generalized WPD that included lumped elements which were able to enable dual band and unequal power division. This particular divider had an edge in the characteristics of isolation and bandwidth of reflection at the output terminals as well as in the choice of inductance. Moreover, this was accomplished with no degradation in the impedance as well as in the power division characteristics. Haijun Fan et al. came up with an innovative, reconfigurable structure with a high dividing ratio power divider [22]. Merely three pin diodes were utilized without any D.C. blocking capacitors. The planar, wide band with tunable power division ratio was the focus of the design in [23]. Artificial transmission line (ATL) technology [24] and double sided parallel strip lines (DSPSL) were used to design a 4:1 unequal WPD [25] so as to reduce the size in the low frequency band and hence minimize the structure in order to suit wireless communication applications. The concept of complete termination for 3 dB power dividers was presented in a Hybrid microstrip form [26]. The accuracy of the design approach was confirmed with the measurement of S-parameters based on their characterization.

Filtering power dividers have been the focus of the multiple studies in recent times. Low pass filter [27] were used to substitute two-quarter wave length transformers in the conventional WPD. In [28], there was a report of a novel class of power dividers that filtered on the basis of quasi-band pass section. It had features like frequency controllable single/multiband operation. Innovative wide band power dividers that include filter function are proposed, searched and presented in [29].

To overcome the limitations observed in the literature survey, this paper proposes a design based on reconfigurable 1:4 WPD which is modified with three sections of

2-port symmetrical power dividers operating in ISM band applications. Z-parameters and S-parameters are analytically convenient to calculate the system performance by cascading individual components of a microwave system.

The paper is systemized as: In Sect. 2, statistical data is analyzed to find out the inter relationship between the S-parameters. Section 3 contains the research methodology where in theoretical analysis of Z-parameters, S-parameters and microstrip lines are explained. Section 4 presents Design and Simulation of the proposed power divider. Section 5 presents the performance comparison of the reported power dividers. The Conclusion based on our research is given in Sect. 6 References are cited at the end of the paper.

2 Statistical Data Analysis

In this, we have summarized all the reviewed literature in Table 1 which compares S-parameters.

Table 1 contains a comparison of the main variables of RF and microwave power dividers such as return loss, insertion loss, isolation and size. Various reference papers are reviewed along with the technology used. The return loss shown in Table 1 varies from 15 [11] to 30.4 dB [3] for power dividers operating at single design frequency. The dual frequency power dividers show excellent return loss from 15 [20] to 50 dB [17].The highest insertion loss observed is 8 dB [13] and the lowest is 0.16 dB [6] due to the CMOS technology used. It is further observed that isolation loss reduces with increasing output ports. In general, the reduction in isolation marked is 30.5 to 27.2% when the output ports go up from 4 to 6.

The main scope of this paper is to present previous statistical data available on microwave power dividers. We have reviewed multiple approaches for the design of multiband power dividers. Moreover, a discussion on varying frequencies and power rations for multiband WPDs using various methods has been conducted. The size reduction marked varies from 57.67 to 91.25%. It is further observed that the isolation loss goes down by 30.5 to 27.2% when output ports vary from 4 to 6.

The main variables of RF and microwave power dividers have a barter relationship, and hence, the selection is made by the designer depending upon the requirements of the applications. In the area of wideband and ultra-band filter design, the technique for co-design of filters and power dividers is a new topic of interest. Using computer aided design tools capable of analysis of microwave circuits, the calculation of the amplitude and phase of the scattering parameters can be done very accurately. The use of a multilayer substrate can be useful for increasing the bandwidth, reducing the size and improving the isolation of power dividers.

Table 1 Performance comparison of S-parameter

Method	Return loss (dB)	Insertion loss (dB)	Isolation (dB)	Size in area mm2	Center Freq. (GHz)	Remarks
Lu et al. [3]	> 30.4	< 0.16	27	0.015	4.5	Limited bandwidth because of high frequency Q factor and insertion loss
Chiu et al. [4]	2	< 0.7	> 25	25.10 cm^2	2	Parallel strip lines introduced a discontinuity and reduction in performance of the circuit
HJiang et al. [5]	Good	4.77	> 30	31.32	0.9	Size is reduced drastically due to bent microstrip lines but it has introduced phase imbalance variations of the order of $\pm$ 0.35
Yuenjin and Menjiang et al. [6]	Notquoted	Not quoted	Not quoted	0.36	1.616	Reflection coefficient of the center frequency was offset slightly toward the low frequency
Ding et al. [7]	< -20	Not quoted	> 17	87.40	2.5	Simple, compact and easy design and fabrication
Mirzavand et al. [8]	27.34	3.34	30	3.78 cm^2	1	Due to approximation in values of the elements, extra circuit loss occurs. Moreover the response of the structure used depends on self-resonant frequency of used lumped inductors

(continued)

Table 1 (continued)

Method	Return loss (dB)	Insertion loss (dB)	Isolation (dB)	Size in area mm2	Center Freq. (GHz)	Remarks
Pribawa et al. [9]	< 12.3	< 7.8	< 12.8	Not quoted	1.5	The isolation of S_{24} and S_{57} are lesser than other ports due to the effect of wirebonds length
Fu et al. [10]	−20	< 0.02	−20	14.4	1.6–2.1 Pass band	The SIR is used to modify the isolation and input port return loss. No serious limitations are mentioned in the published work
	< −20	0.5	> 20		1.7−3.7 Pass band	
Tseng et al. [11]	15	1.75	15	21.16	1.6	Glass-based TFIPD technology was used to compact chip size. Broad band 4 –way PD
Choe et al. [12]	16.2	4.8	24	0.06	2	n-way WPD reduced the insertion loss and circuit size but the process of analysis involved plenty of theory
Wan et al. [13]	−20	8	> 20	Not quoted	1.71–2.5	Planner six-way broad band, 1:6 PD based on Nubuo Nagal theory
Liu et al. [14]	< 10	Small	< −10	0.285	3.1−10.6	Simulated and experimental values of S_{21} at high frequency differ due to the use of open stubs
Xu et al. [15]	< 1.25	−3.5	−22	9.30 cm^2	2–8	Measured results are worse than simulation result due to fabrication technique when the divider is welded SMA adaptors

(continued)

Table 1 (continued)

Method	Return loss (dB)	Insertion loss (dB)	Isolation (dB)	Size in area mm2	Center Freq. (GHz)	Remarks
Afanasieve et al. [16]	1.5	0.4	−18	8.36 cm^2	2–18	Coupled strip lines used to improve isolation and impedance matching at high frequencies. UWB PD. No limitations are mentioned in the published work
Wei et al. [17]	50	–	> 50	Not quoted	$f_1 = 0.9$	There exist some deviations to ideal values due to loss of characteristics of the material it self
	50	–	> 50		$f_2 = 1.8$	
Wu et al. [18]	> 25	> −3.15	> 30	2.36 cm^2	$f_1 = 1.1$	Additional interconnecting transmission lines are required to obtained in phase characteristics in dual band applications
	> 25	> −3.19	> 30	2.36 cm^2	$f_2 = 2.2$	
Wu et al. [19]	Not quoted	1.51	< −21	Not quoted	$f_1 = 0.6$	The discrepancies between measured and simulated result occurred due to reactive components at high frequencies
		1.57	−9.8	Not quoted	$f_2 = 2.45$	
Shie et al. [20]	> 15	−7.36	−26	Not quoted	$f_1 = 2.45$	Physical size of power divider is increased due to multisection impedance transformers
	> 15	−7.37	−25		$f_2 = 5.25$	

(continued)

Table 1 (continued)

Method	Return loss (dB)	Insertion loss (dB)	Isolation (dB)	Size in area mm2	Center Freq. (GHz)	Remarks
Wang et al. [21]	< -20	−4.77	< -20	Not quoted	$f_1 = 1$	There are no remarks on the size reduction on 4 types of power divider developed
	< -20	1.76	< -20		$f_2 = 1.8$	
Fan et al. [22]	> 15.4	1.6	> 18	Not quoted	5	Simulations are executed by HFSS and ADS with pin diode equivalent circuits 0.1:5 power divider
Guo et al. [23]	10	3.5–5.8	> 15	6.0cm^2	0.7–1.4	Ring cavity multiple –way technique is used
Wang et al. [24]	10	1.17	15	20.90 cm^2	0.9	4:1 unequal WPD with artificial transmission line (ATL) are used to get high impedance
Huang et al. [25]	> -20	Not quoted	> -20	12cm^2	3.5	Even–odd mode analysis was used in generalized power divider
Piacibello et al. [26]	> 25	4.75	23	0.02 λ g^2	0.9	Branch directional coupler is used
Tu et al. [27]	Not quoted	0.9	Not quoted	Not quoted	0.8 to 1.2	Reconfigurable, single /multiband Filtering PD
Dimitra Psychogiou et al. [28]	< -15	Not quoted	> 17	0.80 λ_g × 0.50 λ_g	1.55–4.24	Wide band filtering PD operation by embedding transversal filters is realized

(continued)

Table 1 (continued)

Method	Return loss (dB)	Insertion loss (dB)	Isolation (dB)	Size in area mm2	Center Freq. (GHz)	Remarks
Jiao et al. [29]	> 19.5	4.77	19.5	0.34	3.3	Asymmetric 3-way equal wide band filtering PD
Faraz shaikh et al. [32]	−10	Not quoted	−10	66.3 × 100 mm^2	3–8	The modified power divider shows equal power split which is suitable for antenna feeding network
Masrakin et al. [33]	−18	Not quoted	−18	40 × 30 mm^2	2	The two way power divider is used for IoT application

3 Research Methodology

In research methodology, the design procedure is systematically carried out to modify 1:2 WPD into a 4-port WPD. In theoretical analysis a 2 port structure of the WPD is analyzed using Z-parameters and the modified 4 port structure is presented in Sect. 3.1. In Sect. 3.2, the analysis of an S-matrix helps the designer to calculate the critical S-parameters of the proposed system.

3.1 Theoretical Analysis

In the structure of the WPD shown in Fig. 4a, a resistor R is connected between the output ports to improve the matching and isolation. Let, the characteristic impedance Z_0 and load impedance Z_L be 50Ω each that is $Z_0 = Z_L = 50\Omega$.

If all input ports are matched for no reflection then,

$$\frac{1}{Z_0} = \frac{1}{Z_{in\,2}} = \frac{1}{Z_{in\,3}} \tag{1}$$

From Fig. 2a,

$$Z_{in\,2} = Z_{in\,3} = Z_1 + Z_L \tag{2}$$

where $Z1$ is the characteristic impedance of the $\lambda/4$ transmission line.

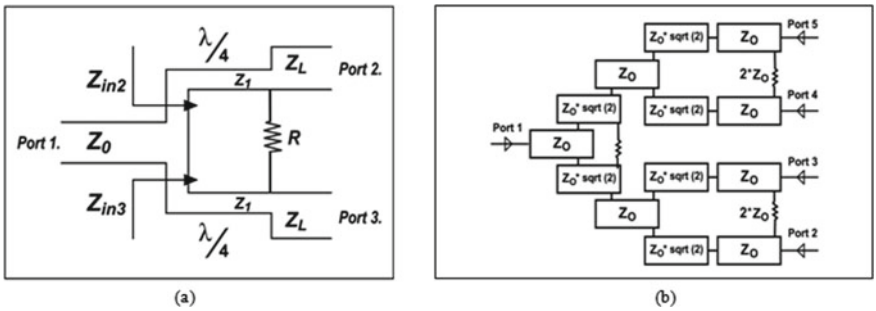


Fig. 4 **a** 2 port structure of WPD, **b** 4 Port Structure of WPD

According to the quarter wave impedance transforming theory,

$$Z_{\text{in}\,2} \times Z_L = Z_1^2 \tag{3}$$

$$Z_{\text{in}\,3} \times Z_L = Z_1^2 \tag{4}$$

By solving Eqs. 2, 3 and 4 Z_1 is expressed as

$$Z1 = \mathbf{2}Z0 \tag{5}$$

From the even/odd mode analysis, the resistor R is obtained as:

$$R = 2Z_0$$

The proposed WPD is modified with three sections of two-port symmetrical conventional power dividers to get 4 output ports as shown in Fig. 4b.

3.2 S-parameters

S-parameters are complex numbers having real and imaginary parts. Here, S-parameters are used to describe the relationship between different ports when it becomes important to describe the network in terms of amplitude and phase versus frequency rather than voltages and currents. They are measured in dB. To derive the fundamental properties of a lossless, reciprocal 3 port junction power divider the general expression of S-parameters in matrix form is derived as under.

The matrix of S-parameters for an ideal three-port WPD is

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix}$$

For reciprocal networks the matrix is symmetric and ideally, to avoid any loss of power, the network would be lossless and matched at all ports. When all ports are matched and the reciprocal matrix reduces to

$$S = \begin{bmatrix} 0 & s_{12} & s_{13} \\ s_{21} & 0 & s_{23} \\ s_{31} & s_{32} & 0 \end{bmatrix}$$

If the network is lossless, [S] must be unitary, which implies the following conditions:

$$\begin{aligned} S_{31}^{*} S_{32} &= 0 \\ S_{21} S_{23} &= 0 \\ S_{12} S_{13} &= 0 \\ |S_{12}|^2 + |S_{13}|^2 &= 1 \\ |S_{21}|^2 + |S_{23}|^2 &= 1 \\ |S_{31}|^2 + |S_{32}|^2 &= 1 \end{aligned}$$

These equations can be satisfied in one of two ways. Either

$$S_{12} = S_{23} = S_{31} = 0, |S_{21}| = |S_{32}| = |S_{13}| = 1$$

OR

$$S_{21} = S_{32} = S_{13} = 0, |S_{12}| = |S_{23}| = |S_{31}| = 1$$

With reference to the above, the S-matrix for 1:4 port network can be developed as

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} & S_{14} & S_{15} \\ S_{21} & S_{22} & S_{23} & S_{24} & S_{25} \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ S_{41} & S_{42} & S_{43} & S_{44} & S_{45} \\ S_{51} & S_{52} & S_{53} & S_{54} & S_{55} \end{bmatrix} \tag{6}$$

For an ideal 1:4 WPD, the following condition must be satisfied.

$$S_{11} = S_{22} = S_{33} = S_{44} = S_{55} = 0 \tag{7}$$

$$S_{12} = S_{21}, S_{13} = S_{31}, \ S_{14} = S_{41}, S_{15} = S_{51} \tag{8}$$

$$S_{23} = S_{24} = S_{25} = S_{34} = S_{35} = S_{45} = 0 \tag{9}$$

Above equations show that at least two of the three parameters ($S_{12}; S_{13}; S_{23}$) must be zero. However, this condition will always be inconsistent with one of equations, implying that a three-port network cannot be simultaneously lossless, reciprocal, and matched at all ports. If any one of these three conditions is relaxed, then a physically realizable device is possible. These results show that $S_{ij} \neq S_{ji}$ for $i \neq j$ which implies that the device must be non-reciprocal.

4 Design and Simulation

Figure 5 depicts the proposed WPD which plays an important role in communication systems because of its characteristics, namely simple configuration, matching of impedance, and isolation at output ports. The structure of the WPD, as displayed in Fig. 5a comprises two $\lambda/4$ transmission lines and the branch impedances $\sqrt{2Z_0}$ and $2Z_0$.

An analysis of the circuit is undertaken using the even–odd mode [10]. The prototype power divider is visible in Fig. 5b which is printed on the FR-4 substrate (εr 4.3, thickness 1.6 mm) having an overall dimension of 157 × 112 mm^2. The WPD acts as a network of corporate feed for feeding the phase shifter with equal phase and amplitude. The output ports have no reflection, and hence, they are perfectly isolated.

The surface current for the power divider is shown in Fig. 6a–e. It can be seen that no coupling of the current is observed between ports when port 1 is excited by terminating the other ports (such as 3,4,5) and observing the current at the testing port

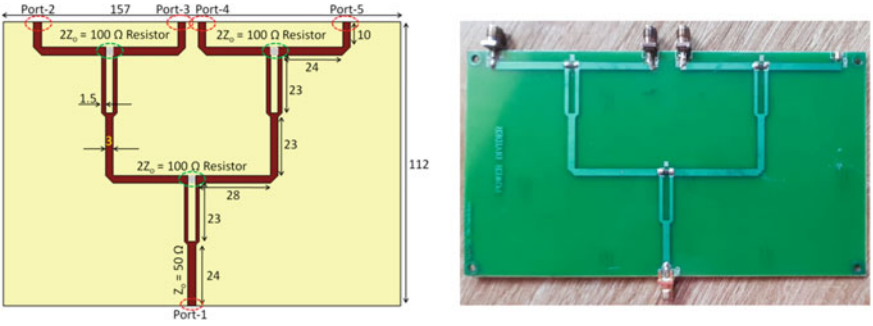


Fig. 5 Wilkinson power divider (WPD) **a** Geometry **b** Fabricated prototype

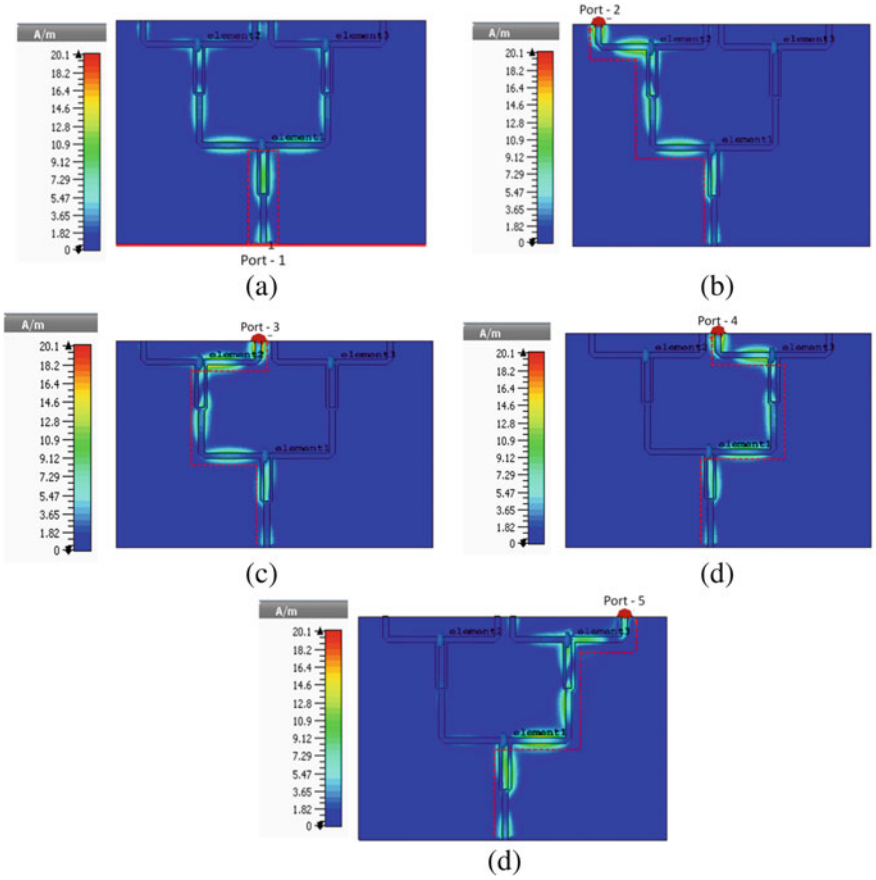


Fig. 6 Surface current at **a** Port 1 **b** Port 2 **c** Port 3 **d** Port 4 **e** Port 5

2. Similarly the same procedure is true for testing port 3, 4 and 5 keeping excitation port as 1.

Figure 7 illustrates that the power divider has a measured fractional bandwidth of 39.69%. It can also be visualized that the measured S21, S31, S41, and S51 are − 6.25, − 6.31, − 6.28, and − 6.31, respectively, which are very low and shows good agreement with the simulated values. It can also be seen that, S21, resembles S31, S41, and S51 which indicates equal power distributed to all the output ports.

From the above analysis, it is verified that the designed 1:4 WPD exhibits a good performance, the insertion loss is low and good VSWR.

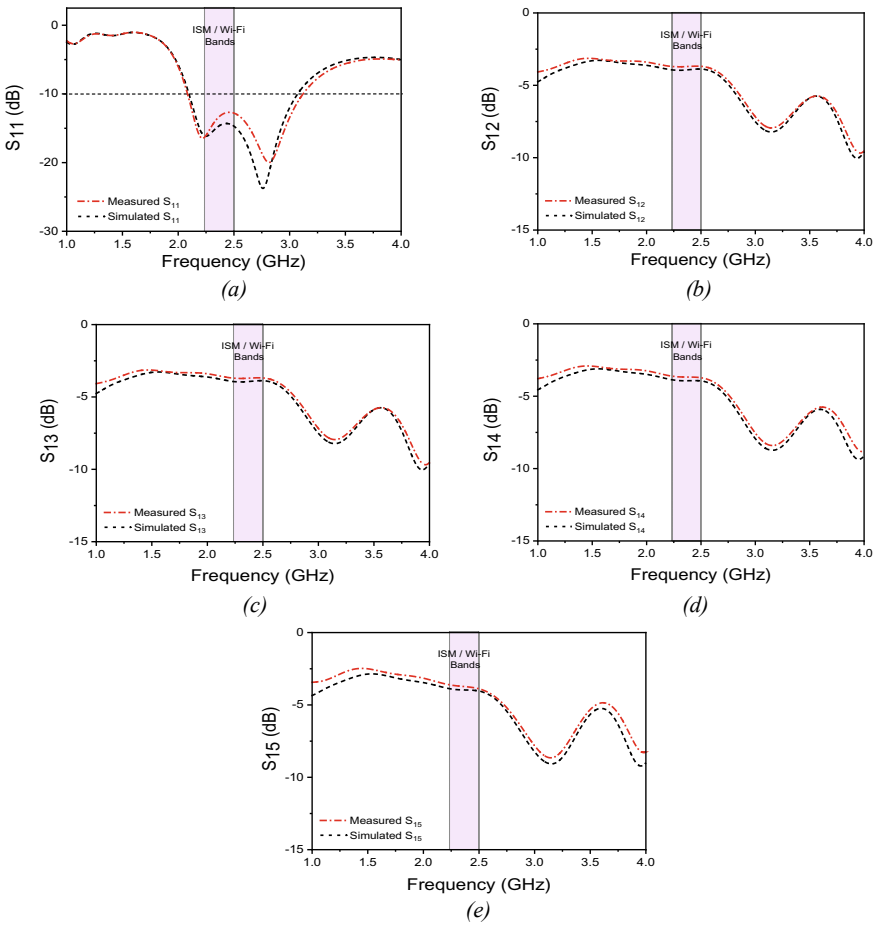


Fig. 7 S-Parameters result of the Wilkinson power divider **a** S11 **b** S21 **c** S31 **d** S41 **e** S51 simulated (Solid)/measured (Dashed)

5 Performance Comparison

For proof of the enhanced performance a comparison of the 1:4 WPD under study, conducted including the size, S-parameters and center frequency is exhibited in Table 2.The experimental and theoretical results are tabulated so as to compare each of the modified WPDs with the proposed work. As per the theoretical calculations the maximum insertion loss expected at 2-port 3-port, 4-port, 6-port of the power dividers are −3, −4.8, −6 and −7.78 dB, respectively. Similarly, the ideal isolation loss is infinity (∞) for 2-port and 3-port power dividers. For 4-port and 6-port power dividers, it is 21.6 and 17.6 dB, respectively.

Table 2 Performance comparison of with recently reported Wilkinson power dividers

Method	VSWR	Insertion loss (dB)	Isolation (dB)	Size in area mm2	Center Freq (GHz)	Comments
HJiang et al. [5]	1.92	4.2	> 25	25.10 cm^2	2	A parallel-strip swap was used to enhance isolation. 1:2 PD
Yuenjin and Xing et al. [6]	–	4.77	> 30	31.32	0.9	Bent micro strip line reduced the size of the divider.1:3 unequal PD
Agma Pribawa et al. [9]	1.75	< 7.8	< 12.8	Not quoted	1.5	Microstrip technology, 1:6 Wilkinson topology
Tseng et al. [11]	1.433	7.75	15	21.16	1.6	Glass-based TFIPD technology was used to compact chip size. Broad band 1:4 PD
Christain Mahardika et al. [30]	1.715	6.1	30	Not quoted	2	Modified 1:4 WPD operating at S band
Choe et al. [31]	1.67	6.3	33	Not quoted	2	Compact modified 1:4 WPD with physical output port isolation
Proposed work	1.614	6.28	> 20	157 × 112 mm^2	2.45	Modified 1:4 WPD

The most critical parameter in the design of the power divider is the insertion loss which decides the mode of power flow in the output ports of multiport power dividers. The maximum insertion loss for the 1:4 power divider is -6 dB. But due to critical design analysis, we have obtained it as -6.28 dB in the proposed work. The same is the case with all the reported 1:4 power dividers. The VSWR of the proposed work is 1.614 which is less than 2 making it suitable for antenna applications. No coupling current is observed between output ports when the input port is excited which indicates that the isolation is perfect between output ports. The only drawback in the proposed power divider is the larger size, which is 174.84 cm2 size because six $\lambda/4$ transmission lines have been utilized and additional lines are used to make a connection between the SMA connectors and the two way power dividers. The proposed power divider is the cheapest and the simplest in design and meets all the

requirements of the so called modified versions of the WPD topology. It is most suitable as a part of feed network used in phase array.

6 Conclusion

Wilkinson topology has been used to design the proposed 1:4 WPD. It is distinguished to exhibit its execution in ISM band frequency range. The power divider that has been proposed in this paper uses the 1:2 power divider in the first stage and makes use of two 1:2 power dividers in the second stage. A FR4 substrate is used for deploying it. The insertion loss of the proposed power divider is −6.28 dB, the VSWR is 1.614 and the isolation is greater than 20 dB at 2.45 GHz. These characteristics accommodate all the requirements at the desired frequency. The isolation at the output port is perfectly matched due to the fact that no reflection is shown at the output ports. The designed WPD meets all the requirements of an advanced modified WPD except the large size.

The technique for co-design of filters and power dividers is a new topic of future interest. Using computer aided design tools capable of analysis of microwave circuits, the calculation of the amplitude and phase of the scattering parameters can be achieved very accurately. The use of a multilayer substrate can be useful for increasing the bandwidth, improving the isolation and reducing the size of power dividers in future design of such power dividers.

References

1. E.J. Wilkinson, An N-way hybrid power divider. IRE Trans. Microwav. Theory Techniq. **8**(1), 116–118 (1960)
2. R.B. Ekinge, A new method of synthesizing matched broad-band tem-mode three-ports. IEEE Trans. Microwave Theory Tech. **19**, 81–88 (1971)
3. L.H. Lu, Y.-T. Liao, C.R. Wu, A miniaturized wilkinson power divider with CMOS active inductors. IEEE Microwav. Wirless Components Lett. **15**(11), (2005)
4. L. Chiu, Q. Xue, A parallel-strip ring power divider with high isolation and arbitrary power-dividing ratio. IEEE Trans. Microwav. Techniq. **55**(11) (2007)
5. J.-B. HJiang, Z.-H. Yan, F.-F. Fan, X. Zhang, A miniaturized three way power divider. IEEE (2008)
6. L. Yuenjin, X. Menjiang, Z. Zhangming, Y. Yintage, Novel compact compass navigation system (CNS) power divider. IEEE (2010)
7. K. Ding, D. Su, F. He, Y. Wang, A novel micro strip line three-way power divider. IEEE (2011)
8. R. Mirzavand, M.M. Honari, A. Abdipour, G. Moradi, Compact microstrip wilkinson power dividers with harmonic suppression and arbirary power division ratios. IEEE Trans. Microwav. Techniq. **61**(1) (2013)
9. N.A. Pribawa, A. Munir, Wilkinson topology-based 1:6 power divider for l-band frequency application. IEEE (2012)
10. B. Fu, X. Wei, A compact wilkinson power divider with LTCC technology. IEEE (2012)
11. T.-G. Tseng, Y.-S. Lin, Miniature broadband four-way power divider in glass-based thin-film integrated passive devise technology. IEEE (2013)

12. W. Choe, J. Jeong, Compact modified wilkinson power divider with physical output port isolation. IEEE Trans. Microwav. Techniq. **24**(12) (2014)
13. L.-B. Wan, Y.-L. Guan, W.-L. Fu, A novel planar six-way power divider. IEEE (2013)
14. W.-Q. Liu, F. Wei, C.-H. Pang, X.-W. Shri, Design of a compact ultra-wideband power divider. IEEE (2012)
15. F. Xu, G. Guo, E. Lie, J. Wu, An ultra-broadband 3 dB power divider. IEEE (2012)
16. P.O. Afanasieve, V.A. Sledkov, M.B. Manuilov, A novel design of ultra-wideband strip-line power divider for 2–18 GHz. IEEE (2013)
17. W. Wei, L. Wencheng, C. Dan, P. Neng, L. Kai, L. Zhongwen, Design of dual-frequency unequal power divider with genetic algorithm. IEEE (2009)
18. Y. Wu, Y. Liu, Q. Xue, An analytcical approach for novel coupled-line dual-band wilkinson power divider. IEEE Trans. Microwave Techniq. **59**(2), (2010)
19. Y. Wu, Y. Liu, Q. Xue, S. Li, C. Yu, Analytical design method of multiway dual-band planar power dividers with arbitrary power division. IEEE Trans. Microwav. Techniques **58**(12) (2010)
20. S. Shie, W. Che, Dual-band power divider based on double-side parallel strip Line (DSPSL). in *CJMW 2011 Proceedings* (2011)
21. X. Wang, I. Sakagami, K. Takahashi, S. Okamura, A generalized dual-band wilkinson power divider with parallel L C, and R components. IEEE Trans. Microwav. Theory Techniq. **60**(4) (2012)
22. H. Fan, X. Liang, J. Geng, R. Jin, X. Zhou, Reconfigurable unequal power divider with a high dividing ratio. IEEE Microwav. Wireless Components Lett. (2015)
23. L. Guo, H. Zhu, A.M. Abbosh Wideband tunable in phase power divider using three –line coupled structure. IEEE Microwav. Wireless Components Lett. (2016)
24. C.-W. Wang, T.-G. Ma, C.-F. Yang, A New planar artificial transmission line and its applications to aminiaturized butler matrix. IEEE Trans. Microw. Theory Tech. **55**(12), 2792–2801 (2007)
25. W. Huang, W. Ruan, F. Tan, A miniaturized 4:1 unequal wilkinson power divider using artificial transmission lines and double sided parallel strip lines. Int. J. Antennas Propog (2017)
26. A. Piacibello, M. Pirola, G. Ghione, Generalized symmetrical 3dB power dividers with complex termination impedances, IEEE Access (2020)
27. W.H. Tu, K. Chang, Microstrip elliptic- function low pass filters using distributed elements or slotted ground structure. IEEE Trans. Microw. Theory Tech. **54**(10), 3792 (2006)
28. D. Psychogiou, R.G. Garcia, A.C. Guyette, D. Peroulis, Reconfigurable single/multi-band filtering power divider based on quasi bandpass sections. IEEE Microwave Wireless Components Lett. **64**(9) (2016)
29. L. Jiao, Y. Wu, Y. Liu, Z. Ghassemlooy, Wideband filtering power divider with embedded transversal signal interference sections. IEEE Microwave Wireless Components Lett. **27**(12) (2017)
30. C. Mahardika, B.S. Nugroho, B. Syihabuddin, A.D. Prasetyo, Modified wilkon son power divider 1 to 4 at S band. in *The 2016 International Conference on Control, Electronics, Renewable Energu and Communicarion* (ICCEREC) (2016)
31. W. Choe, J. Jeong, Compact modified wilkinson power divider with physical output port isolation. IEEE Trans Microwave Techniques **24**(12) (2014)
32. F.A. Shaikh, S. Khan, S. Khan, M.H. Habaebi, Design and analysis of 1×4 wilkinson power divider for antenna array feeding network. in *IEEE International Conference on Innovative Research and Development* (ICIRD), Bangkok Thailand, 11-12 May (2018)
33. K. Masrakin, M.I. Zulkepli, S.Z. Ibrahim, H.A. Rahim, A.A. Dewani, M.N.A. Karim, Compact 2-way power divider for IoT application. in *5th International Conference on Electronic Design* (ICED) (2020)

An OAuth-Based Authentication System for IoT Networks Using LabVIEW

P. Kalpana Devi, M. Manasa, S. Naga Chandra Prakash, and B. Vittal Teja

Abstract OAuth is a significant authentication method used for ensuring the security of Internet of Things (IoT). In Internet of Things (IoT) networks, OAuth is a framework used for incorporating authentication, which has developed the workflow model for web server applications. This framework offers authentication to the web service through which the user's account is hosted, as well as secure API access for external users. To protect server resources and enhance scaling and distribution, a dedicated authorization framework is required. Henceforth, this research work has suggested OAuth2.0 for accessing data from IoT devices with sensors, which allows the users to maintain precise records of temperature and pressure in industries. It is informed that, creating APIs in other operating systems is difficult. For developing the Web services API, the API tokens are used to include the user account into the access token. OAuth2.0 offers authorization without requiring a user account. To address the insecurity issue, we integrate the OAuth2.0 process with the API in order to securely access data. OAuth2.0 grants the client "secure delegated access" to the server.

Keywords OAuth2.0 · LabVIEW · API · Server · Web servers

P. Kalpana Devi (✉) · M. Manasa · S. N. C. Prakash · B. V. Teja
VelTech Rangarajan Dr. Sagunthala, R & D Institute of Science and Technology, Chennai, India
e-mail: drkalpanadevip@veltech.edu.in

M. Manasa
e-mail: vtu10533@veltech.edu.in

S. N. C. Prakash
e-mail: vtu9814@veltech.edu.in

B. V. Teja
e-mail: vtu9945@veltech.edu.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_46

1 Introduction

IoT security has scope for widespread IoT visions such as trusted sensing, communication, computation and security. Now a days most of the users suffers for securing the important data while login/signup with their email or password. In view of overcome, the problem of insecurity of the personal data of the users OAuth2.0 plays an important role in providing such delegated services where in which their username or password are not stored, with the help OAuth API'S it establish the secure connections to users [1]. OAuth 2 is an authorization framework that enables applications to obtain limited access to user accounts on an HTTP service, such as Facebook, GitHub, and DigitalOcean or any application of the users. It works by delegating user authentication to the service that hosts the user account, and authorizing third party applications to access the user account. The proposed research work intends to design a delegated authorization framework for the API called as OAuth2.0, which primarily create LabVIEW web service because as some users will attempt to use the toolkit on other operating system. Also, certain functions are not supported and it is more important to make a design decision for target operating systems and hence it is difficult for the user to design the API in other operating system and so we know that "ALL WEBSERVICES ARE API," so in order to overcome the difficulty we have decided to create a Web services, secondly we are going to enable the OAuth2.0 workflow in the API /web services as the API tokens incorporate the user account in the access token, while OAuth2.0 performs authorization without a user account so in order to overcome the difficult of insecurity [2], we add the workflow of OAuth2.0 with the API in order to access the data more securely OAuth2.0 provides to client a "secure delegated access" to server resources on behalf of a resource owner. In this project, OAuth2.0 plays important role in accessing the data from the IOT device such as Aurdino UNO Device and helps the users to make the precise record of the temperature, pressure in the Industries. OAuth 2.0 provides authorization flows for web and desktop applications, and mobile devices. In this project, we aim to create a web application in LabVIEW where we have enabled the OAuth2.0 work flow in the API'S for enabling the service for the third party applications for the signing into the particular web application created for controlling the industry related problems such as pressure temperature by using Arudino UNO interfacing with the LabVIEW. OAuth 2.0 is an authorization framework for enabling resource sharing in a secured manner through a sequence of steps where resource owner permits a client application to a certain protected resource for a limited time. By using delegation pattern, client applications will be enabled by OAuth framework [3]. From existing resource server, some functionalities can be delegated without repeation. For access the files on a Google Drive is made through OAuth-based authorization. A website provides single sign-on, utilize your existing Facebook, Google or Twitter account through OAuth-based authorization and then continue using that site without signing up with a separate account on that site. We have used the Arduino UNO, Breadboard, Temperature sensors MakerHub linx and NIVISA package and LabVIEW interface with the arduino.

2 Methodology

2.1 Creating the Web Server

HTTP protocols are used in web server and for responding the client requests made with WWW [4]. The website content is displayed in web pages with storing, processing and delivering to users. Client server model is working in a similar process with web server process. Website content is delivered to the requesting user through the web server software [5]. The domain names can be used for accessing the web server software. HTTP server is the key component in software. HTTP and URLs can be handled with HTTP server.

Web server software and website contents are stored in web server. Website contents are HTML documents, images and Javascript files. Webserver should consists of web browser information like google chrome and Firefox files [6]. The request is made through HTTP. When request raised through HTTP to web server, the content requested is identified and sent back to the browser requested. Even more a specific page can also be requested through HTTP by web browser with series of processing steps.

Web browser URL address link to be specified. The domain name IP address will be provided to the web browser.

The domain name is provided with Domain name system or search in Cache to enable the browser. The browser will then make another HTTP request for a specified file. The webserver accepts the request through HTTP and returns the desired page. If that particular page is not available, the web server will issue an error message.

LabVIEW web service request proceeds with refnum by identifying the current HTTP request as mentioned in Fig. 1 Interfacing Makerhub with LabVIEW for LED lighting by using arduino board as an example program to carry out the operation. Include this file in your web resource to allow access to the web server.

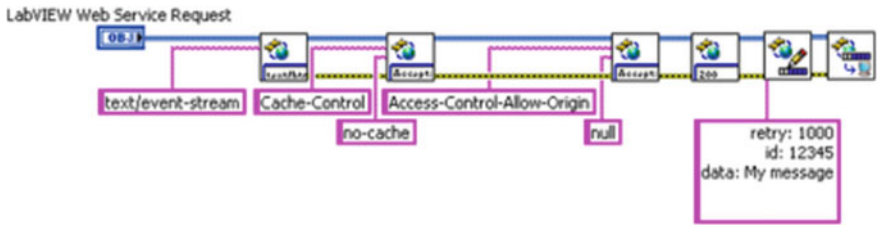


Fig. 1 Creating the web server

2.2 OAuth in LabVIEW

The limited access will be provided to HTTP service by any third party application and further it will be done by OAuth 2.0 application framework.

This access is provided through the approval interaction on behalf of resource owner with HTTP service. The third party can also access the service on its own behalf.

The OAuth2.0 provides the limited access for the application in order to access the protected resources. User does not require their login credentials to access the application. This application sends a request to the server in order to access the protected assets from the owner's resources on mobile devices and standard web applications. The request to access was granted by the owner.

3 Simulation of OAuth2.0 in LabVIEW

Specification

Facebook and Google HR services in intranet are the services offered by the resource server. When OAuth 2.0 token is raised, validation is checked further to process API requests from Apgee edge. The authorization is required for the resource server, especially for the application's protected resources (Fig. 2).

The OAuth 2.0 specification is validating the authorization grants from server and it also issues the access tokens. The token endpoints is required to be configured on Apgee edge. The authorization server is operated by Edge. The application permission is provided by authorization grant for retrieving the access tokens [7, 8]. Four specific grant types are defined by OAuth 2.0. To access the protected resources,

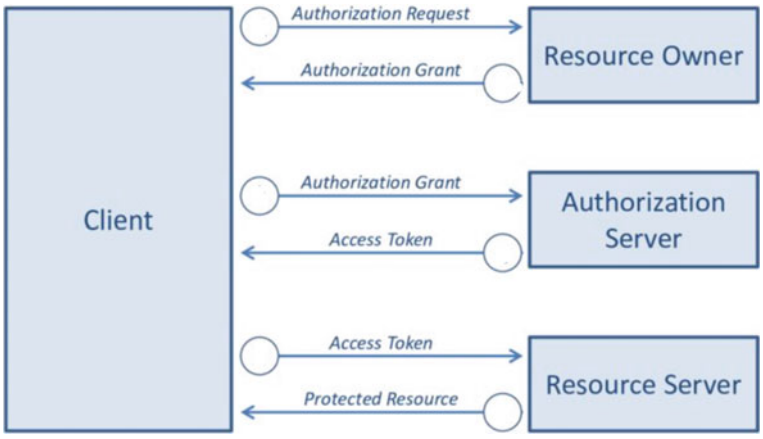


Fig. 2 Flow chart for OAuth 2.0 framework

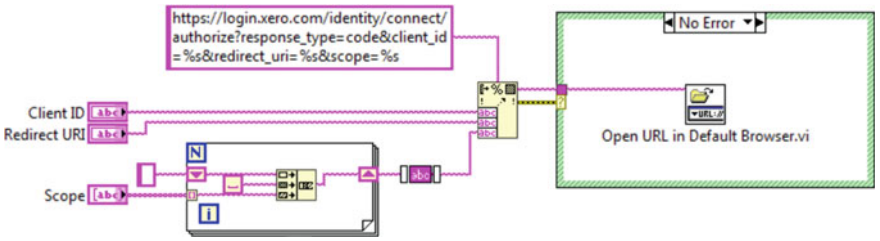


Fig. 3 Overall block diagram of string to Base64url

a long character string called "access token" is used. Resource owner handles the protected data resource such as the user's contact list, account information and other sensitive data. The steps of OAuth 2.0 flow for creating the webserver, getting the url of the web server is as follows:

3 Steps to design OAuth2.0

1. String to Base64url (Fig. 3)

Publishing Web services

A stand alone web services called Application web server is published in LabVIEW for accessing Web server. The stand alone web service applications is hosted by web server in a network. To protect the network data exchange, the multiple security related features is provided along with secure Socket Layer (SSL) encryption. This method is executable for even complex application.

The web service handles the different web servers in stand alone applications. As a result, the application web server publish should be selected and the deployment progress dialog box appears. If the web service is successfully published, the close button will be selected. Now, data exchange from web clients will be established via HTTP method. As mentioned in debugging, select the web service project item and choose the application web server in order to manage the web server. Web browser opens the NI web-based configuration for monitoring. Then, the web services management and tutorial service will be selected from the list. If the list is shown as empty, the process will be refreshed. Also, ensure that the web service status is in running mode. To make any further adjustments, such as resume, restart, or unpublished web service, use the buttons at the bottom of the page. To register the firm name with the Google Console Developers, first connect to the Google Cloud Platform with the appropriate gmail id and accept the required access credentials requirements (Fig. 4).

For credentials, enter our firm name or any other app name under project name, choose the place where it is located, and click on as indicated. Next, go to the OAuth consent screen and enter the company under the external and click on next to update the procedure. Now fill the required things such as the app name and domain name, respectively, in the OAuth consent form and click on save and continue to move forward for the further process and similarly fill the scope and next process by save

Fig. 4 Output readings from sensor ınto the webpage in xml format

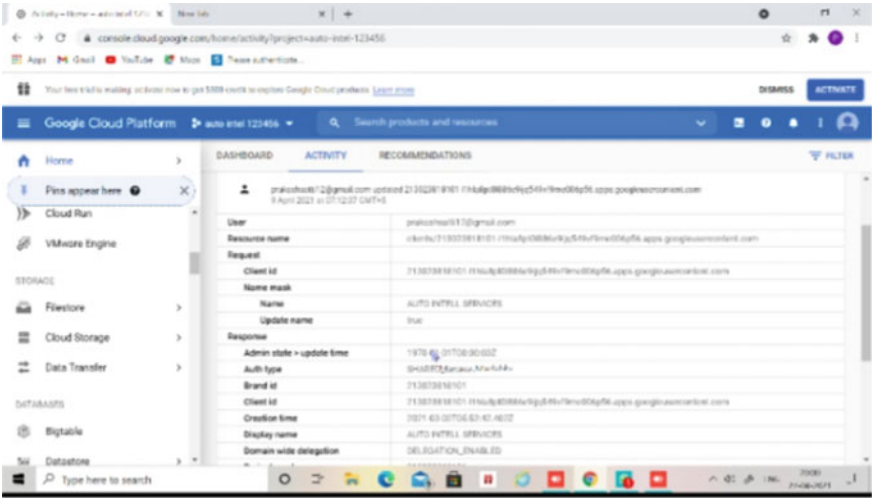


Fig. 5 Client Id and secrete code of the company

and continue. again open the credentials and open on the company name and register the redirect URL, which is created by the users using LabVIEW webserver a new client id and secret code is released for different redirect url you will get different client id and secret copy paste all the required things in the LabVIEW code such as redirect link client id and client password as shown in Fig. 5 and inserting the sample inputs as shown in Fig. 6 .

4 Conclusion

From this research work, it is evident that, nowadays, people transfer most of their day-to-day activities to be computerized. Traditionally, different industrial activities are monitored by using a camera, which can also lead to some errors. To overcome such challenges, this research work has successfully created a webserver with OAuth2.0 authentication. In this, we need to enter the industry and also the user name and password should also be entered; this may also create a fear of stealing important information or accessing important things, or it may cause insecurities.

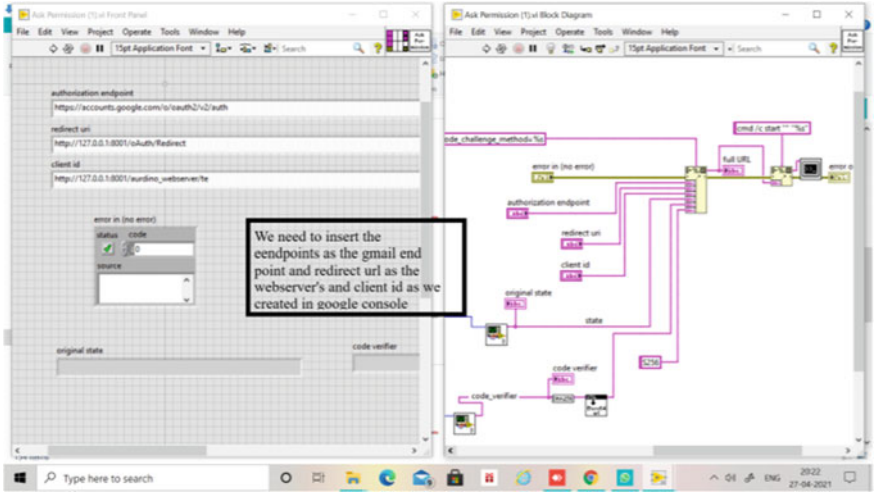


Fig. 6 Inserting the sample inputs

Henceforth, this research work has integrated OAuth2.0 with LabVIEW for security purposes and also efficiently developed a web server, where one may retrieve the data of the sensor or any data from their location.

References

1. The OAuth 2.0 Authorization Framework. https://tools.ietf.org/html/rfc6749. [Online]; Accessed October 2019. User Authentication with OAuth 2.0.https://oauth.net/articles/authentication/. [Online]; Accessed October 2019
2. R. Buyya, A.V. Dastjerdi, Internet of things: principles and paradigms. (Elsevier, 2016)
3. M. Picone, P. Gonizzi, L. Veltri, G. Ferrari, IoT-OAS: an OAuth-based authorization service architecture for secure services in IoT scenarios simone cirani. IEEE Sensor J.
4. K. Yokogi, N. Kitagawa, N. Yamai, Access control model for IOT environment including automated configuration. in *2018 IEEE 42nd Annual Computer Software and Applications Conference* (COMPSAC), Tokyo, Japan, July 2018, vol. 02 (IEEE, 2016), pp. 616–621
5. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov
6. C.-J. Chae, K.-N. Choi, K. Choi, Y.-H. Yae, Y. Shin, The extended authentication protocol us-ing e-mail authentication in OAuth 2.0 protocol forsecure granting of user access. J Internet Comput Services (JICS) 16(1), 21–28 (2015)
7. D. Dolev, A. Yao, On the security of public key protocols. IEEE Trans. Inform. Theory **29**(2), 198–208 (1983)
8. X. Hang, J. Li, Z.Z. Xu, D. Feng, H. Hu, Multiple handshakes security of TLS 1.3 candidates. in SP. (IEEE, 2016)
9. A. Kumar, Using automated model analysis for reasoning about security of web protocols. in *ACSAC* (ACM, 2012)
10. N. Fotiou, G.C. Polyzos, Authentication and authorization for interoperable iot architectures. in *International Workshop on Emerging Technologies for Authorization and Authentication* (Springer, Barcelona, 2018) pp. 3–16

11. A. Niruntasukrat, C. Issariyapat, P. Pongpaibool, K. Meesublak, P. Aiumsupucgul, A. Panya, Authorization mechanism for mqtt-based internet of things. in *2016 IEEE International Conference on Communications Workshops* (ICC). Kuala Lumpur, Malaysia, May 2016 (IEEE), pp. 290–295
12. H. Ning, H. Liu, L.T. Yang, Cyberentity security in the internet of things. Computer **46**(4), 46–53 (2013)

Routing Stability Important Factor in Streaming in VANETS

Pooja Sharma, Ajay Kaul, and M. L. Garg

Abstract VANETs are a subset of MANETs that are essentially vehicular ad hoc networks, a type of Ad hoc Network. By considering the resource requirements of various applications, networks make resources accessible to these applications in order to maximize the network usage. However, when Ad hoc networks are considered, it is highly difficult to offer these resources to applications when there are variable resources, such as bandwidth, which changes every time when Ad hoc networks are considered.

Keywords MANETs · VANETs · QoS · Routing protocols

1 Introduction

Vehicular Ad hoc networks (VANETs) are Ad hoc networks in which nodes move at high speeds in a predetermined pattern, as opposed to MANETs in which nodes move randomly [1]. However, MANET nodes move at random and clearly not as quickly as VANETs. Bandwidth, energy, latency, jitter, and channel capacity are all required for various applications. By considering the resource requirements of various applications, networks make resources accessible to these applications in order to maximize the network usage. However, when Ad hoc networks are taken in consideration, it is extremely challenging to provide these resources to applications, when it has dynamic resources such as bandwidth that changes every time when Ad hoc networks are discussed. Many external factors influence the performance of applications using ad hoc networks.

P. Sharma (✉) · A. Kaul · M. L. Garg
Shri Mata Vaishno Devi University Katra, Jammu and Kashmir, India
e-mail: pooja.sharma@smvdu.ac.in

A. Kaul
e-mail: ajay.kaul@smvdu.ac.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_47

1.1 VANET Characteristics

Ad hoc networks first appeared in the 1970s, when networks were known as packet radio networks. VANETs must prioritize safety and traffic management [2]. Vehicles or nodes can alert other vehicles of impending difficulties such as road conditions, traffic congestion, or sudden halt. IEEE 802.11p mentions a new physical layer as well as a Medium access control (MAC) layer for vehicle communication [3, 4]. Maintaining the quality of service parameters is one of the most difficult tasks in VANETs. Due to the great mobility and complexity of vehicle flow, trustworthy data streaming via vehicular Ad hoc networks is a challenging task when compared to mobile Ad hoc networks. Vehicular Ad hoc networks are always important in smart cities, especially for safety applications and video monitoring services. Vehicles assist cooperative drivers and first aid personnel in sharing real-time information about accidents in a certain region. As a result, we require a robust way to sharing such critical information through the use of resilient routes even in multi hop, multi path, and dynamic environments [5]. Human-shared and viewed videos are described in terms of quality of service (QoS) and quality of experience (QoE), which is entirely based on user acceptance. Connectivity is a measure of the network's reachability, and it is critical to the QoS performance of vehicular Ad hoc networks. When they are in each other's transmission range, they are said to be connected.

There are various approaches, which work on improving the quality of service (QoS) of video streaming applications like clustering, cross layer design, data dissemination, opportunistic forwarding and many more [5, 6]. All these techniques help in improving the performance of VANET applications.

Video streaming is one of most critical and delay sensitive applications, which is highly sensitive to delay, jitter, data rate, loss rate and error rate. Route stability is one of the big factors employed in evaluating the quality of service of streaming applications as the more the routes are stable, where the more packets are transferred per unit time, thus increasing packet delivery ratio. If the packet delivery ratio is increased, quality of service of video streaming applications is increased. Therefore, in this thesis, to address the problem of video streaming applications over VANETs because of unstable routes as route are broken frequently in vehicular networks, the standard protocol Ad hoc On Demand routing Vector (AODV) [7] is modified by incorporating route stability. To make routes more stable, one new metric called minimum lifetime is added to the links by which we can predict and chose a path which has maximum minimum lifetime. Hence, problem is to maximize the minimum lifetimes of each of the link which in turn makes a path. This metric minimum lifetime has been given by Vinod Namboodri et al. it is based on various parameters like range of wireless networks, distance between two nodes and velocities of two nodes. In [8, 9] various routing strategies attempt to provide a stable route among nodes and ensure quality of service. Reliability of links are very important for the stability of a particular route [10]. By incorporating this route stability factor, there is an improvement seen in the throughput of video applications when they are transmitted as CBR and VBR traffic (Fig. 1).

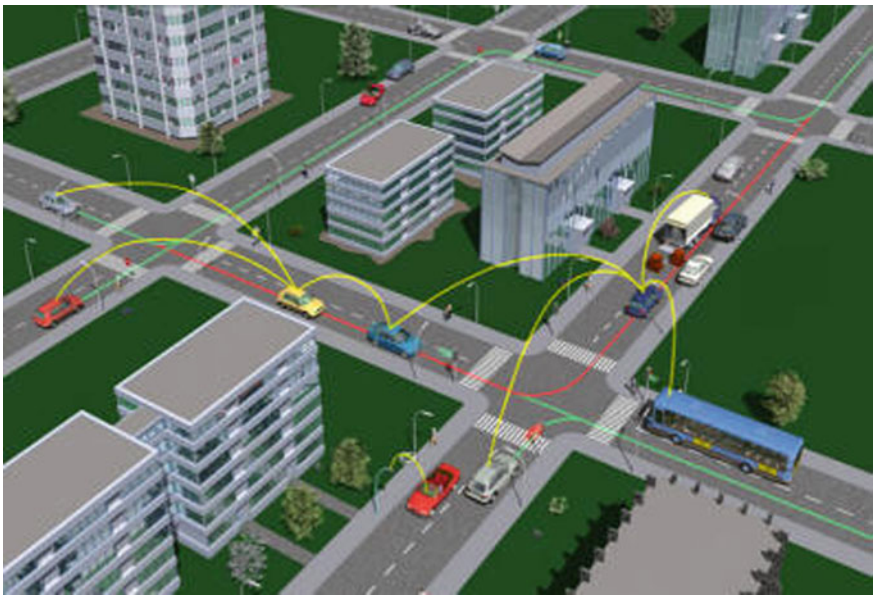


Fig. 1 VANET scenarios

At the same time, when video is transmitted as Voice over Internet Protocol (VOIP) traffic [11], there is a clear increase shown in mean opinion score (MOS) [12] is defined as a subjective metric to compute the quality of video at the application level. PSNR is one of the subjective parameter used to evaluate video quality. PSNR is computed with respect to the input video and received video in order to evaluate quality of video. There is a reduced delay and jitter which in turn improves the performance of an application. Our motive is to optimize the performance of applications over VANETs. Another important contribution of this thesis is that route error messages also decrease due to less route failures which in a way reduces packet losses.

Mean opinion score is a commonly used measure for video, audio and audio video quality evaluation. The MOS is expressed as a single rational number, typically in the range of 1–5. 1 is the lowest perceived quality, whereas 5 is the highest perceived quality. With reference to the multimedia, quality of service-based approaches assess the quality of streaming services through network oriented metrics. By taking the mean opinion score into consideration in simulation, performance of video streaming applications can be evaluated in a better way. Some of the ways for measuring video quality has been considered like packet loss ratio, mean opinion score, link failures and they are compared with its counterpart modified AODV where predicted lifetime has been taken into consideration. We can definitely include other video quality measures like PSNR. Algorithm has been modified with additional parameters like distance and speed of nodes. Comparison with other routing algorithms has been mentioned in future work. We have compared R-AODV with one routing protocol

Location aided routing as it is also location-based protocol. R-AODV has shown better performance than LAR [13].

2 Resource Provisioning for Video Streaming in VANET

The advantage of bit stream is that it allows to conduct simulations, where the quality of video after suffering losses in the network is evaluated [14]. One limitation of bit stream per second that bit stream is large in size [15, 16]. Another limitation is that they are usually proprietary protected by copyright. Therefore, network researchers are limited to bit streams and also limit the exchange of bit streams across the research groups. Generally, it is advised to cover more and more videos in order to cover maximum features in different scenarios. In a study, [17] has discussed many examples for scalable and non-scalable encoders. The video traffic trace is an abstraction of real video stream. It typically gives the frame, frame type (P, B, I) and frame size in a text file to characterize the real video traffic [18]. After decoding the video, performance can be evaluated by metric such as peak signal to noise ratio (PSNR). Video stream can be seen as a constant bit rate traffic (CBR) and variable bit rate (VBR) traffic. CBR traffic is mainly for real-time data dissemination and would be appropriate for voice and video application. CBR carries traffic at constant bit rate. Real-time variable bit rate is used for traffic which carries data at variable bit rate. Example is compressed video. Non-real-time variable bit rate is for VBR traffic where there is no reliance on time synchronization in traffic source and traffic destination. Constant bandwidth has to be guaranteed for CBR type of traffic. VBR traffic is meant for real-time and non-real-time data, i.e., which has changing traffic characteristics. This helps us to compare certain new hybrid protocols with the standard protocols on different parameters such as packet delivery ratio, packet loss, delay and jitter.

We need to provide resources in vehicular Ad hoc networks so as to have smooth execution of network applications, especially video streaming applications. Video streaming applications requirements are studied with an insight to delay, jitter, bandwidth requirements, throughput, etc., these metrics help us in assessing the quality of service for video streaming applications. Other network applications are also to be studied in order to study the effect of routing protocols in VANETs, especially with their characteristics like frequent disconnections because of high mobility of VANET nodes. When we need to provide better resources, we have to find stable routes, increase throughput, decrease delay and jitter so that quality of service of network application is maintained [19]. V2V communication is very important for ensuring safety and reliability of passengers in vehicles. Delay time is decreased by reducing route failures since it is proportional to the time it takes to send route error messages and time for retransmission. However, when a route is already chosen by considering its stability, route error messages and latency will also reduce, resulting in improved application performance. Also, a higher throughput and reduced latency and jitter

has been observed. Henceforth, this research work is very much in compliance to optimize and improve the performance of applications over VANETs.

3 Motivation

When it comes to intelligent transportation systems, VANET is considered as an emerging research area. It also possesses the requirement to have vehicle-to-vehicle communication, vehicle-to-infrastructure communication, and infrastructure-to-vehicle communication. These VANET connections significantly contribute to the safety of drivers and passengers. Traffic congestions and accidents ahead will be known earlier and appropriate action can be taken by driver and passenger. More live updates can also be availed by drivers [20, 21]. Vehicular Ad hoc communications are as critical as the speed of nodes, which is more comparable to vehicle nodes that lead to frequent disconnections of vehicles since the vehicles join the Ad hoc network and leave the network in less time [22]. This comes up with the need to look stable path, i.e., where nodes shall remain connected for more time. When nodes are connected, they form a stable path and lead to better throughput and packet delivery ratio for all kinds of network applications, especially time constrained applications like video streaming applications.

There are various metrics that can be considered for evaluating and assessing the quality of service of video applications like throughput, delay, jitter, bandwidth, response time, packet loss rate, error rate. Various link stability factors have been studied in literature, they are frequency and bandwidth. Various challenges of VANET communications should be considered, they are security, authentication, integrity, confidentiality, accessibility, scalability, reliability, and media access control. The security of message content has been an issue for communication. The message received needs to be verified in a short period of time so as to use the information at the earliest [23, 24].

Considering the link stability, the distance between two nodes play a very important role in connectivity, the lesser the distance between two nodes, the more they are connected. When they are approaching each other, they are connected and when they go away from each other, slowing going out of range and get disconnected.

Route stability is incorporated in standard AODV which is very helpful in making robust video transmission in VANET nodes as route stability is one of the key factors in Ad hoc networks as there is no predefined infrastructure. Route stability is used to determine the time for which route will be stable or the link between two nodes is alive or connected. In VANETs, where nodes are moving with high speed leading to more route breaks and in turn increasing RERR messages.

There has always been a focus on the beginning of VANET research present time. Early work focused on finding feasible routes without considering predicted lifetimes of links or QoS. A preemptive routing has been used in the general context of mobile Ad hoc networks, but never for vehicular Ad hoc networks [14 , 25, 26] PBR.

AODV protocol is widely used reactive routing protocols. In AODV, node broadcasts a RREQ packet, when it wants a route to specific host. Each node that receives route request (RREQ) packet checks whether it is the destination node, if it is, it sends a route reply (RREP) packet, otherwise it rebroadcasts a route request packet. The intermediate node forwards the RREP packet to the source according to their routing tables. One unique point in AODV is that nodes use "Hello" messages to probe their neighbors in order to validate routes Problem to address here is to optimize the performance of video streaming applications in vehicular Ad hoc network by using hybrid routing protocol in turn improving mean opinion score (MOS), packet delivery ratio, thereby reducing delay, jitter and packet loss ratio by reducing route error (RERR) messages, while disseminating video over Ad hoc networks.

4 System and Proposed R-AODV Routing Protocol Model

In the proposed routing protocol, standard AODV protocol is modified with one additional metric which is predicted lifetime of a link. Predicted lifetime is the lifetime for which nodes will remain connected. This lifetime has been calculated on the basis of distance between nodes and speed of the nodes or vehicles. This prediction will help us to predict the time for which nodes are going to be connected before selecting a route, a path is chosen on the basis of minimum predicted lifetime of a link which will form the part of route. Every route is made of certain links from source to destination and each link having predicted lifetime.

Our proposed protocol works for changing coordinates as they are changing frequently in VANETs. These estimates will not be stale as this will take very less time and would not take so long. Moreover, the calculation is updated according to the distance between nodes. Distance will keep changing according to the position of nodes. As long as distance is within the transmission range, nodes will remain connected. Every node's position and speed will be updated in RREQ packet by functions myposition(), myspeed().This info is used. Distance will be calculated for that time period only when two nodes will be connected. Coordinates of both nodes have been taken into account to calculate the distance. Actually both nodes are moving, it is for that particular time period. In real life, it can be deduced when distance is increasing between nodes/ vehicles, they are going away and will lose connection after a particular range. Moving nodes are taken into consideration by their direction and speed.

The route from source to destination is selected where we have maximum value for minimum predicted lifetime. It is a way to obtain the maximum of minimum predicted lifetimes. We are going to maximize the minimum predicted lifetime of various links forming the route. It will predict the stability of route formed from source to destination. In standard routing protocol AODV, RREQ and RREP messages can be seen in figure given below in Fig. 2.

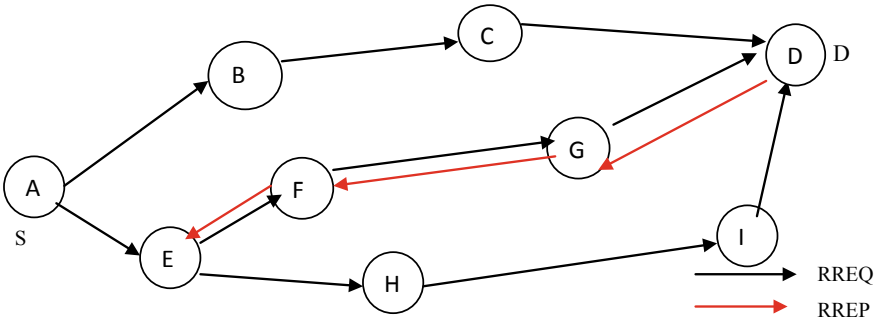


Fig. 2 RREQ and RREP messages in AODV

RERR messages can be seen in figure mentioned below. RERR messages are generated when link breaks. Every time a link is broken, RERR message is generated as shown in Fig. 3.

Now, modified figure for R-AODV is seen here with predicted link lifetimes given and deciding\on the basis of maximizing the minimum lifetime for which route will remain stable. The predicted lifetime is used to give the source a predicted lifetime for the route. This metric will help in forming better and stable routes. The route chosen will be on the basis of predicted lifetime of each link and in turn route lifetime is minimum of all the value of lifetimes of links. Initially, a node puts its location and velocity information and sets a lifetime field in the RREQ packet header equal to some value that is expected to be greater than the minimum of all link lifetimes along the route. The lifetimes of links are represented here as LLT1, LLT2 as shown in Fig. 4.

We can observe changes in routing in these transmissions.

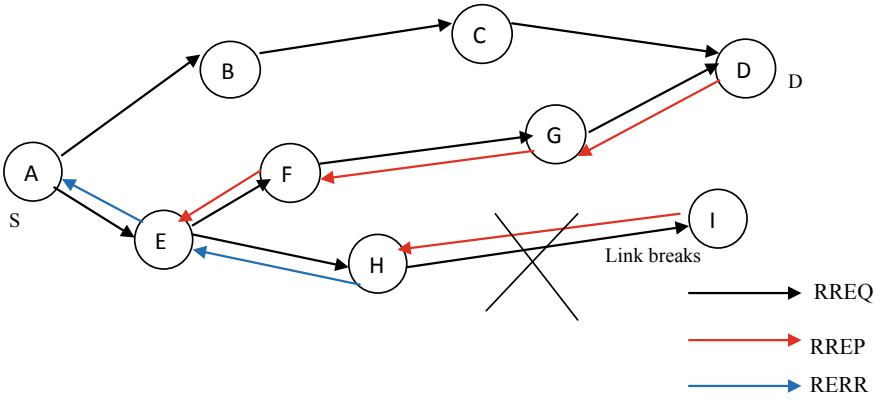


Fig. 3 RERR messages in AODV

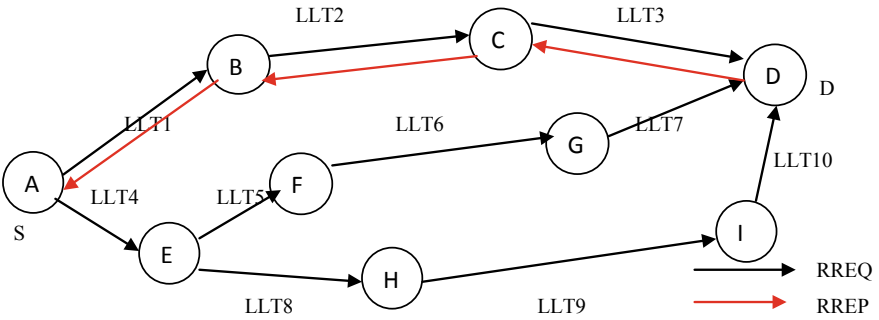


Fig. 4 RREQ messages along with link lifetimes

5 Conclusion

Different network applications have varying resource requirements, which have been carefully examined in this research article. The goal of investigating and evaluating various resource needs in terms of quality of service characteristics is to aid in the intelligent use of resources in networks, particularly in Ad hoc networks such as vehicular Ad hoc networks. Furthermore, a few quality of service (QoS) factors such as constant bit rate (CBR), variable bit rate (VBR), and video traffic have been considered to assess the performance of video streaming applications.

The proposed routing protocols would also reduce the number of route error (RERR) signals issued and link failures, resulting in more resilient routes. As route stability is very much related to number of route error (RERR) messages, a decrease in number of RERR messages for different scenarios of vehicular Ad hoc networks will routinely improve the route stability.

References

1. W. Chen, R.K. Guha, T.J. Kwon, J. Lee, Y.Y. Hsu, A survey and challenges in routing and data dissemination in vehicular ad hoc networks. Wireless Commun. Mobile Comput. **11**(7), 787–795 (2011)
2. R. Yugapriya, P. Dhivya, M.M. Dhivya, S. Kirubakaran, Adaptive traffic management with VANET in V to I communication using greedy forwarding algorithm. in *International Conference on Information Communication and Embedded Systems* (ICICES2014) February, (IEEE, 2014) pp. 1–6
3. M. Wellens, B. Westphal, P. Mahonen, Performance evaluation of IEEE 802.11-based WLANs in vehicular scenarios. in *2007 IEEE 65th Vehicular Technology Conference-VTC2007*, April (Spring, IEEE, 2007) pp. 1167–1171
4. T. Wiegand, G J. Sullivan, G. Bjontegaard, A. Luthra, Overview of the H. 264/AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol. **13**(7), 560–576 (2003)
5. M. Amadeo, C. Campolo, A. Molinaro, Enhancing IEEE 802.11 p/WAVE to provide infotainment applications in VANETs. Ad Hoc Netw. **10**(2), 253–269 (2012)

6. N. Kumar, N. Chilamkurti, J.J. Rodrigues, Learning automata-based opportunistic data aggregation and forwarding scheme for alert generation in vehicular ad hoc networks. Comput. Commun. **39**, 22–32 (2014)
7. I.D. Chakeres, L. Klein-Berndt, AODVjr, AODV simplified. ACM SIGMOBILE Mobile Comput. Commun. Rev. **6**(3), 100–101 (2002)
8. T. Taleb, E. Sakhaee, A. Jamalipour, K. Hashimoto, N. Kato, Y. Nemoto, A stable routing protocol to support ITS services in VANET networks. IEEE Trans. Veh. Technol. **56**(6), 3337–3347 (2007)
9. J. Rak, LLA: a new any path routing scheme providing long path lifetime in VANETs. IEEE Commun. Lett. **18**(2), 281–284 (2013)
10. O.A. Wahab, H. Otrok, A. Mourad, VANET QoS-OLSR: QoS-based clustering protocol for vehicular ad hoc networks. Comput. Commun. **36**(13), 1422–1435 (2013)
11. S. El Brak, M. Bouhorma, M. El Brak, A. Bohdhir, Speech quality evaluation based codec for VoIP over 802.11 P. Int. J. Wireless Mobile Netw. **5**(2), 59 (2013)
12. T.A.Q. Pham, K. Piamrat, C. Viho, Qoe-aware routing for video streaming over vanets. in *2014 IEEE 80th Vehicular Technology Conference* (VTC2014-Fall) September. (IEEE, 2014) pp. 1–5
13. R.S. Raw, D.K. Lobiyal, S. Das, S. Kumar, Analytical evaluation of directional-location aided routing protocol for VANETs. Wireless Pers. Commun. **82**(3), 1877–1891 (2015)
14. G. Van der Auwera, P.T. David, M. Reisslein, Traffic and quality characterization of single-layer video streams encoded with the H. 264/MPEG-4 advanced video coding standard and scalable video coding extension. IEEE Trans. Broadcasting **54**(3), 698–718 (2008)
15. S. Devillers, C. Timmerer, J. Heuer, H. Hellwagner, Bitstream syntax description-based adaptation in streaming and constrained environments. IEEE Trans. Multimedia **7**(3), 463–470 (2005)
16. P. Seeling, M. Reisslein, Video transport evaluation with H. 264 video traces. IEEE Commun. Surv. Tutorials **14**(4), 1142–1165 (2011)
17. S. Li, J. Xu, M. van der Schaar, W. Li, Trend-aware video caching through online learning. IEEE Trans. Multimedia **18**(12), 2503–2516 (2016)
18. J. Greengrass, J. Evans, A.C. Begen, Not all packets are equal, part i: streaming video coding and sla requirements. IEEE Internet Comput. **13**(1), 70–75 (2009)
19. P. Seeling, M. Reisslein, B. Kulapala, Network performance evaluation using frame size and quality traces of single-layer and two-layer video: a tutorial. IEEE Commun. Surv. Tutori. **6**(3), 58–78 (2004)
20. X. Tang, A. Zakhor, Matching pursuits multiple description coding for wireless video. IEEE Trans. Circuits Syst. Video Technol. **12**(6), 566–575 (2002)
21. S. Wenger, M. Horowitz, FMO: flexible macroblock ordering. JVT-C089 2430–2437 (2002)
22. A.M. Vegni, M. Biagi, R. Cusani, Smart vehicles, technologies and main applications in vehicular ad hoc networks. Vehicular Technol. Deploym. Appl. 3–20 (2013)
23. N. Qadri, M. Altaf M. Fleury, M. Ghanbari, H. Sammak Robust video streaming over an urban vanet. in *2009 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications* (IEEE, 2009) pp. 429–434
24. C. Rezende, A. Boukerche, H.S. Ramos, A.A. Loureiro, A reactive and scalable unicast solution for video streaming over VANETs. IEEE Trans. Comput. **64**(3), 614–626 (2014)
25. M. Sepulcre, J. Gozalvez, J. Härri, H. Hartenstein, Application-based congestion control policy for the communication channel in VANETs. IEEE Commun. Lett. **14**(10), 951–953 (2010)
26. B.T. Sharef, R.A. Alsaqour, M. Ismail, Vehicular communication ad hoc routing protocols: a survey. J. Netw. Comput. Appl. **40**, 363–396 (2014)

Evaluation of the Probability of Breaking the Electronic Digital Signature Elements

Lakhno Valeriy, Sahun Andrii, Khaidurov Vladyslav, Panasko Elena, Chepynoha Anatolii, and Ustianovska Nataliia

Abstract This article analyzes the standards, algorithms, and hash functions used in electronic digital signature (EDS). It is determined that most of the modern hash functions and algorithms used in EDS schemes are based on elliptic curves above the field. Some of the collisions causes, methods, and algorithms for hash attacks are covered in the article. A mathematical apparatus for estimating the probability of hash functions breaking based on the "birthday paradox" is formed. The results of the probability of breaking for hash functions used in EDS were obtained. It is confirmed that in case of the same length of the hash, the probability of breaking the hash by the breaking the resistance to collisions method is much lower than using the method of breaking the strong resistance to collisions. It has been suggested that it is dangerous to add a key at the beginning or end of a message when working with key hash functions.

Keywords Birthday paradox · Hash function attacks · Birthday attack · Collision of the second kind

L. Valeriy · S. Andrii (✉)
National University of Life and Environmental Sciences of Ukraine, Kiev, Ukraine
e-mail: avd29@ukr.net

L. Valeriy
e-mail: lva964@nubip.edu.ua

K. Vladyslav
Institute of Engineering, Thermophysics of NAS of Ukraine, Kiev, Ukraine

P. Elena · C. Anatolii · U. Nataliia
Cherkasy State Technological University, Cherkasy, Ukraine
e-mail: lena.pa@ukr.net

C. Anatolii
e-mail: a.chepynoha@chdtu.edu.ua

U. Nataliia
e-mail: Nataustyanovska@ukr.net

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_48

1 Introduction

Today, electronic digital signatures (EDS) are a common technology for data integrity protection. Some of its mechanisms (hash functions) are very widely used in cryptocurrency generation systems. The presence of several different EDS schemes using different hash functions in application systems makes it necessary to assess the level of security of data that is protected using EDS. Many publications have studied the problem of crypto resistance or the probability of breaking various elements of EDS [1–7]. We can single out the study of the probability of collisions of the first or second kind in the generated values of hash functions [3–6]. To facilitate the assessment of the detection of collisions (the appearance of two identical values of hash functions) often use the paradox of birthdays [7, 8].

2 Steps of EDS Elements Reliability Evaluation

2.1 Estimation of the Probability of Break of EDS Elements

In some countries, including Canada, South Africa, the United States of America, Algeria, Turkey, India, Brazil, Indonesia, Mexico, Saudi Arabia, Uruguay, Switzerland, Chile, Russia, Ukraine, and the European Union, electronic signatures have legal significance. Therefore, the assessment of the probability of break of the EDS elements is of great practical importance.

At present, the stability of almost all modern EDS standards is based on the complexity of solving the problem of discrete logarithm in a group of points of elliptic curves [9]. Moreover, the standards for obtaining the main element of information security in EDS-hash functions are very similar. For example, now any EDS scheme must define at least three functional algorithms, in particular, the algorithm for generating a key pair for the signature and its verification, the signature algorithm, the signature verification algorithm.

Consider the EDS standards of Ukraine, Russia, and the United States of America. At present, the text of the specification DSTU 4145–2002 for the EDS algorithm exists in the form of a corresponding document [10, 11]. This algorithm is based on elliptic curves over the field F (2) (non-supersingular curves). The standard uses curves in polynomial and optimal normal bases. Moreover, the curves in the polynomial basis can be of the second and third types. The basis for the development of this standard is the international standard IEEE 1363a. DSTU 4145–2002 is authentic compared to IEEE 1363a, due to the use of its own hashing function. The interstate standard GOST R 34.311–95 (formerly GOST 28,147–89 in the mode of imitation insert) is used as an algorithm of the hash function. In the United States of America, there is a regulatory document FIPS 202 which describes the requirements for the hashing algorithm in EDS systems, whereas a hashing function it is recommended to

use "SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions" [12].

If we compare the state standards for EDS algorithms in countries where there are legal requirements for algorithms and functions in EDS schemes, then, for example, in Ukraine, Russia and the United States of America, they are similar but differ only in some numerical parameters and details of key pair generation, calculation mechanism, and signature verification. Specifically, these three algorithms are variants of the El Gamal scheme [13].

Therefore, taking into account the similarity of the mathematical apparatus of cryptocurrencies when finding the basic in cryptographic algorithms EDS, it is possible to extrapolate the obtained values of estimates of the probability of breaking hash functions in EDS to the above schemes. Thus, in comparison with the existing methods considered in the [7–9, 12], the described approach allows us to support the possibility of using the existing mathematical apparatus for systems with an EDS.

2.2 Mathematical Apparatus for Estimating the Probability of Breakage

The birthday paradox is an imaginary paradoxical statement that the probability of coincidence of birthdays (dates) for at least two members of a group of 23 or more people exceeds 0.5. For 60 or more people, the probability of such a match exceeds 0.99, although the probability of 1.0, it theoretically reaches only when the group is at least 367 people. This statement may seem unobvious because the probability of coincidence of birthdays of two people on any day of the year ($1/365 = 0.0027$) multiplied by the number of people in the group of 23 gives only $23/365 = 0.063$. This reasoning is incorrect as the number of possible couples (253) significantly exceeds the number of people in the group. There is no logical contradiction in this, and the paradox is only the difference between intuitive perception and mathematical calculation. However, under the conditions of using this paradox in cryptanalysis, and especially if the attack is carried out on a set of known open messages, it allows to estimate the probability of breaking a hash function with a fairly high accuracy [14, 15]. This will allow a «birthday attack», in which random values are tried, until two are found that work.

«Birthday attack» is used in cryptanalysis as a method of cracking ciphers or finding collisions of hash functions. The essence of the method is to significantly reduce the number of transmitted hash functions arguments required to detect the collision, because if the hash function generates n—bit values, the number of random arguments of the hash function, which is likely to detect at least one hash function collision, is not 2^n, but only about $2^{n/2}$ [16].

It should be noted that these conditions of "birthday paradoxes" do not necessarily hold true in the real world. In particular, in the real world, people are not born with uniform randomness. There may be accurate statistics on the days people are born.

But this model is not an accurate reflection of the real world, its simplicity makes it much easier to analyze the problem. In particular, in cryptanalysis, this model has satisfactory accuracy. So, for example, this paradox can be applied to assess the cryptographic strength of various elements of the blockchain technology due to the fact that it also uses hash functions similar to those under consideration [17].

2.3 Methods and Algorithms for Hash Attacks

You can consider all possible underlying hash attacks related to value generation collisions:

(1) Frontal attack. Having a convolution of type $H(m_1)$ of the message m_1, the cryptanalyst must use the search method to find the message m_2 ($m_2 \neq m_1$), for which $H(m_1) = H(m_2)$. If this condition is met, the cryptanalyst may declare that the resulting convolution corresponds to the message m_2, but not to the message m_1. If the hash function outputs an n—bit string, then the complexity of this method of cryptanalysis is estimated as $O(2^n)$.
(2) Attack based on "birthday paradoxes." This paradox is used for many applied problems (including cryptanalysis problems). It is based on a well-known statistical problem—"birthday paradox."

Therefore, the general task of cryptanalysis can be reduced to the task of finding collisions, namely how many open messages the cryptanalyst needs to review to find messages with the same hashes [17].

The probability of encountering the same hashes for messages from two different sets containing, respectively, n_1 and n_2 plaintexts will be equal to:

$$P \approx 1 - e^{\frac{n_1 n_2}{l}} \tag{1}$$

If in expression (1) $n_1 = n_2 = 2^{l/2}$, then the probability of success of such an attack will be: $P \approx 1 - e^{-1} \approx 0.63$, and the complexity of the attack can be estimated as $2^{\frac{l}{2}+1}$. In order to determine the collision, it is necessary to generate two pseudo-random sets of messages so that there would be $2^{n/2}$ messages in each set, and then you need to find the corresponding hash values. Then, according to the "birthday paradox," the probability that among them, there will be a pair of messages with the same hash values will be more than 0.5. Such an attack requires a large amount of memory to store sets of plaintext and apply effective methods of sorting them [16].

The scenario of an attack on a hash function based on the "birthday paradox" where the attacker A wants to attack party B and sign a contract that is unfavorable for party B looks like this:

(1) Attacker A is preparing two versions of the contract—a "good" option M_1 and a "bad" option M_2. Then, he makes minor changes to each document, manipulating periods, commas, other punctuation marks, and spaces.

(2) Attacker A calculates the values of hash functions for all versions of the contract created by him, compares sets of hashes, and searches for identical pairs. If the length of the hash is 64 bits, then for this purpose 2^{32} pairs are enough. Attacker A selects a pair of messages for which the value of the hash function is similar.

(3) Attacker A selects the selected pair with similar hashes and passes the "bad" version of the contract M_2 for signature to party B. In this case, the attacker A can prove in court that party B signed an unfavorable contract knowingly.

Thus, for a given hash function H(m), the purpose of the attack is to find a collision of the second kind.

The reliability of EDS is determined by the resistance to cryptological attacks of its components, such as: (1) hash functions, (2) EDS algorithms.

The main task of EDS cryptanalysis is to assess the probability of breaking the hash function under the following attacks: (a) brute-force (rough selection); (b) selection of the hash value, grouped per affiliation.

2.4 Methods and Algorithms for Hash Attacks

Based on the cryptanalysis apparatus formulated above, we estimate the probability of breaking the hash function. To attack with this method (brute-force), we considered two methods:

(1) **The method of breakage resistance to collisions**. According to the known value of the hash function $H(M)$, the attacker creates another document $M^{'}$ such that $H\left(M^{'}\right) = H(M)$. An attacker tries to find two random messages M and $M^{'}$ such that $H\left(M^{'}\right) = H(M)$.

Suppose that a hash function is absolutely reliable and unidirectional, and there is no other method of breaking it than a complete search of its values. At the output of this function, we obtain a number whose bit size is m. Then, the number of output values of the function $H = 2^m$. Let us denote $P(n, k)$ the probability that for a specific X value, at least one Y_i value from the range of values $Y_1 \ldots Y_k$ equality $H(X) = H(Y)$ is satisfied. For one Y, the probability of $H(X) = H(Y)$ will be $1/n$. From here, it is easy to find that the probability of identity of the expression $H(X) \neq H(Y)$ will be $(1 - (1/n))$. If we generate k number of values, then the probability that none of them will be repeated is defined as the product of the probabilities, calculated as $\left(1 - (1/n)^k\right)$. Therefore, the probability of at least one match will be equal $P(n, k) = 1 - (1 - (1/n))$. For the case of absolute break, the probability of such a case is $P(n, k) = 1$.

Using the "birthday paradox" discussed above to estimate the probability of a collision in a hash function, we can estimate the probability of such a case as $P(n, k) = 0.5$, whence the value of the parameter $k = n/2 = 2^{(m-1)}$. Thus, it can be argued that the relative probability of a successful attack by the method

of breaking the resistance to collisions (finding such a hash generated for a given value of an open message) is achieved with $2^{(m-1)}$ random messages. Then, the probability of breaking the hash function is $1/2^{(m-1)}$ [14–16]. The obtained values of the hash function breakage probabilities for different n-bit message lengths (Table 1).

Function graph showing estimation of the probability of breaking the hash function by the method of burglary resistance to collisions we can see in Fig. 1.

(2) **The method of breaking the strict resistance to collisions**. Let $P(n, g)$ be the probability that in a set of g elements, each of which can acquire n number of values, there are at least two elements whose values are identical. Then, the value that must acquire g in order for the condition to be fulfilled:

Table 1 Estimation of the probability of breaking the hash function by the method of burglary resistance to collisions

Hash length, bits	Probability of breaking the hash function
64	1,0842021724855044340074528008699e-19
128	5,8774717541114375398436826861112e-39
192	3,1861838222649045540577607955354e-58
256	1,7272337110188889250772703725601e-77
384	5,0758836746312984465480491116611e-116
512	1,4916681462400413486581930630926e-154
1024	1,1125369292536006915451163586662e-308
2048	6,1886920947651565509603667399424e-617

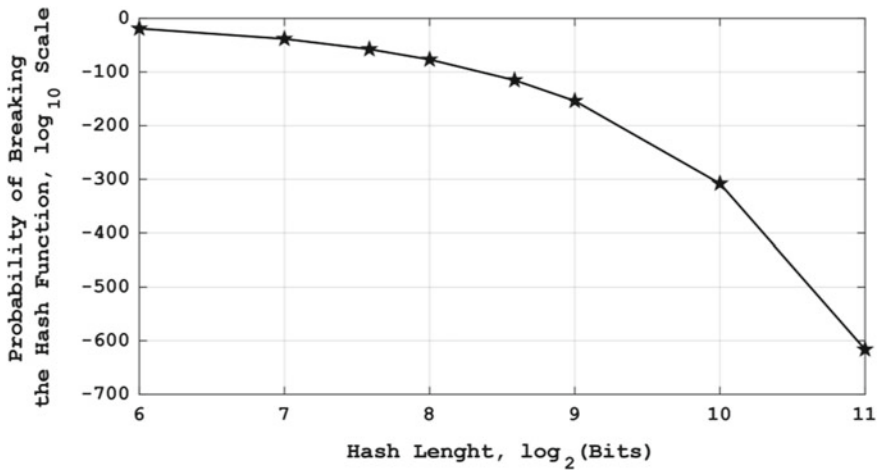


Fig. 1 Estimation of the probability of breaking the hash function by the method of burglary resistance to collisions

$$P(n, g) > 0.5 \tag{2}$$

In (2), condition for estimating the probability of meeting identical values is formulated. From expression (2), it can be understood that the number of ways to select elements such that there were no repetitions is $n!(n - g)!$ The total number of ways to select such elements is equal to n^g. The probability that the elements are not repeated is $n!/[(n - g)!n^g]$. Expression (2) can be written to determine the cases of probability of repetition:

$$P(n, g) = 1 - \left[n!/((n - r)!)n^g\right] \tag{3}$$

In (3), overall probability estimate for cases of repetition of several elements is formulated. If $P(n, g) = 0.5$, then $g = \sqrt{nln2} = ln2\sqrt{n} \approx n^{1/2}$. Let us assume that the length of the hash is m—bit, then the number of values that the hash acquires is equal to 2^m, then $g = \sqrt{2^m} = 2^{m/2}$. Therefore, to receive two messages with the same content (collision), it is necessary to calculate a hash from $2^{m/2}$ random open messages. The probability of breaking the hash function is estimated as $(1/2)^{m/2}$ [14–16]. Using expression (3), we can obtain the values of the probability of breaking the hash function at different values of the bit size of the obtained hash (Table 2).

From the obtained results of comparisons of hash function break probabilities, it is clear that under the same hash value length, the hash value break by collision resistance break method is much lower than with the use of strict collision resistance break method.

From the obtained results of comparisons of hash function break probabilities, it is clear that under (Fig. 2).

On the first (Fig. 1) and second (2) graphs, you can see serious differences in the probability of breaking hashes when using the first and second methods.

Suppose that for a given cryptographic hash function f, the aim of the attack is to find a collision of the second kind. To do this, it is need to calculate function f values for randomly selected blocks of input data until two blocks are found that have the same hash. Thus, if f has N different equally probable output values and

Table 2 Estimation of the probability of breaking the hash function for an attack by the method of strict resistance to collisions

Hash length, bits	Probability of breaking the hash function
64	0,00,000,000,023,283,064,365,386,962,890,625
128	5,421010862427522170037264004349 7e-20
192	1,262177448353618888658765704452 5e-29
256	2,938735877055718769921841343055 6e-39
384	1,593091911132452277028880397767 7e-58
512	8,636168555094444625386351862800 4e-78
1024	7,458340731200206743290965315462 9e-155
2048	5,562684646268003457725581793331e-309

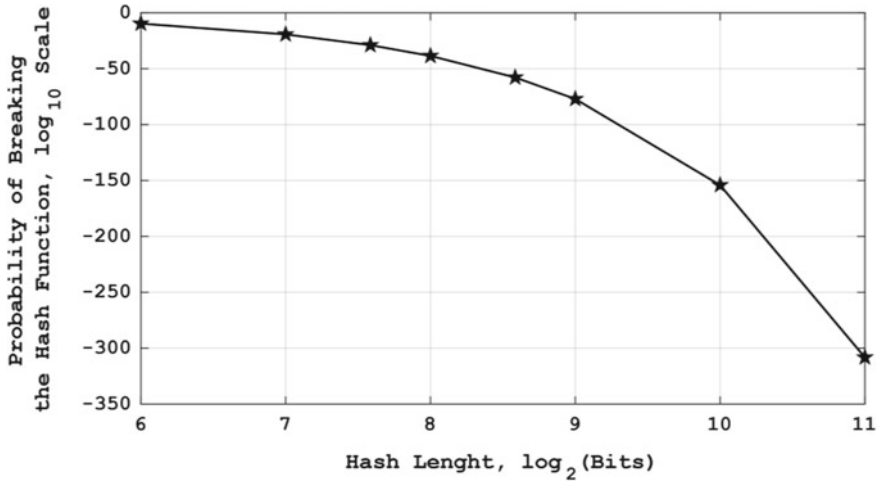


Fig. 2 Estimation of the probability of breaking the hash function for an attack by the method of strict resistance to collisions

N is large enough, then from the birthday paradox it follows that, on average, after enumerating $1{,}25\sqrt{N}$ different input values, the required collision will be found.

If the hash function generates an N—bit value, then the number of random input data for which the hash codes are likely to give a collision is not 2^N, but only about $2^{N/2}$. The light cells show the amount of input data at which a collision will occur with a given probability (analogy with a paradox, the number of different output chains is 365). This why there is no need to create data sets. Therefore, the goal of this research is fully achieved. Thus, without forming data samples for a given cryptographic hash function f, one can easily find a collision of the second kind. For this, it is sufficient to use only expression (3).

3 Conclusion

Estimation of probabilistic modeling demonstrates a systematic approach to cryptographic analysis of the method. The probabilistic approach makes it possible to assess the dependence of input and output data during encryption, as well as the selection of "bad" keys, which may not be reliable in practice. It can be concluded that to increase the resistance of the hash function in the EDS scheme to break at least to the level of 10^{-25}, hashing of at least 192-bit or even greater should be used.

When working with key hash functions, it is dangerous to add a key at the beginning or end of the message. This can be explained as follows:

- Let the key k be added at the beginning of the message, and the convolution function be constructed according to the Mercle-Dogmar scheme. Then, according

to the message M known to the cryptanalyst and its convolution function can be constructed.

- $H = H(k \| M)$, it is possible to determine the convolution value for all messages of the type $M \| M^{'}$, where after the message M any message $M^{'}$ is added. Indeed, due to the iterative nature of the hash function, there is no need to know the key to determine $H^{'} = H\left(k \| M \| M^{'}\right)$. It is enough to use the intermediate value of the convolution H when calculating;
- Let the key k added at the end of the message $H = H(M \| k)$. The collision value for the hash function (pair M_1, M_2) provided that $M_1 \neq M_2$, but $H(M_1) = H(M_2)$ can be calculated through the hashes $H(M_1, k) = H(M_2, k)$ for any k value. From this we can conclude that the complexity of changing the message M_1 is no longer estimated by the value of $O(2^n)$. Due to the use of the paradox of birthdays, the complexity will be significantly reduced and will be $O\left(2^{n/2}\right)$. In order to reduce the probability of successful cryptanalysis, it is recommended to write the key when calculating the convolution in the hash function several times. For example, as shown in an expression $H = (k \| y \| M \| k)$ that can be rewritten like $H = H(k \| y_1 \| H(k \| y_2 \| M))$ where y, y_1, y_2 is a complement of the key to a size multiple of the length of the hashing block.

Keyless functions are built on this principle. They are extremely reliable for attacks of collisions of the first and second kind, but their disadvantage is the large length of the resulting convolution, which significantly complicates their use.

For example, the standards of the Russian Federation and Ukraine have not only one, but several variants of hashes, because the cipher parameter is a set of replacement nodes. In this case, we get our own hash for each set. This nuance has the advantage of cryptanalysis complication, but creates compatibility problems, because each of these algorithms determines the step hashing function, which receives two blocks of data at the input:

- the current hash value from the previous step;
- another fragment of the input data array.

Thus, for a 128-bit hash function, the breakage probability differs for both methods by 9,223,372,036,854,775,808 times. You can also estimate the machine time required to break by the first and second methods. For example, the probable time to find the collision by the first method, provided that the hash has dimension m = 128 it is necessary to calculate $2^{128} = 3,4 \cdot 10^{38}$ variants. Then using the «birthday paradox», the number of these options will be $2^{m/2} = 1,8446744073709551616 \cdot 10^{19}$ by the second method and $2^{m-1} = 1,7 \cdot 10^{38}$ by the first method.

The results obtained in this work can find practical application in solving a number of problems of modern cryptography, for example, in evaluating the effectiveness of parallel implementation of methods for determining multicollisions of hash functions, as well as methods of discrete logarithm, similar to the Pollard ρ-method [18].

References

1. X. Wang, D. Feng, X. Lai, H. Yu, Collisions for hash functions MD4, MD5, HAVAL-128 and RIPEMD. CRYPTO National Center for Biotechnology Information (2006). https://eprint.iacr.org/2006/381.pdf
2. T. Okamoto, M. Tada, A. Miyaji, Efficient 'on the fly' signature schemes based on integer factoring. in *Proceedings of the 2nd International Conference on Cryptology in India*, INDOCRYPT'01, LNCS 2247 (2001) pp. 275–286
3. A. Fiat, A. Shamir, How to prove yourself: practical solutions to identification and signature problems. in *Advances in CryptologyCRYPTO'86*, LNCS 263 (1986) pp. 186–194
4. National Institute of Standards.: FIPS 180–1, Secure hash standard, NIST. US Department of Commerce, Washington D.C., April (1995)
5. P.R. Kasselman, W.T. Penzhorn, Cryptananlysis od reduced version of HAVAL. Electron. Lett. **36**(1) (2000)
6. A.V. Sahun, V.A. Lakhno, P.Y. Kravchuk, S.S. Kosenko, E.M. Kisiliuk, Elliptic curves in modern cryptographic systems. Int. J. Adv. Trends Comput. Sci. Eng. **9**(4). 5949–5955 (2020). http://www.warse.org/IJATCSE/static/pdf/file/ijatcse259942020.pdf
7. M. Bellare, T. Kohno, Hash function balance and its impact on birthday attacks. EUROCRYPT 2004, LNCS 3027 (2004) pp. 401–418
8. D. Coppersmith, Another birthday attack. in *Proceeding CRYPTO '85. Advances in Cryptology*, (1985) pp. 14–17
9. P. Shor, Algorithms for quantum computation: discrete logarithms and factoring. Foundations of Computer Science. in *IEEE 1994 Proceedings. 35th Annual Symposium* (1994) pp. 124–134. ISBN 0-8186-6580-7. https://doi.org/10.1109/SFCS.1994.365700
10. National Ukrainian standard of digital signature based on elliptic curves (DSTU 4145-2002), https://sourceforge.net/projects/dstu4145-2002/
11. GOST 28147-89, Sistemy obrabotki informaczii. Zashhita kriptograficheskaya. Algoritm kriptograficheskogo preobrazovaniya. https://web.archive.org/web/20120226123825/http:/gostshif.narod.ru/
12. National Institute of Standards, FIPS 180–4, Secure Hash Standard (SHS). https://csrc.nist.gov/publications/detail/fips/180/4/final
13. E. Taher, A public-key cryptosystem and a signature scheme based on discrete logarithms. IEEE Trans. Inform. Theory **31**(4), 469–472 (1985). https://doi.org/10.1109/TIT.1985.1057074
14. M.C. Borja, J. Haigh, The birthday problem. Significance. Royal Stat. Soc. **4**(3), 124–127 (2007). https://doi.org/10.1111/j.1740-9713.2007.00246.x
15. F.H. Mathis, A generalized birthday problem. SIAM Rev. **33**(2), 265–270 (1990). https://doi.org/10.1137/1033051.ISSN0036-1445.JSTOR2031144.OCLC37699182
16. A.D Kozhukhivs'kij, A.V. Sahun, O.A. Kozhukhivs'ka, O.Y. Snisarenko, Y.V. Stepanecz, Analiz kriptostijkosti elementiv elektronnogo czifrovogo pidpisu. Visnik ChDTU № 3, (2013) pp. 80–85. ISSN 2306–4455
17. V.V. Khilenko, Application of blockchain technologies for improving the quality of ACS of complex dynamic systems. Cybern Syst Anal **56**, 181–186 (2020). https://doi.org/10.1007/s10559-020-00233-w
18. J.M. Pollard, B.J. Green, Theorems on factorization and primality testing. in *Mathematical Proceedings of the Cambridge Philosophical Society*. vol. 76(3), (Cambridge University Press, 1974) pp. 521-528. ISSN 0305-0041; 1469–8064. https://doi.org/10.1017/S0305004100049252

A Novel Privacy-Preserving and Denser Traffic Management System in 6G-VANET Routing Against Black Hole Attack

Gaurav Soni and Kamlesh Chandravanshi

Abstract Vehicular ad hoc networks (VANETs) establish dynamic connections between cars, vehicles, and RSUs at a 6G data rate of 1KGbps. Communication between or among cars is feasible with the assistance of an RSU or an intermediary vehicle, and also the vehicles convey traffic status to the leading and neighboring vehicles. If any car engages in inappropriate activity with other vehicles, the data privacy is jeopardized. This paper offers a novel privacy-preserving under denser traffic management (PPDM) routing strategy for the 6G-VANET to protect it against malicious black hole attacks in VANET. All key information from traffic status packets supplied by leading cars to the following vehicles is discarded by black hole vehicles. A security system detects and prevents packet drops on a connection through a node. The performance of the present SAODV security system is compared to that of the innovative PPDM. After preventing malicious vehicles operating in the network, the PPDM secures the VANET and improves performance. The performance of the proposed PPDM scheme is compared with the existing SAODV. The PPDM has enhanced the performance and reduced the data dropping when compared to SAODV. The network performance in the presence of attack and secure PPDM performance is measured through the performance metrics like throughput, PDR, and end-to-end delay. PPDM is proven that it improves the data receiving and minimizes data dropping in the network.

Keywords VANET · 6G · PPDM · Black hole attacker · Routing · Security

1 Introduction

In recent years, the research interest in intelligent transportation systems (ITSs) has increased in industry and academia [1]. ITSs' primary goal, in addition to providing entertainment services in automobiles, is to improve road safety and driving conditions [2]. Vehicular ad hoc network (VANET) is built using two forms of

G. Soni · K. Chandravanshi (✉)
Lakshmi Narain College of Technology(Excellence), Bhopal, India
e-mail: kamlesh.vjti@gmail.cm

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_49

communication, namely vehicle-to-vehicle (V to V) communication and vehicles-to-infrastructure (V to I or RSU) communication to transmit the essential driving information [3]. The larger bandwidth and high data rate show better results in traffic control and monitoring since the number of vehicles on roads continually increasing**.** Vehicles interact with adjacent vehicles to replace the data in V to V communication, whereas the vehicles communicate directly with roadside units (RSUs) in V2I communication [4]. For V to V and V to I communications in VANETs, a dedicated short-range communication (DSRC) radio [5] and a handful of IEEE standards can be utilized. VANETs' unique properties, such as their great mobility and instability, have rendered them to be more vulnerable to many types of external and internal assaults [6]. These assaults have given rise security, privacy, and trust challenge in the design of safe VANETs. The major goal of VANETs is to enhance the road safety for drivers. VANETs will transmit different types of information, such as traffic signal violation warnings, precrash sensing, traffic jam warnings, curve speed warnings, and many more. Since mobility is a primary element of VANETs, any relevant simulation model should consider the consequences of mobility with the varied mobility patterns, and a suitable mobility model that depicts various facets of the experienced movement [7]. The VANET uses IEEE 802.11 standard Industrial, Scientific, and Medical band (ISM) band, and it is useful for many real-time implementation with the ad hoc nature. One of the vital features of such networks is flexibility, and it can be implemented without pre-existing infrastructures. Therefore, these networks are appropriate for an emergency application. The example of V to V and V to RSU is mentioned in Fig. 1. Here, the vehicles are also exchanging the traffic information with other vehicles and RSU. The prominent role of RSU is to control the traffic, check the traffic information of vehicles, and monitor the vehicle's activities.

When dealing with mobility models, it is also important to understand mobility, which is divided into macro- and micromobility. Macromobility defines and addresses motion limitations such as road topology, street features, traffic signals and signs, node flow, density, and distribution, i.e., any street element that may limit or impact mobility. On the other hand, micromobility refers to the movement of individual vehicles and their interactions with other vehicles, such as overtaking and acceleration/deceleration [8]. The 6G network will certainly improve the network service when compared to previous generations. The 6G vision is expected to be developed between 2022 and 2023 in order to define the 6G requirement,

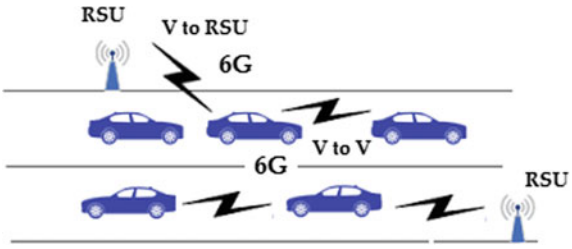


Fig. 1 Example of V to V and V to RSU communication

and it assesses 6G development, technologies, standards, and so on. The International Telecommunication Union (ITU) and the Third Generation Partnership Project (3GPP) are anticipated to develop the standards for 6G by 2026–2027 [9 ,10].

The presence of a black hole attacker causes harmful activity in the network [11]. Without security guarantees, some misbehaving or malevolent vehicles render the system to be susceptible for offering low-quality services or even put user vehicles in unsafe circumstances in 6G-VANET. As a result, identifying the misbehaving or malicious vehicles has become a critical task in VANET security. The security method of finding an attacker's vehicle is based on the information identified about the traffic network. The scheme detects the attacker's presence and estimates the total number of packets dropped by the black hole attacker in the network.

2 Network Model

In VANET, network model for vehicles can be separated into three groups [12]. These groups include servers for application and authorization, facilities on the road side, and nodes/vehicles in 6G-VANET.

2.1 Application and Authorization Servers

These are powerful workstations, which are, respectively, responsible for managing and providing service data. The authority knows all the keys and is accountable for maintenance planning. For cars, device servers provide the operation details. The government or foreign operators will fund them. We assume that there are powerful processing capabilities for authorization and application servers. So, here we ignored the computation time.

2.2 Road Side Infrastructure

Road infrastructure consists of power supplies located near roads, and it is also responsible for the collection and dissemination of data. Through wired networks, RSUs are connected to power and it will communicate via radio signals with vehicle but for vehicle-to-vehicle wireless communication radio or microwave signals are used.

2.3 Nodes/Vehicles

Nodes or vehicles move in the road and communication with the RSU or also their information exchange information is received by RSU in network. Every vehicle is presumed to be fitted with a differential GPS receiver with an order accuracy and On-Board Unit (OBU) responsible for all communication and computing tasks [13].

3 6G Overview

The sixth-generation (6G) communication technology provides faster data rate for communication among or between the vehicles. 6G operates at a higher frequency to obtain a broader bandwidth, i.e., in Terahertz. The huge amount of bandwidth is available for users for sharing traffic information. The responses of RSU to V and V to V also improved. Compared to 5G, 6G can boost data rates by up to a hundred times, sustaining Tb/s highest data rate and 1000 Gb/s data rate experienced by user [9, 14]. Furthermore, 6G may employ flexible frequency sharing technologies to improve frequency reuse efficiency. 6G is a customized intelligent network because broad area for communication here is available to vehicles. 6G should be a ubiquitous and integrated network with more extensive and deeper coverage, covering terrestrial communication, satellite communication, short-range device-to-device communication, and so on. 6G is intellectual mobility management technology that can operate in various situations, including airspace, land, and water, resulting in a worldwide ubiquitous mobile broadband communication system. When combined with artificial intelligence technologies, 6G will enable virtualized personal mobile communication, with the network transitioning from a classic function centralized type for user centralized, data centralized, and completely content centralized. The main drawback in centralization is dependency on centralized unit for communication. It is also supports decentralized networks like WSN, MANET, and VANET. The 6G commutation is also very helpful for sustaining the current location of vehicles. It is same as location information of nodes in MANET [15, 16]. The 6G network will feature an endogenous security scheme or function security integrated design.

Furthermore, by incorporating trust and safety mechanisms, 6G can self-awareness, real-time dynamic analysis, and adaptive risk and confidence evaluation, all of which will aid in the realization of cyberspace security. Furthermore, 6G will combine processing, sensing with communications and navigation. For example, 6G will incorporate with satellite communication systems, satellite navigation, and positioning systems, as well as radar sensing systems. Finally, 6G might have a more efficient structure that includes software-defined core networks and radio access networks. 6G will be capable of rapid and self-intelligent expansion and quick dynamic exploitation of network functionalities. 6G can generate massive amounts of data via the Internet of Everything; 6G can also combine with novel technologies

such as artificial intelligence, block chain, cloud computing, edge computing, and so on [9].

4 Routing in VANET

Routing protocols may be divided into several types based on their characteristics [14, 15]. Most popular method for differentiating VANET routing protocols depends on obtained and preserved routing information. The routing protocols are classified as follows:- (Table 1).

5 Literature Survey

This section discussed the previous work on security with its drawbacks. The different author's contribution provides different security approaches for securing vehicular communication.

Li et al. [11] proposed an approach that maintains the reliable and safe communication between vehicles and improves the functioning of routing protocol in presence of threats in vehicular ad hoc networks. When a vehicle receives an AODV control message requesting that it establishes or modifies a route for a specific destination, it searches its routing database for an entry that leads to the location. AODV control messages are sent to vehicles to establish or change routes for specific destinations. The system produces a route entry with the sequence number specified in the control packet if there is no route entry; otherwise, the sequence number is flagged as invalid. The most significant flaw in the study is that the author restricts his attention to comparing the performance of various routing protocols. As a result, there is nothing unique about this work. The vehicle's speed is likewise indicated in kilometers per hour (km/h), but in milliseconds per second (m/s) is mentioned in Fig. 2 [11].

Soni et al. [17] proposed the PSO method for V to RSU communication to ensure security and communicate information about the attacker's vehicle to all RSU at rest and the vehicles near the attacker. The suggested preventive technique is intended to thwart the harmful acts of the attacker while still ensuring a secure connection in the VANET environment. However, the most significant flaw in this work is that it fails to specify the range of pheromones and the dropping behavior of the attacker.

Dhaya et al. [18] proposed an ACO scheme to find the shortest path among several possible paths. Accordingly, the probability of street consistency can be calculated using the expression given in the previous sentence. The buses are employed as a way of transporting the packages and directing them to their intended destinations. In each path connecting the two ends, streets are used as connecting points, and the relay bus that corresponds to the specified street is picked along with the path. Its primary weakness is that the ACO method by default supports the multipath approach, which is not ideal. Even in the presence of traffic congestion, the multipath routing

Table 1 Routing protocols of VANET

Proactive routing	Reactive routing	Hybrid routing
Proactive routing is also known as table-driven routing protocols. They monitor the networks topology at all the times, and the routes will be continuously evaluated for all the destinations	Reactive routing protocols are often known as on demand routing protocols, and it only performs route determination on demand. In data transfer, reactive protocols only select a route at the start of a connection	The benefits of both proactive and reactive routing systems are combined in hybrid routing strategies. Hybrid protocols make use of hierarchical network designs
Routes are stored by performing periodically interchanging routing tables in network, like the wired network	Once the route is created, the details are kept in the database of routing until the destination is unable to reach or the route expires	Table-driven routing guarantees great quality in static topologies but cannot be used to mobile networks
An advantage of this protocol is that routes must be decided and maintained in the buffer so whenever routing takes place these routes are used immediately so that the overall delay in the system must be reduced. To get the information of route and making a session is not very time consuming	The advantage of routing information is not updated on a regular basis, routing overhead is decreased dramatically during topology changes	The advantage of this protocol is combining the best features of both, a few hybrid routing protocols have been devised, in which routing is launched with certain proactive routes and subsequently satisfies demand from other nodes via reactive floods
This protocol has a disadvantage that whenever the topology changes, it reacts to those changes, though traffic is not affected, and results in the unwanted use of the bandwidth even there is not transfer of data	The delay experienced during route finding is one of these systems' drawbacks. These protocols, however, must be particularly efficient for highly mobile networks	In a dynamic network, typical routing protocols of this category function depends on radius of zone. On-demand routing offers less routing costs than traditional routing, however it suffers from routing latency
The protocols of this type are Optimized link State Routing (PLSR) protocol and Destination Sequenced Distance Vector Routing (DSDV) protocol	Ad hoc On-demand Distance Vector routing protocol (AODV) and Dynamic Source Routing (DSR) are reactive routing protocols	Zone Routing Protocol (ZRP) is the example of zone routing

condition will not function well. ACO approach is based on pheromone values but not mentioned.

Hu et al. [19] proposed a REPLACE trust-based platoon service recommendation approach, to rank platoon head vehicles by developing a trust and reputation system. They also recommended that based on feedback from user vehicles, the server computes platoon head vehicle reputation rankings for the platoon head vehicle. This allows the server to clearly distinguish between well-behaved platoon head vehicles

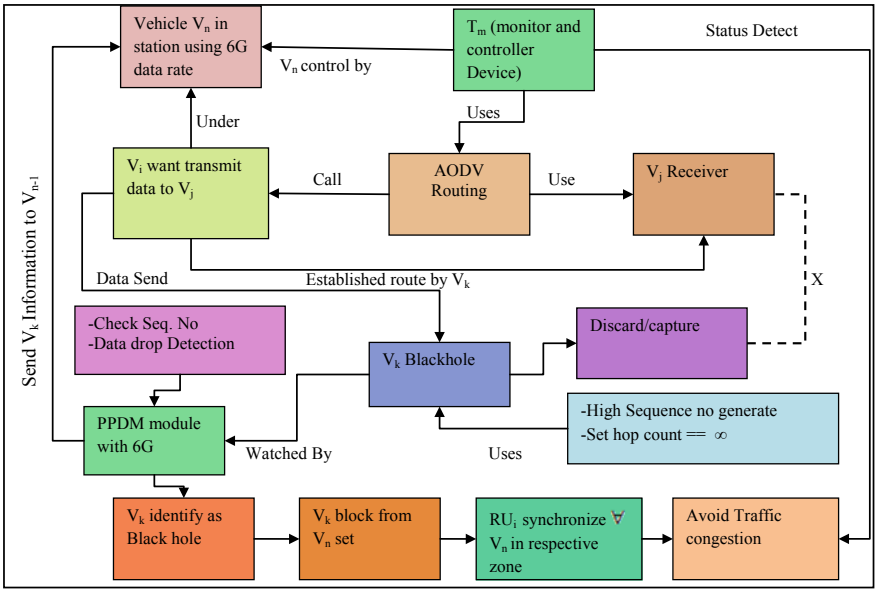


Fig. 2 Security architecture of PPDM

and poorly behaved platoon head vehicles based on their reputation ratings. The malicious functioning of vehicles is based on bad and good behavior. The network's overhead increases because the server-side first verifies the traffic information and then declares it is a normal or malicious vehicle. It increases overhead in the network.

Qureshi et al. [20] proposed an Intelligent Secure Routing Scheme (ISRS) to assure data security during data transfer while transferring large amounts of data. This technology protects traffic information data in the network during data transfer from potentially damaging attacks and guarantees it remains secure. Forging, modification, and replay attacks are the three most serious security concerns that the proposed solution in VANET is designed to address. The main drawback of the paper is that the clustering mechanism is used but it is better for stationary or slow speed vehicles scenarios. That is why the delay is showing in *seconds* but it is better in *milliseconds*.

Memon [21] describes developing an authentication key setup approach based on IPv6 road networks. Using this technique, each neighboring car or RSU provides an unique address to a mobile vehicle, which prevents the mobile vehicle from receiving duplicate addresses. It includes a cryptographic authentication technique based on zero-knowledge proof, which each node may use to convince another node of the presence of a given secret without revealing it. The drawback of research is that the speed of vehicles is considered in meters/seconds but practically, it is measured in km/hour. Duplicate address identification procedure is complex. Not mentioned the specific use of IPv6 in research.

6 Problem Identification

Malicious vehicles might act both positively and negatively. The attackers have the ability to produce false information in the network and delete all traffic packets. Since packets are heavily lost on the network, the attacker cars increase network overhead. Because of the packet dropping attacker, retransmission of packets increases network latency even further. It is impossible to secure vehicle connection and deliver traffic status appropriately. In recent trends, it is observed that continuously the number of vehicles on the road increases. So the main problem is how to manage the traffic efficiently. The 6G communication technology provides a faster data rate for communication. The vehicles are continuously moving on roads but have no information of traffic on routes. The 6G communication support provides larger bandwidth and high data rate, i.e., the primary requirement of VANET. Although no significant research is being conducted in this area, it has been discovered that the offered solutions are not comprehensive in terms of effective and efficient routing security. All solutions have limits. They may have a substantial computational or communication overhead.

7 Privacy-Preserving Under Denser Traffic Management (PPDM)

VANET is highly dynamic and may expose internal and external attackers, posing significant technological problems in dependability and safe routing. Compared to the way roads are built, the attacker driving pattern is a less sustainable and more expensive approach to alleviate network load and increases the risk of routing misbehavior. The novel PPDM algorithm offers a security method against black hole attacks using an RSU unit to gather and evaluate audit data for the whole network. So, based on the above description, we infer that VANET is dispersed in nature and that we cannot trust any of the vehicles, since we cannot manage the network's topology as it changes. The PPDM secures routing from malicious attacks in VANET. The proposed PPDM allows for the safe transmission of data packets in a multihop way for identifying and blocking malicious vehicles.

7.1 Proposed Algorithm

Input:
V_n: nth vehicles in network.
T_m: m traffic monitoring system.
RSU_i: no. of ith road side unit.
B_{la}: black hole attacker.
M_r: monitoring and preventer node $\forall$ RSU_i.

S_i: message sender $\forall V_n$.
R_k: message receiver $\forall V_n$.
M_l: intermediate vehicle $\forall V_n$.
R_{req}: route request.
R_{rep}: route reply.
h_{seq}: higher sequence number.
Y_p: capture packet.
E_p: drop packet.
a_{bh}: abnormal behavior (h_{seq}, Y_p, E_p).
P_{roto}: AODV for routing.
ψ: RU_i control range.
dr: 6G data_rate(Tbps).
Output: Throughput, drop analysis, PDR, end-to-end delay.

Method 1: *Black hole Detection and Network Monitoring*

T_m watch the activity of assigned route
$T_m \leftarrow RSU_m$
If V_k receive R_{req} & not forward to V_{k+1} **Then** // 1Tbps data rate (6G)
Watch V_k activity by T_m node
If V_k generate the false H_{seq} of R_{req} **Then**
T_m trace V_k $a_{bh}(h_{seq})$
If match $a_{bh}(h_{seq})$ **Then**
$V_k \leftarrow B_{la}$ set by T_m
T_m Send feedback to M_r module
Else
In real time watch V_k by T_m
End If
End if
Else If V_k receives message of S_i & V_k != R & not forward **Then**
T_m trace V_k $a_{bh}(Y_p, E_p)$
V_k create loop
If match $a_{bh}(Y_p, E_p)$ Then
$V_k \leftarrow B_{la}$ set by T_m
T_m Send feedback to M_r module
Else
In real time watch V_k by T_m
End If
End If

Method 2: *Black hole Prevention*
M_r take response of route from T_m
M_r re-analysis the activity of detected B_{la}
$M_r \leftarrow RU_m$
If V_k as B_{la} **Then**
M_r check S_i to V_k Hop count , Y_p, E_p, H_{seq} field
If V_k as $a_{bh}(h_{seq}, Y_p, E_p)$ & hop count == ∞ **Then**
M_r take confirmation V_k as B_{la}
Block V_k and Send negative response of V_k to all V_{n-1}
S_i call forward AODV(S_i, R_k, R_{req})
RSU_i provide safe route to S_i $\nexists$ V_k
RSU_i communicate with V_n or RU_j
RSU_i synchronize $\forall$ V_n in respective zone
Avoid congestion through T_m module
End If
End If

The black hole attacker has degraded the network's routing performance, and the attacker infection is measured in forty node density scenarios. After detecting the malicious vehicles, the proposed security mechanism is also prevented from an attacker by denying the possibility of routing through hostile vehicles. The PPDM approach employs V to V and V to RSU communication to preserve security and forwards attacker vehicle information to all RSU and their surrounding vehicles.

7.2 Propose PPDM Architecture

The architecture of privacy-preserving scheme is mentioned in Fig. 2. In this architecture, all vehicle-controlled vehicles are controlled by the Tm base station. If any vehicle wants to communicate with another vehicle, then Vi calls AODV routing and broadcasts routing packets to search the route. At the same time, route is found through the V_k node in between source to destination and using 6G data rate. The step-by-step description of secure routing in 6G mentioned in Fig. 2.

Data privacy is not guaranteed due to the intermediary vehicle that facilitates communication from source to destination. Various man-in-the-middle attacks exist in network communication. The cross (X) in between Vj receiver and discard/capture block means attacker functioning. Black hole attackers generate a high sequence number, and false route information is given and captured and drop the data packet. To handle this situation, a PPDM module is proposed that periodically watches every active node. If any node drops the data by false routing behavior, block the particular vehicle from the communication and provide a new path from source to destination and preserve privacy during transmission. Proposed PPDM mechanism is considered as a more secured routing mechanism when compared to existing mechanism.

8 Result Description

A number of nodes 10, 20, 30, 40, and 50 are taken for simulation having 550 m radio range. The grid layout is about 800 m*800 m and propagation is two-ray ground. The simulator version is employed for simulation is Network Simulator version 2.31 (NS-2.31) [22]. The performance of PPDM security scheme routing is measured with previous SAODV and black hole AODV (BAODV). The number of nodes scenarios is same in all modules.

8.1 Throughput Analysis

When the throughput performance is high, the data reception in the network will be improved. In this graph (Fig. 3), the throughput performance measurement of the existing scheme, black hole attack, and PPDM is measured in high 6G data rate. In SAODV existing approach, the throughput performance reaches about a maximum of 20,000 Kbps in the network. Still, in the case of black hole assault, the throughput performance is negligible or about 500 bits/sec in all different vehicle density but after applying the proposed PPDM scheme, the throughput is enhanced up to 27000Kbps. Thus, the proposed PPDM scheme improves the network performance and provides the attacker-free background of communication between sender and receiver through intermediate vehicles.

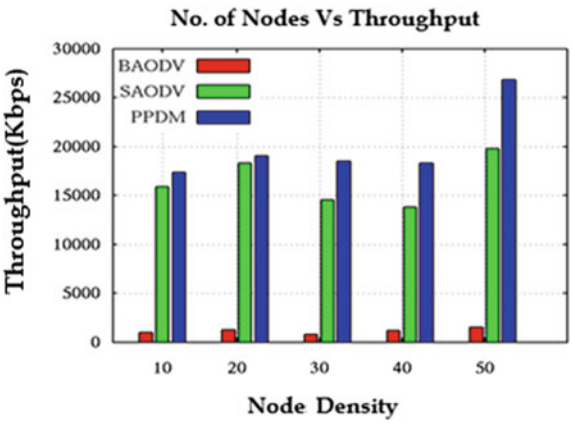


Fig. 3 Throughput performance analysis

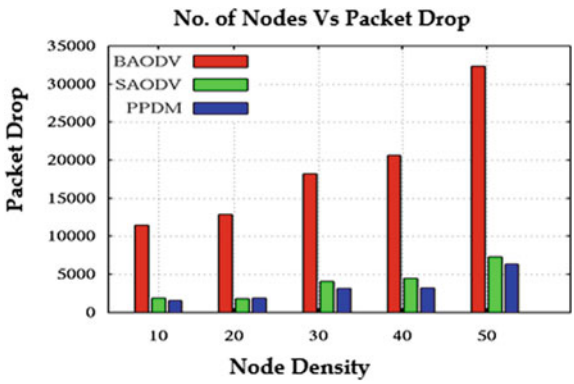


Fig. 4 Packets dropping analysis

8.2 Packet Drop Analysis

The attacker's presence in the network will degrade the routing performance. The vehicles in the attacker's range can send the fake message to the sender or send the wrong information of traffic to the sender vehicle. The number of vehicles directly connected to the assailant and attacker receives the packets from the sender and drops the whole packets mentioned in Fig. 4. The data dropping also increases due to delay in response of leading vehicles. The vehicles OBU used 6G standard for communication. The traffic status sharing using 6G improves the traffic controlling and examines traffic information. The performance data drop will decrease due to malicious vehicles mentioned in all different vehicle density scenarios. The number of attacker vehicles in the network is in different quantities in other vehicles density scenarios. The data dropping in the network is more than 30,000 in the presence of an attacker. That is, the attacker consumes the data packets in the network as well as the remainder of the data received at the destination, but the data received is much less.

8.3 PDR Analysis

The packets percentage performance is measured through of delivery ratio (PDR). This graph (Fig. 5) evaluated the PDR analysis in black hole attack, previous security scheme, and privacy-preserving under denser traffic management (PPDM). The standard routing performance is only evaluated to stint the network performance after applying the proposed PPDM scheme. The PDR of black hole attack in the network is only 20%, which means an 80% data drop in 20 vehicles density and the rest of scenarios PDR performance showing in between 5 and 10%. The PPDM scheme improves the network performance and provides secure routing. The network's performance almost provides 80% PDR minimum in 50 vehicles, 91% maximum in 10 vehicles, and 20 vehicle scenarios. After applying a security scheme against

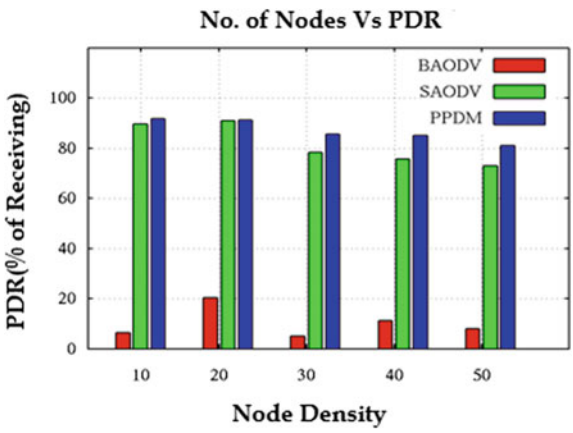


Fig. 5 PDR performance analysis

attack, PDR performance is improved and better than SAODV. The proposed security scheme has enhanced the performance in the presence of malicious vehicles. The performance of the proposed method reduces the overhead that improves receiving.

8.4 End-to-End Delay Analyses

The end-to-end delay analysis is measured in the presence of the attacker, in the existing security scheme and in PPDM proposed scheme mentioned in Fig. 6. The performance is measured in a scenario of five different vehicles density, and 6G high data rate quick response improves communication capability. The main reason

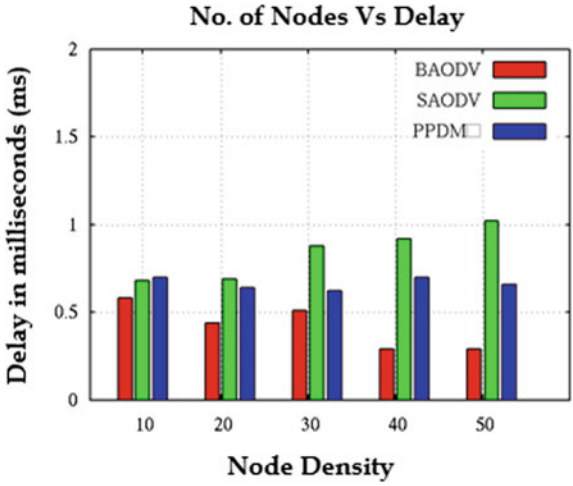


Fig. 6 Delay performance analysis

for the delay is re-establishing the connection frequently due to link breakage or successful not receiving the data at the destination. The end-to-end delay in attacker presence is high and the delay is minimized by applying the existing security scheme. The proposed PPDM scheme minimizes the delay by providing strong connection establishment in between source to destination. The delay in the network degrades the routing performance due to the presence of a black hole attacker.

9 Conclusion and Future Scope

The privacy-preserving under denser traffic management for the 6G-VANET (PPDM) security algorithm provides security from black hole attack/s, which is to be a part of a fake path established for loss data packets. The 6G existence in the network also reduces the possibility of delay in communication because a sufficient bandwidth is available for proper traffic status delivery. The performance metrics show the difference in the performance of attacker, previous SAODV, and novel PPDM. It is clearly shown that proposed scheme is providing secure communication and prevent network from attacker. The PDR is about 92% evaluated in novel PPDM scheme, and it is about to 5% more than the existing SAODV. The presence of an attacker is atrocious because packet dropping is really more than 30,000 packets (50 vehicles scenario). The PPDM reduces the packet dropping and enhances receiving of data packets, i.e., also enhances throughput in 6G communication. The attacker's presence is confirmed by detecting the loss of data and fake route information in a dynamic network. The PPDM minimizes overhead and delay, i.e., the leading cause degrades the routing performance because more packets are drooping that enhancing the delay.

In future, the same concept also simulates with MP-DSR multipath routing protocols and also using a security concept with a location tracker system that traces attackers easily in a dynamic network. The location tracker system maintains the separate table for attacker and normal nodes with its mobility speed.

References

1. M. Alam, J. Ferreira, J. Fonseca, Introduction to intelligent transportation systems. in *Intelligent Transportation Systems.* (Springer, Cham, Switzerland, 2016) pp. 1–17
2. M.N. Mejri, J. Ben-Othman, M. Hamdi, Survey on VANET security challenges and possible cryptographic solutions. Vehicul. Commun. **1**(2), 53–66 (2014)
3. A. Dua, N. Kumar, S. Bawa, A systematic review on routing protocols for vehicular ad hoc networks. Vehicul. Commun. **1**(1), 33–52 (2014)
4. M. Azees, P. Vijayakumar, L.J. Deborah, Comprehensive survey on security services in vehicular ad-hoc networks. IET Intell. Transport Syst. **10**(6), 379–388 (2016)
5. D. Jiang, L. Delgrossi, IEEE 802.11p: towards an international standard for wireless access in vehicular environments. in *Proceeding of IEEE Vehicular Technology Conference* (VTC Spring), May (2008), pp. 2036–2040

6. R.G. Engoulou, M. Bellaïche, S. Pierre, A. Quintero, VANET security surveys. Comput. Commun. **44**, 1–13 (2014)
7. H. Hartenstein, K.P. Lanerteaux, A Tutorial on vehicular ad hoc networks. IEEE Commun. Magazine (2008) pp. 164–171
8. J. Haerri, F. Filali, C. Bonnet, M. Fiore, Vanet MobiSim: generating realistic mobility patterns for VANETs. (California, 2006)
9. M. Wang, T. Zhu, T. Zhang, J. Zhang , S. Yu, W. Zhou, Security and privacy in 6G networks: new areas and new challenges. Digital Commun. Netw. 281–291 (2020)
10. K.B. Letaief, W. Chen, Y. Shi, J. Zhang, Y.-J.A. Zhang, The roadmap to 6G: AI empowered wireless networks. IEEE Commun. Magaz. **57**(8), 84–90 (2019)
11. A.P. Jadhao, D.N. Chaudhari, Security aware routing scheme in vehicular adhoc network. in *IEEE Proceedings of the Second International Conference on Inventive Systems and Control* (ICISC 2018) (2018)
12. Y. Hao, Y. Cheng, C. Zhou, W. Song, A distributed key management framework with cooperative message authentication in VANETs. IEEE J. Selected Areas Commun. **29**(3), 616–629 (2011)
13. G. Li, L. Boukhatem, J. Wu, Adaptive quality-of-service-based routing for vehicular ad hoc networks with ant colony optimization. IEEE Trans. Vehicul. Technol. **66**(4), 3249–3264 (2017)
14. Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G.K. Karagiannidis, P. Fan, 6G wireless networks vision, requirements, architecture, and key technologies. IEEE Vehicul. Technol. Magazine (2019)
15. G. Soni, K. Chandravanshi, A multipath location based hybrid DMR protocol in MANET. in *IEEE 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things* (ICETCE) (2020)
16. G. Soni, M.K. Jhariys, Quadrant base location tracking technique in MANET. in *IEEE 2nd International Conference on Data, Engineering and Applications* (IDEA) (2020)
17. G. Soni, K. Chandravanshi, M.K. Jhariya, A. Rajput, An IPS approach to secure V-RSU communication from blackhole and wormhole attacks in VANET. in *First International Conference on Communication, Cloud, and Big Data* (CCB) (2020)
18. R. Dhaya, R. Kanthavel, Bus-based VANET using ACO multipath routing algorithm. J. Trends in Comput. Sci. Smart Technol. (TCSST) **03**(01), 40–48 (2021)
19. H. Hu, R. Lu, Z. Zhang, J.Shao, REPLACE: a reliable trust-based platoon service recommendation scheme in VANET. IEEE Trans. Vehicular Technol. **66**(2) (2017)
20. K.N. Qureshi, F. Bashir, A.H. Abdullah, Provision of security in vehicular Ad hoc networks through an intelligent secure routing scheme. *in IEEE International Conference on Frontiers of Information Technology* (2017)
21. I. Memon, A secure and efficient communication scheme with authenticated key establishment protocol for road networks. Wireless Pers. Commun. **85**, 1167–1191 (2015)
22. Network NS-2 (The Network Simulator) (2021). https://www.isi.edu/nsnam/ns/. Access from March

Smart Customized Charging of Portable Devices Through an Authorized App

S. Kavitha, S. Hrushikesava Raju, Venkata Ramana Karumanchi, D. Srinivasa Rao, and T. S. Rajeswari

Abstract The proposed research work aims to develop not only portable devices but also huge items that may require charging at times when there is an immediate shortage of power supply. It is anticipated that, every portable device will be IP-based via the DHCP protocol so that it may be accessed via the app, in which users must previously register their usable gadgets. When the user wants to obtain power charging from a remote partner, they may do it by entering their phone number into a phonepe or any other similar app. The user who intends to transfer power in an eco-friendly manner must make a request to the person on their contact list, wherein the state of charging in the percentage form is immediately updated. When the recipient accepted the request with mutual permission, he or she was given the option of selecting the percentage of charging that they were willing to provide to the sender using QR-codes that held the amount of charging. The sender will be notified of the static QR-code via the app. When the sender scans the receiver's QR-code, their charge status is changed to the new level. This procedure comprises technical upgrades as well as enabling plug-ins to be added to the device in order to create an internet-based gadget, as well as supporting long-term batteries with quick charging and fast discharging capabilities.

Keywords App · Fasten charging · Fasten discharging · Customized options · QR-codes and IoT

S. Kavitha (✉) · S. H. Raju
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur 522002, India

V. R. Karumanchi
Department of Electronics and Communication Engineering, MLR Institute of Technology, Dundigal Police Station Road, Hyderabad, India

D. S. Rao
Department of Master of Computer Applications, Medi-VCaps University, A. B.Road, Pigdamber, Indore, Mandhya Pradesh 453331, India

T. S. Rajeswari
Department of English, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur 522002, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_50

1 Introduction

There are times when the need for power in remote locations for small-level management is necessary. Such circumstances necessitate sophisticated power banks, which would aid in the unexpected arrangement of interesting occurrences.

The few modules discussed here are fast charging and fast discharging batteries, designed to create apps such as phone pe app, checking the receiver gadget charging status, and submitting a request, and transfer the precise amount of power charging to the sender as mutual permission. The success is then informed, and the sender device is updated when the receiver QR- code is scanned by auto-mirror scanner, which is a built-in service offered by the developed software, as well as the receiver gadget's power charging. The significance of this app is that many people are busy in their life, but when they are distant from amenities, giving facilities to clients at their door step. With little research, hardware resources may succeed in bringing fast charging and fast discharging batteries, as well as making the devices IP based and detectable through their IP address by the remote user using efficient protocols and practices, designing the app with novelty, putting request, and receiving a QR-code from the receiver with mutual consent. As a last step, it updates both the sender and recipient gadgets, which will be accessed via the intended app. In the future, the charging pe app, like the phone pe revolution, will be a demanding technology due to technological advancements and available hardware resources. If this initiative is successful, it has the potential to earn significant revenue for the application partners. Yearly, the turnover may cross the revenues of the previous year's output. It also may extend to the carrying charging devices like carrying gas cylinders. There are two-time bounds in this project. The first is on small portable devices, and the second is on building up carrying power tanks with limited capacity like gas cylinders. The expectation of income for the first year is 100 Billions, and it may cross the figures for every consecutive year. The development of plug-ins and the app is a one-time expense. When the software is launched at the market, and plug-ins are made available everywhere, the income can be increased in any way. Only plug-in manufacturing will be required every quarter and yearly, based on demand and at a reasonable service cost by the app. That service cost alone could generate a large turnover in each quarter and year. In addition, there is little expenditure on app maintenance and regular additions to the existing app for the time being. The feasibility of the project depends on the success of the seed idea as well as the publicity of the app. The primary objective is to transfer the power charging from one device to another device using QR-codes. The storing of power and grabbing of the power in QR-codes is possible. Furthermore, whatever portable gadgets exist now are designed to be capable of replacing their batteries with unique tailored batteries with ease. This operation also generates money for the project's manufacturing partner. The predicted challenges include batteries that enable fast charging and fast discharging, safety checks to be monitored while charging and discharging, saving the sender's power charge in a QR-code and obtaining the receiver by scanning the receiver's QR-Code. With proper analysis and proper interaction with the industry experts

and pioneered company hardware professionals would experience overcoming these kinds of pitfalls. The objective is to transfer the customized amount of charging to the receiver gadget through the internet as well as through the designated app. It is not only intended for transferring between portable devices, but also for transferring the customized portion of charging between supporting objects such as vehicles, larger entities such as devices that are used to cast movies in halls, provide flame in multi-cuisine restaurant cooking purposes, and run the motors in agriculture. The main goal is to simulate the phone, but it is for battery level changes by asking people in the social circles. The aim is reached by mutual agreement. The order of steps required to be used during the transformation of power from one device to anther device is as below:

(1) Design the app with demanded features.
(2) Users and devices are to be registered in the app, where devices are accessed by putting a request to the friends in their circle.
(3) Once mutual concern is done, sender would ask for specific charging. In turn, the receiver would give response by transferring. That transferred level would be stored in a QR-code and that would be scanned by the sender using auto-mirror scanning option.
(4) Once that QR-code is scanned that requested level would be transferred to the sender.
(5) Update both the sender and receiver gadgets charging.

2 Literature Review

The shortcomings of previous research can be used to develop a new model, where upgrades or innovative structures might be considered. The material offered in the sources listed in [1-8] is explanations and pointers to serve as a guide while seeking eye sight therapy. The key descriptions mentioned in [9–12] are illustrating many diseases for the eye and are also given the ways to resolve such problems. The extraction of data from the sources mentioned in [13-18] are demonstrating certain methodologies and techniques for getting eye sight to be done somehow but not intended to automate the process. With regard to the information given in [19, 21–23] are describing different scenarios using IoT technology in the traditional way but not mentioned toward achieving the full automation. The applications include social problems that are to be eliminated using IoT. As per the data given in [20], the concern is on providing the security and make storage confidentiality. In the view of information provided in [24–26], the information provided in the cloud in terms of storage as well as security using RSA and multifactor authorization. With regard to the article as in [27], the consequences of pervasive environment are described and making awareness of hybrid context studies and its framework is depicted. As per demonstration in [28], fire environment is handled in a smart manner using IoT and its drawbacks, and side effects are also listed under as study. This would be quick responsive approach in the emergency environment. In the view of [29], agriculture

field is controlled and is monitored using a novel indexing approach for decision making. With regard of information given in [30], gathering of temperature using specific type tubes arrays and their impact is noticed in this study. Regarding sources mentioned in [31-35], the descriptions are done using sensors, communication among them, generation of reports for further decision or action to take up, reducing time complexity, and automating the process as extent possible. Regarding the source mentioned in [36], the approach uses an indoor environment as a point cloud, and PCOC provides the sensor device connectivity as a LoS detection technique. The efficiency of this approach is verified by ray-voxel intersection technique. In the view of the source specified in [37], the various devices in terms of sensors, scanners and other useful forms are built to capture customer's eye focus on the items in the retail store and produce the statistics from which Fujitsu Forum increases sales of the items. In the source mentioned in [38], the sight laser detection model is designed that emits light from laser emitter through the CCD Detector, assess the attenuation and turbulence, error rate to be computed by various parameters involved in the double-pass propagation. Also, the MATLAB coding systems are used to detect the mono-static or bi-static type. The parameters used here are for the improvement of the budget link of the model. In the source mentioned in [39], the user's eye focuses on the line of sight on the eyeglasses and detects the angle based on four angles on the eyeglasses. The angle is detected using a CCD camera against the field of view. In this, the error rate is observed as less than 1.5 degrees. The micro-fabricated optical sensors are used for LOS measurement and are helpful in extracting point of sight using specific selection of objects and their tuning their functionality. With regard to the source demonstrated w.r.to [40], the vision sensors use a camera as a primary unit that captures the image for the presence, orientation, and accuracy of that specific part. The two models, such as monochromatic and color models, are demonstrated to process the accuracy of the parts in the image. In the source defined in [41], the diabetic retinal (DR) and STDR are determined using the NM-based approach to determine the sensitivity and specificity parameters. The related PPV and NPV are also determined in assessing the DR and STDR. In the aspect of [42], the approach such as diabetic retinopathy (DR) is to be detected using the techniques such as non-mydriatic fundus photography, 7-field fundus photography, and smart phone fundus photography. These techniques are helpful in evaluating the specific parameters which are of interest. In the view of the source specified in [43], the AI is used as automated software that determines the parameters such as sensitivity, and specificity, and other related terms for models such as diabetic retinopathy and sight threatening DR. Here, the smart phone-based system is used and helpful in ophthalmologist's grading. Regarding the source specified in [44], the impaired vision of the adults is because of direct diabetic retinopathy. The smartphone-based camera is preferred for digital retina images. This device would help determine the factors such as affordable, accessibility, and easier to use than other traditional approaches used.

Regarding the source described in [45], impaired vision is focused on the global population due to diabetic retinopathy and requires life-long monitoring to predict early. The different techniques proposed, such as smartphone photography and other

photography, are applied to compute DR, STDR. The remote interpretation and timely referral of such objects are sent to ophthalmologists. The advances in referrals support multi-modal and ultra-wide imaging creates new avenues for assessing the DR. The various fundus photography are used in screening the DR and management of it in the world. In the source defined in [46], the imaging devices and tools came in the ophthalmologists are used to examine the close interaction between the eye and tester. The high quality of the eye is managed via many parameters. Many newer tools are devised time being in this field in order to examine the eyes. In view of the source demonstrated in [47], smart phone retinal photography is proposed in screening the retina over the eye to determine diabetic retinopathy and other related issues. The primary setting called usability testing (retina scope) is required to improve quality service using medical assistants. They refine iteratively in the rounds based on results. In respect of the source specified in [48], diabetic retinopathy causes severe loss of eyesight in the patients. To prevent the loss early, NPDR is proposed. The advancements in DR would include with and without mydriasis, imaging in multiple filed or single fields, and the use of ultra-wide-field screening or other field screening. In this, many such technological advancements, evidence gaps, and studies enhance the DR screening. With respect to the source specified in [49], the study is on the northeast of Italy area called Padova where DR is screened using telemedicine screening. The report produce based on DR incidence, maculopathy, rate of progression to SDTR, and optimal screening interval with no DR in the patients with first examination.

In respect of source elaborated in [50], the review done over 50 countries in Asia about diabetic retinopathy which is serious cause of eye to be impaired. The screening of DR is a cost-effective approach to reduce severity of loss of sight. The review finally compares the affected population % and its country. After analysis, the affect rate is less compared with other country ICO guidelines. Regarding the source defined in [51], the full-field flicker ERG with RETeval device is used to assess the DR with subjects such as mydriasis free and a tool such as full-field electro retinogram. The ROC is visualized for the type of DR such as mild, moderate, and severe are judged. Regarding the source demonstrated in [52], telemedicine is applied over databases such as PubMed, EMBASE, and another type to determine the levels of DR and DME. The quality of the study is judged by the QUARAS-2 technique. The parameters such as sensitivity and specificity are pooled based on random effects. The group of ROCs is assessed to show the overall performance of the test. The heterogeneity is judged using meta-analysis and subgroup approaches. In the description of the source mentioned in [53], the combination of few measures such as retina structure as well as function that intervene diabetic retinopathy is better than modality state the two approaches such as (ETDRS-DR) severity $\geq$ level 53 (ETDRS-DR +) and RETeval DR Score > 23.5 (RETeval +). Regarding the study of [54], the communication from one device to other device is described sensors and also broadcast the timings of gate to the people in the range as well as to the registered users. The above mentioned every study or work demonstrates the usage of certain functionality or methodology to be adapted in shaping the Internet of things in the actual cause for real-time processing of the applications.

3 Proposed Approach

The physical entity for this is in the form of notes since the money is maintained in soft copy. Similarly, by storing the charge in a soft copy, the physical entity is in the form of the charging level of the device.

The app is titled as charge pe—where the gadget charging level status is displayed at the receiver, when the sender gadget is requested. It is not an infraction, but it does raise our contacts' awareness on charging level. If they have adequate charge, the receiver will make a charging request based on which the sender will approve and transmit, and both the parties' charging will be updated at their levels at the end. This is one method of charging devices automatically. Another idea is to use power banks as power carriers to run large organizations for hours in remote locations or on a regular basis. The emphasis is on the former situation, in which charging has to be transmitted via something similar to a phone or contact number. In this scenario, the ER diagram for this proposed system is shown, as well as the modules specified and also the pseudo procedures and flowchart for this intended system.

In the Fig. 1, the three modules identified such as.

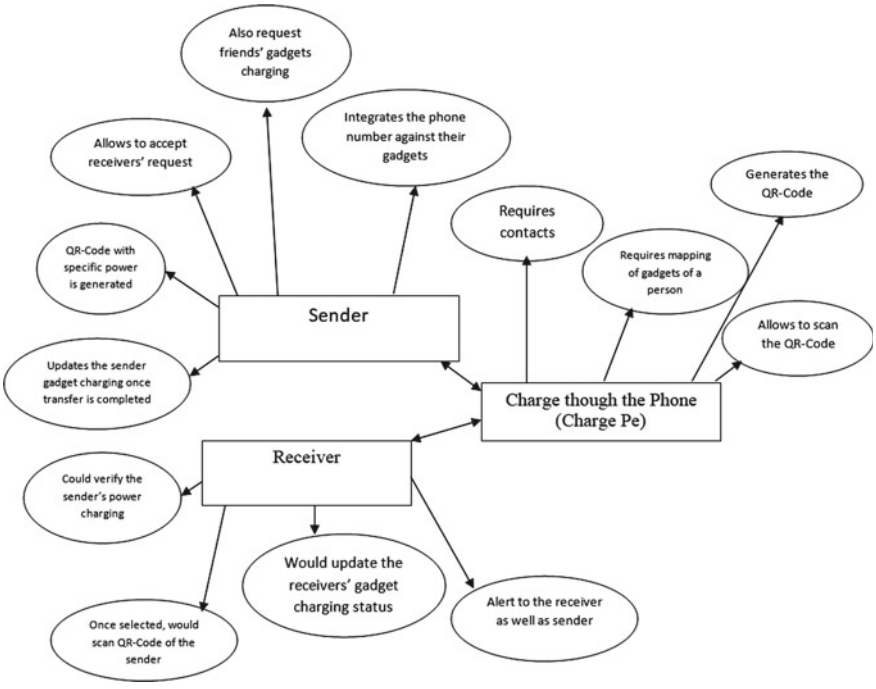


Fig.1 ER diagram of smart customized power charging

- **Charge Pe:** It is the central theme that is to be designed and needs to be installed at both the sender's and receiver's side in their mobiles. Both of them would like to add their gadgets under the phone number.
- **Sender:** It includes the sender phone number and could enter the details of their gadgets uniquely using gadget id per gadget. Once accepted the receivers' request, allowed to change the charging percentage (%) to be transferred, the QR-code is generated with that much amount of charging, and sent that QR-code in the encryption format with a key.
- **Receiver:** It includes receivers' phone number and also would enquire form whom sender planning to ask the charging %. Once select the senders' details would put a request to the sender. Once sender accepts it, encrypted QR-code to be decrypted using the key shared. Generally, it is the OTP would act as the key. Once receiver decrypts, both receiver and sender gadgets status to be updated after successful transfer.

The pseudo procedures of these three modules are defined as below:

(A) Pseudo_Procedure **Charge_Pe**(Text Guide, Object[] gadgets_IDs,Phone_number):

Step 1: Follow the guide, install the charge pe app.
- 1.1 Ask for phone number to enter
- 1.2 Ask for gadgets_ID to add under entered phone number
- 1.3 Mapping is done among the gadgets and phone number

Step 2: Supports to login to the charge pe using passcode
- 2.1 Supports receiver and their activities such as verifying the sender's gadgets and put a request.
- 2.2 Sender would accept, enters the % of charging to be willing to transfer to the receiver, once entered, generates the QR-code that could ne encrypted with key, alerts that key(OTP) to the receiver.
- 2.3 Receiver when received the QR-code would decrypt it, and automatically both the parties' gadgets charging levels are updated.
- 2.4 alert the same updation to both parties for easy future reference.

(B) Pseudo_Procedure **Sender** (request[],receiver[]):

Step 1: Gets a request of specific amount of charging from a specific gadget

Step 2: Selects a specific level of charging or enters willing % of charging to be transferred

Step 3 : Generates the QR-code with specific key, shares the encrypted QR-code ie E-QR-Code

Step 4: Expecting the receiver to scan

(C) Pseudo_Procedure **Receiver**(E-QR-Code, Key):

Step 1: Auto-mirror scanner is a built-in feature defined in the charge pe would scan the E-QR-Code with key

Step 2: Gets the receivers' gadget charging is to updated
- 2.1 R_Charging = R_Charging + S_Willingness_Charging

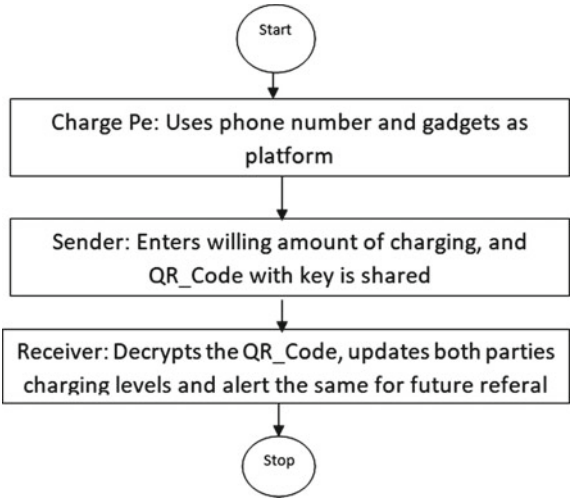


Fig. 2 Flowchart of smart customized charging

2.2 If (S_Willingness_Charging < S_Charging)
S_Charging=S_Charging–S_Willingness_Charging
else
alert("enter the valid % of charging").
where S_Willingness_Charging is sender entered % of charging to the receiver, R_Charging is the receiver gadget available charging and S_Charging is the sender gadget present charging level.

The order of modules to be interacted in a pictorial diagram is defined as below (Fig. 2):

4 Results

For any intended theme, the output screens play a vital role in understanding the complete functionality of proposed approach.

The output screens are specified as below: It describes the order of modules that are to be interacted to perform the transfer of charging in the user friendly manner.

From Fig. 3, the major activities are mentioned in the visual demonstration, where each sub-window is narrated as per the functionality. The user must first supply a phone number, and then the receiver, who would want to take charge from the contacts of friends and pick the devices' charging levels at the time itself in a fast guide, which must make a request to the sender. The sender may have the ability to violate or input the charge amount as desired, and a QR-code with OTP as a key is produced and encrypted. That encrypted code is delivered to the receiver, where

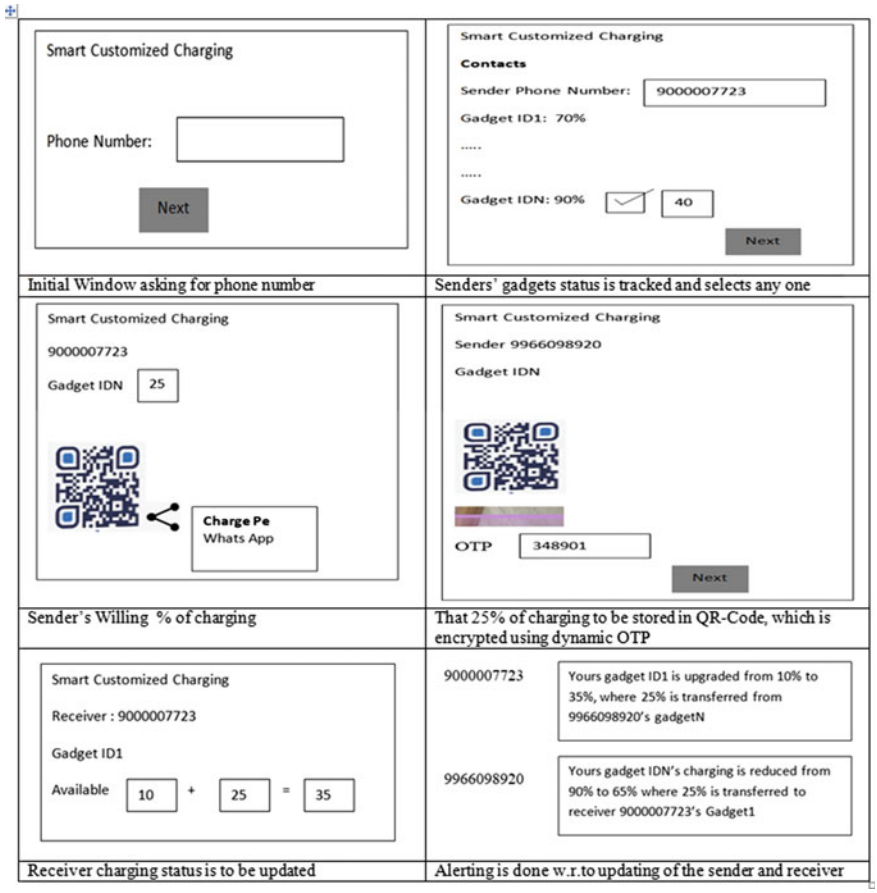


Fig. 3 Order of flow of windows in the smart customized charging system

an auto-web-mirror scanner is employed to decode and update the receivers' and senders' gadget status. That updating details could be communicated by messages.

The following is the diagram that shows the flow of events to be happened to complete the requested the transfer of charging (Figs. 4 and 5):

According to the performance of the desired theme versus existing systems, the intended theme is faster than existing systems

In terms of security, the smart customized charging technique employs encryption and decryption with OTP as the key. The following comparison shows the security of existing versus smart customized charging (Fig. 6):

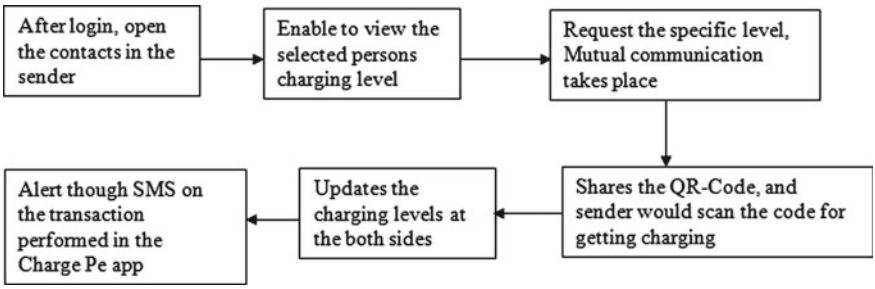


Fig. 4 Flow of activities in the charge Pe app

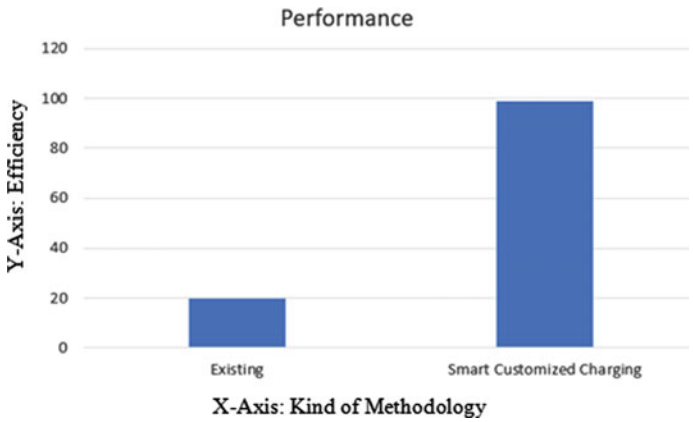


Fig. 5 Performance of smart customizable charging versus existing systems

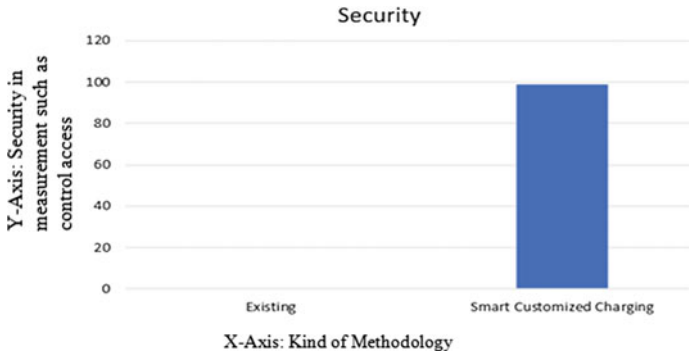


Fig. 6 Security of smart customizable charging versus existing systems

5 Conclusion

To enable charging to be transferred in the same way that Phone Pe does, an automated design is created in which users' gadgets are stored under their unique phone number, and users can send a request to other users in their friends' contacts, another user who receives a request can enter the possibility of charging to the receiver, the QR-code with OTP is encrypted, and that is sent to the receiver. The receiver scans the encrypted data and decrypts it using the OTP. Security is provided between users in this manner. In updating the charging states of both parties' gadgets, where performance is also taken into consideration. The proposed system would be more efficient than currently available systems on the market and would be entirely automated. The automation would be expanded to include large power tanks with dynamic statistics, which would allow for more user friendliness and customization.

References

1. Understanding Vision, Take part in the ZEISS Online Vision Screening Check and test the quality of your vision. ZEISS Online Vision Screening, Oct (2017). https://www.zeiss.co.in/vision-care/better-vision/zeiss-online-vision-screening-check.html
2. A Glossary of Eye Tests and Exams, Eye Health, Reference. https://www.webmd.com/eye-health/eye-tests-exams#1
3. M.X. Repka, D. Gudgel, Home eye test for children and adults. March (2020). https://www.aao.org/eye-health/tips-prevention/home-eye-test-children-adults
4. 6 Eye Tests In A Basic Eye Exam. https://allabouteyes.com/6-eye-tests-basic-eye-exam/
5. D. Turbert, M.D. Brenda Pagan-Duran, Eye exam and vision testing basics (2018). https://www.aao.org/eye-health/tips-prevention/eye-exams-101
6. Working safely with display screen equipment. https://www.hse.gov.uk/msd/dse/eye-tests.htm
7. L. Segre, G. Heiting, What's an eye test? (2021). https://www.allaboutvision.com/eye-test/free-eye-chart/
8. Vision screening, The college of Optometrists, http Eye charts and visual acuity explained, https://www.allaboutvision.com/eye-test/free-eye-chart/s://guidance.college-optometrists.org/guidance-contents/knowledge-skills-and-performance-domain/examining-patients-who-work-with-display-screen-equipment-or/vision-screening/
9. M. Harkin, A.M. Griff, Visual acuity test. (2012). https://www.healthline.com/health/visual-acuity-test
10. Basics of vision and eye health, Basics of Vision and Eye Health, Common Eye Disorders and Diseases, centers for Disease Control and Prevention. https://www.cdc.gov/visionhealth/basics/ced/index.html
11. Top Causes of Eye Problems, Eye Health, Reference. https://www.webmd.com/eye-health/common-eye-problems#1
12. R.E. Gans, The 5 most common vision problems and how to prevent them, health essentials. https://health.clevelandclinic.org/the-5-most-common-vision-problems-and-how-to-prevent-them/
13. R. Constantine, OTR/L, Vision techniques for eye movement disorders associated with autism. ADHD, Dyslexia and other neurological disorders: hands-on assessments and treatments for children and adolescents, RNV063660. https://www.pesi.com/store/detail/26169/vision-techniques-for-eye-movement-disorders-associated

14. D.R. Gold, Eye movement disorders. in *Liu, Volpe, and Galetta's Neuro-Ophthalmology,* 3rd edn. (2019)
15. R.J. Leigh, M. Gross, in *Encyclopedia of Neuroscience*, Eye Movement Disorders (2009). https://www.sciencedirect.com/science/article/pii/B9780080450469010937
16. M.A. Tamhankar, in *Liu, Volpe, and Galetta's Neuro-Ophthalmology* 3rd edn. Eye Movement Disorders: Third, Fourth, and Sixth Nerve Palsies and Other Causes of Diplopia and Ocular Misalignment (2019). https://www.sciencedirect.com/science/article/pii/B9780323340441000158, https://doi.org/10.1016/B978-0-323-34044-1.00015-8
17. C.G. Chisari, A. Serra, in *Reference Module in Neuroscience and Biobehavioral Psychology, 2017*, *Abnormal Eye Movements due to Disease of the Extraocular Muscles and Their Innervation.* https://doi.org/10.1016/B978-0-12-809324-5.01292-X
18. I.V. Ivanov, M. Mackeben, A. Vollmer, P. Martus, N.X. Nguyen, S. Trauzettel-Klosinsk, Eye movement training and suggested gaze strategies in tunnel vision—a randomized and controlled pilot study. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0157825, https://doi.org/10.1371/journal.pone.0157825
19. M. Kavitha, Y. Manideep, M. Vamsi Krishna, P. Prabhuram, Speech controlled home mechanization framework using android gadgets. Int. J. Eng. Technol. (UAE) **7**(1.1), 655–659 (2021)
20. N. Sunanda et al., Light weight secure data sharing scheme for mobile cloud computing. in *Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC).* (IEEE, 2019)
21. M. Kavitha, S. Srinivasulu, K. Savitri, P. Sameera Afroze, P. Akhil Venkata Sai, S.I. Asrith, Garbage bin monitoring and management system using GSM.Int. J. Innov. Explor. Eng. **8**(7), 2632–2636 (2019)
22. M. Kavitha et al., Wireless sensor enabled breast self-examination assistance to detect abnormality. in *International Conference on Computer, Information and Telecommunication Systems (CITS).* (IEEE, 2018)
23. C.S. Kolli, V.V. Krishna Reddy, M. Kavitha, A critical review on internet of things to empower the living style of physically challenged people. in *Advances in Intelligent Systems and Computing*. (Springer, Singapore, 2020). pp. 603–619
24. S. Nalajala, P. Ch, A. Meghana, B. Phani Meghana, Data security using multi prime RSA in cloud. Int. J. Recent Technol. Eng. **7**(6S4), (2019) ISSN: 2277–3878
25. N. Sunanda et al., Data security in cloud computing using three-factor authentication. in *International Conference on Communication, Computing and Electronics Systems.* (Springer, Singapore, 2020)
26. N. Sunanda, N. Sriyuktha, P.S. Sankar, Revocable identity based encryption for secure data storage in cloud. Int. J. Innov. Technol. Explor. Eng. **8**(7), 678–382 (2019)
27. J. Madhusudanan, S. Geetha, Venkatesan, V. Prasanna, U. U. Vignesh, P. Iyappan, Hybrid aspect of context-aware middleware for pervasive smart environment: a review. Mobile Info. Syst. (2018). https://doi.org/10.1155/2018/6546501
28. L.P. Maguluri, T. Srinivasarao, R. Ragupathy, M. Syamala, N.J. Nalini, Efficient smart emergency response system for fire hazards using IoT. Int. J. Adv. Comput. Sci. Appl. (2018)
29. M.S. Mekala, P. Viswanathan, (t,n): Sensor stipulation with THAM index for smart agriculture decision-making IoT system. (2020). https://doi.org/10.1007/s11277-019-06964-0
30. K.G. Kumar, B.S. Avinash, M. Rahimi-Gorji, J. Majdoubi, Photocatalytic activity and smartness of TiO_2 nanotube arrays for room temperature acetone sensing (2020). https://doi.org/10.1016/j.molliq.2019.112418
31. G. Subba Rao, S. Hrushikesava Raju, L. Ramani Burra, V. Naresh Mandhala, P. Seetha Rama Krishna, Smart cyclones: creating artificial cyclones with specific intensity in the dearth situations using IoT, February, 18–19th, SMART DSC-2021, Springer Series (2021)
32. S. Hrushikesava Raju, L. Ramani Burra, S. Faiayaz Waris, S. Kavitha, IoT as a health guide tool. in *IOP Conference Series, Materials Science and Engineering*, vol. 981. (2015) pp. 4. https://doi.org/10.1088/1757-899X/981/4/042015.

33. S. Hrushikesava Raju, L. Ramani Burra, A. Koujalagi, S. Faiayaz Waris, in *Tourism Enhancer App: User-Friendliness of a Map with Relevant Features IOP Conference Series, Materials Science and Engineering*. vol. 981. (2021). pp. 2. https://doi.org/10.1088/1757-899X/981/2/022067
34. N. Sunanda, S. Hrushikesava Raju, S.F. Waris, A. Koulagaji, Smart instant charging of power banks smart instant charging of power banks. in *IOP Conference Series Materials Science and Engineering* vol. 981. (2020). pp. 210. https://doi.org/10.1088/1757-899X/981/2/022066
35. S. Hrushikesava Raju, M. Nagabhushana Rao, Improvement of time complexity on external sorting using refined approach and data preprocessing. Int. J. Comput. Sci. Eng. **4**(11), 82–86 (2016)/ E-ISSN:2347–2693
36. R. Sharma, V. Badarla, V. Sharma, PCOC: a fast sensor-device line of sight detection algorithm for point cloud representations of indoor environments. IEEE Commun. Lett. **24**(6), 1258–1261 (2020). https://doi.org/10.1109/LCOMM.2020.2981058
37. D. Klaus, Line-of-sight detection technology. Demonstration Draws Attention, May (2014). https://blog-archive.global.fujitsu.com/line-of-sight-detection-technology-demonstration-draws-attention/
38. C. Lecocq, G. Deshors, O. Lado-Bordowsky, J.-L. Meyzonnette, Sight laser detection modeling. in *Proceedings SPIE 5086, Laser Radar Technology and Applications* VIII, (21 August 2003). https://doi.org/10.1117/12.486055
39. M. Ozawa, K. Sampei, C. Cortes, M. Ogawa, A. Oikawa, N. Miki, Wearable line-of-sight detection system using micro-fabricated transparent optical sensors on eyeglasses. Sens. Actuators A **205**(1), 208–214 (2014). https://doi.org/10.1016/j.sna.2013.11.028
40. Detection based on "Camera Images" What Are Vision Sensors?. https://www.keyence.co.in/ss/products/sensor/sensorbasics/vision/info/
41. V. Prathiba, R. Rajalakshmi, S. Arulmalar, M. Usha, R. Subhashini, C.E. Gilbert, R.M. Anjana, V. Mohan, Accuracy of the smartphone-based nonmydriatic retinal camera in the detection of sight-threatening diabetic retinopathy. Indian J. Ophthalmol. **68**(Suppl 1), S42–S46 (2020).https://doi.org/10.4103/ijo.IJO_1937_19
42. M.E. Ryan, R. Rajalakshmi, V. Prathiba, R.M. Anjana, H. Ranjani, K.M. Narayan, T.W. Olsen, V. Mohan, L.A. Ward, M.J. Lynn, A.M. Hendrick, Comparison among methods of retinopathy assessment (CAMRA) study: smartphone. Nonmydriatic. Mydriatic Photograp. **122**(10), 2038–2043 (2015). https://doi.org/10.1016/j.ophtha.2015.06.011 (Epub 2015 Jul 16)
43. R. Rajalakshmi, R. Subashini, R. Mohan Anjana, V. Mohan, Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye (Lond) **32**(6), 1138–1144. https://doi.org/10.1038/s41433-018-0064-9. Epub 2018 Mar 9
44. C.H. Tan, B.M. Kyaw, H. Smith, C.S. Tan, L.T. Car, Use of smart phones to detect diabetic retinopathy: scoping review and meta-analysis of diagnostic test accuracy studies. J. Med. Internet Res. **22**(5), e16658 (2020). https://doi.org/10.2196/16658
45. R. Rajalakshmi, V. Prathiba, S. Arulmalar, M. Usha, Review of retinal cameras for global coverage of diabetic retinopathy screening. Eye (Lond) **35**(1), 162–172 (2021). https://doi.org/10.1038/s41433-020-01262-7
46. A. Pujari, G. Saluja, D. Agarwal, H. Selvan, N. Sharma, Clinically useful smartphone ophthalmic imaging techniques. Graefes. Arch. Clin. Exp. Ophthalmol. **259**(2), 279–287 (2021). https://doi.org/10.1007/s00417-020-04917-z (Epub 2020 Sep 11)
47. P. Li, Y.M. Paulus, J.R. Davila, J. Gosbee, T. Margolis, D.A. Fletcher, T.N. Kim, Usability testing of a smartphone-based retinal camera among first-time users in the primary care setting. BMJ Innov. **5**(4),120–126. https://doi.org/10.1136/bmjinnov-2018-000321. Epub 2019 Sep 19
48. E. Pearce, S. Sivaprasad, A review of advancements and evidence gaps in diabetic retinopathy screening models. Clin. Ophthalmol. **14**(14), 3285–3296 (2020). https://doi.org/10.2147/OPTH.S267521
49. S. Vujosevic, P. Pucci, M. Casciano, AnnaRita Daniele, S. Bini, M. Berton, F. Cavarzeran, A. Avogaro, A. Lapolla, E. Midena, A decade-long telemedicine screening program for diabetic retinopathy in the north-east of Italy. J. Diabetes Compl. **31**(8), 1348–1353 (2017). https://doi.org/10.1016/j.jdiacomp.2017.04.010 (Epub 2017 Apr 13)

50. L.Z. Wang, C.Y. Cheung, R.J. Tapp, H. Hamzah, G. Tan, D. Ting, E. Lamoureux, T.Y. Wong, Availability and variability in guidelines on diabetic retinopathy screening in Asian countries. r J. Ophthalmol. **101**(10),1352–1360 (2017). https://doi.org/10.1136/bjophthalmol-2016-310002. Epub 2017 Mar 14
51. Y. Zeng , D. Cao, D. Yang, X. Zhuang, H. Yu, Y. Hu, Y. Zhang, C. Yang, M. He, L. Zhang, Screening for diabetic retinopathy in diabetic patients with a mydriasis-free, full-field flicker electroretinogram recording device,. oc Ophthalmol. **140**(3), 211–220. https://doi.org/10.1007/s10633-019-09734-2. Epub 2019 Nov 12
52. L. Shi, W. Huiqun, J. Dong, K. Jiang, L. Xiting, J. Shi, Telemedicine for detecting diabetic retinopathy: a systematic review and meta-analysis. Br. J. Ophthalmol. **99**(6), 823–831 (2015). https://doi.org/10.1136/bjophthalmol-2014-305631 (Epub 2015 Jan 6)
53. M.G. Brigell, B. Chiang, A.Y. Maa, C.Q. Davis, Enhancing risk assessment in patients with diabetic retinopathy by combining measures of retinal function and structure. Transl. Vis. Sci. Technol, **26**, 9(9), 40. https://doi.org/10.1167/tvst.9.9.40. eCollection 2020 Aug
54. A. Koujalagi, S. Hrushikesava Raju, CH.M.H. Sai Baba, L.R. Burra, K. Sreekantha, Smart railway gate timings app: monitoring of the Indian railway gates through a novel app. in *International Conference on Smart Data Intelligence* (ICSMDI 2021), June (2021). https://ssrn.com/abstract=3852494

Evolving Chaotic Shuffled Frog Leaping Memetic Metaheuristic Model-Based Feature Subset Selection for Alzheimer's Disease Detection

C. Dhanusha, A. V. Senthil Kumar, G. Jagadamba, and Ismail Bin Musirin

Abstract The progress in the medical field has resulted in a significant increase in the scientific understanding about Alzheimer's disease (AD), with a distinct description by analyzing the demographic, clinical, and MR structural imaging data. Using the clinical profile, it is possible to diagnose Alzheimer's disease in early stages and its existence can be detected with high accuracy. However, due to the enormous volume of the continuously expanding AD dataset, data management is becoming a serious challenge, and using raw dataset without eliminating unnecessary and duplicated information would have a significant influence on prediction/classification accuracy. Hence, the proposed research work aims to develop an enhanced memetic metaheuristic behavioral-based feature selection model known as chaotic shuffled frog leaping algorithm (CSFLA). Unlike, the standard version of SFLA, the proposed model is enhanced by adapting the chaotic mapping, when the candidate frog involved in searching process produces worst result, instead of swapping it with a random population, the chaotic mapping is used. CSFLA has produced better subsets of features for AD prediction due to its simplicity, reduced parameters, optimal performance, easy realization, and fast computation while comparing it with ant colony optimization and firefly algorithm. This work has used the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset for discovering the potential features involved in Alzheimer's disease detection. The deep neural network has been used to validate the performance of all three feature subset

C. Dhanusha (✉)
Research Scholar, Department of MCA, Hindusthan College of Arts and Science, Coimbatore, India

A. V. Senthil Kumar
Director, Professor, Department of MCA, Hindusthan College of Arts and Science, Coimbatore, India

G. Jagadamba
Department of Information Science and Engineering, Siddaganaga Institute of Technology, Tumakuru, Karnataka, India

I. B. Musirin
Faculty of Electrical Engineering, Universiti Teknologi Mara, Mara, Malaysia

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_51

selection models, and the results demonstrated that CSFLA achieves higher accuracy than other models.

Keywords Alzheimer's disease · Deep learning · Chaotic mapping · Memetic metaheuristic · Shuffled frog leaping · Feature selection · Local optima · Convergence

1 Introduction

When the cognitive impairment worsens over time, it results in a neurodegenerative disorder known as Alzheimer's disease (AD) [1]. Medical professionals have worked hard to discover and improve therapeutic options that can prevent this degenerative process at its early stages. Initially, the symptoms of cognitive impairment tend to be minor, and this condition is termed as mild cognitive impairment (MCI). Further, it is reported that, not all the mild cognitive impairment will be under the risk of developing AD but it has a tenfold increased risk for AD development [2]. To identify the presence of AD, psychometric tests of memory function, demographic details, and clinical profile are used as the dataset. The deep learning, machine learning, and data mining models play a vital role in early detection of AD [3]. Nonetheless, dealing with a large number of irrelevant and redundant data has a significant impact on the performance of prediction models. As a result, it is critical to address this issue by employing optimized feature subset selection models with the goal of maximizing relevancy and minimizing redundancy in AD datasets.

The latest advancement in the field of metaheuristic science is mainly based on the mimetic models and optimization in handling voluminous medical dataset is remarkable. The importance of significant feature subset selection is the main factor used for improving the disease prediction models as it results in increasing its accuracy through predominant computational issues. For the performance enhancement of support vector machine neural networks in Alzheimer's disease prediction, feature selection models based on optimization algorithms such as particle swarm optimization, ant colony optimization, and firefly algorithm were utilized. The findings of these research explain the functioning of mimetic metaheuristic approaches that perform better than standard models.

Despite the wide application of metaheuristic science in dealing the continuous growing nature of Alzheimer's disease dataset, there is a noticeable gap of knowledge in potential feature selection involves deep neural network-based classification optimization, which was the primary motive of this present research work. It is accomplished by employing chaotic shuffled frog leap algorithm (SHFLA) for achieving optimized feature subset selection which impressively influences the prediction accuracy of deep neural network in Alzheimer's disease detection. The proposed model chaotic shuffled frog leap algorithm (CSHFLA) fine-tunes the process of feature selection for analyzing and involving the reduced feature subset for Alzheimer's disease detection.

2 Related Work

Chtioui et al. [4] developed a nearest neighbor method-based genetic algorithm for finding the feature subsets, which maximize the classification accuracy. The performed discrimination of seeded is achieved by using artificial vision.

Zhang et al. [5] designed a bare bones particle swarm optimization-based feature selection. The quantitative results are obtained using eight different datasets. This binary feature selection solves the feature selection issue.

Amira et al. [6] in their work to perform Alzheimer's disease detection, the feature selection is done using the support vector machine. The features of cortex difference are used to detect the AD.

Firouzeh et al. [7] constructed a two-stage method for predicting the AD in brain images, and scattered filtering along with neural network is used to extract the significant features of the data. The softmax based on regression is used to detect the Alzheimer's disease.

Jack [8] devised a novel model which compares each conceivable pairs of time-based data instances. The model cast-off a learning approach which is accomplished on managed data and estimated with testing dataset separately. An artificial neural network model was developed to determine the cognitive stage of the patients as mild to severe cognitive impairments.

Erguzel et al. [9] used the ant colony-based feature selection for pretreating the EEG dataset which utilizes the frequency of theta and delta as features. The backpropagation model is used for classification.

Li et al. [10] in their work used discriminative feature selection known as firefly algorithm. The supporting decision is done using regression model. The simulated annealing is integrated along with firefly algorithm which is used for data-based learning process.

Zhou et al. [11] used TrAdaboost to predict and classify the severity of Alzheimer using ADNI dataset. The feature selection is done by using information gain method. The knowledge inferred from the ADNI was moved to local hospital's AD samples.

3 Methodology: Chaotic Shuffled Frog Leaping Memetic Metaheuristic Model-Based Feature Subset Selection for Alzheimer's Disease Detection

This work has used ADNI dataset for predicting the Alzheimer's disease, where the dataset comprises of 1409 subjects with 10 demographic and clinical attributes [12]. The raw dataset is normalized using min–max process to convert the range of attributes to fall under the same range of intervals. To improve the irrelevant and redundant attributes in ADNI dataset [12], this work developed a chaotic shuffled frog leaping algorithm (CSFLA) which is a kind of memetic-based nature-inspired metaheuristic model for selecting the potential features. The CSFLA has the ability

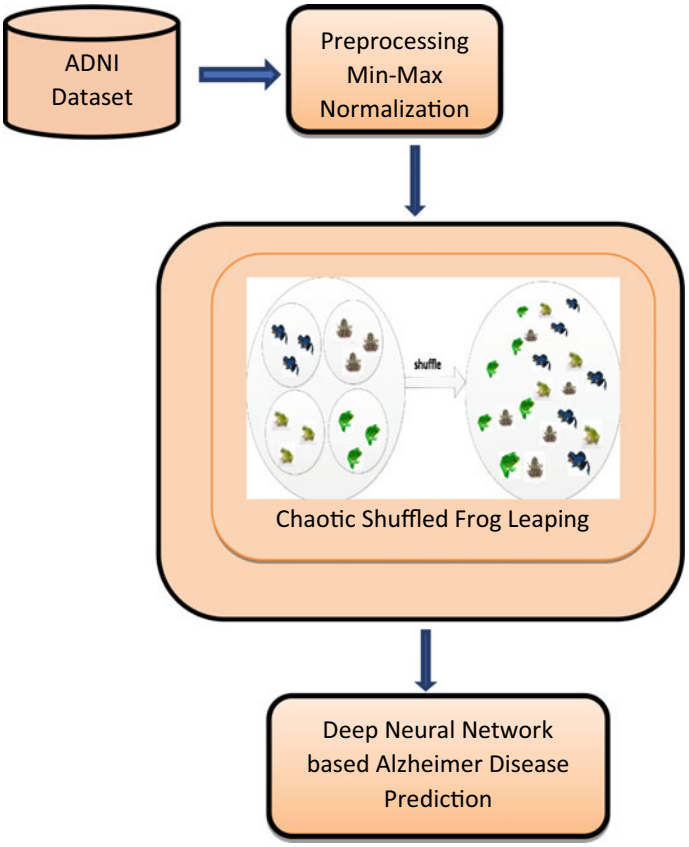


Fig. 1 Workflow of chaotic shuffled frog leaping feature subset selection for AD prediction

to balance both the local search and global search with its leaping behavior and shuffling subgroups to overcome the local optima. The chaotic mapping theory is used for selecting the population so that it increases the optimization of best feature selection process. To validate this model, deep neural network is used for predicting the accuracy of AD detection (Fig. 1).

3.1 Chaotic Shuffled Frog Leaping Algorithm

One of the innovative memetic metaheuristic models is shuffled frog leaping algorithm (SFLA) which is developed based on the nature of the frogs leaping behavior for searching its food [13]. This algorithm is a kind of population-based algorithm which is used for solving significant optimization issues. In this work, significant feature selection in AD dataset is done by inheriting the genetic algorithm structure in order

to enhance the aqueous performance of potential search intensification process. The SFLA to perform local search among subgroups integrates both particle group optimization and memetic process. The SFLA becomes more popular and highly efficient due to its simplicity and high convergence speed to act as a global optimizer.

The entire population of the frog is split into small subgroups, each group denotes variant types of frogs which are scattered in different places of the solution space, in this work complete features of AD dataset. Let us assume that the population of frogs is N with R as problem size and the location of j th frog is mathematically represented as

$$Yi = (yi1,\ yi2, \ldots, yiR) \tag{1}$$

The N frogs are alienated into k memeplexes, where memeplexes are set of frogs with same structure, but their adaptability differs among each other. By calculating each frog fitness value, they are arranged in descending order. The 1st frog of the first list belongs to the first memeplex and the 2nd frog belongs to the second memeplex, and this process is followed until there are no more frogs to be assigned. It should be noted that the 1st memeplex receives the $k + 1$th at the end, and this procedure is continued until all the frogs are subsidized. It is illustrated in Fig. 2.

During each iteration 't', the SLFA method uses the fitness value of each individuals in assigning them to different groups in turn where the worst frog ($N\_w\hat{}t$) has learned from the best individual ($N\_b\hat{}t$) in a subgroup. If there is no progress in its learning behavior for searching optimal solution, then it learns from the global best individual's ($N\_g\hat{}t$). Still, if it does not produce any progress, then ($N\_w\hat{}t$) will be substituted by an arbitrary individual from the population. The grouping of individuals is depicted in Fig. 3.

$$Dt^t = PX\left(N_b^t - N_w^t\right). \tag{2}$$

$$N_w^{t+1} = N_w^t + Dt^t(Dt_m \geq Dt \geq -Dt_m). \tag{3}$$

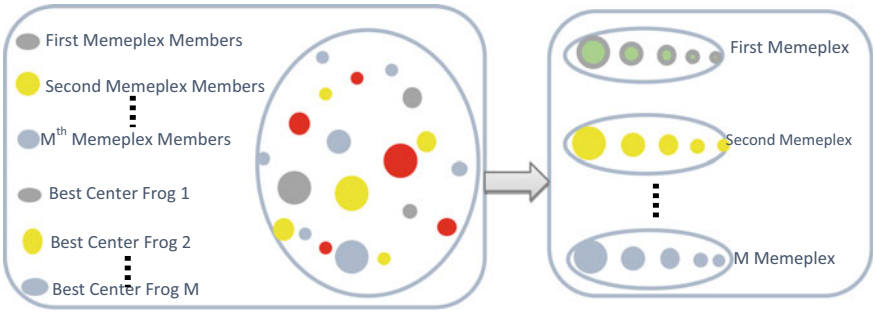


Fig. 2 Shuffling frog leaping algorithm behavior

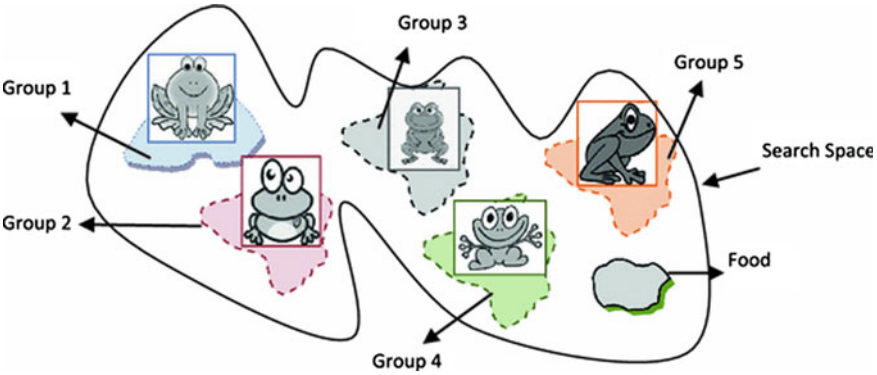


Fig. 3 Virtual frogs searching behavior

Where $N\_w\hat{}(t+1)$, $N\_w1\hat{}(t+1)$, $N\_w2\hat{}(t+1)...N\_wm\hat{}(t+1)$ is a new individual produced ny the updating policy represented as in Eq. (2). $[\![Dt]\!]$ $\hat{}t$ refers to each individual moving step size and P is the arbitrary number with a variation of [0,…,1]. The leaping step value ranges between $-Dt_m$, Dt_m].

After updating if the newly produced individual $N\_w\hat{}(t+1)$ is improved over its previous value then $N\_w\hat{}t$ is replaced by $N\_w\hat{}(t+1)$, else $N\_b\hat{}t$ will be replaced by $N\_g\hat{}t$. Still there is no progress then $N\_w\hat{}t$ will be replaced in arbitrarily by new individual. This is repeated as an iterative process till it reaches the number of subgroups. When subgroups processing is finished, then all the subgroups will be sorted in a random fashion and resplit them into new subgroups is accomplished until it met the predetermined termination.

3.2 Chaotic Mapping-Based Shuffling Frog Leaping Algorithm Behavior

In the standard SFLA, the initial population or while swapping the members of memeplex, it is done in a random manner. This may have the higher chance of not considering the most promising frogs which will produce more optimum feature selection search results. To overcome this problem, this present research work used chaotic mapping instead of random selection, which overcome the local optima and earlier convergence termination. The chaotic mapping [14] is mathematically characterized as shown in the equation.

$$b_{i+1} = \begin{cases} 4\mu b_i(0.5 - b_i), & 0 \le \mathrm{b}_i < 0.5 \\ 1 - 4\mu b_i(0.5 - b_i)(1 - b), & 0.5 \le b_i, \quad \le 1 \end{cases} \tag{4}$$

where $\mu = 4$, *bi* rand $\in (0,1)$ are the predefined controlling parameter that allows to reach global optimization.

Algorithm: Chaotic Shuffled Frog Leaping Algorithm-based feature subset selection for AD Prediction Input: ADNI Dataset.

Output: Reduced Feature Subset of ADNI Dataset

Procedure

1. Initialize number of memeplexes—mp and nf—number of member (Frogs) in each memeplex.
2. Pond population N = mp*nf
3. Select initial virtual frogs VF Y(1), Y(2), …, Y(VF)
4. For I = 1 to VF

   4.1 Compute the fitness value of each virtual frog Y(i)
   //distance of each virtual frog from mean value

$$\text{Fitness} = \text{dist}(\text{Y(i)}, \text{Mean}(\text{Y(i} =_{1\ldots\text{VF}})) \tag{5}$$

   4.2. Sort the virtual frogs in descending order based on their fitness value

5. End for
6. Do
7. Divide the virtual frogs $\text{Y}_{(\text{i=1..VF})}$ into mp memeplexes
8. Construct submemeplex (nf) for each memeplex
9. Compute the new position of the worst frog in each submemeplex using the equation

$$Dt^t = PX\left(N_b^t - N_w^t\right) \tag{2}$$

$$N_w^{t+1} = N_w^t + Dt^t(Dt_m \geq Dt \geq -Dt_m) \tag{3}$$

10. For i = 1 to nf
    if (N_w^{t+1} is better than N_w^t) then
    then N_w^t =. N_w^{t+1}.,
    $if(N_b^t$. not better than N_w^t) then.
    $N_w^{t+1} = N_g^t$.
    Else if $N_g^t isnotbetterthan N_w^t$ then.
    N_w^t will be replaced by swapping with a new frog using chaotic mapping.

    End for

11. Shuffle entire population of frogs, go to step 6
12. Until the stopping condition is not met

3.3 *Deep Neural Network*

A type of artificial neural network with multiple layer as intermediator among the input and output layers is known as deep neural network (DNN) [15]. In general, neural networks are designed based on the inspiration of human brain functionality, and thus, it comprised of neurons such as node, weights, and bias as parameters which provides signal to the link between one layer to another layer. DNN is used to design a multifaceted nonlinear association. This architecture produces compositional method in which the objects are represented as layer composition. DNN works as a feedforward network, and the data is transferred from the input layer to the intermediate layers and finally to the output layers without reversing as shown in the figure (Fig. 4).

The DNN constructs a virtual neurons map and allocates random weight values to the connection between the preceding layer and the following layer. Using the activation function, the weights and input values are multiplied and summed to produce the output between the range of 0 and 1. The produced output which is known as observed output is compared to the expected output, and if there is a difference, then it is considered as error, an algorithm like backpropagation is used for adjusting the weight values, so that the parameters become more influential until it discovers the accepted rate of accuracy. The figure shows the DNN with input, multiple intermediate dense layers, and the output layer.

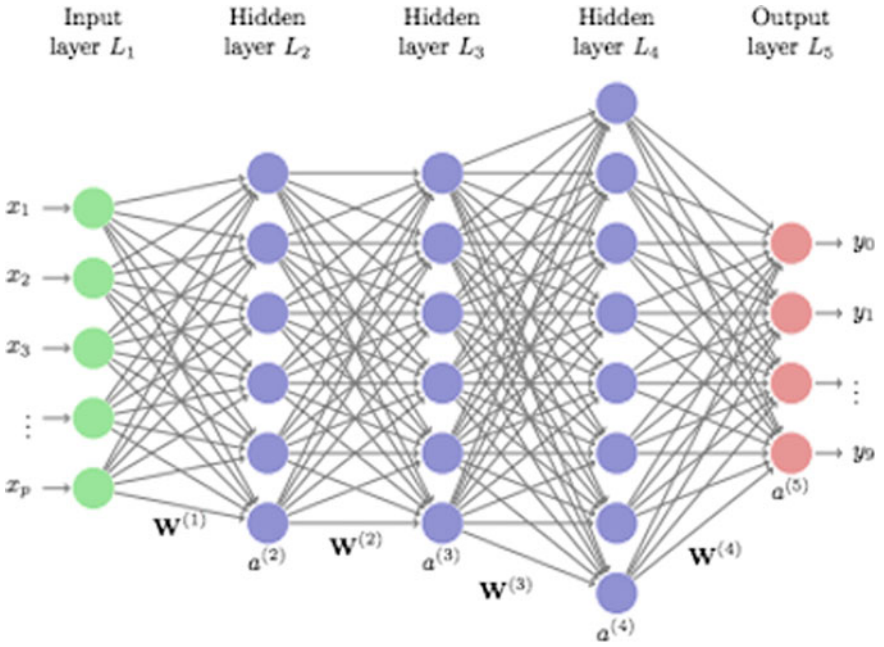


Fig. 4 Architecture of deep neural network

The CSFLA is used for potential feature subset selection in ADNI dataset prediction; using the fitness value of each virtual machine, the best feature is selected and the performance of it is validated by passing the reduced feature subset as the input data to the deep neural network to discover the accuracy of the prediction model.

4 Results and Discussion

This section discusses about the performance evaluation of the three different feature subset selections, such as ant colony optimization (ACO), firefly algorithm (FFA), and proposed chaotic shuffled frog leaping algorithm (CSFLA). The feature subsets are generated with these three feature selection models along with the full dataset (FDS), which are validated by using deep neural network for Alzheimer's disease detection. The proposed chaotic shuffled frog leaping algorithm is deployed using MATLAB code, and the dataset used here is ADNI dataset [12], which consists of 1409 subjects with 10 demographic and clinical attributes Patient ID, Visitor code, Examined Date, Age, Gender, Education, APOE4, Mini Mental State Examination (MMSE), APOE Genotype, and Clinical Dementia Rating. The suggested CSFLA model makes use of 80% training dataset and 20% testing dataset.

Table 1 and Fig. 5 show the number of features selected by different feature

Table 1 Number of features selected

Feature selection methods		No. of feature selected
ACO		8
FFA		6
CSFLA		4
WDS		10

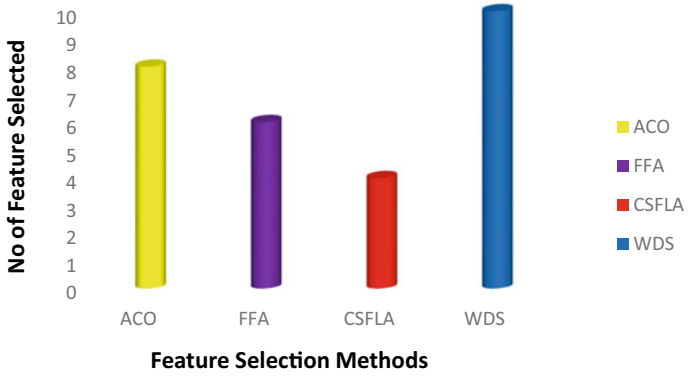


Fig. 5 Number of feature selected by different feature selection models

Table 2 Comparison of feature selection models

	Accuracy	Precision	Recall	F-measure	Error rate
WDS	0.816	0.829	0.819	0.824	0.0193
ACO	0.835	0.831	0.824	0.827	0.0172
FFA	0.861	0.852	0.859	0.855	0.0167
CSFLA	0.973	0.986	0.981	0.983	0.0028

selection models. The ACO selects 8 attributes, FFA selects 6 attributes, and the proposed CSFLA selects 4 attributes as the potential for AD detection using ADNI dataset. WDS means Whole Dataset, which has 10 attributes.

Table 2 illustrates the performance comparison of three different metaheuristic feature subset selection algorithms namely ant colony optimization (ACO), firefly algorithm (FFA), and chaotic shuffled frog leaping algorithm (CSFLA). Along with the whole features of dataset, the results explore that the performance of deep neural network using chaotic shuffled frog leaping algorithm has produced optimized performance compared to other models as it has the ability for balancing both global and local optimization while performing feature selection for Alzheimer's disease prediction.

Figure 6 depicts the performance evaluation based on accuracy while utilizing the entire dataset and three alternative feature selection models for Alzheimer's disease prediction using deep neural network. While comparing ACO and FFA, the CSFLA produced highest accuracy rate of 0.973, whereas ACO and FFA obtain 0.835 and 0.861, respectively. This is due to the proposed CSFLA's well-balanced local and global searching for key features, which entails enhancing the accuracy of DNN for AD prediction (Fig. 7).

The performance comparison based on precision rate is obtained by three different feature subsets along with the whole dataset for AD prediction which is shown in

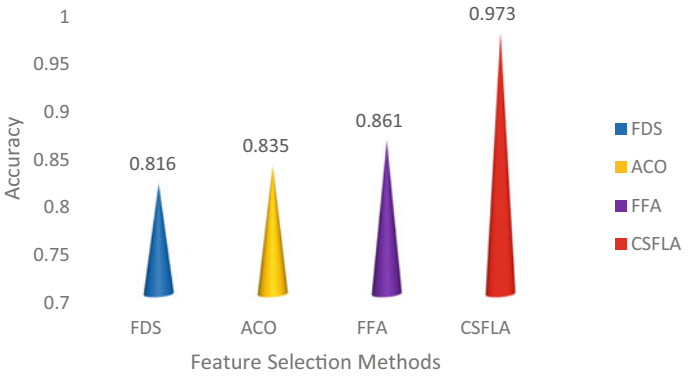


Fig. 6 Evaluation of performance based on accuracy

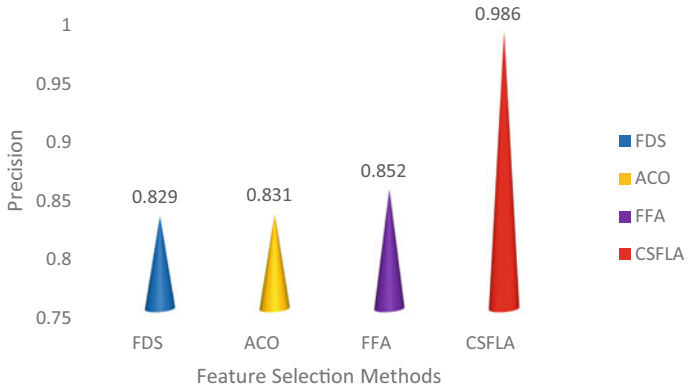


Fig. 7 Evaluation of performance based on precision

the figure. The chaotic mapping optimizes the global search of shuffled frog leaping algorithm in performing significant feature selection. Using that feature subset, deep neural network produces highest precision rate of 0.986 while comparing with ACO, FFA, and whole dataset (FDS). This is because the standard models result in early convergence due to local optima, while performing feature selection.

Figure 8 shows the recall rate produced by DNN for performing AD prediction using whole dataset and feature subsets generated by ACO, FFA, and proposed CSFLA. The recall rate of proposed CSFLA is 0.981%, FFA is 0.859%, ACO is 0.824%, and the whole dataset is 0.819%. The global exploration of CSFLA improves the global optimization challenge, and local search is also optimized by periodically shuffling virtual frogs for discovering potential features in ADNI dataset. Thus, the proposed CSFLA achieves better recall rate compared to other feature selection models.

Figure 9 illustrates the output of harmonic mean which is termed as F-measure to examine the performance of three different feature selection models along with complete dataset-based AD prediction using deep neural network. The F-measure

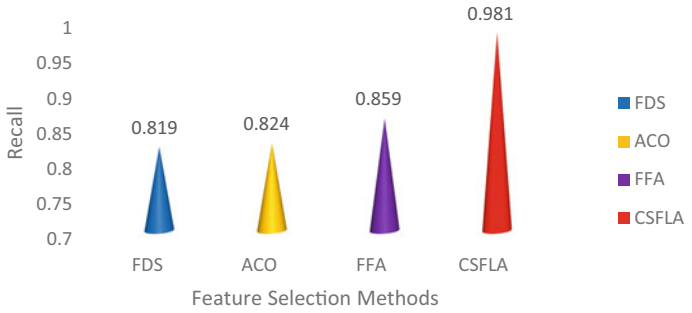


Fig. 8 Evaluation of performance based on recall

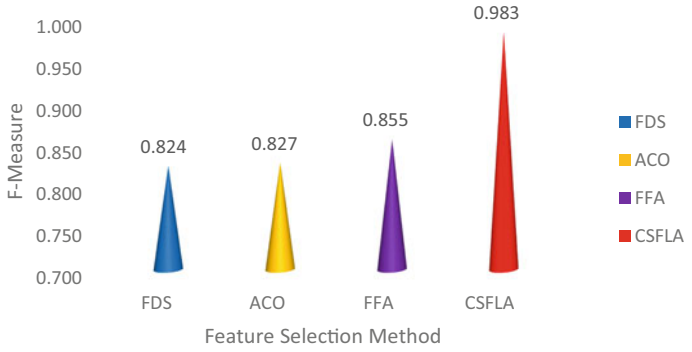


Fig. 9 Evaluation of performance based on F-measure

value of proposed CSFLA is 0.0983%, which is the highest comparative dataset and the feature selection is done by using ACO and FFA. Since this virtual frogs utilize chaotic mapping to search the optimal feature subset without compromising early convergence, the deep neural network outperforms the CSFLA-selected features in Alzheimer's prediction.

Figure 10 illustrates the performance of DNN using complete dataset and three different feature subset selection models-based error rate. The proposed CSFLA has produced least error rate compared to complete dataset, ACO and FFA. The reason is CSFLA with chaotic theory and shuffling of virtual frogs periodically improves the process of feature selection, and it balances local and global searching which results in reduced error rate for DNN in AD prediction.

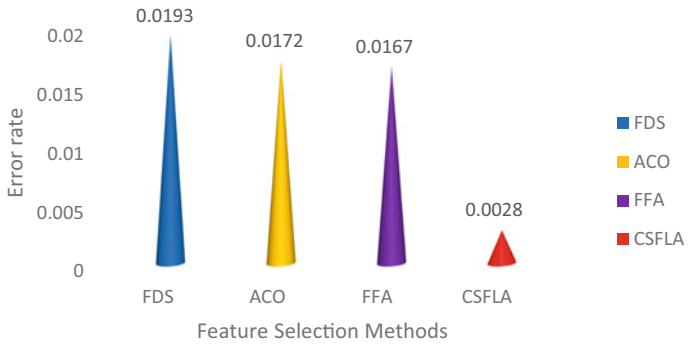


Fig. 10 Evaluation of performance based on error rate

5 Conclusion

The optimized feature subset selection-based Alzheimer's disease prediction is the primary objective of this research work. In this work, ADNI dataset comprises of both clinical and demographic details of the subjects, which are used for performing memetic and metaheuristic-based feature selection. The proposed chaotic shuffled frog leaping algorithm [CSFLA] uses the food searching behavior of the frogs to search the optimized feature subset, which will greatly influence the accuracy of prediction model. With its chaotic mapping capabilities, the CSFLA selects the most promising candidates from the population to participate in the feature selection process. Its leaping behavior, fitness value of each frog, and shuffling approach balance both local and global searches, producing the best potential feature subset when compared to ant colony optimization and firefly algorithm. Since typical metaheuristic models suffer from local optima and unpredictability in agent selection, deep neural network is used to assess the resultant reduced feature subset from these three feature selection models. According to the simulation findings, the feature subset created by CSFLA has produced the most promising accuracy in Alzheimer's disease identification.

References

1. G. Abate, M. Marziano, W. Rungratanawanich, M. Memo, D. Uberti, in *Nutrition and ageing: focusing on Alzheimer's disease*. (Oxid Med Cell Longev, 2017)
2. X. Long, L. Chen, C. Jiang, L. Zhang, Prediction and classification of Alzheimer disease based on quantification of MRI deformation. PLoS ONE **12**, 1–19 (2017)
3. D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage **55**, 856–867 (2011)
4. Y. Chtioui, D. Bertrand, D. Barba, Feature selection by a genetic algorithm application to seed discrimination by artificial vision. J. Sci. Food Agricul. **76**, 76–86 (1998)
5. Y. Zhang, D. Gong, Y. Hu, W. Zhang, Feature selection algorithm based on bare bones particle swarm optimization. Neurocomputing **148**, 150–157 (2015)
6. A.B. Rabeh, F. Benzarti, H. Amiri, Diagnosis of Alzheimer diseases in early step using SVM (Support Vector Machine). in *13th International Conference Computer Graphic, Imaging and Vision*, CGVI (2016) pp. 1–5
7. F. Razavi, M.J. Tarokh, M. Alborzi, An intelligent Alzheimer's disease diagnosis method using unsupervised feature learning. J. Big Data **6**, 32 (2019)
8. J. Albright, Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm. Alzheimer's Dementia: Transl. Res. Clinical Intervent. **5**, 483–491 (2019)
9. T.T. Erguzel, S. Ozekes, S. Gultekin, N. Tarhan, Ant colony optimization based feature selection method for QEEG data classification. Psychiatry Investig. **11**(3), 243–250 (2014)
10. Z. Li, K. Mistry, C. Lim, S. Neoh, Feature selection using firefly optimization for classification and regression models. Decision Support Syst. **106** (2017)
11. K. Zhou, W. He, Y. Xu, G. Xiong, J. Cai, Feature selection and transfer learning for Alzheimer's disease clinical diagnosis, MDPI. Appl. Sci. **8**, 1372 (2018)
12. http://adni.loni.usc.edu/

13. M. Eusuff, K. Lansey, F. Pasha, Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization. Eng. Optim. **38**(2), 129 (2006)
14. B. Alatas, E. Akin, A.B. Ozer, Chaos embedded particle swarm optimization algorithms. Chaos Solitons Fractals **40**, 1715–1734 (2009)
15. S. Christian, T. Alexander, E. Dumitru, Deep neural networks for object detection. Adv. Neural Info. Process. Syst. 2553–2561 (2013)
16. C. Dhanusha, A.V. Senthil Kumar, ıntelligent ıntuitionistic fuzzy with elephant swarm behaviour based rule pruning for early detection of alzheimer in heterogeneous multidomain datasets. Int. J. Recent Technol. Eng. (IJRTE) **8**(4), pp. 9291–9298 (2019). ISSN: 2277–3878
17. C. Dhanusha, A.V. Senthil Kumar, Enriched neutrosophic clustering with knowledge of chaotic crow search algorithm for alzheimer detection in diverse multidomain environment. Int. J. Scientif. Technol. Res. (IJSTR) **9**(4), 474–481 (2020). ISSN:2277-8616
18. C. Dhanusha, A.V. Senthil Kumar, I.B. Musirin, Boosted model of LSTM-RNN for alzheimer disease prediction at their early stages. Int. J. Adv. Sci. Technol. **29**(3), 14097–14108 (2020)
19. C. Dhanusha, A.V. Senthil Kumar, Deep recurrent Q reinforcement learning model to predict the alzheimer disease using smart home sensor data. in *International Conference on Computer Vision, High Performance Computing, Smart Devices and Network, IOP Conference Series: Materials Science and Engineering*, (CHSN 2020) 28th–29th December, Kakinada, India 2021, vol 1074. (2020) pp. 012014
20. D. Sayantan, A. Banerjee, Highly precise modified blue whale method framed by blending bat and local search algorithm for the optimality of image fusion algorithm. J. Soft Comput. Paradigm(JSCP) **2**(04), 195–208 (2020)
21. J.S. Manoharan, Population based metaheuristics algorithm for performance improvement of feed forward neural network. J. Soft Computing Paradigm **2**(1), 36–46 (2020)
22. G. Ganesan @ Subramanian, I.A. Chidambaram, J.S. Manoharan, Flower pollination algorithm based decentralized load-frequency controller for a two-area interconnected restructured power system considering gas turbine unit. Middle East J. Scient. Res. **25**(6), 1308–1314 (2017)

Smart Traffic, Ambulance Clearance, and Stolen Vehicle Detection

S. V. Viraktamath, Arfaali B. Naikar, Sharan N. Nargund, and Manjunath Gayakawad

Abstract The proposed system is to develop a density-based dynamic traffic system where certain signal timing varies intelligently depending on the quantity of traffic at each intersection. Because traffic congestion could be a big problem in most cities across the globe, it is time to change from a manual or fixed-timed system to one that may make decisions. The present traffic signal system operates on a fixed schedule, which can be wasteful if one lane is active while the others do not seem to be. A framework for an intelligent system has been designed to deal with the difficulty. Because higher traffic density on one side of the junction frequently needs additional green signal timing than the standard given time, a system is proposed wherein the fundamental measure of green and red light is assigned in support of the density of traffic existing at that time. Proximity infrared sensors (PIR) are commonly employed. Once the density has been determined, the microcontroller is employed to allocate the green light's glowing duration. The sensors on the road's edges detect the presence of automobiles and communicate this information to the microcontroller, which determines how long a lane will remain open or when the signal light's timing is adjusted. The framework's method was developed in later stages. This framework is employed to implement traffic congestion control, ambulance clearance, and stolen vehicle detection.

1 Introduction

India is the world's second-most populated country, with a rapidly expanding economy. Its cities are experiencing severe traffic congestion. Due to space and economic limits, infrastructure expansion is slower than that of the number of vehicles [1]. It requires traffic control systems, and efficient traffic flow management can

S. V. Viraktamath · A. B. Naikar (✉) · S. N. Nargund
Department of Electronics and Communication Engineering, SDMCET, Dharwad, Karnataka, India

M. Gayakawad
WORMBOTICS Technologies Simplified LLP, Bagalkot, Karnataka, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_52

help to mitigate the detrimental effects of traffic congestion [2]. The difficulty with the traffic system is that the automobiles on the four-way road will be heavy every minute and the traffic lights will be switched to each side for a set amount of time. Even if there are no automobiles on that side, the traffic signals will illuminate for a set period. As a result, on other lanes the automobiles must wait for the process to be completed. So, to prevent the waste of time, this system manages traffic depending on the heavy flow of automobiles on any given side. The density of cars on either side of the intersection gives the path to the side with the most traffic and maintains the stop position.

The main aim is to pass vehicles smoothly and to avoid wasting time. There is less need for human intervention to manage traffic. Traffic congestion can be controlled automatically and saves plenty of time. It is possible to identify stolen vehicle. On-time arrival of emergency vehicles could be a must. Wireless networks are increasingly popular in road transportation in recent years, thanks to their lower costs [3]. The proposed system mainly focuses on these concerned problems with the standard control system management, and this method is predicated on sensor networks and embedded systems (microcontroller 89C51). The issues faced by the manual working of traffic light management systems are resolved efficiently.

The system's design goal is to resolve the matter of traffic congestion while also minimizing violations of traffic laws. This is often concerned with the look of a secure and effective traffic flow system that aids within the reduction of road congestion. Dynamic flexibility in green light timing supported density at each lane appears to be favorable in terms of minimizing signal wait times and fuel consumption. Technologies like the amplitude-shift keying (ASK) transceiver and radio frequency identification (RFID) are samples of technologies that will be employed in control to supply cost-effective solutions. RFID is a wireless technology that sends data between an RFID tag and an RFID reader using radio-frequency radiation. Some RFID devices can only operate within some inches or centimeters, but others can operate at distances of up to 100 m (300 feet). During this case, if an ambulance approaches a traffic intersection, the traffic signals automatically halt and supply a green light to the ambulance. This program's main purpose is to ensure that emergency vehicles, like ambulances, get to hospitals on time, eliminating delays caused by traffic congestion. Each vehicle is provided with an RFID tag. When it comes into contact with the RFID reader, it will send a symbol to that. It compares to a database of stolen RFID tags. If the match is detected, the buzzer will sound, and therefore, the light will turn red. As a consequence, the vehicle is compelled to stop within the traffic intersection, allowing local authorities to take appropriate action.

2 Literature Survey

Traffic congestion is a serious problem in developing countries due to growth in the city population and the rising number of vehicles in the cities [4]. Traffic congestion on roads leads to passive traffic, which lengthens travel times and is thus one of the

most serious challenges in urban areas. The green signal system was used to grant permission to any emergency vehicle by converting all red lights in the emergency vehicle's path to green, giving the requested vehicle a complete green wave. Green waves have the drawback that when they are disrupted, they might cause traffic issues, which are compounded by the synchronization. In such instances, the green wave's line of cars builds until it becomes too big and some cars are unable to reach the green lights in time and must halt. This is referred to as "over-saturation" [5]. The use of RFID traffic control is to circumvent difficulties that generally come with regular traffic control systems, notably those relating to image processing and beam interruption techniques as described [6]. This RFID technology is used in locations with many vehicles, multiple lanes, and many traffic junctions. The work's main flaw is that it does not go into detail on the communication channels employed linking the emergency vehicle and the traffic signal controller [7]; the goal of this work is to shorten the time it takes for an ambulance to get to the hospital by clearing the lane automatically. At traffic intersections, the technology is automated and does not require human interaction. The downside of this method is that it requires full information on the start point and destination of the journey. It would not operate if the ambulance has to take a special route for a few reasons or if the start place is unknown. In Bangalore, a video traffic surveillance and monitoring system have just been installed. The traffic management crew must manually analyze data to calculate the length of traffic lights at each intersection. It will inform local police officials of the case to require the suitable action [6]. In some models an outside parking lot to work out the systems effective functioning on the terms of detection accuracy, false positive and also the false negative, true positive and therefore the true negative rates together with the typical speed of the in identifying the parking slot [8]. The terminals have certain limitations that crib their capability of processing cross-application and diversified services. Within the past, researchers have experimented with several technologies [9]. The advancements within the technologies associated with the wireless communication systems have made the vehicular networks a prominent area of research within the industry [10]. The following may be a quick overview of some traffic congestion remedies.

2.1 Embedded System Affiliations

Rotake and Karmore proposed "Intelligent Traffic Signal Control System" [1]. This system employs an AVR 32 microcontroller with configurable non-volatile storage, integrated 8-channel digitizer, and IR sensors. The IR sensors detect the presence of an emergency vehicle, and therefore, the microcontroller is programmed to send a red signal to any or all sides except the one where the emergency vehicle is parked.

2.2 Wireless Sensor Networks

"Road Traffic Congestion Monitoring and Measurement Using Active RFID and GSM Technology" by Mandal et al. [4] is mentioned in this methodology. Active RFID tags, a wireless coordinator, a wireless router, GSM modems, and monitoring station software are all employed in this system. The wireless devices are installed on each side of the road and collect data from active RFID tags. The monitoring station will gather all of the data through GSM and reply to the appropriate traffic signal.

2.3 Image Processing

Image processing can also help to alleviate traffic congestion difficulties to a larger extent. This is stated in Dangi et al. "Image Processing Based Intelligent Traffic Controller" [11]. A camera is utilized in this approach, and it is mounted on poles or another tall structure so that it can cover the whole traffic scene. The images retrieved from the videos are then evaluated and utilized to detect and count the vehicles, and the time allocated for each side is determined by the density.

3 Block Diagram

In this, the IR Tx and IR Rx sense the traffic density, three sensors kept beside a road, all IR Tx and Rx are connected to a microcontroller input port, the microcontroller receives signal from three input ports. Depending on the input port, the signal controller will give the signal to the output port to activate the driver circuit to control traffic lights and delay is provided by the microcontroller depending on the input port. With a carrier frequency of 433 MHz, ASK sends the bits in serial mode. This is kept in the ambulance. The encoder converts the parallel data into serial data and feeds it to the ASK transmitter for serial transmission. The digital data broadcast by the ASK Tx is received by the ASK receiver and routed to the decoder. The ASK receiver is connected to the traffic signal. The decoder converts the serial data into parallel data and feeds it to the double-pole double-throw (DPDT) relay to which LEDs are connected and the automatic signal will change; at the same time, vehicles will move automatically to give the path to the ambulance. The microcontroller is connected with a crystal oscillator along with a power-on reset circuit for the working of the microcontroller. The speed of execution of the microcontroller depends on the crystal connected to the microcontroller (Fig. 1).

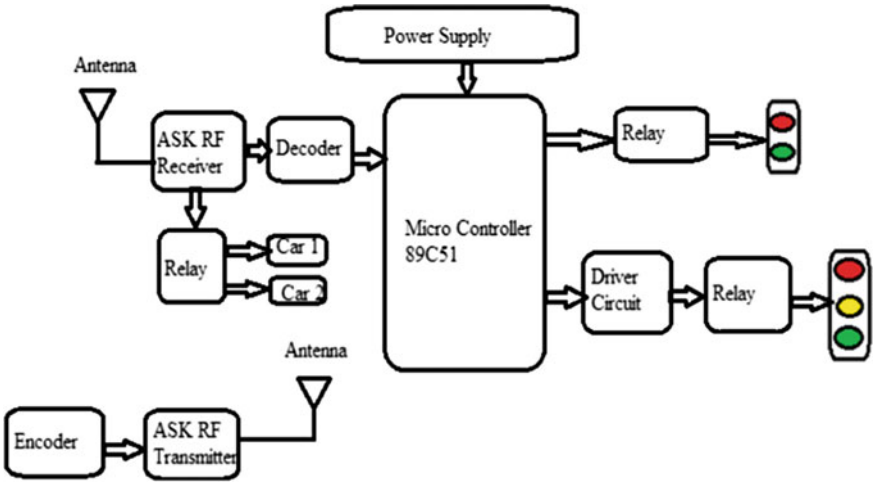


Fig. 1 Block diagram

4 Proposed Work

Existing technologies are insufficient to manage problems such as traffic control, emergency vehicle clearing, stolen vehicle identification, and so on as shown in the current problem section. To solve these problems, the proposal is to implement the Intelligent Traffic Control System.

This model contains congestion control, ambulance clearance, and stolen vehicle detection (Fig. 2).

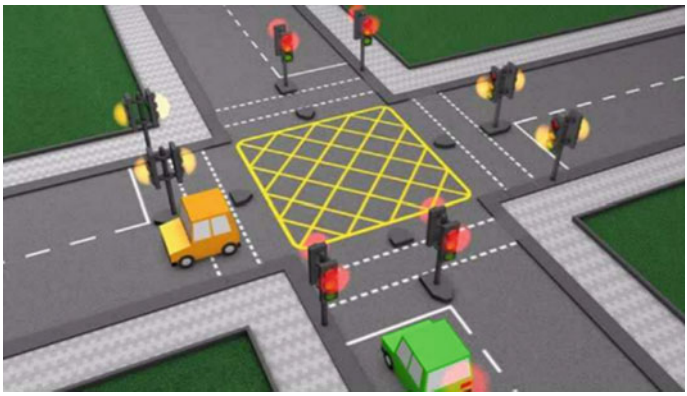


Fig. 2 Traffic signal junction [3]

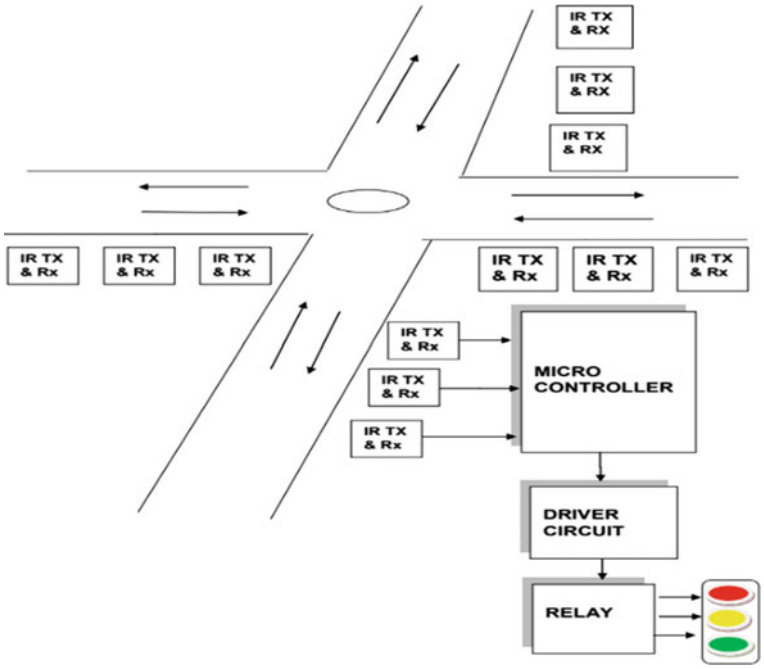


Fig. 3 Block diagram of congestion control [3]

4.1 Congestion Control

The IR transmitter (Tx) and IR receiver (Rx) are used to sense the traffic density, three sensors are placed on the roadside, and all IR Tx and Rx are connected to the microcontroller input port. There are a total of six proximity sensors which are used to detect traffic density, three sensors output directly fed to the input port of the microcontroller and three proximity sensors are connected in parallel to monitor traffic density of road. The microcontroller will get continuous input signal from the sensors and give output to 74LS245 IC and amplify signal level which is sufficient to drive LEDs at the output, to make red, green, and yellow LED ON/OFF depending on the input signal; the traffic signal is controlled using 5 V DC power which is sufficient to run the microcontroller (Figs. 3 and 4).

4.2 Ambulance Clearance

There are two pieces to the module: the first is the ASK transmitter, which is installed in the emergency vehicle. When the switch is pushed, it will broadcast the signal. An HT12E encoder is included in the transmitter. The commands and data are sent to the

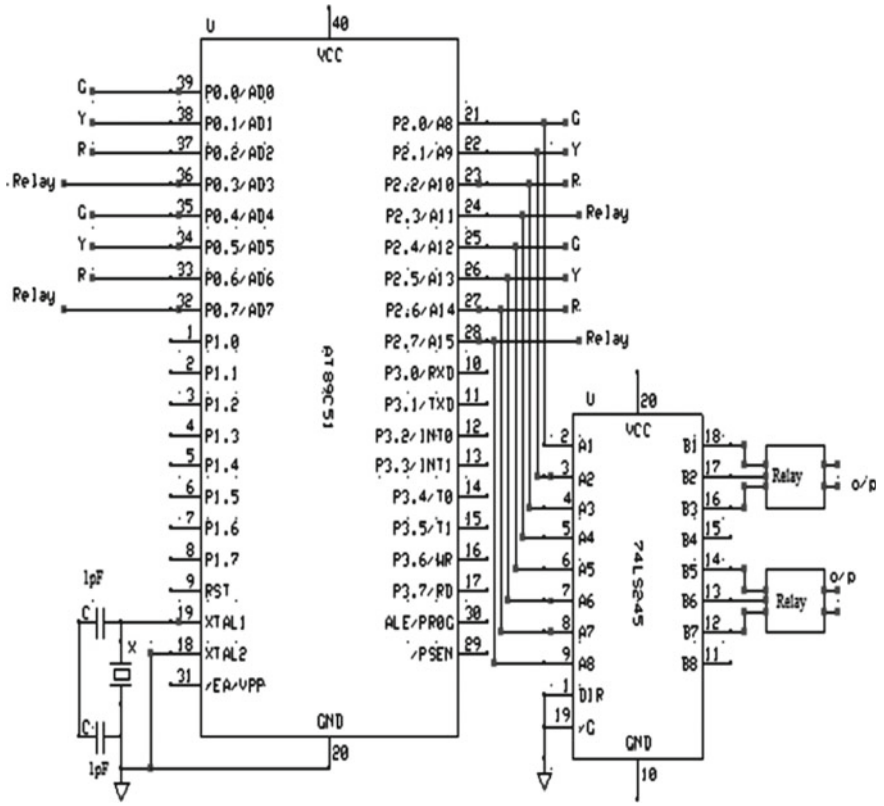


Fig. 4 Schematic of congestion control [3]

ASK Rx through a serial connection from the microcontroller. The receiver, which is mounted on the traffic pole, is the second component. The ATmega 16 microcontroller and the HT12D decoder are also included. The security code received by the receiver is compared to the serial bits stored in its database. The green light will turn ON if it is a match.

4.2.1 ASK RF Transmitter

ASK transmits the bits in serial mode with a carrier frequency of 433 MHz. This is kept in the ambulance. The transmission of RF waves happens at a particular frequency from an allowable set of frequencies (bandwidth) depending on the values of the crystal oscillator and air-core inductor/capacitor used in the circuit of the transmitter. Because 12-bit data is communicated, the first eight bits, known as address bits, must match on both the sending and receiving sides. As a result, these address bits serve as a kind of password until they match, at which point no data is allowed.

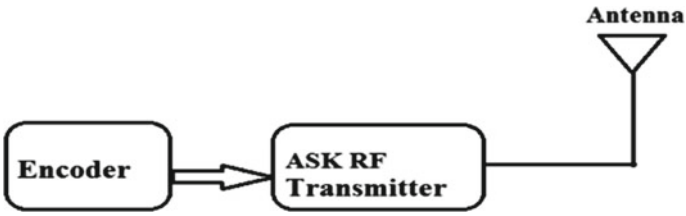


Fig. 5 ASK RF transmitter [12]

Assigning a common clock to both the transmitter and receiver circuits and changing their address bits in a predetermined sequence, or even using a sequence generator, is the safest and most high-level implementation. The data out pin of the HT12E encoder IC is received by an ASK-type transmitter set to 433 MHz; there are 8 number address lines and several data lines are present in HT12E. The data on the D0-D3 pin is transmitted serially and available at the data O/P pin. An RF transmitter of 433 MHz has a pin of data that is now connected to the serial data pin from HT12E. The respective Vcc (5 V) and ground pins are connected. A telescopic antenna has an antenna termination attached to it. The D0-D3 bits are broadcast serially and the 433 MHz transmitter transmits these bits in the form of RF waves through a transmitting antenna if the I/P data for the HT12E is 0001 (Binary) (Fig. 5).

The complete circuit of the transmitter encoder is shown in Fig. 6 . Here, the receiver address to be transmitted is set to 00 h (BCD), by connecting all the address line A0-A7 (pins from 1–7) to GND. The timing resistor 1 M is connected across OSC1 (pin 15) and OSC2 (pin 16), which produces a frequency of 3 kHz. The TE pin (14) is permanently connected to the ground to transmit the data continuously. Finally, the address of the receiver and the data to be sent are serially transmitted through the pin "DOUT" (pin 17) and then transmitted to the "DATA IN" (pin of the Tx433). In the RF transmitter module, the transmitter uses the ASK modulation algorithm to modify the digital data coming in the line "DATA IN" and sends it into the air through the antenna. Alternatively, the HT12E converts parallel inputs to serial outputs. It translates 12-bit parallel data to serial for radio frequency transmission. The 12 bits are divided into 8 address bits and 4 data bits.

4.2.2 ASK RF Receiver

The ASK receiver receives the digital data transmitted from the ASK Tx and is fed to the decoder. The ASK receiver is connected to the traffic signal. The data broadcast by the ASK RF transmitter is received by the ASK RF receiver. The HT12D decoder converts serial data into 4-bit parallel data D0–D3. For data transmission, the condition of these address pins A0-A7 should match the condition of the address pin in the HT12E at the transmitter. When successful data transfer happens from transmitter to receiver, the LED attached to the above circuit flashes. The 51 K

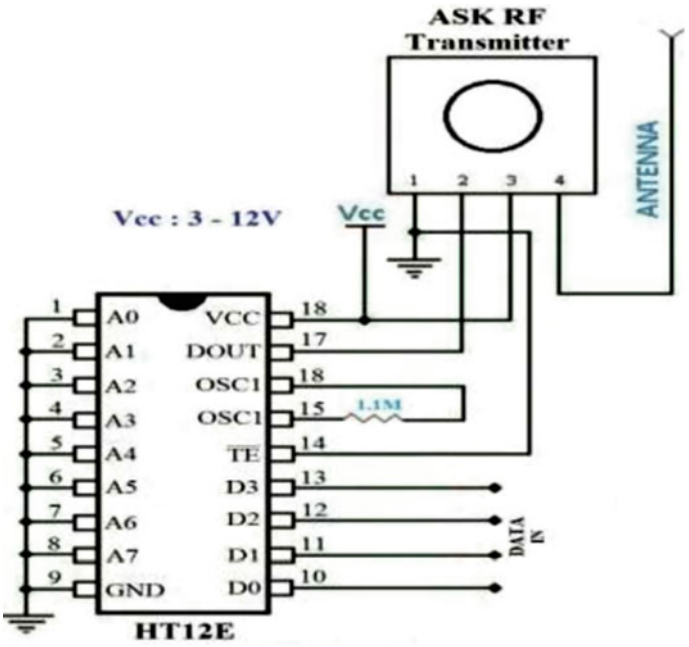


Fig. 6 Schematic ASK RF transmitter [12]

resistor will supply the appropriate resistance for the HT12D's internal oscillator (Figs. 7 and 8).

HT12D decoder decodes the serial data to parallel, and it is available at D1-D4 data bits; the oscillating resistor of 51 K is connected to its oscillating pins (pin15 pin16). Pin 17 is considered as going high when transmitter starts generating and sending RF wave, and the pin 17 is connected to a transistor and an LED which

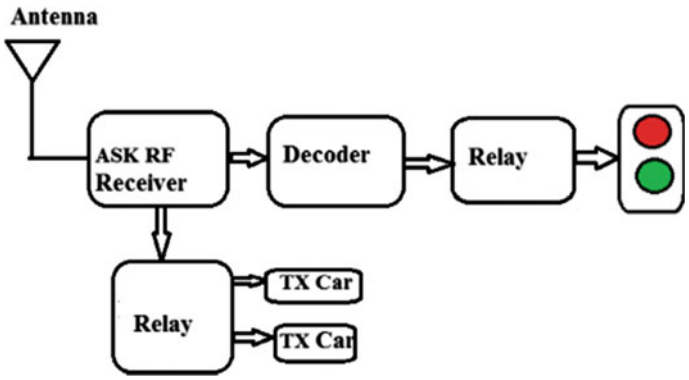


Fig. 7 ASK RF receiver [3]

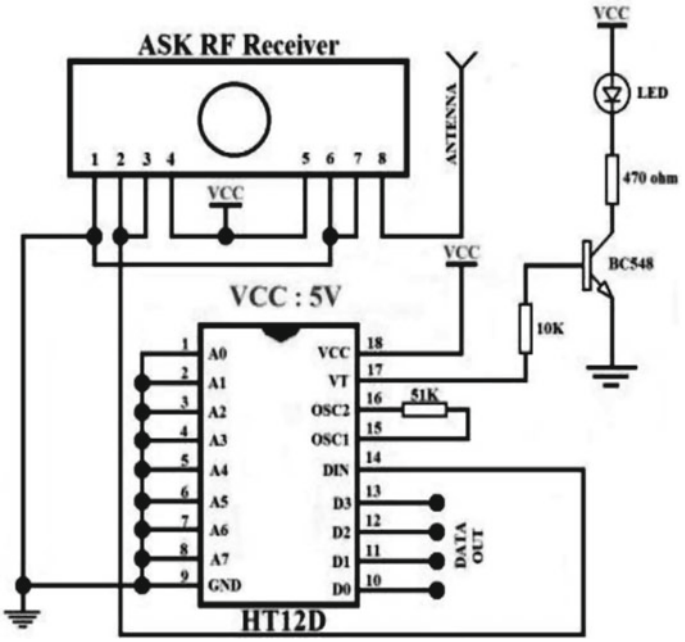


Fig. 8 Schematic of ASK RF receiver [12]

glows and indicates the link between transmitter and receiver. If a data bit of 1010 is applied to the D0-D3 of HT12E, then the data is transmitted serially. The frequency of 433 MHz is transmitted through the transmitter and the receiver receives the data bits and HT12D decodes the serial data bits to parallel and the same data, i.e., 1010 is available at D1-D4 pins of HT12D. The decoder outputs are connected to the microcontroller.

4.3 Stolen Vehicle Detection

Non-contact technologies that employ radio waves to automatically identify persons or things are referred to as radio frequency identification (RFID). The most frequent way of identifying is to record a "unique serial number" that identifies a person or thing on a microchip that is connected to an antenna. An "RFID transponder" or "RFID tag" is a combination antenna and microchip that works in conjunction with an "RFID reader" (sometimes called an "RFID interrogator").

In Fig. 9, "RFID reader" is used to detect the stolen vehicle. The "RFID tag" will be attached to the vehicle, and the reader is kept near the traffic signal. The stolen vehicle number will be stored in the microcontroller as a blacklist. When the vehicle is near the traffic signal, the "RFID reader" transmits the signal and the "RFID reader"

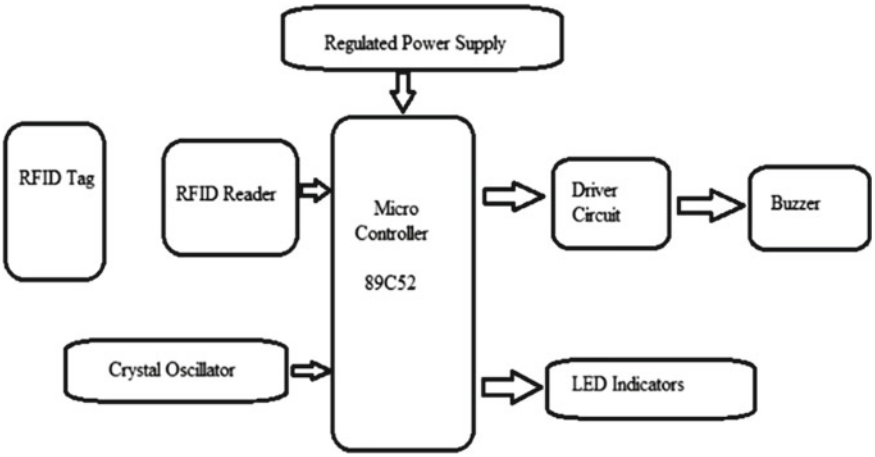


Fig. 9 Stolen vehicle detection [13]

reads the "RFID number" and feeds it to the microcontroller. Microcontroller verifies the number; if it is in blacklist, then it will give output signal to the driver circuit to make buzzer ON for indication and then red led is switched ON (Fig. 10).

5 Results and Discussion

5.1 Traffic Lane

Figure 11 represents the traffic light control which is an automated way of controlling signals by the density of traffic on the roads. At set distances from the junction signal, IR sensors are deployed throughout the whole intersecting route. The traffic signal's timing delay is determined by the number of cars on the route. IR sensors are used to measure traffic density [14]. The microcontroller alters the glow time of the green LED of the appropriate junction to control the traffic light effectively.

5.2 Ambulance Clearance

Figure 12 represents the transmitter module placed on the ambulance. It consists of an ASK RF transmitter module of 433 MHz and encoder HT12E. When the ambulance is entered into the junction, the switch placed in the transmitter is pressed.

Figure 13 represents the RF receiver used to receive the 433 MHz RF signal that has been sent by the RF transmitter in the ambulance. When the RF receiver receives

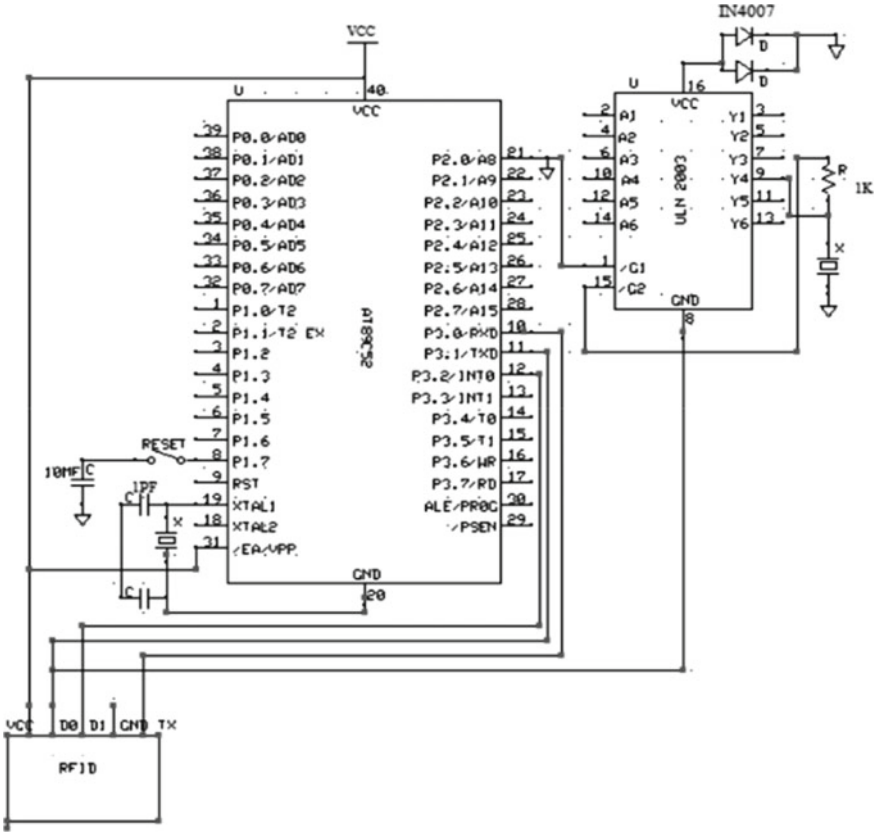


Fig. 10 Schematic of stolen vehicle detection [13]

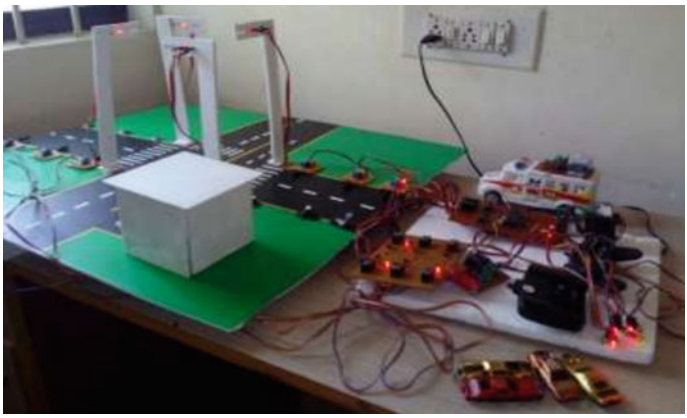


Fig. 11 Initial condition at traffic junction

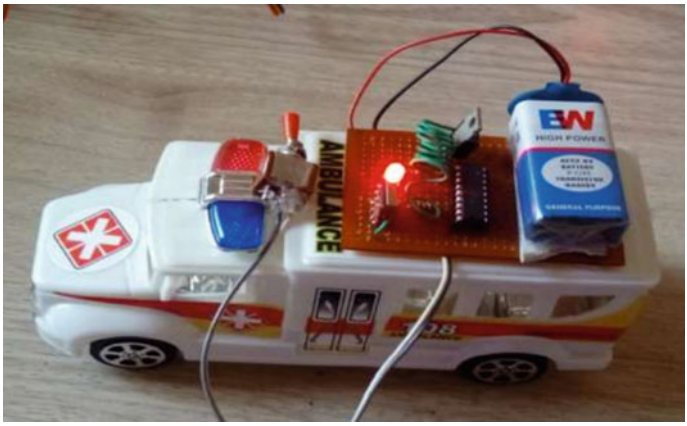


Fig. 12 Ambulance transmitter part

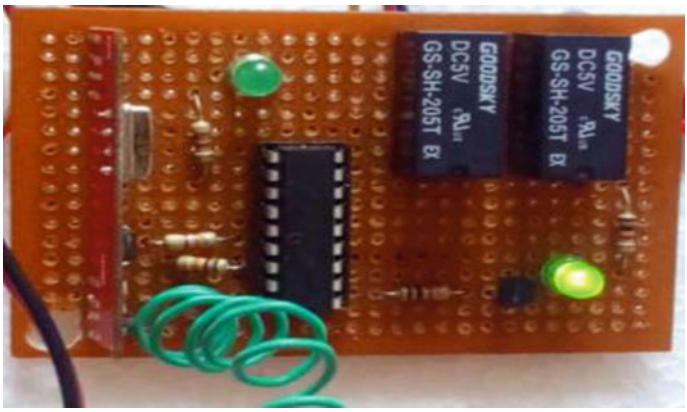


Fig. 13 Receiver part placed at traffic pole

the RF signal, then it will decode the signal and send it to the microcontroller which will alert the traffic unit to make way for the ambulance.

Figure 14 shows the emergency vehicle while crossing the intersection. As said, it consists of an ASK RF transmitter which includes an HT12E encoder and the receiver which is preset at pole has an HT12D decoder. If the ASK RF transmitter is in the range of the ASK RF receiver, then the traffic light will change to green till the receiver receives the signal.

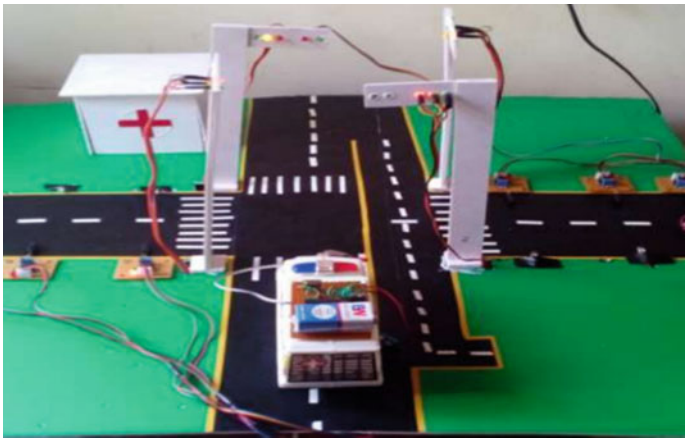


Fig. 14 Traffic signal changes during presence of ambulance

5.3 Stolen Vehicle Detection

In Fig. 15, for testing purposes, the system will compare the unique "RFID tag" read by the "RFID reader" to the stolen "RFIDs" stored in the system. Each vehicle is equipped with an "RFID tag." It will transmit a signal to the "RFID reader" then it is scanned, verified, and checks the status in the database to see if it is stolen or not. If a match is identified, then the traffic light is instantly turned red for 30 s. The buzzer will be activated if the vehicle is stolen.

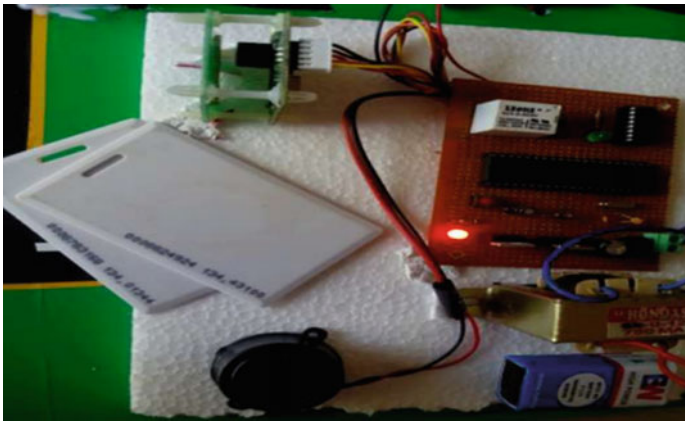


Fig. 15 Stolen vehicle detection

6 Conclusion and Future Work

A traffic light system had been conceived and implemented with correct hardware and software integration. The microcontroller was connected to the infrared sensors. This interface is in sync with the entire traffic system operation. Knowledge of sequential systems, electrical and electronics applications was demonstrated here. It can be programmed in any way to regulate the traffic light model automatically, and it will be valuable for road system design. This module's design and execution are specifically geared at traffic management so that emergency vehicles such as ambulances and fire engines on the road have a clear path to their destination in less time and with no human interference. When there is heavy traffic on the road, this mechanism will assist traffic cops in allowing the ambulance to pass. The buzzer will be activated when a stolen vehicle is detected, allowing the police officer to take proper action. The prototype may be improved further by testing it with longer-range "RFID scanners." Additionally, GPS may be integrated into the stolen car detection module, allowing for the precise position of the stolen car to be determined. Currently, it is using a method that only considers one of the traffic junction's roads. It can be made better by extending it to all of the roads in a multiroad intersection.

Acknowledgements The authors wish to thank K-TECH INNOVATION CENTER regional office SDMCET, Dharwad for their support.

References

1. D. Rotake, S. Karmore, Intelligent traffic signal control system using embedded system. Innov. Syst. Design Eng. **3**(5) (2012). ISSN 2222–1727 (paper) ISSN 2222–2871 (online),
2. Traffic Congestion in Bangalore-A Rising Concern, http://www.commonfloor.com/guide/traffic-congestion-in-bangalore-a-rising-concern-27238.html
3. K.R. Shruthi, K. Vinodha, Priority based traffic lights controller using wireless sensor networks. Int. J. Electron. Signals Syst. (IJESS) **1**(4) (2012) ISSN 2231-5969
4. B.P. Gokulan, D. Srinivasan, Distributed geometric Fuzzy multi-agent urban traffic signal control. IEEE Trans. Intell. Transp. Syst. **11**(3), 714–727 (2010)
5. A.K.R. Mittal, D. Bhandari, A novel approach to implement green wave system and detection of stolen vehicles. in *Proceedings IEEE Conference on Advance Computing*, (2013), pp. 1055–1059
6. S. Sharma, A. Pithora, G. Gupta, M. Goel, M. Sinha, Traffic light priority control for emergency vehicle using RFID. Int. J. Innov. Eng. Technol. **2**(2), 363–366 (2013)
7. R. Hegde, R.R. Sali, M.S. Indira, RFID and GPS based automatic lane clearance system for ambulance. Int. J. Adv. Electr. Electron. Eng. **2**(3), 102–107 (2013)
8. K. Khaled, S. Smys, A. Bashar, Tenancy status identification of parking slots using mobile net binary classifier. J. Artif. Intell. **2**(03), 146–154 (2020)
9. S. Shakya, Collaboration of smart city services with appropriate resource management and privacy protection. J. Ubiquitous Comput. Commun. Technol. (UCCT) **3**(01), 43–51 (2021)
10. R. Bestak, Intelligent traffic control device model using Ad Hoc network. J. Inf. Technol. **1**(02), 68–76 (2019)

11. M. Abdoos, N. Mozayani, A.L.C. Bazzan, Traffic light control in non-stationary environments based on multi agent Q-learning. in *Proceedings IEEE Conference on Intelligent Transportation Systems*, (2011), pp. 580–1585
12. K. Mandal, A. Sen, A. Chakraborty, S. Roy, Road traffic congestion monitoring and measurement using active RFID and GSM technology. in *IEEE | Annual Conference on Intelligent Transportation Systems* (2011)
13. V. Dangi, A. Parab, K. Pawar, S.S. Rathod, Image processing based intelligent traffic controller. Undergrad. Acad. Res. J. (UARJ) **1**(1), (2012) ISSN 2278–1129
14. B. Prashanth Kumar, B. Karthik, Microcontroller based traffic light controller. (Department of Electrical and Electronics Engineering Gokaraju Rangaraju Institute of Engineering and Technology, 2011)
15. S. Kumar, P.G. Abhilash, D.G. Jyothi, G. Varaprasad, Violation detection method for vehicular Ad Hoc networking. ACM/Wiley Security and Communication Networks (2014)https://doi.org/10.1002/sec.427

Competitive Analysis of Web Development Frameworks

Priyanka Jaiswal and Sumit Heliwal

Abstract In day-to-day working, when we are dealing with collecting and managing information for business, designing of interface plays a very important role. For designing an interface, different web designing frameworks are required. This review paper is based on discussion and comparison of different frameworks which are currently used in the industry and are in demand in the industry. It will help one to choose between the different frameworks and get an idea about their uses and where to use them. Anyone can understand and can choose between different frameworks. It will help newbies to get started and make choices about the framework. This review paper focuses on the pros and cons and their requirements for deployment. This paper is also focused on the time requirement, access, what they provide, where they are used, when to use, and how to choose upon it. It also includes the starting requirements and what goals user can achieve using different frameworks. It also includes the basic structure of a Web application or the basic requirements of a web application to start. The comparative study broadly consist of Bootstrap, Django, and .NetBootstrap, Django, Net and other web development framework.

Keywords Web development · Bootstrap · Django · .net · Python · C# · HTML · CSS

1 Introduction

This section describes the need, use, and application of different Web designing framework. The subsection of introduction is based on the discussion of front-end frameworks and full stack framework. Nowadays, the technology evolves rapidly and has numerous technologies for Web development and today's industry is Web-driven which is used to attract customers and gives an idea of our company or product or services which a consumer is wishing to buy or subscribe to. This paper

P. Jaiswal (✉) · S. Heliwal
Department of IT, Yeshwantrao Chavan College of Engineering (YCCE), Nagpur, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_53

is discussing the different types of frameworks and their examples where we can use them according to our needs. Here are many options to choose from.

To run any business smoothly, good Web designing is very important, as per statistical survey, it is observed that 57% Internet users refuse to use or recommend that business application because of poor designing. As compare to designing desktop Website, it is equally important to give emphasis on mobile Website. It is very important to crease a responsive Website which can be applicable to all types of devices including mobile and desktop.

1.1 Front-End Frameworks

Front-end frameworks are used to design the front-end of a Website for the attraction of users. It is a client-side part of the application. It is made up of HTML, CSS, and JavaScript. It is composed of different colors, animations, images, icons, and fonts. It is responsive according to the screen size and with the help of media queries. It is used for building static Websites.

E.g.- Bootstrap, Angular (Fig. 1).

1. ***HTML***

   HTML is an acronym for hypertext markup language which is the for the most familiar part markup language for creating Web pages. It is the most basic to start with for web development. It provides all the basic elements for the development of a basic Website. It can be written in any editor or a notepad itself. It can be run in any browser.
2. ***CSS***

   CSS is cascading style sheets used to decorate the pages with more flexibility. It gives us the power to use and reuse the styling over the page. We can use different libraries for this as well as we can declare it custom for ourselves. It is also used for the description of the presentation of the Web pages.
3. ***JavaScript***

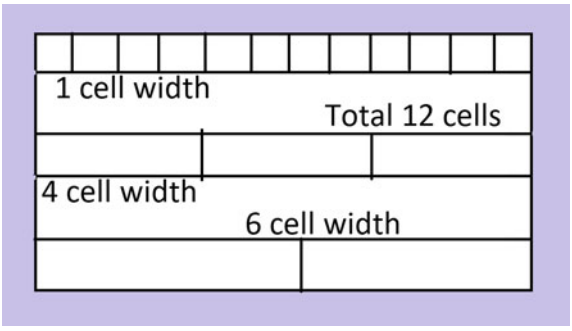


Fig. 1 Bootstrap columns

Fig. 2 Bootstrap design

JavaScript or JS is a high level and just in time compiled language. It is used to make pages interactive to communicate with the user and develop a connection between the user and pages. It provides animation and provides structure to the pages. It also helps pages to connect with the server-side and can be used for server-side programming (Fig. 2).

4. ***Bootstrap***

Bootstrap is a familiar and accepted framework for HTML, CSS, and JS for creating responsive, leading Web projects. It is good to start, and of course, it is open-source. It is good to go for static sites and easy to deploy in any hosting cloud. It is used in Sublime editor, Visual Studio code, Bootstrap Studio, Codepen, etc.

Bootstrap grid system—It is used to design pages where width is dived into 12 equal columns. We use different sets or groups of columns as per need and it is easy to design. We can divide the columns as per device width, i.e., different numbers of columns for mobile and different for tablets, and different for desktops.

Bootstrap consists of font awesome icons, Google fonts, and Glyph icons for designing. It has many libraries for animations and jQuery for event handling and CSS animation. It has extensive prebuilt components and powerful JavaScript plugins. It also provides themes for a start.

1.2 Full Stack Frameworks

Full Stack frameworks are used to design the back end as well as front-end of a Website for the working and processing of user data for the required action. It is a server-side part of the application. It is made up of HTML, CSS, JavaScript, and in addition to this, it also includes programming languages such as Python, Java, PHP, and ASP. It also has database and APIs included in it. Here, developer is responsible for the look as well as for the working of the Website. It is used for building dynamic Websites.

Full stack frameworks provide us with the ease of developing an application with minimum effort and it provides us with the tool to develop the applications. With

the help of this, we do not need to start with scratch, we can start to modify the samples as per our need. It gives us the ease to develop each and everything as per global standards and with minimum efforts. It increases the output and reduces the development time to a great extent. Using this, we save time and deliver the project to the client in less time.

1. ***Django***

   Django is a Python framework for full stack development. All the back end is developed using Python, and the front-end is developed using a front-end framework and uses HTML pages for the processing of the front-end data from the client. It follows the model-view-template (MVT) architecture. It has a high demand for AI/ML-powered sites and database-driven sites. It is open-source and cheap to host.

   It is fully loaded with familiar Web development tasks such as user authentication, content administration, site maps, and RSS feed which are out of the box. It can be written in Spyder, Sublime, Visual Studio Code, PyCharm, etc.

   Django runs on the DRY concept, i.e., do not repeat yourself. It is a fast framework to help developers to take applications from perception to conclusion as speedily as achievable. Sometimes launch is a matter of house as you do not need to focus on reinventing the wheel but writing your app (Fig. 3).

It is easily scalable as it can quickly and flexibly scale.

It takes security seriously and helps developers to avoid many common security mistakes. It provides security itself in the framework and reduces the time of development.

You take care of security and take care of SQL injection, site-through application fraud, and key clicks. It is user authentication system provides a secure way to manage accounts and passwords.

As a Web framework, it needs to have a convenient way to generate HTML dynamically. So it uses templates for the front-end development which is written in

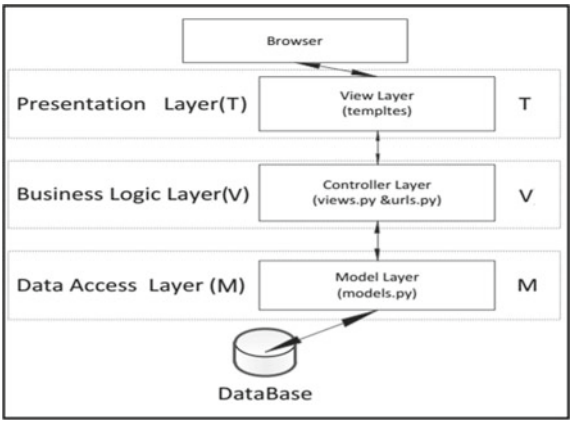


Fig. 3 Django architecture

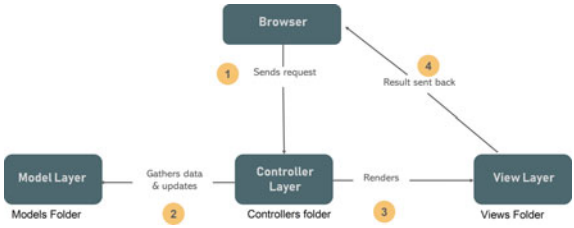


Fig. 4 NET Architecture

HTML itself and has its template system called Django template language (DTL). We can also create our custom template back end.

DTL uses its syntax for accessing the data from the back end. It helps to access the variables using curly brackets {{}} and logic using tags {% %} (Fig. 4).

2. ***.NET***

.Net is a full stack development framework. It supports many languages such as C#, F#, and basic languages. Used for Web design, mobile, desktop, game, and IoT.

.NET core is for cross-platform use of Websites, servers, and console applications. The .NET framework supports Websites, services, and desktop applications on top of the windows platform. Xamarin / Mono is a .NET implementation application for all major mobile applications.

Similarly, the .NET standard is the basic set of standard APIs for all .NET applications.

It has libraries to keep the environment healthy. NuGet is a built-in package manager that contains more than 90,000 packages. It is hard to take care of. It can be recorded in Visual Studio.

C# is a simple, modern, object-oriented programming, and F# is a cross-platform, open-source, functional programming language. F# also has object-focused and important programs. In the same line, Visual basic is an accessible language with a simple syntax for building genre-focused, object-focused applications.

In Web development, options are ASP.NET. Framework for building Web applications and services with .NET and C#. It is fast and scalable and helps build secure applications. It also provides a built-in database. The syntax of a Web page structure known as Razor is used to create dynamic Web pages.

It also follows common Web patterns such as model view Ccntroller (MVC). Its authentication system includes libraries, databases, and multifactor login management template pages, and authentication pages. Editor extensions provide syntax highlighting, coding, and other Web page optimization functionality.

The front code generated on the client-side is written in JavaScript. ASP.NET integrates with JavaScript and includes preconfigured single-page templates (SPA) frameworks such as React and Angular.

After discussion of front end and full stack framework, now we will discuss how these frameworks are useful for developing different applications which are initiated by different authors given in the following Table 1 after 2015.

Table 1 Application of technology in Web designing applications with its usefulness

Authors	Application of web designing framework	Its usefulness in web application
Nen et al. [1]	Use of AngularJS and Bootstrap for data driven web application	(1) It is used to optimize web developers productivity in web application development as the tool is straightforward and simple, and (2) To enhance end user browsing experience with an interactive table in the data-driven web application
Sundresan et al. [2]	Development of efficient time table system using AngularJS and Bootstrap	The developed system has a flexible representation and appropriate methods to create a feasible, automatic timetable and avoid clashes
Wang et al.[3]	Designing of instant messaging system an web	This paper proposes a solution of IMSW (Instant Messaging System based on Web). It introduces the system architecture of IMSW. illustrates the function of the system, and describes the realization of three key technologies (such as Bootstrap responsive design, MD5 +salt encryption algorithm and video call) in detail

2 Related Work

As we have very vast information, this paper will gather some from past papers which give us a brief description of past technical advances and their related work.

In [4] and [5], author discussed the QR code and barcode system. It can be concluded that Django has simpler communication between the layers and coupling is low. It has one architecture, .NET framework-based application has used two platforms, i.e., NET framework and .NET compact framework. It is somewhat difficult to develop an application in this manner.

According to [4], Django uses object-based mapping (ORM) structure. It helps the developer, so he does not need to directly face the database programming, and it only is required that each data model defines a class.

According to [6], Django can be used for a face recognition system and can be used for user authentication and accuracy checks. Here, face image is uploaded for login in an app.

According to [7], a force .Net framework can be used for making mobile robots. A C++ -based OpenCV library is used for object recognition. Image processing techniques such as Blob counter and HSL color space were used to distinguish the object from the background and determine the position of the target object on the image.

According to [8], .NET CF is good for building the mobile application but there are some limitations yet efficient if limitations are overcome. It has a low RAM requirement. Mobile computing is easy using this.

According to [9], ADO.NET should be used for ORM in .NET framework.

According to [10], authors discuss the responsive Web designs have received a lot of attention in recent years because they can meet a variety of online terminal solutions. This paper will discuss how to use Media queries, Bootstrap navigation, and streaming technology programming to achieve responsive Web design. Also make a rational analysis of responsive Web development in this current section.

According to [11], Bootstrap is a popular HTML, CSS, and JavaScript framework for creating responsive and mobile Websites. There is no doubt that one of the great benefits of using Bootstrap is the speed of development. Bootstrap has a huge community support so you often get help when you get into trouble. In addition, the Bootstrap itself is continuously updated and the creators have been very good at releasing timely updates. The Bootstrap framework allows for rapid responsive growth that is consistent and well supported by the development and construction community.

According to [12], responsive Web design tackles the compatibility difficulties of online sites shown in different resolutions, platforms, and screen sizes, while also providing users with a superior quality experience. This paper discusses design ideas and essential accountable design technologies for a responsive business Website based on responsive Web design research and HTML5 and CSS3 associated technology. Response Web design has been shown to be both viable and successful.

Web design since the 1990s is very different from current Web pages. The main purpose of the design is not the design itself, but to enable the transmission of Web content in a readable and understandable way, any device used to present Web content, desktop or hybrid computer, various mobile or portable devices. This paper shows Web design from its inception to modern architecture, and what we can expect in future.

The author [13] states that Web development has evolved dramatically over the years. The field of development is fast evolving, owing to the emergence of types such as mobiles. Web applications are becoming increasingly significant as the Internet grows in popularity. The Web application has served as the company's interface and one of the methods by which they show themselves. On the other hand, the Web has become a component of the file management system that serves as a replacement for paper. Due to the high demand for online applications, engineers must create a low-cost, secure, and well-written Web system. Rather than employing library style additions, the framework is offered to construct application. The framework assists the developer in designing Web design automatically. This paper compared the benefits once the evils of Web development using a framework that deals with the development of library style. These comparisons are based on the previous research, a paper that focuses on two main indicators, the impact of which management and impact on the engineer.

According to [14], Web technology plays a vital role in transforming the developed world into a knowledge society and semantic Web technology allows for the editing

Table 2 Comparative analysis of Bootstrap, Django, and .Net

Property	Bootstrap	Django	.NET
Type	Front-end	Full stack	Full stack
Language	HTML, CSS, JavaScript	Python	C#, F#, Visual basic
Time-required	Less	Less	More
Security	Not Required	Best	Best
Bundle	All included for basic	All included for advanced	Only basic available Advance to be added separately
Administration	Not required	Available by default	Have to create
Platform	Web	Web	Web, desktop, mobile app

and coding of meaningful information creating an advanced information management system. It enables active communication within machine to machine, machine to people, and vice versa.

After analyzing different frameworks, with the help of literature review, the following results are obtained which are given in Table 2. It is consisting of study of different frameworks, which represent the comparative study between Bootstrap, Django, and .Net. After study, it is observed that Bootstrap is only used for the front-end and responsive for all devices of different widths.

Django has all the functionalities wrapped in it in one place and is easy to understand and work upon. It is very dynamic and very fast to develop. .NET framework is a statically typed framework, and all functionalities are distributed in different frameworks which we must include for each part as per need. For example, for ORM, we need ADO.NET, for Web development, ASP.NET, etc.

3 Conclusion

The motive of this study is to understand exiting frameworks available for Web designing. In this, the discussion of front-end and full stack framework is elaborated with the various platforms present in it. All frameworks have their pros and cons, and we should decide which is to be used for our site according to our need and our knowledge and industrial requirements. It should be decided based on time requirement and scalability and maintenance. As discussed above, we can choose Bootstrap for front-end and static Websites and the development of dynamic Websites, Django for faster development and better scalability whereas .NET for different platforms and need different requirements. Django has only one language, whereas .NET has multiple. We can go with Django for more versatile and quick development and faster results.

References

1. S. Al Perumal, M. Tabassum, N.M. Norwawi, G. Al Narayana Samy, S.A. Perumal, Development of an efficient timetable system using angular JS and bootstrap 3. in *2018 8th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, (2018), pp. 70–75. https://doi.org/10.1109/ICCSCE.2018.8685002
2. J. Wang, B. Zhang, J. Guo, H. Wang, The design of instant messaging system based on web. in *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, (2017) pp. 288–291. https://doi.org/10.1109/ICSESS.2017.8342916
3. V.S. Pawar, P. Jaiswal, Use of web services for integrating geospatial system with other systems. CHEMIK 99–105 (2020). Retrieved from http://www.chemikinternational.com/index.php/chemik/article/view/39
4. Q. Yu, W. Yang, The analysis and design of system of experimental consumables based on django and QR code. https://doi.org/10.1109/IICSPI48186.2019.9095914
5. O. Krejcar, M. Macecek, Development of bar code scanner application for windows mobile devices using a symbol mobility developer kit for .NET compact framework. https://doi.org/10.1109/ICMEE.2010.5558524
6. Z. Jin, J. Wang, S. Zhang, Design and implementation of WA face recognition system based on small wechat program. https://doi.org/10.1109/CSEI47661.2019.8938841
7. K. Kungcharoen, P. Palangsantikul, W. Premchaiswadi, Development of object detection software for a mobile robot using an aforce. Net Framework. https://doi.org/10.1109/ICTKE.2012.6152407
8. Hammad-ul-Hasnain, Building mobile application with .NET compact framework. https://doi.org/10.1109/SCONEST.2005.4382871
9. W. Wiphusitphunpol, T. Lertrusdachakul, Fetch performance comparison of object relational mapper in .NET platform. https://doi.org/10.1109/ECTICon.2017.8096264
10. W. Jiang, M. Zhang, B. Zhou, Y. Jiang, Y. Zhang, Responsive web design mode and application. in *2014 IEEE Workshop on Advanced Research and Technology in Industry Applications* (WARTIA) (Ottawa, ON, 2014), pp. 1303–1306.https://doi.org/10.1109/WARTIA.2014.6976522
11. S.S. Gaikwa, P. Adkar, A review paper on bootstrap framework. Iconic Res. Eng. J. **2**(10) (2019)
12. N. Li, B. Zhang, The design and implementation of responsive web page based on HTML5 and CSS3. in *2019 International Conference on Machine Learning, Big Data and Business Intelligence* (MLBDBI), Taiyuan, China (2019), pp. 373–376. https://doi.org/10.1109/MLBDBI48998.2019.00084
13. S. Syafiq, M. Daud, H. Hasan, A. Zairi, S. Imri, E. Akmar, N. Rahim, Comparison of web development using framework over library
14. R. Bastola, P. Campus, Developing domain ontology for issuing certificate of citizenship of Nepal. J. Inf. Technol. **2**(02), 73–90 (2020)

A Comprehensive Review on Security and Privacy Preservation in Cloud Environment

Rajesh Bingu, S. Jothilakshmi, and N. Srinivasu

Abstract With the widespread growth of the Internet of Things (IoT) devices and emergent data generated at the edge computing network, the conventional cloud computing (CC) model encounters various bottleneck issues like resource constraints and bandwidth constraints. Thus, edge computing devices help in processing and storing data at the edge level and intended as a promising solution for the past few years. Moreover, the unique nature of computing devices like parallel processing, content perception, and real-time computing have introduced various challenges in privacy preservation and data security. These two factors are considered as the key factors in this review. In this survey, an empirical analysis of the privacy and security threats, protection methodologies, and related countermeasures are inherited in cloud computing. Specifically, an overview of basic CC definitions, applications, and architectural models are discussed. Subsequently, an extensive analysis is performed specifically with data security, privacy, challenges, requirements, and mechanisms in CC. Various cryptographic-based methods based on privacy preservation and security are discussed in this review. The existing issues related to security and privacy in CC are reviewed and finally, various open research directions related to privacy and security in the CC field are also discussed.

Keywords Cloud Computing · Resource constraints · Bandwidth constraints · Privacy · Security · Edge computing · Threats · Cryptography methods

1 Introduction

In the fast-growing technological advancements, the Cloud Computing concepts are raised from the distributive software architectural model [1]. Cloud Computing

R. Bingu (✉) · S. Jothilakshmi
Department of Information Technology, Annamalai University, Chidambaram, Tamilnadu, India

N. Srinivasu
Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation, Vijayawada, India
e-mail: srinivasu28@kluniversity.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_54

(CC) technology attempts to offer host services on the Internet. Recently, CC is considered as the technological advancements with the rise of various industrial markets and user communities. CC services are offered from small/huge data centres provided in various regions of the world. For instance, Google applications and Microsoft are some common examples that offer CC services where security plays a substantial role during the establishment of CC services [2]. Various prevailing literature concentrates on diverse security solutions that include the execution of security policies and technological advancements. Some researchers concentrate on various security issues due to attacks over the CC environment from the attacker's perspective [3]. The solutions anticipated by the researchers to handle these attacks rely on the security theories to protect the computing environment. Aluvalu et al. [4] describe various security issues that influence the CC attributes. Also, the author intends to provide the solution to the problems identified and directly related to cloud security. Some security-based guidelines are provided by the author and facilitate the computing organizations needs to be aware of vulnerabilities and various approaches are provided to handle these vulnerability in an efficient manner [5].

Various threats and related challenges are grown from the wider use of CC services. At present, CC models are the preliminary source of these vulnerabilities and challenges [6]. The attackers exploit the feebleness of the computing models while the service providers initiate to access the users' private data and influence the computer systems' processing power. Recently, cloud-based intrusion detection and prevention (CIDS) mechanisms are anticipated to handle the problems discussed above. Before the design of the cloud intrusion detection model, the 'Network Intrusion detection and prevention (NIDS)' mechanism is used and it handle the risks over the CC virtual networks [7]. When compared to the NIDS, CIDS is finest during the mitigation of risks and security challenges towards the network model.

Recently, there is a rising growth of CC techniques in Information Technology. Moreover, various services are still unwilling to completely adapt towards the CC as appropriate security technologies which are not matured yet [8]. Therefore, various literature provides the preliminary requirements to invest the amount for the CC-based device security. Few studies concentrate on offering a solution to the evaluation of the CC-based security model. One research model initiates a novel 'attack tree map' to examine the security threats and vulnerabilities. Xue et al. [9] discuss various CC facets merged with the trusted computing environment to offer security services like integrity, authentication, and confidentiality.

Owing to the distinctive characteristics and benefits of the CC paradigm like mobility support requirements, location-awareness, parallel computation, heterogeneity-based distributive architecture, and huge data processing shows some conventional privacy preservation and data security mechanisms in CC [10]. It is no longer appropriate for protecting massive data in the CC paradigm. Specifically, secure data computation, secure data storage, privacy protection, and authentication access control factors are specifically pre-dominant [11]. For instance, CC is a distributed computing system with huge trust domains where multiple functional entities, the authentication mechanism not only needs the identity validation for all entities in the trusted domain; however, it needs every cloud entity need to authenticate

mutually among various trusted domains. However, for certain resource-constraint devices, it is not possible to preserve the huge amount of data or execute the complex security algorithm [12]. To be specific, the privacy preservation and data security in CC specifically concentrates on the following confronts [13]:

(1) **Fine-grained and Lightweight model**: Some novel requirements for fine-grained data sharing systems and lightweight data encryption approaches are based on multiple authorized parties in CC.
(2) **Distributed access control**: The heterogeneous secure data management and multi-source data dissemination issues in the distributed environment.
(3) **Resource constraints**: Security confronts resource-constrained devices and large-scale edge services.
(4) **Efficient privacy-preserving**: Some novel requirements of effectual privacy preservation methods for diverse cloud computing models and edge services faced in the CC environment.

The above-mentioned privacy preservation and security challenge of the CC paradigm motivates us to offer comprehensive research analysis. The significant contributions of this review are summarized as follows.

An extensive analysis is performed for constructing a factor in CC that is discussed holistically. A comprehensive analysis of CC definition and architectural model is provided. Some promising CC applications are also discussed [14]. The privacy preservation and data security requirements are provided based on the critical metrics include privacy requirements, access control, authentication, integrity, confidentiality, and availability. Some comprehensive analysis of the potential privacy and security challenges in CC is pinpointed [15]. Specifically, the prevailing privacy preservation and data security methods are discussed in detail. The architectural level towards the security model is also anticipated.

A comprehensive analysis of the cryptographic-based model for resolving privacy and data security issues is discussed. It includes proxy re-encryption, identity-based encryption, searchable encryption, attribute-based encryption, and homomorphic encryption. Moreover, a comprehensive analysis and comparison of various privacy and data security solutions are provided and the solutions related to these issues are discussed.

Some open issues are also discussed with the future research challenges like the modelling of fine-grained privacy preservation model, dynamic data processing, cross-domain authentication, multi-authority access control mechanism, and distributive data encryption.

The rest of the work is organized as: Sect. 2 discusses the CC-based privacy preservation framework. Section 3 elaborates the security threat mechanism with secure outsourcing computation and computational requirements. Section 4 explains the privacy threat mechanism encountered in the cloud that includes intrusion, public disclosure, appropriation, and false light. Section 5 discusses the regulations of privacy preservation which are followed by the privacy factors established in Sect. 6. Here, factors like security, data access and utilization, data location, and data status are discussed. Sections 7 and 8 discusses privacy-enhancing technologies and data

encryption technologies. Section 9 discusses the research gaps with the research questions that need to be resolved in future. Section 10 provides the summary of the comprehensive survey.

2 CC-Based Privacy Preservation Framework

In CC, the user's privacy preservation data needs to be more confidential. Some data outsourcing mechanisms like storage, multi-tenant and virtualization, big data storage, and other technologies are considered as the major factors of risk privacy disclosure [16]. Privacy preservation mechanisms are considered to eliminate sensitive information and personal information exposure in the cloud environment, i.e., computing, data sharing, delete, integrity verification, and other functionalities. Various solutions are considered to protect personal privacy and user's information like trusted technology, encryption, access control, or various other combinations of these methods [17]. Storage services in the CC platform show substantial variations in the IT field and give a tremendous impact on the privacy and security factors [18]:

(1) User data is completely centralized and security methods need to be fulfilled during the process of CC requirements.
(2) The cloud users' do not have any trustfulness in the CC platform, therefore, the end-users do not believe that the data is stored in a secured manner.
(3) Generally, the user data is highly confidential and centralized, therefore, security must be fulfilled during the data processing in the cloud environment.

Security and privacy are the major factors in cloud where the efficient access control and secure data resources are the key factors to be achieved in an efficient manner [19]. After predicting the legitimate user identity, the access control mechanism denies the request generated from the data access and generally, it is used for preserving the data resources and avoids the illegal functionalities of the intruders. There are diverse access control mechanisms like Mandatory Access Control (MAC), Attribute-based Access Control (ABAC), Discretionary Access Control (DAC), Task-based Access Control (TBAC), Usage Control (UCON), and Role-based Access Control (RBAC). The performances of these mechanisms are evaluated. Some set of procedures and rules can facilitate the authentication of legitimate users to access data by fulfilling the data privacy, data confidentiality, and integrity in the security model [20]. When compared to the conventional network environment, the access control mechanism is extremely essential in the CC environment. While using the computing and storage services, the users need to provide authentication to the Cloud Service Providers (CSP) and consider the essential policies to access services and data. Access control and mutual authentication among the CSP are essential to fulfil the CC security [21]. The user not only requires a side-channel attack; however, it requires appropriate methods to fulfil privacy and security.

At present, CSP uses various access control methods to offer security protection and privacy. But, various issues need to be resolved [22]:

(1) In cloud security, both objects and subjects are not distinguished clearly and the access control method needs to be expanded from the user's authority to protect the data, and access virtual resources in the cloud environment. Additionally, there is some differentiation among the conventional centralized resource management system and distributed access mechanism in the CC and certain access policies are generated as a test for security management.
(2) In the security mechanism, there are diverse applications that come under various security management issues and the user needs to access the certification services and resources in the domain boundaries to deal with the public policies. When the tenants encounter a side-channel attack, it leads to privacy leakage.
(3) In the access control mechanism, the idea behind the object and the subjects in the cloud are extremely influenced by the conventional access control mechanism and it is unable to fulfil the cloud requirements. In the CC environment, the relationship among the various mechanisms is complex and the user scalability frequently varies, the number of administrators is huge and complex, and the authoritative assessment of the cloud authority varies from the conventional models.

In the conventional model, the user needs to deal with the data in a trusted and reliable manner; however, it is complex in the CC environment. Both the server and the users are not credible completely and the user needs to fulfil the data security and avoids CSP malicious activities of the private data. It is essential to merge the diverse technological concepts to build the privacy preservation concept.

3 Security Threat Mechanism

Generally, outsourcing computation is not completely satisfying; there is some preliminary cause for handling the privacy and security factors during the data analysis like user privacy and revealing the data content. It should predict and handle security threats. Thus, it reduces the risk of avoiding privacy and data security [23]. Two diverse threats are considered in the CC environment. First are the external attackers that include threats of hardware attacks and remote software in CC. It uses various kinds of approaches like malware attacks and network overlapping to evaluate unauthorized data and server intrusion. Secondly, security threat is identified from the internal participants. When the user submits the computational tasks, the computational process and data are generally controlled. Therefore, cloud server mortality plays a substantial role in data privacy and security [24]. Moreover, the server does not as a trustful environment. It shows that the server can handle the computational tasks inappropriately and pretends to learn the realistic features. Based on the server's nature, some users need to categorize the adversarial models into two diverse levels: semi-honest and honest level [25]. In the former model, the server carries out the functionality with the essential computational process, however, crucial regarding the

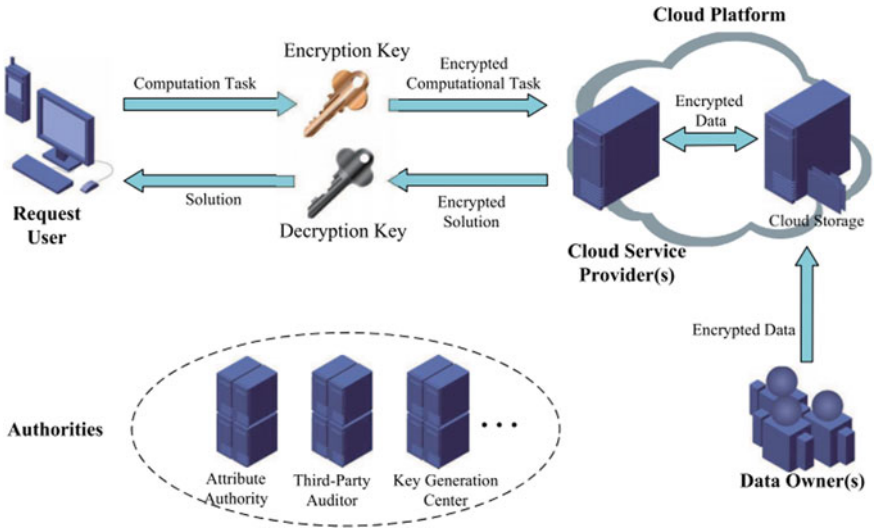


Fig. 1 Security-based cloud system model

sensitive user's information. The malicious servers move against the computational power and attain certain benefits [26]. There are two kinds of threats like internal and external attacks that destroy the user's integrity and confidentiality. Figure 1 depicts the security-based cloud system model.

(a) **Need for secure outsourcing computation**: The security information is also substantially essential in various computational schemes and the efficiency specifies the communication and computational cost for performing the task outsourcing. The evaluation process includes three diverse factors like access controllability, data integrity, and data confidentiality [27]. Here, confidentiality specifies the competency to preserve the user's data to be exposed in the CC environment or to unauthorized parties. Data integrity specifies the fulfilment of data correctness and completeness. It is also termed as the data check ability and verification. Finally, access controllability is a property that executes the permission process towards the valid users and restricts the set of users to attain the computational outcomes and predict the data sources from the data owners. It is a fine-grained process that offers feasible access control mechanisms and thus shows higher significance towards reality and theory.

(b) **Essential requirements**: Another pre-dominant requirement for computational outsourcing is the system performance. It is due to the higher computational cost and it selects the outsourcing of various complex tasks over the cloud environment [28]. The significant factors to evaluate the system efficiency are computational overhead before and after the outsourcing process, i.e., verification, processing, encryption and decryption, and communication cost during the transmission process.

Moreover, outsourcing needs to attain the requirements essentially and shows huge effect over the CC environment which is more impractical. Generally, security needs added operational support and diminish efficiency [29]. However, an excess process with high efficiency generally causes insufficient security factors. Therefore, superior and feasible performance is generally considered to offer a superior trade-off between efficiency and security. Also, it is a huge inspiration to model the design algorithm efficiently. It provides an extensive scope on the certain computational process with the application-specific and fundamental process as the outsourcing process shows limited functionalities.

4 Privacy Threat Mechanisms

The need for privacy arises when the user's data is accessed without any knowledge or consent of the data owners. It happens during the attack, data breaching, eavesdropping, and other forms. It is classified in various forms:

(a) **Public disclosure**: Realising the private or unknown information to the external members is more determined as a public disclosure [30]. The information is offensive while exposed to unauthorized users. Thus, if the data does not offer any public concern, the individual who is accountable for releasing the data is more liable for privacy invasion. Some typical forms of examples are publicly disclosing the private information by the officers, politicians, and other celebrities.

(b) **Intrusion**: The intrusion towards the privacy data includes the process of direct/indirect access towards the institutional or individual private data, i.e., telephonic conversations that are recorded devoid of the knowledge of the owners and trespassing the private data sources [31]. The injections of malicious activities are also identified in the CC environment.

(c) **Appropriation**: It specifies the appropriation of organizational or individual identity. Usually, it happens with the use of the individual name or personal features devoid of knowledge and authorization. It is generally noted in the reference, media cases, marketing, and stories. It is most probable to occur and the issue is more recurrent and happens in the online profiles.

(d) **False light**: It is like public disclosure. It is a form of malicious statements and public disclosure of false statements. Usually, it is performed with trust distortion and fictional factors.

There exist some added privacy threats, for instance, fine-grained taxonomy like information processing, information collection, invasion, and information dissemination [32]. An attack is a process of attempt, alters, expose, disable, destroy, gain unauthorized access, steal, or unauthorized access over a specific environment as stated by ISO. In the privacy preservation process, various forms of consensus specify the form of attack. The attackers evaluate the kind of information that is available and reflects the other form of information methods.

- **Marketer—**The attacker is not interested in re-identifying the individuals.
- **Journalist—**The attacker does not have any prior knowledge of the attack that needs to be held.
- **Prosecutor—**The attacker is aware of the form of data available in the target region and sometimes data is available in the form of a dataset.

Also, it is necessary to ensure certain attack models that operates in the specific environment. There are some forms of attack identification process [33]. They are:

- **Probabilistic attack—**It relies on the uninformative principle that concentrates on actual records. It is known that access to data does not change significantly.
- **Table linkage—**It happens when the attack derives the presence or absence of the data successfully in the targeted record of the table owner.
- **Record linkage—**It happens when the attacker matches the record owner successfully towards the sensitive attributes from the published dataset.
- **Attribute linkage—**It happens when there is no specific record identification process; the attacker infers some sensitive information with a group of data ownership.

Table 1 depicts some sample instances of the privacy invasions with the exposed private data and compromised organizational privacy. The credit card credentials, political interest, and some personal details are disclosed publicly. In certain cases, like Uber and Yahoo, data breaching is identified due to some security reasons. Moreover, in the case of AOL and Netflix, incorrect utilization of anonymization is considered [34]. Similarly, linkage attacks and cross-references are identified with a high risk for data anonymization. These security risks have to be reduced and avoided by considering the attack model.

5 Regulations for Privacy Preservation

Various countries adopt certain laws and regulations for privacy preservation, data handling and sharing, and data access [35]. In European countries, certain essential directives are enforced and they should be adopted by the European Union. Data Protection Regulation needs to be enforced for some businesses or services that handle data from the citizens and enforced to execute it. In the USA, an act known as the Gramm-Leach-Bliley Act is adopted and the Personal Information Protection and Electronic document act are adopted by Canada. In the Russian Federation, the Personal data protection act is used, while in China, the Personal Information Protection act and Computer processed personal data are some examples of regulations [36]. It may vary from one country to the other with some common objective for provisioning legal regulation and protection to the user's private and personal data.

Table 1 Samples of exposed data

Organization	Description	Threat invasion model	Year	Exposed data
Microsoft	Roughly 250 million organizational data are exposed	Security problems	2020	Appropriation and intrusion
Facebook	Roughly 540 million organizational data are exposed	Third-party security problems	2019	Appropriation and intrusion
Cambridge Analytica	Unauthorized way of profiling	Scrapped personal information from the personal user account	2018	Only intrusion
Equifax	Exposed some sensitive and confidential data of around 140 million users	Happens due to hacking	2017	Appropriation and intrusion
Uber	Driver details and 57 million uber customer details are exposed	Happens due to hacking	2016	Appropriation and intrusion
JPMorgan	Driver details and 87 million customer details are exposed	Happens due to hacking	2014	Appropriation and intrusion
Yahoo	Roughly 500 million user account details are exposed	Happens due to hacking	2014	Appropriation and intrusion
Netflix	The exposed dataset helps to identify the user's information	Cross-referring	2006	Disclosed publicly
AOL	The exposed dataset helps to identify the user's information	Cross-referring	2006	Disclosed publicly

6 Establishing Privacy Over the Cloud Environment

The provisioning of cloud services varies based on conventional Internet services. The data processing over the servers or the server's location is concerned with the privacy policies. The preliminary requirements of privacy are discussed below [37]:

(a) **Security—**For establishing security, some essential features like infrastructure, communication, and security features play a substantial role in securing the data. Some general aspects like anti-virus, strong password, and certain regular software updates can improve the security in both the service providers and users side.

(b) **Data status—**The aspect related to CSP needs to consider the method for disclosure of some specific method utilized for preserving data. The data

status during the handling and processing stages needs to be specified, i.e., pseudonymized, anonymized, plain text, and encrypted form.

(c) **Data access and utilization—**It is completely essential to fulfil the appropriate handling and accessing of data. The system service might not be compromised even in case of the least security policies and measures. Fulfilling appropriate cloud utilization policies like logical and physical access are not provided due to diligence. Some specific form of data processing is required. It is necessary to disclose the usage policies in two diverse directions: service provider to user and user to service provider. It is completely crucial for efficiently defining the access rules. Some sets of questions need to be resolved while considering data access and usage like how, who can access the data, where, why, and time.

(d) **Data location—**The privacy regulations and laws differ from one country to other. Thus, compliance in various locations is more challenging. Some organizations that process data from International users face some constraints, i.e., the servers with computing power and databases are distributed in various countries. Some added aspects need to be handled in an efficient manner: customer data management and local laws; and laws based on the origin of the country. Failure leads to some losses in the organization.

As depicted in Fig. 2, there are some requirements like availability, portability, and audibility that needed to be considered. However, there is still some vulnerability in association with CSP to fulfil the higher privacy standards. Generally, data owners with these services have no physical access over the system mode. Thus, indeed of complete trust establishment, there is a semi-trusted relationship [38]. Generally, CSP is used for some common outsourcing data purposes. Users need to consider some proactive approaches like data anonymization. Similarly, various privacy-based approaches and tools are restricted with the exposure of sensitive information.

7 Privacy-Enhancing Technologies

The privacy-enhancing technologies carry out data transformation and operations that help to deal with a higher level of data privacy, i.e., encryption or data anonymization. Various technologies are accessible in handling privacy preservation. Some of the evaluation indicators are provided based on the quality of the technology [39]. They are discussed below:

- **Readiness evaluation**: There are six diverse stages of the readiness evaluation process. They are product, idea, outdated, proof-of-concept, pilot, and research.
- **Quality evaluation**: It is based on nine diverse indicators with varying weights. They are: scope, protection, transferability, trust assumptions, maintainability, side effects, performance efficiency, reliability, and operability

It is chosen based on the score offered by the indicators. The scores are essential for providing a systematic model, i.e., selection of appropriate combinations of the

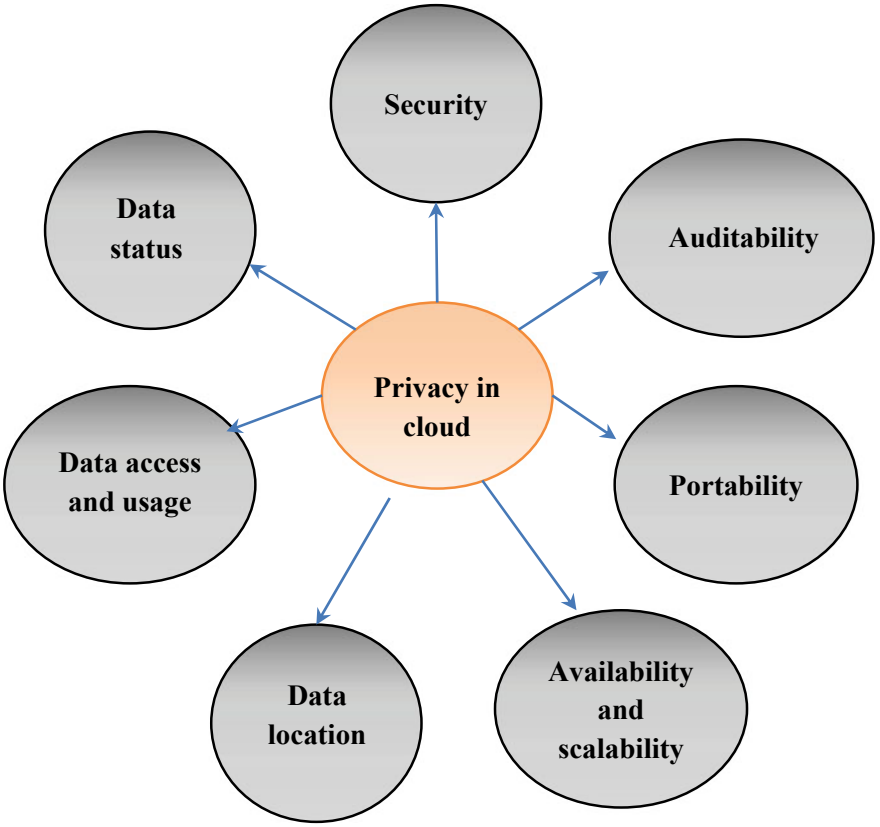


Fig. 2 Privacy preservation consideration in CC

enhancing technologies, for instance, data protection, authorities, online users and researchers, and software developers [40]. Some common types of privacy provisioning technologies are provided in-between the services and the users. Onion routers and the corresponding network provide increased online privacy using web traffic anonymization. It is improved and provided as a modern version of Firefox browser with improved usability. However, it is extremely possible to predict the client data by performing browser attacks [41]. In the communication and network domain, Domain Name Server and encryption models are utilized to improve privacy. However, for the improved privacy preservation process, the technology is complementing one another. Data publishing provides challenges to researchers and authorities owing to the trade-off among the disclosure risk and data utility and inherits the re-identification process by linkage or cross-reference attacks. Attacks rely on the published data or matches with the data anonymization. Chen et al. [42] perform hybridization of Internet Protocol (IP) and timestamp using some distinctive methods and the outcomes provide protection against the re-identification process. In the

timestamp process, multiplicative noise and enumeration are used for preserving the structural flow. Some preservation techniques are suited for applications like unstructured and structured data, real-time, offline application, and reversible data. Based on these applications, data owners need to characterize suppressiveness to operate the unstructured or structured data.

8 Data Encryption Technologies

The security is vulnerable when the provided data is outsourced to the cloud environment. Thus, encryption is considered an efficient technique to preserve data. The significance of the encryption process is to convert the original form of plain text to a string format with unreadable code. It is known as ciphertext [43]. The garbled code attains the original content which significantly preserves the data confidentiality and avoids the data being tampered with. The authorized users can access that encrypted data with the appropriate private key and later update or modify the content. The encryption process is further divided into asymmetric and symmetric encryption processes. In symmetric encryption, the process makes use of the secret key for encrypting and decrypting the data. Moreover, before symmetric encryption, users have to determine the consensus key. However, it is extremely inefficient for multi-user sharing files. Subsequently, asymmetric encryption is also known as the public key encryption process and it is extremely convenient. It has a pair of keys and it is disclosed with the encrypted files. However, the private key is utilized for ciphertext decryption. There are diverse encryption technologies that are extensively adopted in cloud storage systems.

(a) **Identity-based encryption (IBE)**: In the conventional Public key Infrastructure, the identity information is more constant with the public key and used for the encryption process. The sender has to authenticate the identity information of the receiver via a trusted third-party certificate [44]. It may lead to the increase of the sender's workload when the sender needs to share confidential data with multiple receivers. The idea behind the identity-based cryptographic model is anticipated to associate the identity information with the available public key. Therefore, there is no need to validate the receiver's certification before performing the encryption process. The identity-based encryption (IBE) model is designed by Franklin et al. to establish a security model using bilinear mapping to design a secure scheme. For example, 'A' is a sender who needs to transfer an encrypted message to the receiver 'B'. For this purpose, a trusted third party and Private Key Generator (PKG) is needed for generating the corresponding private key and public key. Here, 'A' uses unique identity information for generating the public key and sends the encrypted message to 'B'. Then, 'B' needs to communicate with the key generator for establishing authentication and to attain related private keys (See Fig. 3).

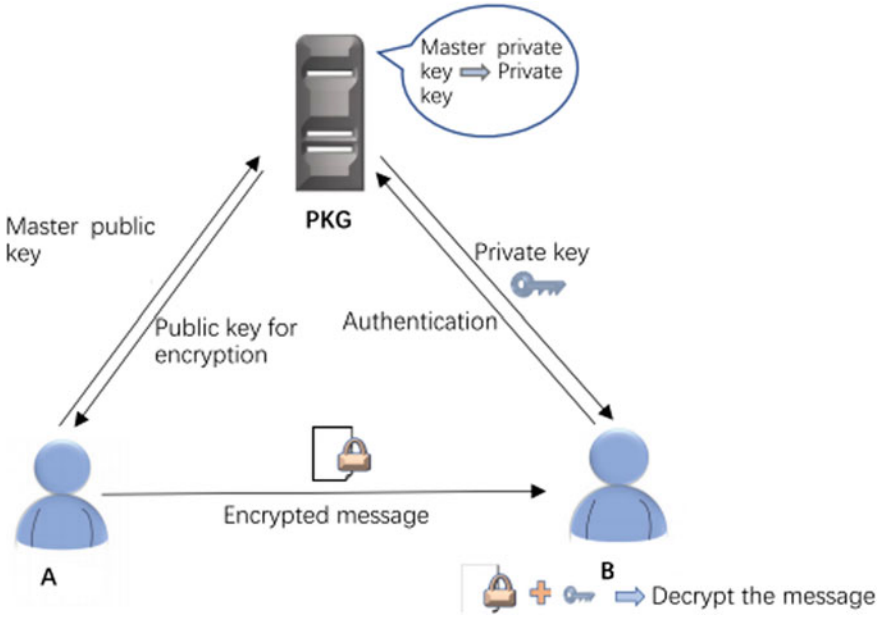


Fig. 3 Identity-based encryption process

(b) **Attribute-based encryption (ABE)**: In the existing IBE scheme, identity is measured as the string which is divergent from one another. Moreover, the feasibility of the model shows some bottlenecks when the ciphertext intends to be legally accessed by the users [45]. The identity of the Integrated Development Environment (IDE) model is substituted by the set of attributes in the ABE model. Here, the user's attribute access policy has the competency to access the encrypted data. It is composed of four diverse parts:

(1) **Setup phase**: It is also known as the system initialization phase with certain security parameters. They are input data, public parameters and the master key.
(2) **KeyGen phase**: Here, the data owner needs to submit their attributes for attaining private key related to the attributes.
(3) **Encryption phase**: The data owners encrypt the data using the public key and attain ciphertext and transfer it to the receiver or the public cloud environment.
(4) **Decryption phase**: here, the users get ciphertext and decrypt with the private key.

The attribute-based encryption model is considered to be a promising solution to offer fine-grained access control towards the encryption files in data sharing applications. The data owner needs to specify who can access these encrypted data. It is subdivided into two phases: Ciphertext-policy attribute-based encryption model

(CP-ABE) and Key-policy attribute-based encryption model (KP-ABE). In the KP-ABE, the ciphertext is related to the attribute set while the private key is associated with the access policy of the attributes. While in CP-ABE, the policy is merged with the ciphertext and the data owner needs to determine the access policy. Here, the private key is associated with the corresponding attributes. The user's attribute can be changed based on various reasons. The malicious nature of some authorized users is disclosed to spoil the privacy and confidentiality of the data [46]. Thus, secure and protective revocation is highly essential in the ABE model. Some prevailing revocation models shows indirect revocation and some may have direct revocation. The trusted party interacts with the non-revoked users periodically and updates the decryption key. Here, the decryption key is composed of two diverse parts. They are the update key and long-term secret key where the update key needs to be regularly updated. The difference among the attributes are partitioned into two diverse disjoint sets and merged with the master key for generating the secret key. These secret keys are diverse and show the re-randomization property. Therefore, resistivity is achieved with the decryption key. Later, a tree-based model is designed for diminishing the computational complexity and key generation phase.

In the direct revocation process, the trusted authority needs to generate the revocation list that includes the revoked users with the public key. The data owner represents the revoked users and cannot decrypt the ciphertext even in the case of matched attributes. Here, the authority revokes the user by updating the revocation list and interacts with the non-revoked users. After receiving the list, the third party needs to update the ciphertext with some public information and ensure the ciphertext is not decrypted by the revoked users. At last, the authorized user has the pleasure to validate the update of the third party appropriately. This model does not forbid the revoked users to decrypt the ciphertext; however, it provides verifiable functions to fulfil ciphertext under the revocation list.

(c) **Homomorphic encryption model**: The introduction of attribute-based encryption (ABE) and Identity-based encryption (IBE) is provided for fulfilling the data confidentiality; there exist certain drawbacks in those models. When the user intends to update the encryption files over the cloud storage, there are two options. One is the modification of ciphertext in the cloud. After this decryption of the modified ciphertext, it is generally considered as the meaningless garbled code that causes data damage. Next is to update the decrypted file and transfer it to the new cloud location. However, these two modifications are complex. Also, the data transmission from the cloud storage system to the cloud storage leads to data leakage risk [46]. The homomorphic encryption model is used for handling these issues with higher significance (See Fig. 4). It is a kind of public key encryption process that facilitates the users to carry out certain algebraic functions and acquires encrypted text and results in the consistent form of decrypted data. The data owner performs file encryption using this homomorphic model and transfers it to the cloud server. The authorized user needs to decrypt the ciphertext with the appropriate private keys. This homomorphic encryption model is more efficient to protect outsourced data. Table 2 depicts

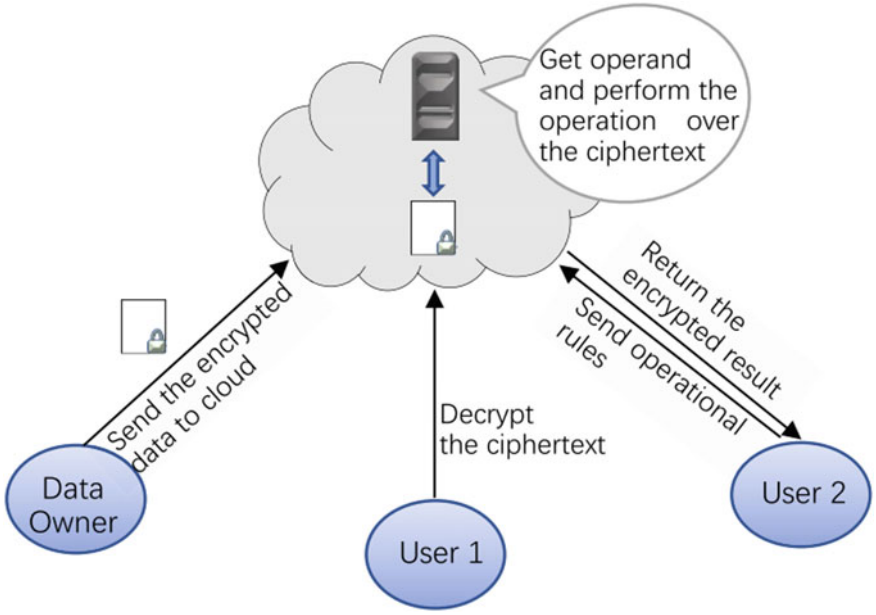


Fig. 4 Homomorphic encryption process

the comparison of the ABE model.

(d) **Searchable encryption:** It is a symmetric key encryption approach that outsources the confidential data from one person to another using searching process. It adopts proxy re-encryption that shares data from the cloud with E2E encryption and confines the authentic recipients. It serves as a privacy and security factor with various cryptographic approaches like key search model and time-dependent proxy re-encryption. It helps to search the unique keyword preserved over the server and uses non-iterative and iterative schemes. These schemes assist in index generation, searching and updating while non-iterative scheme helps in hash chain. It gives superior security model over the huge database and reduces the storage overhead.

(e) **Proxy re-encryption:** It is a cryptographic model provides semi-trusted server (proxy) to re-encrypt the provided ciphertext (See Fig. 5). It is encrypted by one public key to another ciphertext. The proposed proxy re-encryption model facilitates decryption of user's data over the cloud server and credentials devoid of secret key disclosure. A time-based proxy re-encryption is proposed that facilitates users to access the user's record in a pre-defined time interval. The target is to achieve access control, time-based revocation, efficiency, and user revocation.

Table 2 ABE model comparison

Category	Sub-category	Model	Techniques	Benefits
Revocation	Direct	The key-policy attributes are revocable directly with the ciphertext delegation verification process	(1) Linear secret key sharing (2) KP-ABE	(1) Appropriate security (2) Universal construction (3) Update verification (4) Direct revocation
		ABE-based verifiable key concept and direct revocation	(1) Linear secret key sharing (2) KP-ABE (3) Sub-set difference method	(1) Outsource decryption process verification (2) Partially hidden policy (3) Direct revocation
		Partial secure delegation and policy-hidden attributes	(1) Linear secret key sharing (2) Broadcast encryption (3) Outsourcing model (4) CP-ABE	(1) Decryption key resistive exposure (2) Ciphertext delegation (3) Randomized key generation (4) Indirect revocation
	Indirect	Ciphertext delegation and revocable ABE with exposed decryption key	(1) CP-ABE (2) ABE (3) Tree-based revocation	(1) Complete security (2) Ciphertext delegation (3) Storage revocation
Revocable storage system		Ciphertext delegation and dynamic credentials for ABE	(1) CP-ABE and KP-ABE (2) Linear secret sharing model	(1) Least computational complexity (2) Effectual keyword searching process (3) Secured CPA
Computational overhead reduction	Outsource computation	Outsourced ABE with key search function	(1) KP-ABE (2) secret key sharing with interpolation	(1) Reduced redundancy (ciphertext) (2) Lower computational overhead and storage
	Compact policy	Compact ABE	(1) CP-ABE (2) greedy compact model (3) Policy compact	(1) Reduced computation overhead (2) Reduced ciphertext redundancy

(continued)

Table 2 (continued)

Category	Sub-category	Model	Techniques	Benefits
	Enhanced policy management	Scalable ciphertext policy-based ABE	(1) CP-ABE (2) Linear secret key blocking	(1) Collision resistivity (2) Reduced computational overhead (3) Reduced storage cost

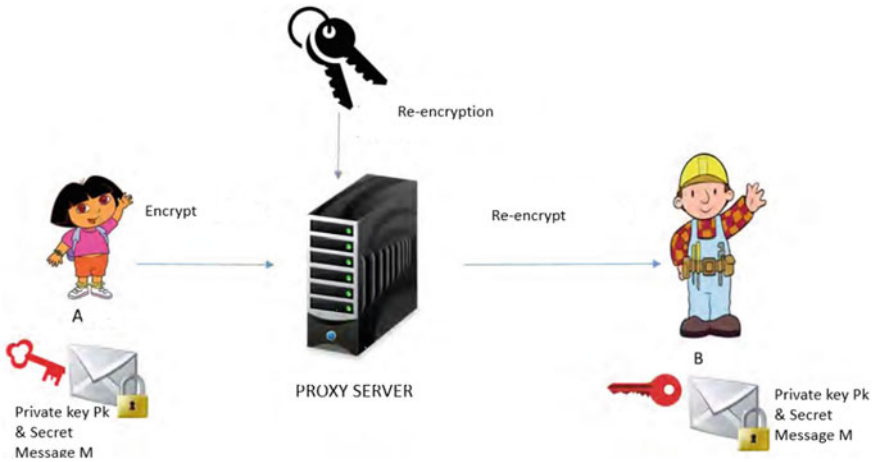


Fig. 5 Proxy re-encryption scheme

9 Research Gaps

Based on the extensive research analysis, it is observed that some research constraints need to be resolved in future. The trust evaluation criteria are insufficient and lack standardization. The access control methods are extremely complex with the selective constraint towards flexibility and scalability. Generally, the access control policies are formulated by the end-users. Thus, it increases the risk of service access and strictly restrictive to third-party access [47, 48]. There is a diverse issue that needs to be resolved efficiently and it is related to policy and key management, unauthorized access, attribute management, and so on. When dealing with encryption-based challenges, ABE is effective but has various constraints. How efficiently the ciphertext can be searched and control the privacy leakage is extremely complex. The construction of traceable, revocable, and feasible attribute encryption is challenging.

(a) **Research questions**

There is a need for a trustful, secure, and reliable service to eliminate illegal access to cloud systems. The researchers need to concentrate on the below-given research

questions to improve the standard of privacy preservation and data security over the cloud [49].

(1) How to compute and store the privacy preservation data efficiently in the cloud environment?
(2) Which standard access control method is efficient for secure data transmission in the cloud?
(3) How efficiently the data is shared with multiple Cloud Service Providers?
(4) Which encryption scheme is used for protecting the data in a secured and private manner?
(5) How to handle the computational complexity efficiently?
(6) How to integrate the encryption scheme with parallel cloud architecture in case of emergency conditions?
(7) How to construct an efficient keyword search mechanism in storage services?
(8) How to offer identity and location-based privacy for the CC environment?

10 Conclusion

This survey provides a comprehensive analysis of privacy preservation factors and security establishment over the cloud storage system. Here, a baseline analysis of the cloud performance at the organizational level, digital economy, and transformation are discussed with the privacy preservation framework. An extensive analysis of the data security elements in cloud storage like confidentiality, integrity, fine-grained access, availability, access control, data sharing, leakage resistivity, delegation, and privacy protection methods are performed. Some encryption concepts like ABE, IBE, and homomorphic encryption models are evaluated by analysing the significance of the modelling and validating whether the security is achieved notably. However, some research constraints need to be resolved which is a hot research topic in the field of security and privacy preservation in the cloud. In future, this research is extended to provide a solution to ensure the security and privacy of the data hierarchically while data sharing occurs. Thus, a constructive hierarchical data sharing (CHDS) method is proposed with the adoption of symmetric encryption over the rooted hierarchical graph structure. The hierarchical graph model deals with the features of incoming data to establish the privacy and authenticity of the model.

References

1. Y. Zhang, Research on the security mechanism of cloud computing service model. Autom. Control Comput. Sci. **50**(2), 98–106 (2016)
2. P.G. Shynu, K.J. Singh, A comprehensive survey and analysis on access control schemes in the cloud environment. Inf. Technol. **16**(1), 19–38 (2016)

3. D. Stevenson, J. Pasek, Privacy concern, trust, and desire for content personalization, in *Proceedings of Research Conference on Communication, Information and Internet Policy* (2015), pp. 1–30
4. R.K. Aluvalu, L. Muddana, A survey on access control models in cloud computing, in *Proceedings of 49th Annual Convention of Computer Society of India (CSI)*, vol. 1 (2015), pp. 653–664
5. J. Li, Y. Zhang, X. Chen, Y. Xiang, Secure attribute-based data sharing for resource-limited users in cloud computing. Comput. Sec. **72**, 1–2 (2018)
6. R. Zhang, H. Ma, Y. Lu, Fine-grained access control system based on fully outsourced attribute-based encryption. J. Syst. Softw. **125**, 344–353 (2017)
7. A. Beimel, A. Ben-Efraim, Multi-linear secret-sharing schemes, in *Proceedings of Theory of Cryptography Conference*. Lecture Notes Comput. Sci. **8349**, 394–418 (2014)
8. M. Bellare, D. Hofheinz, E. Kiltz, Subtleties in the definition of IND-CCA: when and how should challenge decryption be disallowed? J. Cryptol. **28**(1), 29–48 (2015)
9. L. Xue, Y. Yu, Y. Li, M.H. Au, X. Du, B. Yang, Efficient attribute-based encryption with attribute revocation for assured data deletion. Inf. Sci. **479**, 640–650 (2019)
10. A. Balu, K. Kuppusamy, Ciphertext-policy attribute-based encryption with user revocation support, in *Proceedings of International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness* (2013), pp. 696–705
11. J. Wei, W. Liu, X. Hu, Secure and efficient attribute-based access control for multiauthority cloud storage. IEEE Syst. J. **12**(2), 1731–1742 (2018)
12. Q. Li, J. Ma, R. Li, X. Liu, J. Xiong, D. Chen, Secure, efficient and revocable multi-authority access control system in cloud storage. Comput. Sec. **59**, 45–59 (2016)
13. K. Yang, X. Jia, Expressive, efficient, and revocable data access control for multi-authority cloud storage. IEEE Trans. Parallel Distrib. Syst. **25**(7), 1735–1744 (2014)
14. J. Hao, C. Huang, J. Ni, H. Rong, M. Xian, X.S. Shen, Fine-grained data access control with attribute-hiding policy for cloud-based IoT. Comput. Netw. **153**, 1–10 (2019)
15. J. Li, W. Yao, Y. Zhang, H. Qian, J. Han, Flexible and fine-grained attribute-based data storage in cloud computing. IEEE Trans. Serv. Comput. **10**(5), 785–796 (2017)
16. N. D. Hua, M.J. Feng, Enhanced cloud storage access control scheme based on an attribute. J. Commun. **34**(Z1) (2013)
17. Z. Wang, D. Huang, Y. Zhu, B. Li, C.-J. Chung, Efficient attribute-based comparable data access control. IEEE Trans. Comput. **64**(12), 3430–3443 (2015)
18. J. Li, X. Huang, J. Li, X. Chen, Y. Xiang, Securely outsourcing attribute-based encryption with checkability. IEEE Trans. Parallel Distrib. Syst. **25**(8), 2201–2210 (2014)
19. Z. Liu, Z. Cao, D.S. Wong, White-box traceable ciphertext-policy attribute-based encryption supporting any monotone access structures. IEEE Trans. Inf. Forensics Sec. **8**(1), 76–88 (2013)
20. H.-J. Seo, H.-W. Kim, Attribute-based proxy re-encryption with a constant number of pairing operations. Int. J. Inf. Commun. Eng. **10**(1), 53–60 (2012)
21. H. Li, L. Pang, Efficient and adaptively secure attribute-based proxy re-encryption scheme. Int. J. Distrib. Sens Netw. **12**(5) 2016. Article No. 5235714
22. H. Wang, L. Wu, Unbounded anonymous hierarchical identity-based encryption in the standard model. J. Netw. **9**(7), 1846–1853 (2014)
23. G. Wang, Q. Liu, J. Wu, M. Guo, Hierarchical attribute-based encryption and scalable user revocation for sharing data in cloud servers. Comput. Sec. **30**(5), 320–331 (2011)
24. Z. Wan, J. Liu, R.H. Deng, HASBE: a hierarchical attribute-based solution for flexible and scalable access control in cloud computing. IEEE Trans. Inf. Forensics Sec. **7**(2), 743–754 (2012)
25. S. Chenthara, K. Ahmed, H. Wang, F. Whittaker, Security and privacy-preserving challenges of e-Health solutions in cloud computing. IEEE Access **7**, 74361–74382 (2019)
26. J. Zhang, B. Chen, Y. Zhao, X. Cheng, F. Hu, Data security and privacy-preserving in edge computing paradigm: survey and open issues. IEEE Access **6**, 18209–18237 (2018)
27. M. Adjedj, J. Bringer, H. Chabanne, B. Kindarji, Biometric identification over encrypted data made feasible, in *Proceedings of 5th International Conference on Information Systems Security* (2009), pp. 86–100

28. D. Cash, J. Jaeger, S. Jarecki, C.S. Jutla, H. Krawczyk, M.-C. Rosu, M. Steiner, Dynamic searchable encryption in very-large databases: data structures and implementation, in *Proceedings if NDSS Symposium* (2014), pp. 23–26
29. S. Kamara, C. Papamanthou, T. Roeder, Dynamic searchable symmetric encryption, in *Proceedings of ACM Conference on Computer and Communications Security* (2012), pp. 965–976
30. B. Zhu, B. Zhu, K. Ren, PEKsrand: providing predicate privacy in public-key encryption with keyword search, in *Proceedings of IEEE International Conference on Communications*, June 2011, pp. 1–6
31. W. Sun, S. Yu, W. Lou, Y.T. Hou, H. Li, Protecting your right: verifiable attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud. IEEE Trans. Parallel Distrib. Syst. **27**(4), 1187–1198 (2016)
32. M.R. Clark, K. Stewart, K. Stewart, Dynamic, privacy-preserving decentralized reputation systems. IEEE Trans. Mob. Comput. **16**(9), 2506–2517 (2017)
33. N. Busoma, R. Petrlic, F. Sebé, C. Sorge, M. Valls, A privacy-preserving reputation system with user rewards. J. Netw. Comput. Appl. **80**, 58–66 (2017)
34. Y. Lai, Z. Liu, Q. Pan, J. Liu, Study on cloud security based on trust spanning tree protocol. Int. J. Theor. Phys. **54**, 3311–3330 (2015)
35. Q.I. Xia, E.B. Sifah, K.O. Asamoah, J. Gao, X. Du, M. Guizani, MeDShare: trust-less medical data sharing among cloud service providers via blockchain. IEEE Access **5**, 14757–14767 (2017)
36. V.V. Rajendran, S. Swamynathan, Hybrid model for dynamic evaluation of trust in cloud services. Wirel. Netw. **22**(6), 1807–1818 (2016)
37. C. Xu, J. Wang, L. Zhu, C. Zhang, K. Sharif, PPMR: a privacy-preserving online medical service recommendation scheme in eHealthcare system. IEEE Internet Things J. **6**(3), 5665–5673 (2019)
38. Y. Dou, H.C.B. Chan, M.H. Au, A distributed trust evaluation protocol with privacy protection for intercloud. IEEE Trans. Parallel Distrib. Syst. **30**(6), 1208–1221 (2019)
39. F.A.M. Ibrahim, E.E. Hemayed, Trusted cloud computing architectures for infrastructure as a service: survey and systematic literature review. Comput. Sec. **8**(2), 196–226 (2019)
40. L. Chen, R. Urian, DAA-A: direct anonymous attestation with attributes, in *Proceedings of 8th International Conference on Trust and Trustworthy Computing (TRUST)*, Heraklion, Greece, August 2015, pp. 228–245
41. I. Khalil, A. Khreishah, M. Azeem, Consolidated identity management system for secure mobile cloud computing. Comput. Netw. **65**(2), 99–110 (2014)
42. M. Chen, W. Li, Z. Li, S. Lu, D. Chen, Preserving location privacy based on distributed cache pushing, in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Istanbul, Turkey, April 2014, pp. 3456–3461
43. Z. Brakerski, V. Vaikuntanathan, Efficient fully homomorphic encryption from (standard) LWE. SIAM J. Comput. **43**(2), 831–871 (2014)
44. M. Sookhak, F.R. Yu, M.K. Khan, Y. Xiang, R. Buyya, Attribute-based data access control in mobile cloud computing: taxonomy and open issues. Fut. Gener. Comput. Syst. **72**, 273–287 (2017)
45. G. Ateniese, K. Fu, M. Green, S. Hohenberger, Improved proxy re-encryption schemes with applications to secure distributed storage. ACM Trans. Inf. Syst. Sec. **9**(1), 1–30 (2006)
46. J. Ni, X. Lin, X.S. Shen, Toward edge-assisted internet of things: from security and efficiency perspectives. IEEE Netw. **33**(2), 50–57 (2019)
47. S. Bragadeesh, U. Arumugam, A conceptual framework for security and privacy in edge computing, in *Edge Computing*. (Springer, 2019), pp. 173–186
48. W.Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, A. Ahmed, Edge computing: a survey. Futur. Gener. Comput. Syst. **97**, 219–235 (2019)
49. D. Liu, Z. Yan, W. Ding, M. Atiquzzaman, A survey on secure data analytics in edge computing. IEEE Internet Things J. **6**(3), 4946–4967 (2019)

Nature-Inspired Feature Selection Algorithms: A Study

D. Mahalakshmi, S. Appavu Aalias Balamurugan, M. Chinnadurai, and D. Vaishnavi

Abstract In this digital era, the amount of data generated by various functions has increased dramatically with each row and column; this has a negative impact on analytics and will increase the liability of computer algorithms that are used for pattern recognition. Dimensionality reduction (DR) techniques may be used to address the issue of dimensionality. It will be addressed by using two methods: feature extraction (FE) and feature selection (FS). This article focuses on the study of feature selection algorithms, which includes static data. However, with the advent of Web-based applications and IoT, the data are generated with dynamic features and inflate at a rapid rate, thus it is prone to possess noisy data, which further limits the algorithm's efficiency. The scalability of the FS strategies is endangered as the size of the data collection increases. As a result, the existing DR methods do not address the issues with dynamic data. The utilization of FS methods not only reduces the load of the data, but it also avoids the issues associated with overfitting.

Keywords Dimensionality reduction · Feature selection · Feature extraction · Optimization · Machine Learning

D. Mahalakshmi (✉)
Department of Information Technology, AVC College of Engineering, Mayiladuthurai, India

S. A. A. Balamurugan
Department of Computer Science, Central University of Tamil Nadu, Thiruvarur, Tamil Nadu, India

M. Chinnadurai
Département of Computer Science and Engineering, EGS Pillay Engineering College (AUTONOMOUS), Nagapattinam, Tamil Nadu, India

D. Vaishnavi
Department of CSE, SRC, SASTRA Deemed University, Thanjavur, Tamilnadu, India
e-mail: vaishnavi@src.sastra.edu

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_55

1 Introduction

Due to the exponential rise of data, the quality of information (data) is critical for gradually incorporating the data processing by pattern recognition, data mining, image processing, and different ML techniques. Bellman calls this situation as "Curse of Dimensionality." Large-dimensional data may lead to the predominance of noisy, immaterial, and redundant data, which causes overfitting in the ML model. To deal with these issues, dimensionality reduction methods could be appropriate, and it should be a preprocessing step to apply ML algorithms. Feature selection (FS) and feature extraction (FE) are the commonly used dimensionality reduction methods, where FS scrubs up the noisy, redundant, and immaterial data, thus it might boost the efficiency of an algorithm [1]. Hence, this article focuses on the study of various feature selection algorithms.

Primarily, the features are uniquely chosen based on feature redundancy and relevance. The relevancy of the function is measured based on the traits of the data. Statistics is one such approach, which reveals the connection between the options and their importance. The distortion of insignificant and redundant characteristics is not due to the existence of useless data; it is due to the feature that does not have a statistical link with other features. Any characteristic might be inconsequential on its own, but it is undoubtedly linked when combined with other features [2]. Due to technology improvement, conventional algorithms (CAs) have failed to satisfy the criteria of the complex challenges specified below [3]:

- Issues that are nonlinear, discontinuous, and multimodal are tough to solve.
- CA primarily confines native search.
- Most often CAs are problem-specific as a result of they typically use sequel information for solving.
- In CA, the ultimate options will likely be too same as when equivalent preliminary factors are considered. Thereby, CA lacks a variety of the obtained solutions.

To overcome these limitations of the CA, up-to-date algorithms, viz. heuristic and metaheuristic, have been proposed. The strategy of heuristic algorithms is to make use of the trial-and-error methodology in producing new solutions, whereas metaheuristic algorithms use recollection of a historical past of options and improvise the elucidation, i.e., learning approach. Therefore, metaheuristic algorithms are additionally known as high-level heuristic algorithms. Within the current past, more importantly of those metaheuristic algorithms are nature-inspired algorithms that toil on the swarm intelligence (SI) [4]. Nature has all the time been an amazing inspiration and motivation for among the researchers which led to the event of algorithms primarily based on the swarm habits of several social organisms loves fireflies, bees, cockroaches, mosquitoes, bacteria, pigeons, etc., to find options to optimization issues [5].

Generally, SI is a department of artificial intelligence expressing the collective habits of organisms that are self-organized. To imitate such behavior, the SI algorithm generates inhabitants which consist of randomly deployed agents. Thus, on this

methodology, these agents work together domestically with one more and likewise with the encompassing surroundings. The intention of deploying these agents is to create a certain amount of random behavior. Thereby, there is not a centrally managed construction dictating the person's behavior [5]. Moreover, these deployed agents hold clever habits whereas interacting with different to same as agents. It is noticed that within the continuous 25 years, a lot of nature-inspired algorithms have been developed. More importantly, those algorithms possessed some distinctive skills akin to randomness throughout every execution, independence of preliminary values, and likewise uniqueness in the trail it travels to search out the solution.

2 Recent Works: 2009–Present

It is demonstrated from the state-of-art that the FS can increase the generalization capability and performance of scalability, prediction rate of the classifier. In the discovery of knowledge, FS takes a basic part in decreasing the storage, cost, and computational complexity [1]. In this section, we are going to discuss the various FS algorithms from the period of the year 2009 that can be applied in various fields of applications in chronological order, and it is precisely displayed in Table 1.

Yang [6] very firstly formulated the Firefly algorithm (FA), which is a meta-heuristic algorithm within the period of 365 days, 2007 and through 2008 and 2009, it has been improvised by the flashing behavior of fireflies and the manifestation of bioluminescent statement. FA is developed with the subsequent assumptions: (1) A firefly might be attracted to one another no matter their sex as a result of their epicene. (2) Desirability is proportionate to their brightness, whereas a reduced amount of vivid firefly will be interested in the brighter firefly. Nevertheless, the desirability dropped when the space of the two fireflies enlarged. (3) If the brightness of each firefly is the same, the fireflies will transfer randomly. The generations of recent options are by random stroll and desirability of the fireflies [7]. The intensity of the fireflies ought to be related to the target performance of the associated problem. Its desirability has been successful by splitting up into smaller teams and every sub-team swarm across the native models. Thus, FA is appropriate for optimization issues as said by [8]. Similarly, in the same year, Cuckoo search and algorithms are formulated [9, 10], respectively. These algorithms are innovated by the inspiration of bird's behavior.

In the year 2010, the Dove swarm optimization echolocation habits of bats have been designed as Bat algorithm (BA), and this is preliminary research that illustrates that this algorithm could be very hopeful [11]. The potential of echolocation of microbats is attractive as these bats can discover their victim and discriminate against various kinds of bugs even in total dimness. For simplicity, we now use the next imprecise protocols: All bats observe echolocation to feel the distance, and so they additionally "know" the distinction of background boundaries and foods in a mystic aspect. Bats fly arbitrarily in relation to rapidity v_i at place x_i with a hard and fast frequency $f_{\min}$, various land wavelength loudness A_0 to seek for edibles. They can

Table 1 Nature-inspired algorithms for feature selection

Year	Name of the algorithm	Inspired by	Advantages	Applications areas
2009	Firefly Algorithm Cuckoo search algorithm Dove Swarm Optimization	Insects and reptiles Birds Birds	High Convergence speed Efficient global optimum	Image processing and compression Scheduling Feature selection Data Fusion
2010	Bat Algorithm Cockroach Swarm Optimization Marriage-in-honey Bees Optimization Eagle Strategy Hunting Search	Insects and reptiles Insects and reptiles Insects and reptiles Birds Animals	Very quick convergence at a very initial stage	Clustering Image matching
2011	Stochastic Diffusion Search Modified Honey-Bees Mating Optimization	Insects and reptiles Insects and reptiles	Provides facility of dynamic search space	Feature selection Recruitment search resource allocation object recognition
2012	Fruit Fly Optimization SBA Superbug Algorithm BSO Bees Swarm Optimization BCO Bacterial Colony Optimization Krill Herd Algorithm	Insects and reptiles Insects and reptiles Insects and reptiles Insects and reptiles Animals	Low convergence precision and easily relapsing into local extremum	Remote sensing data fusion Clustering Data hiding
2013	Green-Herons Optimization Algorithm	Birds	adaptability for dimensionless	Clustering dynamic and adaptive problems
2014	Pigeon Optimization Chicken swarm Optimization Dispersive Flies Optimization	Birds Birds Insects and reptiles	Fast convergence	Feature extraction Image fusion Optimization in robotics
2015	Dragonfly algorithm Ant Lion Optimizer Elephant Herding Optimization	Insects and reptiles Insects and reptiles Birds	Few parameters used for tuning Convergence time is nominal	Segmentation Image Data hiding Classification Classifier design Feature selection
2016	Lion Optimization Algorithm Camel Algorithm	Animals Animals	Convergence time is nominal	Feature selection

(continued)

Table 1 (continued)

Year	Name of the algorithm	Inspired by	Advantages	Applications areas
2017	Human behavior-based Optimization Gartener Snake Optimization Laying Chicken Algorithm	Humans Insects and reptiles Birds	Increase the diversity of population Flexible Simple	Feature selection Classification Engineering problem
2018	Crow Search Algorithm Queuing Search	Birds Humans	Efficient convergence and accuracy Increase speed of computation	Feature selection, cloud, neural network and machine learning
2019	Butterfly Optimization Algorithm Emperor Penguins Colony Artificial Feeding Birds	Insects and reptiles Insects and reptiles Birds	Avoid local optima	Feature selection Mechanical design optimization

mechanically regulate the wavelength of their emitted pulses and alter the charge of pulse emission r with the range of [0, 1], relying on the relation of their target. Even though it can differ in a lot of perceptions, we adopt the loudness diverges into a big (positive) A_0 to a small fixed rate $A_{\min}$. Similar to bat algorithm, Cockroach Swarm Optimization, Marriage in honey Bees Optimization, Eagle Strategy, and Hunting Search are innovated by the inspiration of insets and reptiles, birds, and animals, respectively.

By the inspiration of bugs, Stochastic diffusion search (SDS) was initially referred as a population-based algorithm in the year 2011 with impression of Bishop [12] which makes use of straightforward communication method equivalent to cooperative transport discovered among social bugs, to carry out analysis of the hunt and optimization hypothesis. SDS is, more importantly, relevant to optimization and hunt issues, and to do this optimization, the place element features may be computed along by utilizing the swarm agents, which preserves speculation with regards to the optima [13]. In comparison with different nature-inspired methods, SDS has a robust mathematical framework defining the behavior of the algorithm in the period of convergence to linear complexity, marginal convergence, and global optimum conditions. Basically, SDS is relevant to altering goal features thus scattering to progressively altering scenarios. This function makes SDS helpful to unravel varied issues throughout different fields similar to eye monitoring in facial pictures through the use of a mix of a stochastic search community and an n-tuple community, Web site resolution for transmission gear for Wi-Fi networks and mouth finding in human

facial pictures [14]. Modified Honey-Bees Mating Optimization was also developed in the same year to enhance the performance of Honey bee optimization [15].

The Fruit Fly Optimization Algorithm (FOA) was invented in the year 2012, which locates world optimization problems primarily based on the meals discovering conduct of the fruit fly [16]. The fly itself is superior to different species in sensing and notion, for the most part in vision and osphresis. The osphresis fruit flies can discover all types of scents floating within the air; it will probably even scent meals supply from 40 km aside. Then, later, it will get near the location of the meal, it will probably additionally custom its delicate imaginative and prescient to search out meals and the company's assembling location and fly in the route of that route too. Bees Swarm Optimization [17], Bacterial Colony Optimization [18], Migrating-Birds Optimization Algorithm [19], and Krill Herd Algorithm [20] are derived in the same year.

Green Heron Swarm Optimization (GHOSA) Algorithm was launched in the year 2013, which is inspired by the fishing expertise of the bird [21]. The algorithm is principally suited to graph primarily based issues similar to combinatorial optimization etc. Total sequence of the GHOSA might be divided into the next fundamental operational steps which can carry out a different search based on mostly variations for a heuristic sequence or path technology and can thus set up an answer for the issues like graph-based problems and combinatorial optimization issues. Anyhow, this algorithm might be prolonged for discrete equations but with the restricted capability of the algorithm. In this context, the drawback of combinatorial optimization is that every one of the people is an occasion, whereas in equations, every one of the people is parameter values. However, the individual restraints are not enough distinctive for every type of issue required to be established into the computation, and the operations are only pointers of what must be occurring with the answer set individually or as an entire.

Pigeon-inspired optimization (PIO) is proposed by Duan and Qiao in 2014 based on a swarm intelligence technique [1]. The unique objective of the algorithm is to simulate pigeons' homing behavior utilizing attractive fields and landmarks as inputs. Two operators are intended to utilize the rues to be able to idealize a few homing traits of pigeons and they are: (1) Map and compass operator: pigeons can identify the world discipline through the use of magnetoreception to form the route map into their brains and compass used to regulate the orientation of fly. (2) Landmark operator: The pigeons can fly to their destination by acquainted to nearby landmark. If they are removed from the vacation spot and exotic landmarks, they'll observe the pigeons who are accustomed to the landmarks. It produces a very feasible solution to engineering problems and specifically on robotics. Chicken Swarm Optimization (CSO) is introduced by mimicking the hierarchal order within the rooster swarm. As it works based on the behaviors of the hen swarm, together with roosters, it can efficiently reach the swarm intelligence to solve the optimize problems [22]. Dispersive Flies Optimization (DFO) whose inspiration is the swarming behavior of flies over meals sources [23].

Dragonfly Algorithm (DA) was primarily inspired through the looking and migration habits of the dragonfly during the year 2015. DA is imitating the swarming habits

of a dragonfly. The explanation for his or her swarming is either migration (dynamic swarm) or looking (static swarm, respectively). In a static swarm, limited teams of dragonflies transfer over a microscopic space to search different flies and the behaviors of any such swarming exemplify native actions and hasty deviations. Conversely, in dynamic swarming, a large variety of dragonflies produce a single group and transfer in the direction of one path for an extended distance [20]. The aforesaid swarming behaviors are reckoned as the principal inspiration of DA. According to the exploration and exploitation phases of the metaheuristic optimization algorithm, static and dynamic swarming behaviors are classified, respectively. To direct manmade dragonflies to numerous paths, five weights have been used, that is, alignment weight, cohesion weight, separation weight meals weight, inertia weight, and the enemy weight. To discover the search space, low-cohesion and high-alignment weights are used. Furthermore, to switch amid exploration and exploitation, the neighborhood range overstated regularly to the variety of iterations that have been used. Likewise, Ant Lion Optimizer [24] and Elephant Herding Optimization [25] algorithms have been inspired by the insects and the animals, respectively.

The Lion Optimization Algorithm (LOA), was launched during the days of 2016, in which the lifestyle of lions and its teamwork traits has been progressed. A preliminary inhabitant is fashioned by a set of randomly generated options referred to as Lions. The preliminary inhabitants (%N) are classified into nomad lions and relaxation inhabitants are classified as resident lions which can randomly categorized into a points of pride %S. The members of pride are regarded as female and relaxation is regarded as male. For every lion, one of the best-obtained resolutions during the repetitions is named greatest visited position.

In 2017, Human Behavior-based PSO (HPSO) was introduced by mimicking human habits by announcing the worldwide worst speck, who's of the worst health in the complete inhabitants at every repetition. All particles solely are taught from one of the best particles Pbest and Gbest. It is an excellent social condition [26]. However, contemplating human behavior, there exist some individuals who possess bad habits or behaviors around us. On a similar time, as all of us identified that these bad habits or behaviors will deliver some results on individuals around them. If we grasp warning from these bad habits or behaviors, it is helpful to us. Conversely, if we are taught from these bad habits or behaviors, it is dangerous to us. Therefore, we should provide a goal and rational view on these bad habits or behavior. To simulate human habits, the worldwide worst speck was launched, and the educational coefficient which adopts the usual steadiness, exploitation and exploration skills by altering the flying route of specks. It is known as impelled leaning coefficient, if it is positive and it can be useful to augment the exploration ability. Otherwise, it is named penalized studying, and it can be useful for enhancing the exploitation ability. Also, during this same year, Gartener Snake Optimization [27] and Laying Chicken Algorithm [28] are introduced by mimicking the habits of reptiles and birds correspondingly.

A brand new population-based algorithm referred to as Crow Search Algorithm (CSA) was recommended by Askarzadeh in 2018, which simulates the hiding of foods habits of crow. Crow is a clever fowl that may bear in mind faces and tip-off its species in risk. Probably the most proof of their intelligence is hiding foods

and bear in mind their location. Within the optimization theory, the crow is the seeker, the encompassing surroundings are the search space, and arbitrarily storing the placement of meals is a possible resolution. Among all foods locations, the place where essentially the most foods is accumulated is taken into account to be the worldwide optimum solution, and the quantity of food is considered as an objective function. By simulating the clever habits of crows, CSA tries to seek out optimum decisions to numerous optimization problems. It has extended a substantial curiosity due to its advantages in terms of flexibility, ease of employment, and a variety of other factors. A novel metaheuristic algorithm referred to as queuing search (QS) is also proposed in the same year, which is inspired by human actions in queuing process [29]. QS must not preset the supplementary parameters excluding the population dimension and stopping criterion.

Butterfly Optimization Algorithm (BOA) is a nature-stimulated meta-heuristics procedure that mimic the butterflies' pure hunting and mating habits [30]. BOA works on the basis of perfume emitted by them, which in turn helps to look out meals in addition to mating partners. BOA performs a global search in addition to local search while on the lookout for the optimum resolution for the problem. In this algorithm, the butterfly's traits are concerned: (1) Unique perfume emission of butterflies helps to draw in directions of every other. (2) Hunting behavior is randomly or directly the perfume emission. (3) Goal function's panorama is utilized in figuring out the stimulus intensity of butterflies. Another brand new metaheuristic algorithm named Emperor Penguins Colony (EPC) is proposed with an impression of emperor penguins [31]. Artificial Feeding Birds (AFB) are inspired by the very insignificant habits of birds looking for food [32]. AFB may be simple, well-organized, and will be simply tailored for the diverse optimization glitches.

3 Conclusion

Since we are in the digital world, each second generates millions, billions of bytes of data, which may increase the load for data processing and, as a result, have an influence on the application of choice. It makes it difficult for researchers to provide you with the ideal function selection mannequin that suits any software regardless of restrictions. Dimensionality reduction delivers many advantages: it leads to a low-dimensional model, less reminiscence space, moderates the danger of overfitting, higher accuracy, and decreases the time complexity. Researchers have demonstrated it by using numerous experimental assessments performed on the static dataset. Certainly, many applications create dynamic and real data, which frequently drives the ideas. As a result, there is a scope to comprehend the idea and propose relevant dimensionality reduction models.

References

1. B. Venkatesh, J. Anuradha, A review of feature selection and its methods. Cybern. Inf. Technol. **19**, 3–26 (2019)
2. I.A. Gheyas, L.S. Smith, Feature subset selection in large dimensionality domains. Pattern Recogn. **43**, 5–13 (2010)
3. V.R. Balasaraswathi, M. Sugumaran, Y. Hamid, Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms. J. Commun. Inf. Netw. **2**, 107–119 (2017)
4. A. Varun, M. Sandeep Kumar, A comprehensive review of the pigeon-inspired optimization algorithm. Int. J. Eng. Technol. **7** (2018). www.sciencepubco.com/index.php/IJET
5. S. Colaco, S. Kumar, A. Tamang, V.G. Biju, A review on feature selection algorithms, in *Advances in Intelligent Systems and Computing*, vol. 906 (Springer Verlag, 2019), pp. 133–153
6. X.-S. Yang, Nature-inspired mateheuristic algorithms: success and new challenges. J. Comput. Eng. Inf. Technol. **1** (2012)
7. X.S. Yang, S. Koziel, Computational optimization: an overview. Stud. Comput. Intell. **356**, 1–11 (2011)
8. X.-S. Yang, Firefly algorithm, stochastic test functions and design optimisation. Int. J. Bio-Inspired Comput. **2**, 78–84 (2010)
9. X.S. Yang, S. Deb, Cuckoo search via Lévy flights, in *2009 World Congress on Nature and Biologically Inspired Computing, NABIC 2009—Proceedings* (2009), pp. 210–214. https://doi.org/10.1109/NABIC.2009.5393690.
10. M.C. Su, S.Y. Su, Y.X. Zhao, A swarm-inspired projection algorithm. Pattern Recogn. **42**, 2764–2786 (2009)
11. X.S. Yang, Bat algorithm: literature review and applications. Int. J. Bio-Inspired Comput. **5**, 141–149 (2013)
12. H. Williams, M. Bishop, Stochastic diffusion search: a comparison of swarm intelligence parameter estimation algorithms with RANSAC. Algorithms **7**, 206–228 (2014)
13. S. Maroufpoor, R. Azadnia, O. Bozorg-Haddad, Stochastic optimization: Stochastic diffusion search algorithm, in *Handbook of Probabilistic Models* (Elsevier, 2019), pp. 437–448. https://doi.org/10.1016/B978-0-12-816514-0.00017-5
14. M.M. al-Rifaie, J.M. Bishop, Stochastic diffusion search review. Paladyn J. Behav. Robot. **4** (2015)
15. T. Niknam, S.I. Taheri, J. Aghaei, S. Tabatabaei, M. Nayeripour, A modified honey bee mating optimization algorithm for multiobjective placement of renewable energy resources. Appl. Energ. **88**, 4817–4830 (2011)
16. W.T. Pan, A new fruit fly optimization algorithm: taking the financial distress model as an example. Knowledge-Based Syst. **26**, 69–74 (2012)
17. M. Talo, O. Yildirim, U.B. Baloglu, G. Aydin, U.R. Acharya, Convolutional neural networks for multi-class brain disease detection using MRI images. Comput. Med. Imaging Graph. **78**, 101673 (2019)
18. B. Niu, H. Wang, Bacterial colony optimization. Discret. Dyn. Nat. Soc. **2012** (2012)
19. E. Duman, M. Uysal, A.F. Alkaya, Migrating birds optimization: a new meta-heuristic approach and its application to the quadratic assignment problem, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6624 LNCS (2011), pp. 254–263
20. A.H. Gandomi, A.H. Alavi, Krill herd: a new bio-inspired optimization algorithm. Commun. Nonlinear Sci. Numer. Simul. **17**, 4831–4845 (2012)
21. C. Sur, A. Shukla, Green heron swarm optimization algorithm-state-of-the-art of a new nature inspired discrete meta-heuristics
22. X. Meng, Y Liu, X. Gao, H. Zhang, A new bio-inspired algorithm: Chicken swarm optimization. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) **8794**, 86–94 (2014)

23. (6) (PDF) Dispersive Flies Optimisation. https://www.researchgate.net/publication/267514160_Dispersive_Flies_Optimisation
24. S. Mirjalili, The ant lion optimizer. Adv. Eng. Softw. **83**, 80–98 (2015)
25. G.G. Wang, S. Deb, L.D.S. Coelho, Elephant herding optimization, in *Proceedings—2015 3rd International Symposium on Computational and Business Intelligence, ISCBI 2015* (Institute of Electrical and Electronics Engineers Inc., 2016), pp. 1–5. https://doi.org/10.1109/ISCBI.2015.8
26. H. Liu, G. Xu, G.Y. Ding, Y.B. Sun, Human behavior-based particle swarm optimization. Sci. World J. **2014** (2014)
27. M. Naghdiani, M. Jahanshahi, GSO: a new solution for solving unconstrained optimization tasks using garter snake's behavior, in *Proceedings—2017 International Conference on Computational Science and Computational Intelligence, CSCI 2017* (Institute of Electrical and Electronics Engineers Inc., 2018), pp. 328–333. https://doi.org/10.1109/CSCI.2017.55
28. E. Hosseini, Laying chicken algorithm: a new meta-heuristic approach to solve continuous programming problems (2017).https://doi.org/10.4172/2168-9679.1000344
29. J. Zhang, M. Xiao, L. Gao, Q. Pan, Queuing search algorithm: a novel metaheuristic algorithm for solving engineering optimization problems. Appl. Math. Model. **63**, 464–490 (2018)
30. S. Arora, S. Singh, Butterfly optimization algorithm: a novel approach for global optimization. Soft Comput. **23**, 715–734 (2019)
31. S. Harifi, M. Khalilian, J. Mohammadzadeh, S. Ebrahimnejad, Emperor Penguins Colony: a new metaheuristic algorithm for optimization. Evol. Intell. **12**, 211–226 (2019)
32. J.B. Lamy, Artificial feeding birds (AFB): a new metaheuristic inspired by the behavior of pigeons, in *EAI/Springer Innovations in Communication and Computing* (Springer Science and Business Media Deutschland GmbH, 2019), pp. 43–60. https://doi.org/10.1007/978-3-319-96451-5_3

High-Speed Antenna Selection for Underwater Cognitive Radio Wireless Sensor Networks

S. Sankar Ganesh and S. Rajaprakash

Abstract The underwater acoustic sensor network (UCASN) is used to monitor amphibian climate. The submerged hubs request for continual and ceaseless inspection. The submerged hubs are placed at various depths to stimulate the information transmission from the water depth sensor to floats hub. Regardless, the submerged organization has limited organizational assets. For example, battery, data transfer capacity and helpless in portability, heat-graphic conditions and spread postponement. The resulting delay and heat-graphic condition drags out the packet node's preparation and delays network operation. Deferred network activity increases hub battery usage and overhead. As a result, a bio-powered Dolphin Swarm Optimization Algorithm (DSA) is executed to find the optimal receiving wire for information transmission. DSA selects the ideal radio wire for information correspondence by assessing affirmation parcel obtained for the information sent through accessible receiving wires. The DSA was recreated in NS2-Aquasim climate and test bed to assess its presentation. The DSA improved information possess unwavering quality.

Keywords UCASN · Thermographic condition · DSA · SST

1 Introduction

The upper sea partners on a surface level by constraining the wind, heat, and new water by the extended calm sea, where this warmth and new water are separated and released through time and on a global scale. The surface layer consists of a higher blended layer influenced by the climate and a profoundly separated zone underneath the blended layer. The surface layer may be exposed to air sooner or later; the water is mostly legally exposed inside the blended layer. As a result, they reflect that the vertical structure of surface layer does not change rapidly to drive the climate but rather evolves due to earlier constraints. These driving occasions may have happened either locally in the area, or distantly at different areas and moved by sea flows. This

S. Sankar Ganesh (✉) · S. Rajaprakash
Department of CSE, Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation, Chennai, Tamil Nadu, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_56

upper ocean vertical design portrays the traits of upper-ocean vertical construction and discusses what impacts and how looks after them. We show numerous instances of shapes fluctuation by environmental compelling. This upper-ocean vertical design portrays the traits of upper-ocean vertical construction. These upward constructions can be taken by the assortment.

Author proposed the mechanism of the multiple input and multiple output (MIMO) communication for underwater cooperative/distributed. It is used to apply the techniques for attractive induction and mixture acoustic. Attractive induction used to tackle the synchronization problem for inter node, and reduce the errors and time of synchronization [1]. Authors discuss about the communication for MIMO; and it is used to analyze the underwater visible light imaging for spatial correlation; and studied about the performance of bit error rate and capacity with the help of MIMO [2]. They have proposed the orthogonal frequency division multiplexing (OFDM) system. The results are compared with the performance of OFDM with repetition coding, MIMO, and Alamouti [3]. They studied about the survey about the previous research work for MIMO with OFDM and proposed the underwater acoustic. This communication still not have perfect algorithm from the research view [4]. The Authors have studied and proposed a communication method for underwater MIMO. The results are compared with the same receiver type, which doesnt use the joint sparse mechanism [5]. The proposed algorithm is power allocation, and it is used to minimize the spatial modulations and bit error rate, and investigate the performance of the MIMO for underwater wireless optical and combined with both flag dual-amplitude pulse position and spatial modulations [6]. Author discusses about the system for multi-carrier and coded MIMO in communication of the underwater. It is used to detect the minimum mean square rate and zero forcing, and compare the performances [7]. They analyzed the On–Off keying module system with underwater wireless optical communication, and it is used to find the absent and present for MIMO, MISO, SIMO, and SISO [8]. Author genetic program used to design the controller of the MIMO with the help of underwater vehicles, and provide the robust UUV operation and stable UUV operation [9]. They mainly discuss about the underwater vehicle and to find the problem of the ideal perturbation signal, and considered the two signals like multi-level pseudorandom and random noise, and compared the performance for different signals [10]. Authors investigate about the theorems 1 and 2. These theorems are used to evaluate with the help of simulation by using remotely operated vehicle and compare the performances [11]. They detect the targets by using underwater acoustic waveguide with the help of MIMO model [12]. Author mainly focuses MIMO and proposed method is reducing the time complexity and space complexity from the receiver [13]. Author mainly investigates about the MIO technology. This technology used to brought the data rate gain for communication of UWV and compare the performance between SIMO and MIMO for data rate gain [14]. The proposed scheme is MIMO OFDM for communication of UWV. This scheme transmission has two modulations like QPSK and 16QAM and compares the performance [15].

2 Methodology

The proposed methodology is Dolphin Optimization Algorithm. This method is mainly focused on hunting behavior of dolphins. This dolphin has been used to search the environment location, generate sounds like shape, direction, and distance of prey and shown in Tables 1 and 2.

Table 1 Optimal Dolphin solution

Optimal Dolphin solution
$\text{Dol}_i = [X_1, X_2, X_3, \ldots, X_D]^{\text{T}}$ (1) where "$i = 1, 2, \ldots, N$," 'N' addresses several radio wires present in the hub for information correspondence, 'i' addresses the distinctive ideal choice of receiving wire for various water segment
$DK_i = \text{Dol}_i - K_i, i = 1, 2, \ldots, N$(2) where K_i addresses numerous arrangements accessible to an issue
$DKL_i = L_i - K_i, i = 1, 2, \ldots, N$(3) where L_i represents Boolean
$X_{\text{ot}} = \text{Dol}_i + V_j t$(4) $E_{ijt} = \text{Fitness}(X_{\text{ot}})$(5) where V_j addresses the arbitrary information transmission through various receiving wires, t addresses time taken to accomplish ideal arrangement, and the wellness capacity of ideal radio wire for information transmission
$L_i = X_{iab}$(6) where L_i is the initial data transmission
$\text{Fitness}(K_i) > \text{Fitness}(K_j)$(7) where K_i and K_j are evaluating the fitness function and the fruitful information transmission from source hub to objective hub

Table 2 Algorithm

Step 1: The 50% of available antennas used to send the data from source to destination node
Step 2: To determine the most successful data pack from destination node for data reception pack rate
Step 3: To calculate vector of the distance between neighborhood ideal and ideal arrangement
Step 4: The 100% of antenna used to build search span and communicate information
Step 5: To calculate the ideal receiving wire for information correspondence

3 Result and Discussion

The determination of reception apparatus for MIMO correspondence was re-enacted in NS2—network test system. Table 3 gives the organization boundaries of NS2. The exhibition of DOA assess for various number of receiving wires. At first, the receiving wire pooling starts with 25% of all out number of reception apparatuses. The pooling of receiving wire in steps limits calculation over-burden at the transmitter. The best receiving wire for information transfer is selected based on the bit blunder declared by the receiver hub. The transmitter reception device delivers a 1500-byte query information bundle. The touch blunder pace of information bundle increases because of sea thermographic properties, for example, water thickness, consistency and electrical conductivity. The adjustment in water thickness happens at various water sections, for example, Mesopelagic, Bathypelagic, and Abyssopelagic cause higher piece mistake rate because of water flow as per profundity. Moreover, the information transmission speed at shallow water is higher and as the profundity increases, the information transmission speed decreases significantly by leveraging spread postponement and touch mistake rate. Figure 1 shows the touch mistake rate variety for various reception apparatus choice. The information communicated through reception apparatus is selected by DOA from all out mix of radio wire, which causes negligible piece mistake rate contrasted with receiving wire mix selected from lower mix of reception apparatuses. The spot mistake rate is low for receiving wire selected from complete radio wires due to the reception apparatus positioning from affirmation bundles.

The bit error rate is low for radio wire chosen from all receiving wire mixes because of the decrease in charge bundle transmission and preparing among source and objective hub as in Fig. 2. The control bundle transmission works through the extended network, which will help in attaining the expected packet transmission. The receiving wire chosen from under 90% of all-out radio wires have higher SNR due to under intermingling on optimal radio wire by DSA. The source of the intermingling is a lack of space in the distance vector between the obtained perfect radio wire and the neighborhood arrangement reception wire. The optimal radio wire for information transmission is determined by the receiving wire determination from 100% of all-out radio wire.

Table 3 Network parameters

Network parameter	Value
Amount of nodes	100
Area of the total sensor	1000 m × 1000 m
Size of the packet control	12
Size of the packet data	30
Nodes of the energy	5j
Range of the communication node	200 m
Speed of the node	10 m/s

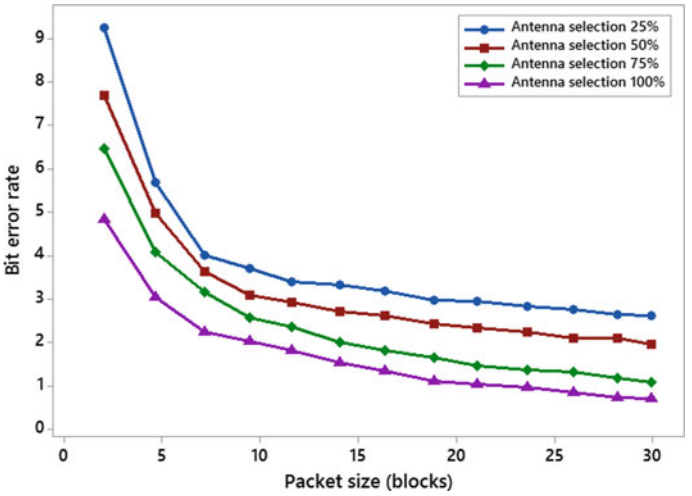


Fig. 1 Bitter error rate versus packet size

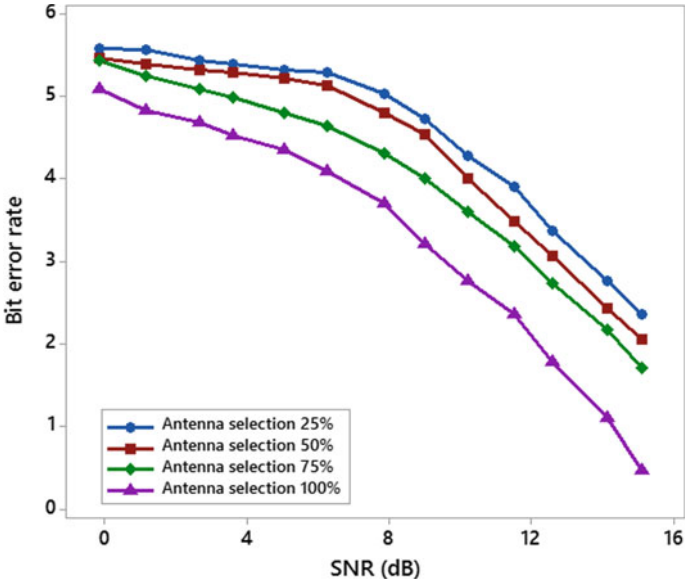


Fig. 2 Bit Error rate versus SNR

4 Conclusion

The detected information of submerged hubs acquire bit mistake due to the sea heat-graphic conditions. The heat-graphic conditions are dynamic at various water

segments of sea. The submerged hubs convey at various water sections in order to encourage information correspondence from fixed hubs at lower part of sea in order to float hubs. Be that as it may, due to the dynamic heat-graphic conditions, the touch blunder increment fundamentally causes bundle misfortune and better quality end-to-end delay. Henceforth, a bio-roused DSA was actualized to select the ideal receiving wire for achieving data transmission to improve the information unwavering quality. The DSA has selected the ideal receiving wire with negligible intermingling time, and it provides better information transmission to objective hub. The DSA has performed better as far as PDR, throughput, and energy utilization contrasted with conventional radio wire determination calculations, for example, BAB, Greedy, and Random inquiry.

References

1. Z. Li, S. Desai, V.D. Sudev, P. Wang, J. Han, Z. Sun, Underwater cooperative MIMO communications using hybrid acoustic and magnetic induction technique. Comput. Netw. **173**(1), 107191 (2020). https://doi.org/10.1016/j.comnet.2020.107191
2. Y. Li, H. Qiu, X. Chen, J. Fu, M. Musa, X. Li, Spatial correlation analysis of imaging MIMO for underwater visible light communication. Opt. Commun. **443**(2), 221–229 (2019). https://doi.org/10.1016/j.optcom.2019.03.048
3. Y. Song et al., Experimental demonstration of MIMO-OFDM underwater wireless optical communication. Opt. Commun. **403**(3), 205–210 (2017). https://doi.org/10.1016/j.optcom.2017.07.051
4. G. Qiao, Z. Babar, L. Ma, S. Liu, J. Wu, MIMO-OFDM underwater acoustic communication systems—a review. Phys. Commun. **23**(2), 56–64 (2017). https://doi.org/10.1016/j.phycom.2017.02.007
5. Y. Hai Zhou, W. Hua Jiang, F. Tong, G. Qiang Zhang, Exploiting joint sparsity for underwater acoustic MIMO communications. Appl. Acoust. **116**(1), 357–363 (2017). https://doi.org/10.1016/j.apacoust.2016.10.010
6. U.M. Prakash, K. Kottursamy, K. Cengiz, U. Kose, B.T. Hung, 4x-expert systems for early prediction of osteoporosis using multi-model algorithms. Measurement **180**, 109543 (2021)
7. A. Saranya, K. Kottilingam, A survey on bone fracture identification techniques using quantitative and learning based algorithms, in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. (IEEE, 2021), pp. 241–248
8. R.D. Rajagopal, S. Murugan, K. Kottursamy, V. Raju, Cluster based effective prediction approach for improving the curable rate of lymphatic filariasis affected patients. Clust. Comput. **22**(1), 197–205 (2019)
9. K.P. Vijayakumar, K.P.M. Kumar, K. Kottilingam, T. Karthick, P. Vijayakumar, P. Ganeshkumar, An adaptive neuro-fuzzy logic based jamming detection system in WSN. Soft. Comput. **23**(8), 2655–2667 (2019)
10. R. Arul, G. Raja, K. Kottursamy, P. Sathiyanarayanan, S. Venkatraman, User path prediction based key caching and authentication mechanism for broadband wireless networks. Wirel. Pers. Commun. **94**(4), 2645–2664 (2017)
11. G. Raja, K. Kottursamy, A. Theetharappan, K. Cengiz, A. Ganapathisubramaniyan, R. Kharel, K. Yu, Dynamic polygon generation for flexible pattern formation in large-scale uav swarm networks, in *2020 IEEE Globecom Workshops (GC Wkshps)*. (IEEE, 2020), pp. 1–6
12. R. Mehra, K. Pachpor, K. Kottilingam, A. Saranya, An initiative to prevent Japanese encephalitis using genetic algorithm and artificial neural network, in *2020 International Conference on Computational Intelligence (ICCI)*. (IEEE, 2020), pp. 142–148

13. A. Shanthini, G. Manogaran, G. Vadivu, K. Kottilingam, P. Nithyakani, C. Fancy, Threshold segmentation based multi-layer analysis for detecting diabetic retinopathy using convolution neural network. J. Ambient Intell. Humanized Comput. 1–15 (2021)
14. K. Kottursamy, G. Raja, J. Padmanabhan, V. Srinivasan, An improved database synchronization mechanism for mobile data using software-defined networking control. Comput. Electr. Eng. **57**, 93–103 (2017)
15. V. Nallarasan, K. Kottilingam, Spectrum management analysis for cognitive radio IoT, in *2021 International Conference on Computer Communication and Informatics (ICCCI)*. (IEEE, 2021), pp. 1–5

UTPDS-ML: Utility Techniques for Plant Disease Identification Using Machine Learning

N. Mahendran and T. Mekala

Abstract Agriculture plays a significant and critical role in the Indian economy, employing over half of all Indian workers. India is often regarded as the world's leading producer of rice, pulses, spices, wheat, and spice crops. Farmers routinely examine grasses to discover illnesses, even using specialized fertilizers. Recent machine learning algorithms illustrate the monitoring of crops database and assist botanists in disease diagnosis with high precision. The suggested work combines the cross-central filter (CCF) method for image noise removal with the k-means algorithm for predicting the suspicious zone from the original region.

Keywords Agriculture · Machine learning · CCF · K-means

1 Introduction

Horticulture is a branch of agriculture that focuses on the advanced art and science of plant gardening. Seed germination, reproductive phase, panicle starting, blooming, maturing, ripening phase, and senescence are the plant budding phases. Quality of service (QoS) is an important function of agricultural products and is evolving from the knowledge of providing healthy food. By employing the quality auditing method of nutrition foods, quality assurance is important to authorize the quality of the agricultural product that monitors the natural defects [1].

Agriculture is the backbone of Indian economy, wherein 17% of India's GDP is based on the agriculture. Unfortunate changes in the atmospheric conditions, insufficient irrigation methods, inadequate familiarity in current expertise and pesticides make a variance in agriculture, irregular harvest of plants and harmful crops thereby threatening the global food safety. [2] Based on the details of Food and Agriculture Organization (FAO), the humanity population reaches 9.1 billion in 2050. Accordingly, agricultural production requirements to be enhanced up to 75% to convince the food needs of a progressively growing people. According to [3], the humanity

N. Mahendran (✉) · T. Mekala
M. Kumarasamy College of Engineering, Karur, Tamilnadu, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_57

Fig. 1 Example of plant diseases in various plants

population is anticipated to hit 9.1 billion in 2050. Consequently, farming manufacture needs to be enlarged up to 75% to complete the food necessities of a gradually increasing people.

Sometimes, the person is not competent to discover the plant infection exactly as shown in Fig. 1. The technology like advanced image classification and image processing to be helpful to identify the infections accurately. The premature disease recognition in plants reduces the hazard of maximum crop malfunction and enlarges the yields. Plant diseases prediction method is to formulate crops that are maximum healthful, and minimize the health-related problems for people. In this paper, a plan is prepared in a competent and accurate manner which might be used to find the infected crops and also to compute the strength of disease in the plant. The proposed work combines the CCF method to remove the noise in the image and k-means algorithm to identify the objects in the image to predict the suspicious region from the original region.

2 Related Work

Author [4] describes smart horticulture and techniques used for prediction and detection of diseases in crops. The digital image processing and sensor networks play a vital role, in which plant leaf image is captured for under consideration of analysis[5–7] through modern communication technology too [6, 13–21]. The author describes

the general flow of the image processing technique shown in Fig. 2. The first phase describes noise removal process, in which the input images are transformed to the necessary layout. The second phase is to segment the preprocessed image samples to allocate the image into dissimilar partitions. The important features are obtained from the segmented image and classified accordingly. The author finds the infectious crops by image segmentation and soft computing techniques as described in [4].

Author [8] presents a method for image processing which can be used to find the plant diseases using mobile app. The scheme isolates the perceptible that at dissimilar parts of the crops. The scheme tests the grape vineyard diseases based on leaves' image. The method extracts the injury features such as the particular gray level area, different spots, and histogram image pixels that can indicate RGB colors like red, blue, and green. A fixed threshold value is considered for an individual leaf disease. The RGB range is less than threshold, the particular pixel range is blackened else whitened. In this manner, the output image consist of a pattern of black spots, which can be predicted by evaluating the input data set and calculating whether the crops endure from the different disease or not. The technique attained an accuracy of 80%. The step by step process shown in Fig. 3.

Author [9] proposed an SVM algorithm for the identification of crop infections. The SVM is a type of nonlinear classifier that is mostly utilized in texture partition and pattern identification. The RGB feature pixel-counting method is to detect the affected areas like plant stem, leaf, and fruit diseases. The SVM nonlinear input data are matched into linearly divide data in high-dimensional space. The impure part is computed concerning percentage and the affected area is noticed based on SVM classifier. The algorithm step is shown in Fig. 4.

The above-discussed schemes are undemanding and do not have any systematic processes. Though, in the recent technology where systematic plays a vital role, there is a need to concern and to maximize the accuracy. Authors [10, 11] made up with a KNN and GLCM procedure for plant disease identification. After loading the input

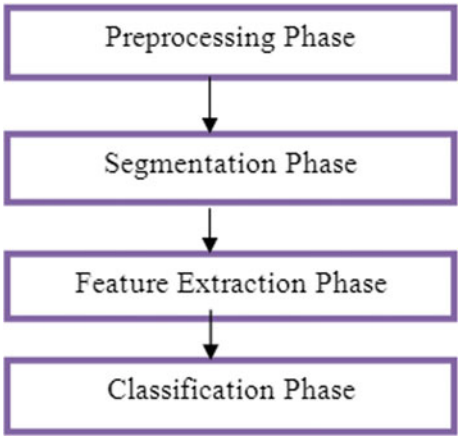


Fig. 2 Phase of İmage Processing

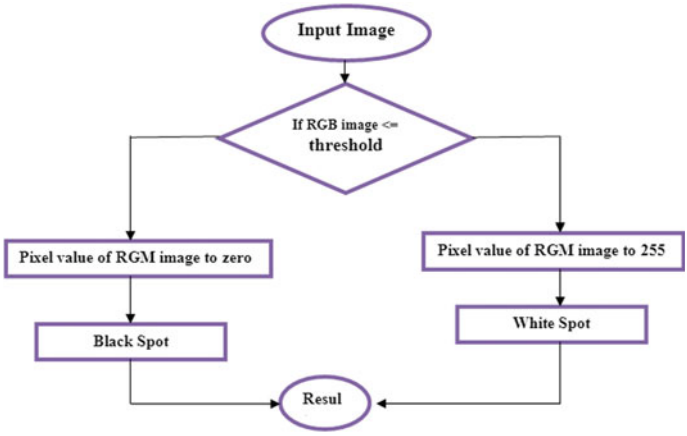
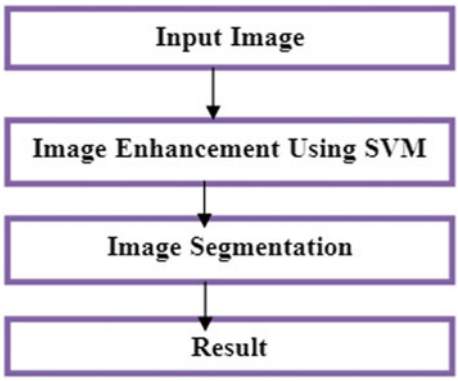


Fig. 3 Steps of threshold algorithm

Fig. 4 Steps in SVM classifier



plant leaf dataset, preprocessing of the image from data set observed by segmentation using k-means algorithm. GLCM helps to extract the feature and clustering is performed through KNN.

3 UTPDS-ML System Design

The scheme applies image processing method for crop disease identification; image segmentation and feature extraction have been observed through k-means algorithm as well as SVM-classification to categorize the crop disease as shown in Fig. 5.

The noise removal from the input images and transforming to the necessary layout is the first step called preprocessing, and it provides an insight to the process that has been observed by the enhancement of crop leaves. The next phase handles the

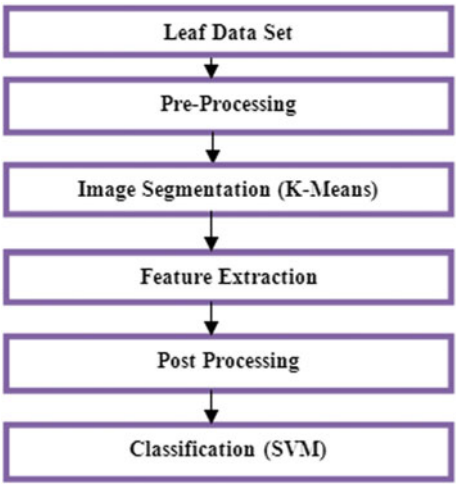


Fig. 5 Steps in proposed system

feature extraction followed by post-processing; false features are removed from the collection of obtained image.

4 Image Enhancement

The plant image enrichment is predictable to progress the dissimilarity between affected images and shrink noises in the leaf images. Maximum eminence leaf image is most significant for infection confirmation or classification to work accurately. In actual life, the feature of the plant image is unnatural by noise.

4.1 Binarization Process (B_P)

B_P is the procedure that decodes a grayscale input image into binary (1's and 0's) image. This improves the dissimilarity among the precious places in a given input image and formulates the potential feature extraction. Hence, B_P involves evaluating grayscale values of every pixel, if values is larger than the inclusive threshold set binary value as 1, else 0.

4.2 ROI Extraction (R_E)

In the image dataset, the region of interest (ROI) is the part of an input image, which is significant for the extraction of feature points. R_E has two different morphological operations like OPEN (OP) and CLOSE (CL). The OP operation can enlarge images and eliminate peaks initiated by background noise. The CL operation can minimize and reduce small differences.

4.3 Post-processing (P_P)

The segmented images may have numerous specious points. This might happen due to the presence of lines could not be improved even after enhancement results in negative points. The redundant points are eliminated in the P_p stage[12].

5 Classification and Clustering

Clustering with k-means scheme divides the input image into different partitions with more similarity along with the substance of awareness. Structures of the clustered image support in a peaceful identification of dynamic or diseased illustrations.

Algorithm: K-means clustering

Input: set of leaf images $\{S_{L1}, S_{L2}, \ldots..S_{Ln}\}$
Output: cluster the similar objects into k groups

do
Arbitrarily choose k objects from S_{Ln} as the initial cluster centers;
Repeat
Divide the input dataset into similar groups
Calculate the Euclidean distance (d)
Every pixel is referred to its closest K-centroid
while (no change);

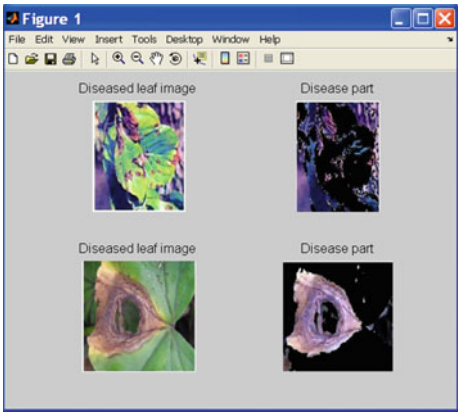


Fig. 6 Image segmentation

Algorithm: SVM Classification

Input: cluster data C_d = {group of similar data}
Output: classification according to their relevance

do (violating point)
Find disease name
cluster data C_d =U C_d
Repeat
violator
while (until pruned);

6 Results and Discussion

Figure 6 shows the feature extraction-based segmentation process using K-means algorithm. The algorithm computes the centric point through Euclidian distance and images are grouped through similarity. After clustering, the classification processed by SVM method.

Clusters formation completed based on a k-means clustering. This formation is transformed into a different feature vector (F_V) and transformed to the support vector machine for identification of the crop disease shown in Fig. 7.

7 Conclusion

The proposed work mainly focuses on identifying the plant disease and to construct an effective method. Based on the SVM classification, the corrective measures are reported. The proposed process entirely removes the necessity to check with

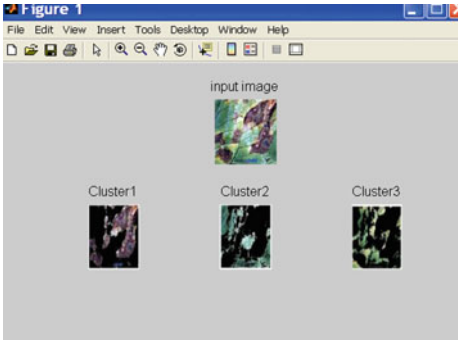


Fig. 7 Formation of clusters

a specialist to resolve whether a crop is disease affected or not. The humanity is changing near an age based on knowledge. Farmers frequently look on the grasses to identify diseases, even using exclusive fertilizers. The proposed machine learning methods demonstrate the crop monitoring database and support botanists in disease diagnosis with a large deal of precision.

References

1. A.B. Mankar, Datamining—an evolutionary view of agriculture. Int. J. Appl. Innovation Eng. Manage. **3**(3), 102–105 (2014)
2. Food and Agriculture Organization (FAO). Declaration of the World Summit on Food Security. Rome: Food and Agriculture Organization; 2009
3. K. Golhani, S.K. Balasundram, G. Vadamalai, B. Pradhan, A review of neural networks in plant disease detection using hyperspectral data. Inform. Process. Agric. **5**(3), 354–371 (2018). https://doi.org/10.1016/j.inpa.2018.05.002
4. V. Singh, A.K. Mishra, Detection of plant leaf diseases using image segmentation and soft computing procedia computer science. Inform. Process. Agric. **65** (2016)
5. N. Mahendran, T. Mekala, An efficient optimization technique for scheduling in wireless sensor networks: a survey. Stud. Comput. Intell. 223–232. https://doi.org/10.1007/978-981-10-8797-4_24
6. N. Mahendran, T. Mekala, A survey: sentiment analysis using machine learning techniques for social media analytics. Int. J. Pure Appl. Math. **118**(8), 419–423 (2018)
7. M. Natarajan, S. Subramanian, A cross-layer design: energy efficient multilevel dynamic feedback scheduling in wireless sensor networks using deadline aware active time quantum for environmental monitoring. Int. J. Electron. **106**(1), 87–108 (2019).https://doi.org/10.1080/00207217.2018.1501615
8. N. Petrellis, A smart phone image processing application for plant disease diagnosis, in *2017 6th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, Thessaloniki, pp. 1–4. https://doi.org/10.1109/MOCAST.2017.7937683
9. S. Varshney, T. Dalal, A novel approach for the detection of plant diseases, Int. J. Comput. Sci. Mobile Comput. **5**(7), 44–54 (2016). ISSN 2320-088X
10. Ch.R. Babu, D. Srinivasa Rao, V.S. Kiran, N. Rajasekhar, Assessment of plant disease identification using GLCM and KNN algorithms. Int. J. Recent Technol. Eng. (IJRTE) **8**(5) (2020). ISSN 2277-3878

11. B.J. Sowmya, C. Shetty, S. Seema, K.G. Srinivasa, Utility system for premature plant disease detection using machine learning. Hybrid Comput. Intell. 149–172 (2020).https://doi.org/10.1016/b978-0-12-818699-2.0
12. T. Mekala, M. Natarajan, in *Proceedings of IEEE International Conference on Computer Communication and Systems ICCCS14* (2014), pp. 151–155
13. T. Mekala, N. Mahendran, Improved security in adaptive steganography using game theory. Int. J. Pure Appl. Math. **118**(8), 111–116 (2018)
14. N. Mahendran, S. Shankar, T. Mekala, LSAPSP: load distribution-based slot allocation and path establishment using optimized substance particle selection in sensor networks. Int. J. Commun. Syst. (2020). https://doi.org/10.1002/dac.4343
15. N. Mahendran, S. Shankar, T. Mekala, EMA-PRBDS: efficient multi-attribute packet rank based data scheduling in wireless sensor networks for real time monitoring systems. Int. J. Electron **106** (2019). Taylor and Francis.https://doi.org/10.1080/00207217.2019.1692244
16. N. Mahendran, T. Mekala, Energy-distance and degree: multi-objective based optimized clustering and route discovery in wireless sensor network. IOP Conf. Ser. Mater. Sci. Eng. **872**(1) (2020)
17. R. Gomathi, N. Mahendran, An efficient data packet scheduling schemes in wireless sensor networks, in *Proceeding 2015 IEEE international Conference on Electronics and Communication Systems (ICECS'15)* (2015), pp. 542–547. ISBN 978-1-4799-7225-8
18. R. Gomathi, N. Mahendran, S. Shankar, A survey on real-time data scheduling schemes in wireless sensor networks. Int. J. Appl. Eng. Res. **10**(9), 8445–8451 (2015)
19. N. Mahendran, T. Mekala, Improving energy efficiency of virtual resource allocation in cloud datacenter. Indian J. Sci. Technol. **11**(19), 1–8 (2018)
20. S. Vanithamani, N. Mahendran, Performance analysis of queue based scheduling schemes in wireless sensor networks, in *Proceeding 2014 IEEE international Conference on Electronics and Communication Systems (ICECS'14)* (2014), pp. 1–6. ISBN 978-1-4799-2320-5
21. N. Mahendran, Collaborative location based sleep scheduling with load balancing in sensor-cloud. Int. J. Comput. Sci. Inform. Sec. **14**, 20–27 (2016)

Forecasting the User Prediction from Weblogs Using Improved IncSpan Algorithm

P. G. Om Prakash, A. Abdul Rahman, J. Nagaraj, and N. Sivakumar

Abstract A weblog is a collection of patterns that appear every time a person visits a website. When the buyer retrieves information from the server, the server will collect the log. The record contains a stream of transactions that will aid in recognizing the individual buyer behavior. The proposed framework can be utilized to identify the buyer behavior. It consists of preprocessing, navigation pattern modeling, and classification to identify the buyer behavior. The preprocessing module includes the three-stage process for identifying the buying behavior. The buyer pattern modeling contains the sequence of the process of website navigation. Here, the classification process has improved IncSpan algorithm; it classifies the recurrent, semi-recurrent, and in-recurrent sequence. The improved IncSpan algorithm utilized previous navigation results; from that, user behavior is analyzed.

Keywords User navigation · Mining · Path traversal · Prediction · Web mining · Weblog · Pattern classification

1 Introduction

In the current situation, the car items are currently sold online, the number of consumers who intend to acquire an automobile online has increased. The organization must examine customer demands; therefore, they must research buyer intent, and only then the vehicle sales will increase. The customer will analyze the product and product delivery; if they are satisfied, then only they will purchase the product. Identifying the buying interest of individual buyers is a difficult task in online automobile shopping.

P. G. Om Prakash (✉) · A. Abdul Rahman · J. Nagaraj
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

N. Sivakumar
Department of Computer Science and Engineering, Jain University, Bangalore, Ramanagara District 562 112, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_58

Here, automobile organization is chosen; the automobile organization must know the demand of the vehicles. The organization makes a study of buyer intention based on their choice. The major buyer intention is the overall price, buying price, maintenance of a vehicle, technical characteristics of a vehicle, the comfort of a vehicle, the number of people to occupy, and safety measures of the vehicle. So, the organization concentrates on the above attributes to identify the user interest to increase the sales of the vehicle.

The process of identifying the buyer behavior is a hard task; the buyer behavior is analyzed from buyer navigation pattern that is recognized from weblog. The buyer navigation pattern converts the set of sequence patterns into a subsequence pattern based on IP address and session. Here, the improved IncSpan algorithm identifies the above pattern as recurrent pattern, in-recurrent pattern, and semi-recurrent pattern, from the above-generated pattern buyer behavior, is analyzed.

2 Review of Literature

Web usage mining is a technique that is to recognize the buyer behavior from buyer pattern; from that, buyer behavior is analyzed from the log file.

Zlio et al. [1] suggested that the adaptive-based preprocessing technique helps to recognize the buyer behavior and accuracy also improved. Prakash et al. [2] identified buyer behavior helps to classify the navigation behavior of buyer. Buyer behavior identifies by the total value of the clustered pattern. If the total value of the above buyer pattern attains the standard level, then the recognized buyer pattern as buyer interested pattern (BIP); if it does not attain the benchmark, then it is buyer not interested pattern (BNIP); from that, the above buyer behavior pattern is recognized. Patel et al. [3] suggest that weblog has an unprocessed log, so the altering the unprocessed log to meaningful information by using the preprocessing phase. Hong et al. [4] suggested, converting the sequence of transaction patterns to segment in to number of sequences by navigation pattern modeling. The algorithm has converted into segment sequence by pattern generation; from that recurrent, semi-recurrent, and in-recurrent sequence is identified.

Kerana et al. [5] identify the recurrent pattern, semi-recurrent pattern, and in-recurrent pattern by classification analysis. It classifies the recurrent, semi-recurrent, and in-recurrent sequences; from that, optimal buyer pattern is identified using clustering technique. Mobas et al. [6] suggest that user personalizes make use of clustering technique. After received the series of pattern from preprocessing, the clustering algorithm gathers the similar pattern; from that, buying interested behavior and not interested buying behavior is identified. Vino et al. [7] utilize RBFN classification model to improve the prediction rate to 99.02%. Kousar et al. [8] suggested that optimization of text document is gathered by using cluster with distance approach. Babu et al. [9] suggest a rapid web miner tool, this tool suggests the customer to yield high profit and reduce the risk of the customers.

Gummadi et al. [10] suggested that performance is increased by using LEACH algorithm; it will track the automobile by sensing wireless networks. Karthik et al. [11] utilized the routing technique to increase the accuracy during packet delivery; initially, it will check the signal strength; then routing table helps to identify the neighbors. Bommadev et al. [12] suggest a clustering technique to predict the heart disease like heart attack derived by using K means. Angel et al. [13] suggested a classification technique, and this technique predicts the growth of farmers in agriculture development. The farmer will cultivate any corps, based on the soil condition and land; this technique is suggested to the farmers to get more profit. Anjali et al. [14] suggested a technique to improve the placement status of every university with the help of every ward academic record.

Arvind et al. [15] utilize ML techniques; the model will suggest more features to predict the thyroid disease accurately. Chanint et al. [16] discussed that the forecasting process will be derived by regression technique and produce optimal results.

Wang et al. [17] discussed that the preprocessing has two steps, first, it will eliminate unwanted similar patterns in datasets, and then it calculates the ranking. Xin et al. [18] suggested two behaviors; one is extroversive and the second is introversive behavior. In extroversive behavior recognize the users in online and assembly the information in online by introversive behavior.

Ruili et al. [19] identified the pattern based on user behavior, and it analyzes the usage pattern from database and number of pattern registered for the user. Surbi et al. [20] suggested back navigation pattern approach to identify the sequence of recurrent pattern, semi-recurrent pattern, and in-recurrent pattern.

3 Research Gap

The research gap identified from the above issues is.

- It is hard to capture the individual buying behavior from weblogs.
- Need to minimize the mining time.
- Lack of identifying prediction accuracy of users is too difficult.

4 User Navigation Framework

Figure 1 shows user navigation system framework; it consists of.

- Generation of weblogs
- Preprocessing
- Navigation pattern modeling
- Classifier.

A. **Generation of Weblogs**

The log will be generated in the server when the user will access the website.

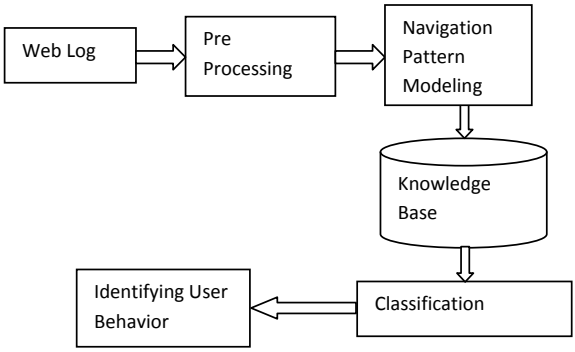


Fig. 1 User navigation system framework

1. User address
2. Date and time
3. Categories
4. Bytes transmission
5. Status code

B. **Preprocessing**

In data preprocessing, initially the log as the unstructured file transactions, preprocessing has a three-stage process.

- Cleaning the data
- User identification
- Session identification.

a. **Data Cleaning**

The logs are produced from automobile portal; the period is recorded from 23/Nov/2018 to 02/Oct/2018. The log file has 3667; each log contains the following details. The cleaning of data process will remove the unwanted information such as.jpg,.gif and the status code greater than 200 and less than 400 has to be processed. After the process of data cleaning, the obtained log in 1987 is generated.

E.g.: Fig. 2 represents "178.19.3.9 - [23/Nov/2018] "GET 9.gif HTTP/1.0" 200 442"; the above stats code is considered as a successful transaction of the webpage of the user.

b. **User Identification**

From the above data cleaning process, 1987 records are generated; with the help of the user IP address, the corresponding user is identified for the further process.

E.g.: Fig. 2 represents the IP 178.19.3.9 which is recognized from the above datasets.

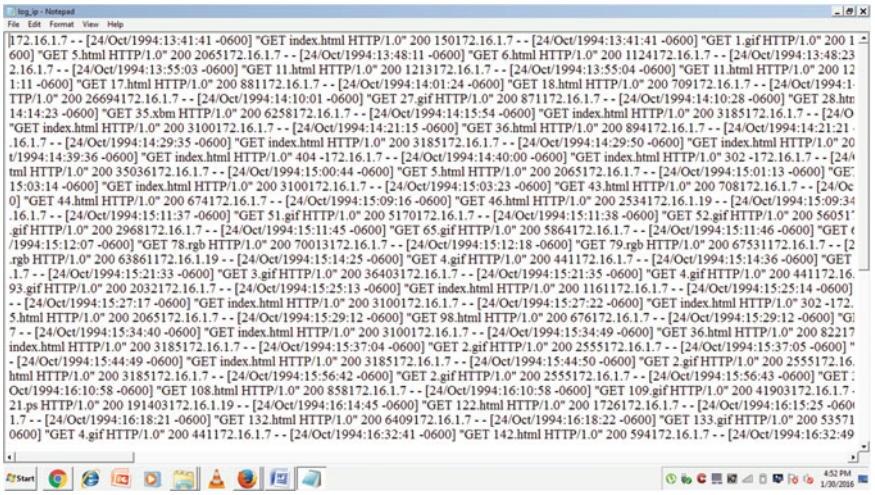


Fig. 2 Structure of log

c. **Session Identification**

The session is identified based on the time sessions for the corresponding user session of a different timestamp.

E.g.: Fig. 2 represents the similar IP of different time stamp 12:43:45 to 12:49:46 and 13:56:32 to 15:09:59 is considered as two different session.

d. Navigation Pattern Modeling

The user and session identification are generated by preprocessing. From Fig. 3, the tree generation helps to identify the IU and NIU from user traversal patterns.

The tree pattern is generated from the user transaction in log; it has a sequence of web transactions from the vehicle website. The navigation pattern modeling classified the interested user pattern and not interested user pattern based on the sequence and subsequence of the pattern. It supports the organization in the decision-making process to improving their business.

IV. **Classification**

Using the improved IncSpan algorithm, the pattern is converted into recurrent sequence pattern, semi-recurrent sequence pattern, and in-recurrent sequence pattern. The improved IncSpan algorithm classifies the above pattern count based on the value count. The count is above the trusted value and then it has an interested user pattern; if it is less than the value, then the user have a not interested user pattern.

Table 1 shows the user browsing pattern. Table 2 shows the pattern total and Table 3 shows the interested user pattern and not interested user pattern.

- add the new sequence in V'

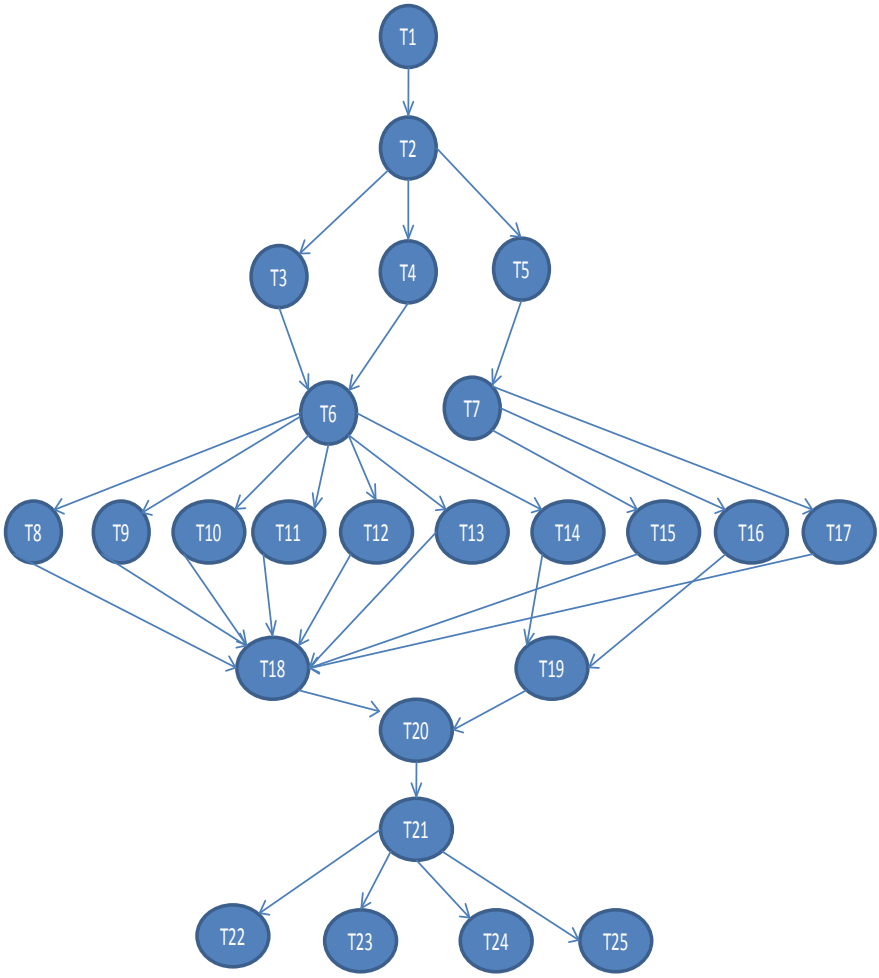


Fig. 3 Tree generation process of IU and NIU

Table 1 User pattern

P Nos	Patterns
1	< V1,V2,V5,V6,V8,V18,V20,V21,V22 >
2	< V1,V2,V5,V6,V10,V18,V21,V22,V23 >
3	< V1,V2,V4,V7,V17,V7,V16 >
4	< V1,V2,V3,V5,V7,V16,V19,V20,V21,V25 >
5	< V1,V2,V5,V5,V7,V15,V19,V20,V21,V24 >
6	< V1,V2,V5,V7,V11,V6,V9 >
7	< V1,V2,V3,V5,V6,V4,V7 >

Table 2 Pattern count

P No	Pattern	Count
1	< V1,V2,V5,V6,V8,V18,V20,V21,V22 >	16
2	< V1,V2,V5,V6,V10,V18,V21,V22,V23 >	17
3	< V1,V2,V4,V7,V17,V7,V16 >	11
4	< V1,V2,V5,V7,V16,V19,V20,V21,V25 >	16
5	< V1,V2,V5,V5,V7,V15,V19,V20,V21 >	2

Table 3 Behavior patterns

P No	Pattern	IU / NIU
1	< V1,V2,V5,V6,V8,V18,V20,V21,V22 >	IU
2	< V1,V2,V5,V6,V10,V18,V21,V22,V23 >	IU
3	< V1,V2,V4,V7,V17,V7,V16 >	IU
4	< V1,V2,V5,V7,V16,V19,V20,V21,V25 >	IU
5	< V1,V2,V5,V5,V7,V15,V19,V20,V21 >	NIU

- if V' in IUP_SE then go to step 3 else step 9
- cmp [V' in Seq_D_H]
- if V' is GreathanSucc_count[i] Seq_D_H
- then it is "IUP"
- IUP_SE = V'
- else "NIUP"
- NIUP_SE = V'
- if V' in NIUP_SE then goto step 3
- Seq_D_H = V'

5 Results and Comparative Discussion

The implementation will perform by automobile dataset, and it is executed by the tool weka; the compared performance is measured by recall, f measure, and precision.

Fig. 4 shows the comparison chart of existing model and proposed model. The proposed improved IncSpan algorithm model attains the 0.8343 precision, 0.8236 recall, and 0.8574 f measure, respectively.

6 Comparison Evaluation Parameters of Existing and Proposed System

The discussion Table 4 compares the existing and proposed models took from automobile dataset. The proposed model gives better results, while compared with the

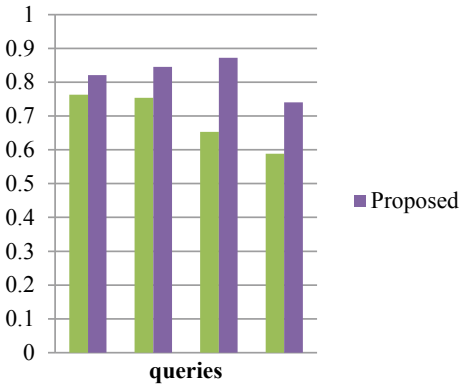


Fig. 4 Comparison chart

Table 4 Automobile dataset for existing and proposed method

Dataset	Methods	Precision	Recall	*F* measure	Prediction %
Automobile Dataset	Existing	0.7903	0.7396	0.6784	0.5934
	Proposed	0.8343	0.8236	0.8574	0.7534

existing methods. The proposed method achieves the 0.8343 precision, 0.8236 recall, and 0.8574 f measure.

The proposed model is implemented using an improved IncSpan algorithm; in that, 9 attributes were compared such as buying, maintenance, doors, person capacity, baggage space, safety, and a class of automobile. Figures 5 and 6 show the category and different class of automobile.

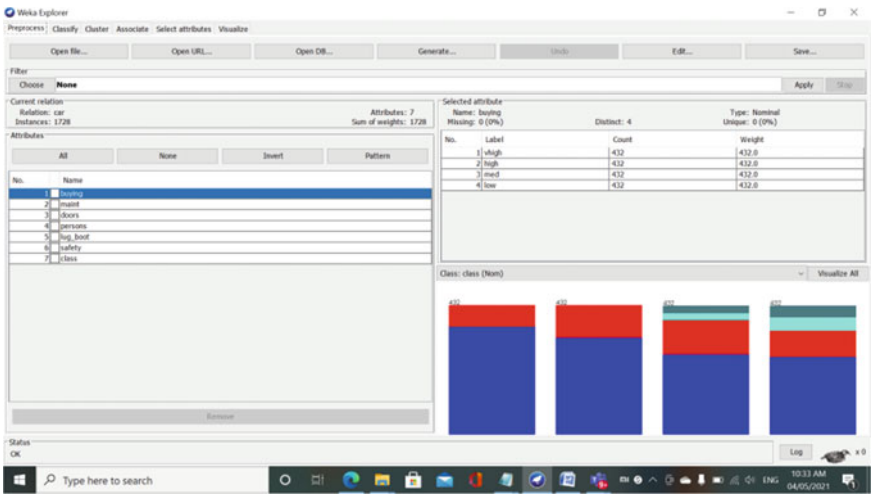


Fig. 5 Category of automobile

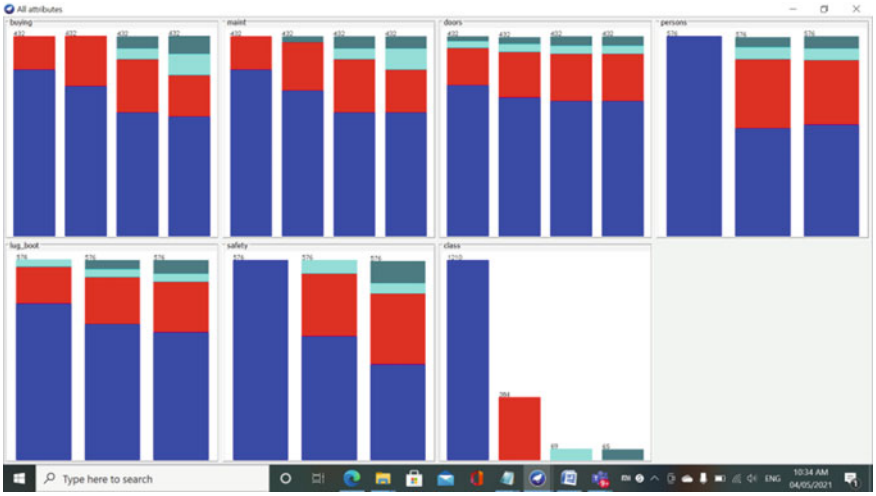
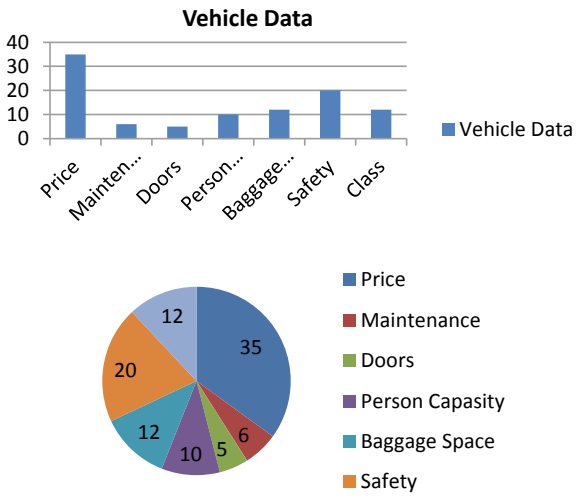


Fig. 6 Different class of automobile

Figure 7 describes the user prediction in automobile industry, and it compares major attributes such as price, maintenance, doors, person capacity, baggage space, safety, and class of vehicle; the system compares with those attributes; from that, buyer prediction will be increased.

Fig. 7 User prediction in automobile



7 Conclusion

The proposed methodology is implemented by using an improved IncSpan classification technique, where the system will predict the buyers interest, and the attributes that are taken into consideration are price, maintenance, doors, person capacity, baggage space, safety, and class of the vehicle from which the buying prediction rate will be increased. In future, based on the prediction rate, the automobile industry will mainly focus to increase or decrease the products.

References

1. I. Zliobaite, B. Gabrys, Adaptive preprocessing for streaming data. IEEE Trans. Knowl. Data Eng. **26**(2), 309–321 (2014). https://doi.org/10.1109/TKDE.2012.147
2. P.O. Om Prakash et al., Analyzing and predicting user behavior pattern from weblogs. Int. J. Appl. Eng. Res. **11**(9), 6278–6283 (2016)
3. P.G. Om Prakash, A. Jaya, Analyzing and predicting user behavior pattern from weblogs. Int. J. Appl. Eng. Res. **11**, 6278–6283 (2016)
4. K. Patel, A.R. Patel, Process of web usage mining to find interesting patterns from web usage data. Int J Comput Technol **3**, 144–148 (2004). https://doi.org/10.24297/ijct.v3i1c.2767
5. H. Cheng, X. Yan, J. Han, IncSpan: Incremental mining of sequential patterns in large database. Available in http://hanj.cs.illinois.edu/pdf/kdd04_IncSpan.pdf (2004)
6. D. Kerana Hanirex et al, Efficient algorithm for mining recurrent itemsets using clustering technique. Int J Comput Sci Eng **3**(3) (March, 2011)
7. P. Rajesh, N. Srinivas, K. Reddy, G. Vamsipriya, M. Dwija, D. Himaja, Stock trend prediction using ensemble learning techniques in python. Int J Innovative Technol Exploring Eng **8**, 150–154 (2019)
8. R. Vinothkanna, T. Vijayakumar, Using contourlet transform based RBFN classifier for face detection and recognition. Int Conf ISMAC Comput Vision Bio-Eng **30**, 1911–1920 (2019). https://doi.org/10.1007/978-3-030-00665-5_176
9. K. Nikhath et al., Feature selection, optimization and clustering strategies of text documents. Int J Electr Comput Eng **9**(2), 1313–1320 (2019)
10. B.S. Babu et al., Customer data clustering using density based algorithm. Int. J. Eng. Technol. (UAE) **7**(2), pp. 35–38 (2018). https://doi.org/10.14419/ijet.v7i2.32.13520
11. A. Gummadi et al., EECLA: Clustering and localization techniques to improve energy efficient routing in vehicle tracking using wireless sensor networks. Int. J. Eng. Technol. (UAE) **7**, 926–929 (2018)
12. T. Karthikeyan et al., Investigation on maximizing packet delivery rate in WSN using cluster approach. Wireless Pers. Commun. **103**(4), 3025–3039 (2018)
13. G. Akhila et al., A survey on health prediction using human activity patterns through smart devices. Int. J. Eng. Technol. (UAE) **7**(1), 226–229 (2018)
14. A. Prathyusha et al., A survey on prediction of suitable crop selection for agriculture development using data mining classification techniques. Int. J. Eng. Technol. (UAE) **7**(3.3), 107–109 (2018)
15. M. Arshad et al., A real-time LAN/WAN and web attack prediction framework using hybrid machine learning model. Int. J. Eng. Technol. (UAE) **7**(3.12), 1128–1136 (2018)
16. S.S. Arvind Selwal et al., A Multi-layer perceptron based intelligent thyroid disease prediction system. Indonesian J. Electr. Eng. Comput. Sci. **17**(1), 524–533 (Jan 2020)
17. W. Liu, S. Liu, Q. Gu, J. Chen, X. Chen, D. Chen, Empirical studies of a two-stage data preprocessing approach for software fault prediction. IEEE Trans. Reliab. **65**(1), 38–53 (2016)

18. X. Ruan, Z. Wu, H. Wang, S. Jajodia, Profiling online social behaviors for compromised account detection. IEEE Trans. Inf. Forensics Secur. **11**(1), 176–187 (2016)
19. R. Geng, J. Tian, Improving web navigation usability by comparing actual and anticipated usage. IEEE Trans. Hum. Mach. Syst. **45**(1), 84–94 (2015). https://doi.org/10.1109/THMS.2014.2363125
20. S. Huria, J. Singh, Implementation of dynamic association rule mining using back navigation approach, in *2015 Fifth International Conference on Communication Systems and Network Technologies* (Gwalior, 2015), pp. 1048–1050

Machine Learning-Based Algorithmic Approach for Enhanced Anomaly Detection in Financial Transactions

Sivakumar, Mariyappan, and P. G. Om Prakash

Abstract Governments all over the world aim to encourage digital money transaction to minimize administration cost, avoid customer's physical present while purchasing, avoid money loss from theft, and to save time. So nowadays, people have started to use cashless transaction. Due to the increasing nature of cashless transaction, credit card becomes common nowadays, and many people have adapted to this payment mode. Customers are expanding on a regular basis as a result of the flexibility afforded by credit card transactions. Because of their ubiquity, credit card payments will not come without any risk. There are many types of credit card frauds, for instance, skimming, phishing, card lost or stolen, etc. and by using some machine learning algorithms, many fraud detection systems are available in recent years but the efficiency of those algorithms is not appreciable. Genetic algorithm, *k*-means algorithm, artificial neural network, logistic regression, SVM, etc. are few algorithms used to develop the system to detect credit card fraud. This paper makes use of an algorithm called isolation forest, which helps to detect the outlier data points in a better way. It performs well in large dataset, and also, it provides better results and accuracy when compared to SVM and local outlier factor.

Keywords Support vector machine (SVM) · Outlier · ANN · Logistic regression · Isolation forest

1 Introduction

A cashless transaction is similar to an online transaction that takes place between two parties, one from the merchant and another from the consumer, mainly for the purpose of acquiring products. This transaction will not involve any money. Customers do

Sivakumar · Mariyappan
CSE (Specialization), Jain Deemed To Be University, Bengaluru, Karnataka, India

P. G. O. Prakash (✉)
Assistant Professor, School of Computing, Department of Computational Intelligence, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_59

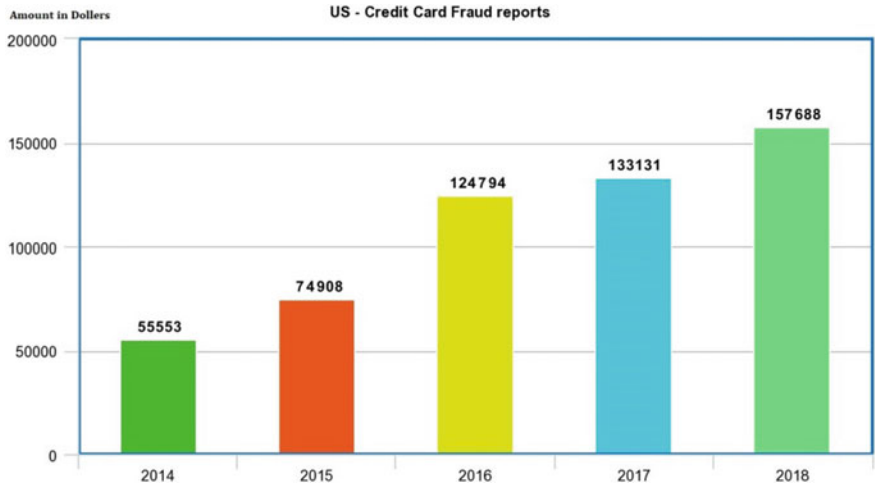


Fig. 1 Credit card fraud reports recorded in us

not need to bring cash with them while using this method of payment. This kind of transaction is very easy and accurate, and it consumes only less time. So people have started using credit card, and this necessitates the need to develop a secure credit card system to protect the cash amount from fraudster. The online digital transactions are done through internet. Due to the fraud nature in online system, it is highly required to develop an efficient fraud detection system. Since big companies, banks, and organizations commonly use this payment method, they are confronted with the difficulty of detecting fraud in such systems. Credit card fraud has surpassed identity theft as the most common type of identity theft fraud. Nearly, 24 billion dollars have been lost globally as a result of credit card theft.

Figure 1 shows year and amount losses in dollars particularly in the USA. USA uses more than 1 billion credit cards and 2.8 billion all over the world. In the year 2018, the recorded number of credit card fraud is 130,928. USA is the most affected country with 38% of reported fraud losses in 2018, and still it is increasing. Credit card fraud is a kind of identity theft, suppose your credit card is stolen by someone and that person can make purchase or that person can do online transaction without the knowledge of the credit card user. Account number, pin number, and security code can be taken by fraudster to make purchase without any authentication. Skimming is one type of credit card fraud; when you swipe your credit card through the card reader machine at cash point, the reader machine will illegally take a copy of the information, which resides in your credit card. As a result, the fraudster can make clone cards.

1.1 Types of credit card fraud:

- Card lost or stolen by someone
- Skimming

- Phishing
- Card not present
- Counterfeit
- Fraudulent credit application
- Ac takeover.

1.2 **Techniques available to detect fraud**:

- Artificial neural network
- Genetic algorithm
- Hidden Markov model
- SVM
- Bayesian network
- Decision tree (DT)
- Expert system.

2 Related Works

The implementation of DT, RF, SVM, and LR on highly skewed data and the efficiency of this techniques is entirely depends upon (1). accuracy, (2). sensitivity, (3). Specificity, and (4). precision. The results indicate about the optimal accuracy for logistic regression, decision tree, random forest, and SVM classifiers are 97.7%, 95.5%, 98.6%, and 97.5%, respectively [1]. A metaheuristic-based detection technique is used to overcome the problems and provide effective detections. This work proposes a parallel PSO-based hybrid technique that incorporates simulated annealing to provide time effective predictions that are also more accurate compared to conventional techniques [2]. Genetic algorithm is used in detection of credit card fraud mechanism and examined the result based on the principles of this algorithm. Moreover, it is used to execute credit card fraud how credit card fraud impact on financial institution as well as merchant and customer [3]. Two segments are used to find fraudulent activity in credit card. In the first segment, clustering technique is used for partitioning the data points. On the other hand, second segment deals with two different outlier detection techniques, which are applied in the partitions separately for finding the outliers. Finally, the outliers are combined, and fraudulent cases are found. These algorithms are used for developing a hybrid algorithm for detecting fraud in online transactions [4], 5. The objective of the research is to develop a new outlier detection method for streaming data transaction, with an objective, to analyze the past transaction details of the customers and extract the behavioral patterns. Where customers are grouped into many different groups depends on their amount they spent. By using a sliding window, the clusters are extracted from transactions based on the behavior pattern of the customers. Data mining methods such as classification which is based on ID3 DT (decision tree) and cryptographic are used to detect fraud in credit card-based problem. So, the implementation of these techniques for detecting anomalies, the unauthorized payment transaction can be reduced [6]. The

technique which is present is based on AI, DM, fuzzy, ML, genetic algorithm, and so on to catch the fraud. The investigation was done on Bayesian classification, KNN on highly skewed data to detect fraud. The logistic regression finds outcome based on 2 values that are yes or no, otherwise true or false. This system uses genetic algorithm and optimization technique to detect fraudulent activities [7]. The various classification algorithms such as ANN, KNN, and C5.0 are used to detect the fraud in credit card. In the first phase, neural network was used and obtained 92.86% accuracy, and it was pretty good after that *K*-nearest neighbor algorithm was used to find the fraud, and it gave 91.8% accuracy; finally, C5.0 was tested and got 77.8% accuracy. Finally, it was concluded that ANN provides better result compared to the rest [8].

Complete past data of the customers including normal or fraud was analyzed to support the system to find fraud; here, 2 ML algorithms were used one is SVM, and another one is ANN. Finally, the comparison study was carried out to identify the best method to detect fraudulent activity in such financial transaction. In the first approach, SVM was used and the accuracy obtained is 82%, and then ANN was applied and got 92.86% accuracy. So it was concluded that ANN shows much better result [9]. Random forest was used to increase the accuracy of prediction, dataset classification was done using random forest, and evaluation of this technique depends on specificity, sensitivity, precision, and accuracy. Even though the random forest method works better with huge amount of data, it was found that it suffers in terms of speed while testing [10]. Few research scholars analyzed and applied some data mining technique to detect fraud in credit card system. The dataset was divided into tiny subsets and then applied distributed DM to create method of customer behavior. Finally, the result oriented methods are combined to create metaclassifier to increase the accuracy of prediction [11]. Unsupervised and supervised learning methods are combined to detect fraud. Here, behavior and rule-based techniques were used for credit card fraud detection [12].

3 Challenges Faced by Researchers in Credit Card Fraud Detection System

Since the increasing nature of cashless transaction, credit card becomes so famous now days, and many people are start adapting to this payment mode. The researchers find it difficult to find one system to prevent the money from credit card fraud. They face lot of challenges including.

3.1. Unavailability of dataset—The researchers need the customer's data for research purpose but the banks and some other commercial organization are not ready to provide their customer's data for some security issues. Without the proper data, it is the big challenge for the researchers to find credit card fraud detection system.

3.2. Imbalanced dataset—Another major issues faced by researchers are imbalanced data. Data imbalance in the sense the dataset is not balanced one. Suppose the dataset having 50% of genuine transaction and 50% of fraudulent transaction, then we can say that the dataset is balanced one. But in credit card transaction case, the genuine transactions are more but very less fraudulent transactions are available. Maximum 99% of transactions are normal and less than 1% of fraudulent transaction. Due to this reason, it is difficult for the researchers to develop one model and make the model train well. See in this imbalanced case, the model can get trained well in terms of genuine transaction but the model will work poor in terms of fraudulent transactions.

3.3. System speed—This is another issue for the researchers because daily the credit card is used by millions of people. The system which predicts the fraudulent activity should be very fast enough to respond. If it takes more time, it will become the benefit for the fraudster.

It is not possible check transaction of each customer is genuine or fraud. It is a time waste process. Asking too much of confirmation with client while transaction is also not good and it makes unsatisfaction for the customers.

3.4. Novel Technique—It is mandatory to find a novel machine learning technique which performs efficiently on imbalanced data and should be processed in less time.

4 Methodology

Isolation forest is the efficient technical method for detecting novelties in dataset. Isolation forest (IF) is the new method used to find the fraud in financial transaction. Sometimes, we may think like the isolation forest algorithm works exactly similar to decision tree algorithm. In case of decision tree algorithm, it start doing its work (partitioning) from root node and the partitioning method depends upon information. On the other hand, in IF algorithm, the partition depends upon randomness. Even though lot of anomaly finding methods available in the present technological world, the accuracy and the efficiency are not up to the level, but IF algorithm provides better results and provides better accuracy. Moreover, it performs well in large set of data. Even though lots of fraud detection systems exist, for example, *K*-nearest neighbor algorithm, support vector machine, random forest, decision tree, etc. but efficiency and adaptability of that algorithm with the dataset are in question only. IF algorithm creates a concept called isolation, i.e., it isolates the anomalies. This IF algorithm provides a better effectiveness and efficiency to find the fraud compare to the common general methods such as distance and density measures, moreover IF method requires less memory requirement and low time complexity (Fig. 2).

The IF method is completely based on fraud or outlier, i.e., the outliers are the data points which are very few in counts, and it will be totally different when comparing with normal data points. IF separates each and every data point, and it splits the data points into two; one is outliers and another is inliers. This split totally depends

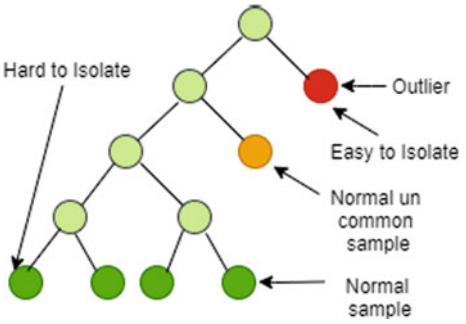


Fig. 2 Isolation forest

upon how long it takes to separate the data points. For instance, there is a need to separate data point (normal data point), where the normal data point can have many data points around it, and the process is very difficult to separate or isolate because normal data point needs more condition for its separation.

In Fig. 3, we are isolating a normal point, and many other data points are there around the particular normal data point **x.** It requires more than 10 splits to isolate the particular normal point.

On the other hand, if you choose outlier data point, then the process is very easy because the outlier data points are very less, so the condition here is very less; moreover, not more data points will be available around outlier data point. Actually the outlier data is alone. In Fig. 4, isolation of an abnormal point is identified. Here, only 5 splits are needed to isolate the abnormal point. In comparing with Fig. 3, Fig. 4 requires very less splits. The algorithm may face difficulties due to the randomness; to overcome this issues, the process has to be done again and again and then we need to calculate the average path length and should be normalized.

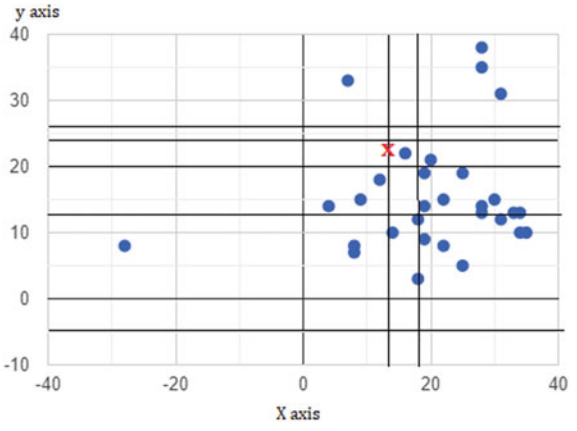


Fig. 3 Isolation of a normal point

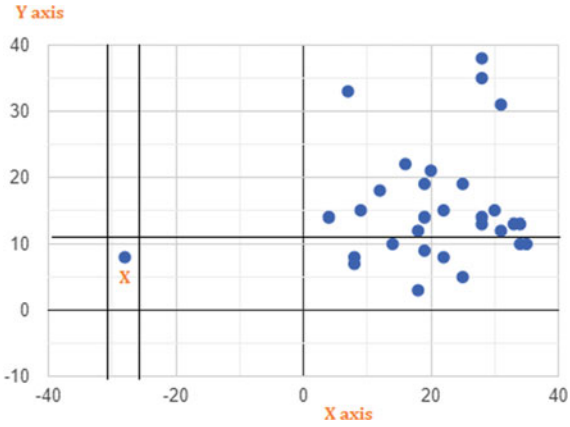


Fig. 4 Isolation of an abnormal point

4.1 Procedure for Isolation and Outlier Detection

Normal values are not extremely different values rather its nearer values. We cannot get the leaf node quickly. Since normal values are very huge in numbers, automatically the depth will be increased and score becomes high. Depth of the node is the length from root node to the particular node (normal or outlier). Here, the outliers are spited initially because outlier is a completely a different value. Since it is a different value and also it is rare, depth will be less. We quickly get the leaf node and score becomes less. Based on the score, we can identify the outlier (Fig. 5).

Let $X = \{\times 1, \times 2, \times 3 \ldots \text{xn-1,xn}\}$ is a d-dimensional data points. T is a node in the tree, and T may be external node without child; otherwise, T may be internal node with only one 'test' or two node (Tl, Tr) exactly. Moreover, the test node T consists of q(attribute)and p(split value), and $q < p$ indicates the traversal of data point to Tl or Tr. For the isolation of a point, the isolation tree algorithm recursively generates partitions on given sample; it can be done by selecting features randomly; after that, the algorithm randomly selects a split value for the feature or attribute, between the min and max values allowed for that feature. To develop the isolation tree, the method has to divide the data points by randomly selecting a feature; it has to be continuing until the node of the tree exactly having only one instance (or) all the data at the node having same value. We need to find the path length also.

$$\text{Path length} = \text{total number of partitions for isolation}$$

Generally, normal data points in the sample having many data points near to it. It is huge in size and here conditions also many. So, it is not easy to isolate the data points which having lot of condition. By contrast, the data points, which are very less in size, totally vary from normal data points, and have less conditions to isolate, are called abnormal data points that is outliers. We need to find the depth of the isolation,

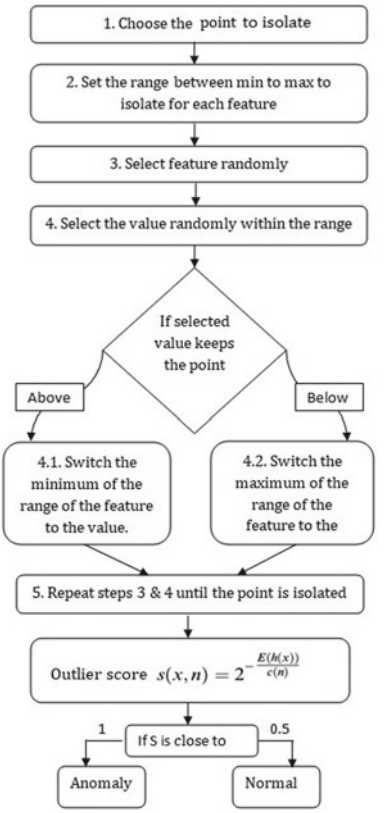


Fig. 5 Steps to isolate data point

and after this step, we need to normalize it. So that the data points lie between the range 0–1.

Finally, we need to find the anomaly score for the data points. It can be achieved by the following formula,

S(x, n) = 2.
c (n) = 2 H (n – 1) – (2 (n - 1) / n).
Here E(h(x))- Avg. path lengths, **n**- Number of data points, **c(n)**- avg. path length of unsuccessful search.
Note: if score = around 1 - anomaly.
If score < 0.5 - normal.
If score = around 0.5 - outlier.

4.2 Dataset Description

The total number of observations is 284 k credit card transactions, out of which 492 are fraudulent or 0.17% of all transactions. As a result, the dataset is unbalanced. Because of security concerns, the characteristics of the observations are modified using PCA. There are 32 PCA applied features and rest that is, time, class (1—fraud, 0—genuine), and amount are non-PCA applied features.

Parameters are listed below:

1. **Age** of the customer
2. Current working **position** of the customer
3. Monthly **income** of the customer
4. **Maritalstatus** of the customer
5. Credit card number
6. Customer profession
7. Credit grade for the card
8. Frequency of card usage
9. Total working experience of the customer
10. Housing type of the customer
11. Type of the credit card
12. Credit line
13. Balance
14. Number of times of overdraft
15. Number of cards the customer has
16. Average number of days per overdraft
17. Average amount of the customer
18. Number of times bad debt
19. Times of overdraft but not bad debt
20. Daily spending amount of the customer
21. Average of daily overdraft
22. Present residence
23. Transaction area
24. Religion of the customer
25. Overdraft rate
26. Average amount per transaction
27. Growth rate of shopping
28. Overdraft rate
29. Identification for each customer
30. Merchant category code
31. Status of the transaction
32. Type of the credit card used.

5 Result and Discussion

This technique (isolation forest) is implemented using python and the experiments were conducted on a DELL Precision T7600 Workstation with Intel Xeon E5 2680 CPU with 32 CPUs of 2.7 GHz each and 32 GB RAM. The results obtained by isolation forest techniques were compared with local outlier factor and support vector machine to exhibit the efficiency toward accuracy. Accuracy obtained from isolation forest technique and support vector machine is compared in Fig. 6.

It could be observed that the accuracy obtained by isolation forest method is much higher than the accuracy exhibited by support vector machine. By using IF technique, we got the accuracy of 99.74%, and we got 70.09% accuracy by using support vector machine.

Accuracy obtained from isolation forest technique and local outlier factor is compared in Fig. 7. It could be observed that the accuracy obtained by isolation forest method is little higher than the accuracy exhibited by local outlier factor. By using IF technique, we got the accuracy of 99.74%, and we got 99.65% accuracy by using local outlier factor.

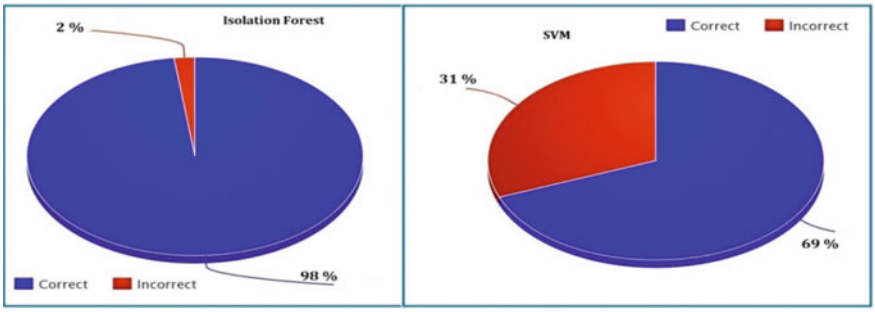


Fig. 6 Accuracy between IF and SVM

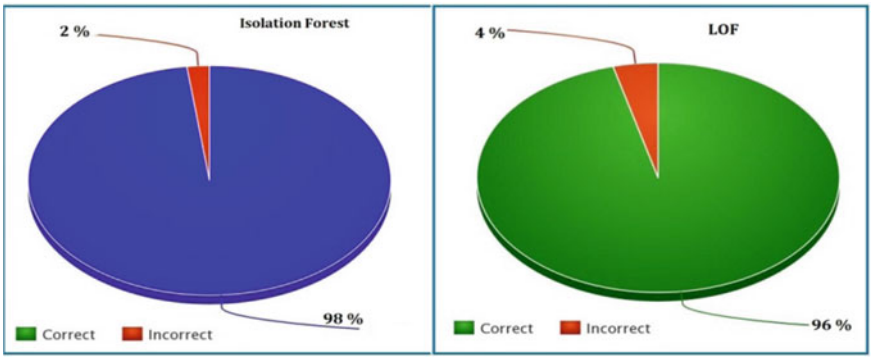


Fig. 7 Accuracy between IF and LOF

Accuracy obtained from isolation forest technique vs. support vector machine and local outlier factor are compared in Fig. 8. It could be observed that the accuracy obtained by isolation forest method is higher than the accuracy exhibited by LOF and much higher than the accuracy exhibited by SVM.

Accuracy of incorrect prediction obtained from isolation forest technique vs. support vector machine and local outlier factor is compared in Fig. 9. It could be observed that the accuracy obtained by isolation forest method is much lower than the accuracy exhibited by SVM and local outlier factor.

The below is the accuracy comparison table of few techniques (Table 1)

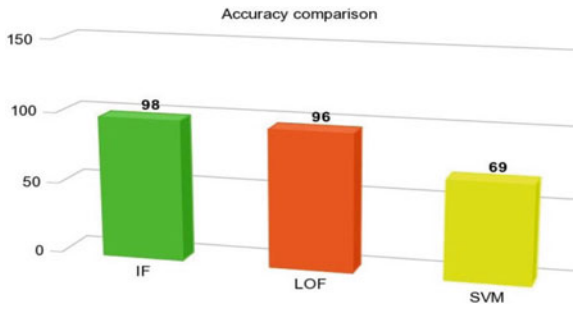


Fig. 8 Accuracy between IF, SVM, and LOF

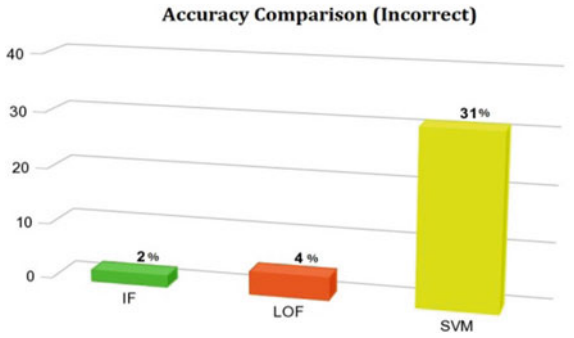


Fig. 9 Incorrect prediction between IF, SVM, and LOF

Table 1 Accuracy comparison among several techniques

Technique	Accuracy
Isolation forest	98%
Local outlier factor	96%
SVM	69%

6 Conclusion

From the proposed research study, it is evident that a large number of people, including customers and merchants, are affected by the growing nature of fraud in the financial transaction system. One example is credit card fraud. The primary issues with credit card transactions are data confidentiality and the unbalanced nature of the dataset. This paper uses an algorithm called Isolation forest. This algorithm works well for huge set of data. It isolates the data points in the observation. Since, the outliers are few and different; we can easily identify the outlier. Finally, this technique is compared with local outlier factor and support vector machine and found to be effective in terms of accuracy. Still, we can improve the accuracy of prediction of isolation technique by extending the sample size.

References

1. N. Khare, S.Y. Sait, Credit card fraud detection using machine learning models and collating machine learning models. Int. J. Pure Appl. Math. **118**(20), 825–838 (2018)
2. S. Nadarajan, Fast and effective credit card fraud detection in imbalanced data using parallel hybrid PSO. Int. J. Adv. Res. Sci. Eng. Technol. **3**(9) (Sept 2016)
3. K. RamaKalyani, D. UmaDevi, Fraud detection of credit card payment system by genetic algorithm. Int. J. Sci. Eng. Res. **3**(7) (July 2012) ISSN 2229-5518
4. N. Sivakumar, A hybrid algorithmic approach for fraud detection in online transactions. Int. J. Adv. Res. Inform. Commun. Eng. **2**(5) (May 2014)
5. V.N. Dornadulaa, S. Geetha, Credit card fraud detection using machine learning algorithms, in *International Conference on Recent Trends in Advanced Computing 2019*, ICRTAC 2019 (2019)
6. J.R. Gaikwad, A.B. Deshmane, H.V. Somavanshi, S.V. Patil, R.A. Badgujar, Credit card fraud detection using decision tree induction algorithm. Int. J. Innov. Technol. Exploring Eng. (IJITEE) **4**(6) (Nov 2014) ISSN: 2278-3075
7. P. Sai Gowtham Kumar, P.A. Sumanth Reddy, A. Mary Posonia, Credit card fraud detection using machine learning. Int. J. Eng. Adv. Technol. (IJEAT) **9**(2) (Dec 2019) ISSN: 2249-8958
8. A. Rohilla, Comparative analysis of various classification algorithms in the case of fraud detection. Int. J. Eng. Res. Technol. (IJERT) **6**(09) (Sept 2017) ISSN: 2278-0181
9. S. Shamshida, M. Basthikodi, F. Zohara, Thameeza, M. Mumthaz, Evaluation of credit card fraud detection using SVM and ANN. J Emerg. Technol. Innov. Res. (JETIR) **6**(5) (May 2019)
10. B. Devi Meenakshi, B. Janani, S. Gayathri, N. Indira, Credit card fraud detection using random forest. Int. Res. J. Eng. Technol. (IRJET) **06**(03) (Mar 2019) e-ISSN: 2395-0056
11. P.K. Chan, W. Fan, A.L. Prodromidis, S.J. Stolfo, Distributed data mining in credit card fraud detection, in *Proceedings of the IEEE Intelligent Systems* (1999), pp. 67–74
12. M. Krivko, A hybrid-model for plastic-card fraud detection systems. Expert Syst. Appl. **37**, 6070–6076 (2010)

Regular Expression-Based Sentence Pattern Generation and Embedding for Convolution Classification Model RE-Sent2VecCNN

Velayutham Valli Mayil and Palaniappan Sharmila

Abstract Semantic representation of sentences in text analysis plays vital role in construction of classification model. The technique of embedding or vectorization of sentences includes additional semantic meaning of text, which makes the process of sentence classification straightforward in convolution neural network. Sentence patterns from text document are designed using regular expressions, and it is generated from sequence of tokens. In this paper, we propose a framework which promotes the semantic representation of sentences in convolution classification model. The framework also encourages cascaded regular expression to generate sentence sequences from text documents. The semantic feature of sentences is modeled by the vectorization process which converts the text into dimensional vector. In this paper, we present the proposed model with three phases and its contribution are (i) sentence pattern generation using regular expression model (ii) embedding or semantic representation of sentences using dimensional vector (iii) train the dimensional vector model in convolution classification model.

Keywords Semantic embedding · Sentence embedding · Regular expression in NLP · Text classification · Convolution neural network

1 Introduction

Deep learning models have produced remarkable output in Natural Language Processing (NLP). Improving unsupervised learning in NLP is the key importance of machine learning methods. Semantic representation of word in the documents would facilitate the process of document collection in NLP. Several research studies have been conducted to promote the word embedding model and then it was trained on very large collection of raw data in NLP [1–3]. The invention in turn creates the

V. Valli Mayil (✉)
Palanisamy College of Arts, Perundurai, India

P. Sharmila
Navarasam Arts and Science College for Women, Arachalur, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_60

challenges to produce semantic embedding for large text or sentences, paragraph or entire documents. Many studies reveal that semantic embedding based on selected keyword or sentences which improve the performance of convolution model [4, 5].

Large collection of text documents in corpus must have been clearly processed and converted into feature extraction model for NLP. The success of NLP is based on the factor of number and quality of training sample. Large as well as entire collection of text documents increases the complexity in classification model as it slows the training process. The selection of training data is possible by having preprocessed text in an effective way. A constraint based selection of sample would be the key advantage in unsupervised learning.

The characteristics or rule of word and sentences can be depicted by a user using Regular Expression (RE). The text preprocessing phases generates the specific sentences that have the same characteristics depicted in RE. Our work explored the use of cascaded RE to generate the sentences from text documents. The combination of RE, sentence embedding and convolution network exploit the effectiveness of sentence classification [6, 7].

This paper presents a novel approach to combine cascaded RE, sentence embedding with CNN at different levels. The proposed model consists of three phases. After having preprocessed text document, group of specific word is selected through cascaded RE which forms sentences in the first phase of the work. The features of sentences are extracted through sentence embedding technique in second phase of the model.

Sentence embedding make sure that the related sentences are closer and unrelated sentences are farther. Employing the sentence embedding will keep the semantic relationship between sentences. The semantic of sentences are represented in real valued vector through the process of vectorization. Third phase of the model is proposed for sentence classification model using CNN.

The main contribution of this work is summarized as follows:

(1) We propose a sentence embedding model which allows selecting certain sentences, properly modeled by Regular expression. Selected sentences are semantically represented by Sent2Vec model [8] MatteoPagliardini et al., 2017, which is simple unsupervised model allowing to compose sentence embeddings along with *n*-gram embeddings. Our model RE-sent2vec focused on specified sentences by RE, and it is converted into vector format.
(2) The distributed representations of sentences [9] in the form of vectors are classified over convolution neural network.

2 Proposed Architecture

The proposed model architecture is depicted in Fig. 1, which combines the unique representation of (a) cascaded Regular Expression for sentence extraction (b) model for sentence embedding (c) CNN for sentence classification.

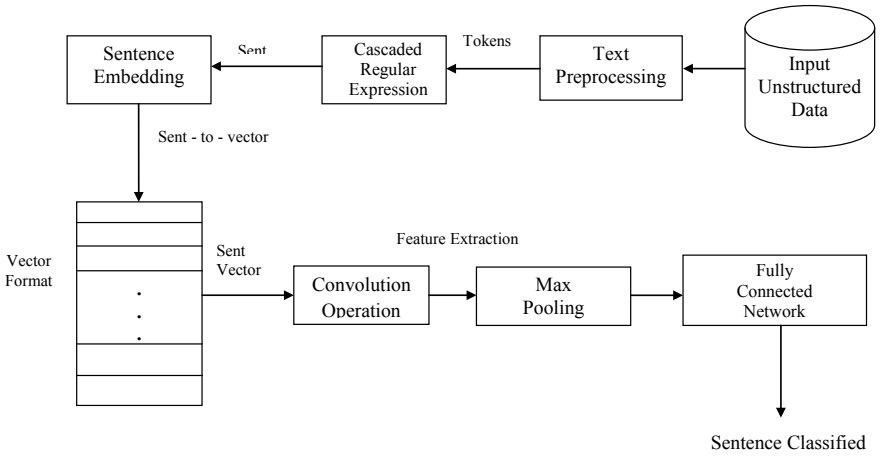


Fig. 1 Architecture of proposed system

2.1 Text Preprocessing

The major task of text classification is text preprocessing, which perform various task to convert document into sequence of words. In general, text preprocessing mainly includes word segmentation, word stemming, stop words removing, and low-frequency words removing.

The supervised learning on unstructured data can be performed by converting the document representation into numeric value semantic representation. The vectorization model is known as document term matrix which represents a text document in terms of rows and columns. In order to prepare vectorization model, the text corpus is required to perform preprocessing tasks such as stop word and whitespace removal, tokenization, case normalization, stemming, and lemmatization. The preprocessing task starts with tokenization. Keras supports a function such as "text to word sequence()" to split text into words. Keras provides the text to word sequence() function to split text into a list of words. It splits words by space, filters out punctuation and converts text to lowercase. Stemming is the process of reducing each word to its root or base. Porter stemming algorithm is used to perform stemming.

2.2 Model RE-Sent2vec

We propose a new unsupervised model, RE-Sent2Vec, for learning universal sentence embeddings. The model is the extension of sent2vec which consider the entire document for sentence embedding, whereas the new approach extracts only the relevant sentences satisfied the constraints depicted in RE. The sentences are optimized toward the combination of regular expression and embedding.

A regular expression is a modern language or notation for characterizing a set of strings. It returns all pattern of string which will be searched over the document or corpus. The rule or constraint can be characterized using regular expression. The cascaded regular expression formation can also be possible by having the sequence of regular expression which will give the sequence of strings called sentences. In this work, we propose a model to apply the RE in order to find the sentences which will be relevant for our search so that the search space of entire document can be reduced. Cascaded RE is used to match multiple regular expressions which extract the sentences contains sequence of strings. The multi-stage regular expression builds cascaded finite state automata which can be effectively used to find part of speech.

A framework, called TOKENSREGEX [7] Angel X. Chang, Christopher D. Manning, is popularly used to define cascaded regular expressions over token sequences. It is currently assembled in Stanford CoreNLP software package which defines various task on tokens as well as sequence of tokens.

The rules of sequences of tokens and grouping can be given by separate notations. TOKENSREGEX provides an extraction pipeline for matching against multiple regular expressions in stages. The pipeline is similar to a cascade of finite automata.

The extraction rules of TOKENSREGEX can be applied for various pattern identification such as string or tokens, compositional tokens, and filtering pattern. Composition of tokens can be identified by applying expression over previously identified tokens. Different tokens from expressions are combined to form composite tokens which extract sequence of tokens. Each token is represented by [<expression >] where < expression > uniquely extracts the strings matched with expression.

2.3 RE-Sent2vec-Distributed Representation of Sentence Embedding

The resulted sentence generation is considered for the next process called sentence embedding. In this phase, semantics of sentences are analyzed and the relative sentences are expressed by real values in the vector format [3]. Continuous bag-of-words and count approach of embedding techniques are not supported to find semantic or context of word. In this approach, a group of words in sentences are analyzed based on context and meaning. Sentence embedding converts the vector of words into vector of sentences. This process map sentences from high-dimensional space to low-dimensional space, and the position vector of different sentences reflects the relationship between them [3, 9]. The similarity and dissimilarity of sentences are determined by having the measure of centroid distances of sentences. Based on the result of similarity, the groups of sentences are classified. Sentence classification tasks can be improved by having better sentence embedding. Semantic representation is applied on machine learning by various studies [10–12], and it was experimented for sentence classification.

The author Pagliardini, Gupta, and Martin Jaggi [8] discussed unsupervised learning of sentence embeddings using compositional *n*-gram features and proposed a model Sent2Vec. The sentence embedding is defined as the average of the source word embeddings in the sentences. The collection of word for sentences is taken from output of the regular expression constructions.

Pennington, Jeffrey, et al. [13] proposed GloVe Averaging method, which is a supervised learning method that finds the average of word vector of the word in a sentence but it does not consider the order of word. Skip-Thought Vectors [7] is an unsupervised learning method which finds the neighboring phrases using NN. It uses encoder-decoder model where a sentence is encoded into a vector, and it is decoded for neighboring or surrounding text. FastSent [14] in this method, a sentence is represented as sum of word vectors. Methods discussed above have predicted the sentence from the neighboring words around it.

In our approach, a sentence is predicted from the representation of cascaded Regular expression. Each sentence is modeled by n-gram representation. Sentence embedding '*S*' is modeled as with length "*n*" is represented as Sent = (*s*1, *s*2,…,*s*n) *n*-gram model. Sentences are converted into vector format as sent2vect(*s*1), sent2vect(*s*2),….,(*s*n). The resulted vectors of a given sentence are mapped into a matrix "*M*" of size $N \times K$, as follows in Sent-Matric = sent2vec(*s*1), sent2vect(*s*2),….,sent2vec(*s*n).

2.4 *Re-Sent2vecCNN- Convolution Neural Network Model for Text Classification*

After having obtained vector representation of sentences, which is in the form of $W_1 + W_2 + ,\ldots, + W_n$ belongs to R^k, "*K*" dimensional vector with sentence length "*n*", W_i represents the *i*th word in the sentence, it is fed to the next phase of the proposed framework called Convolution Neural Network model. The input sentence vectors are trained with convolution operations consisting of filter/kernel and pooling functions.

A convolution operation entails a filter "*f*" $\in R$, which is applied to a window of "*h*" words to produce a new feature. For example, a new feature map is generated from a window of words $f\text{map} = \text{fn}(W_{i:h} + b)$ where $b \in R$ is a bias term and 'fn' is a nonlinear function.

The filter is applied to all windows of word in the sentences (*W*1:*h*,*W*2:*h* + 1,…,*W*n − *h* + 1:*n*) to generate new feature map feaMAP = [$f\text{map}_1$,$f\text{map}_2$,…$f\text{map}_\text{n}$],

Kernal operation can be applied on vector of sentences with different filter width to extract different feature of sentences. Pooling layer is applied over the resulted feature maps at the convolution layer. Pooling operations such as minimum, maximum and average extracts the most vital features and reduces the attributes size. The final component of connected layer combines all the feature maps from the previous

layers. It is capable to generate the output of the CNN network by learning the complex nonlinear interactions.

Convolutional neural network (CNN) model applies convolution operations on sent2vec vector and extract high-level abstract features of documents. Convolution operation applies a simple application called filter on the input matrix format to obtain a feature map which indicates the location and strength of detected features of the input document.

For example, a feature "feaMap" is generated from a window of words of sentences $S = (\text{word1}, \text{word2}, \ldots \text{wordn})$ by the following function. $F(\text{weight} * (\text{word1} \ldots \text{wordn}) + b)$, b is a bias term and F is a nonlinear function. Continuous mapping $C = (S1, S2 \ldots Sn)$ C is feature map obtained from filter operation.

The next operation which is to be applied on resultant feature map C is pooling function. In this work Max pooling operation is applied over the feature map and feaMap = maximum feature of $\{c\}$. Above described operations such as convolution and pooling are developed to find the single feature from the text document. Multiple filter operations are applied to obtain multiple features. These features are passed to fully connected softmax layer whose output depends on maximum probability features from output layer.

3 Experiment Setup and Dataset

The performance of different phases of proposed work has been evaluated with 2 datasets such as 20 News Group and REUTERS dataset. Various phases shown in Table 1 called text preprocessing. Sentence classification model is evaluated with the datasets. The 20 newsgroup dataset contains 20,000 documents and are divided into 20 categories and 8 categories such as misc.forsale, rec.motorcycles, re.sports.hockey, sci.space, sci.electronics are used for evaluation. Training and testing documents are separately available in the dataset.

20 newsgroups dataset for document classification, Source: http://people.csail.mit.edu/jrennie/20Newsgroups, #train 13,180, #test 5648, #labels 20, the second dataset REUTER with the version ModApte consists of 21,578 documents with 135 categories and 10 categories are used for evaluation.

Text preprocessing is the process of converting unstructured data into structured format called document term matrix (DTM) which represents the text in the form of rows and columns. Before preparing DTM, the irrelevant data from document must be eliminated. The process of tokenization stopword removal and stemming is applied on documents. Regular expression context is applied to refine the documents further. Cascaded regular expression is applied to bring the sequence of tokens called sentence.

CNN model is designed with sequence of layers for various components like convolution, dropout, pooling and connected network. In programming model, the architecture is initiated by calling sequential() function in keras, and it is proceeded with constructing convolution layer with number of filters, kernel size and activation

function. The model is initiated with the parameter such as 32 filters, kernel size = 8, activation "relu." The model is followed by pooling process with pool size-2, which reduces the dimensionality of the features, and it is got flatten in a 2D vector format. The resulted vector format is fed to the dense or connected network for the compilation of CNN with activation function and optimizer. Output layer of CNN uses a softmax activation function. The network is fit on training data using back propagation Adam stochastic gradient descent for the loss function binary cross entrophy.

Performance of the proposed work is reported by measuring factors such as accuracy, precision, recall, F1-score, and specificity. The performance of the algorithm with dataset 20 Newsgroup is given in the table. The accuracy measure is evaluated for the balanced data for both training (60%) and testing documents (20%). Training set is used to predict the model and testing dataset is used to estimate the accuracy of model. Recall, F score are measured with imbalanced dataset where the samples are selected randomly.

In order to prove more the recognition rate of the model, the random subsampling method called k-fold cross validation is incorporated. In this, initial data are randomly partitioned into k mutually exclusive subset with equal size. The one part of subset is selected for testing phase and other part are used for training the model, and it is performed "k" times. We experimented the proposed model with 9 different categories of dataset by having fivefold and tenfold validation. We achieved the best results in all categories with training and test set.

The second dataset that we used is Reuters dataset, and it is a multiclass and multi label dataset. It has 90 classes, 7769 training documents and 3019 testing documents. The experiment for proposed model has been conducted, and the result is depicted in table. The result in table shows that our proposed work performs well for different categories of documents. The incorporation of regular expression reduces the search time and overall classification time will be quite minimal. Similarly, the semantic vector inclusion is also reduces the time complexity.

4 Conclusion

Our proposed work mainly focuses on combining regular expression and sentence embedding in CNN. The model of RE-Sent2vec can be effectively combined with CNN based text classification. The inclusion regular expression constructs reduces the search space of documents and the sentence embedding will add semantic or context of sentences in vector format which improves the text classification process in simple way. The performance of proposed work has been evaluated with two different dataset. The experiments are conducted with varying filter size with convolution neural networks built on top of RE-sent2vec. Various metrics are estimated to evaluate the performance of proposed model. The experiments are showing the result of more than 85% in all metrics.

Table 1 Text preprocessing

RE-Sent2Vec-CNN	20 Newsgroup					Reuter				
Category	Accuracy	Precision	Recall	Fscore	Specificity	Category	Acc	Prec	Recall	fscore
rec.sport.hockey	0.98	0.96	0.99	0.99	0.98	Earn	0.98	0.88	0.93	0.90
ci.crypt	0.98	0.96	0.96	0.94	0.95	Acquisition	0.98	0.89	0.94	0.89
sci.electronics	0.87	0.87	0.89	0.88	0.87	Money-fx	0.87	0.91	0.93	0.90
sci.med	0.88	0.83	0.97	0.78	0.95	Grain	0.96	0.94	0.92	0.91
sci.space	0.97	0.96	0.97	0.97	0.88	Crude	0.97	0.88	0.89	0.90
soc.religion.christian	0.88	0.95	0.85	0.75	0.96	Trade	0.96	0.87	0.92	0.89
talk.politics.guns	0.93	0.95	0.95	0.94	0.95	Interest	0.93	0.92	0.91	0.89
talk.politics.mideast	0.96	0.96	0.90	0.86	0.86	Ship	0.96	0.87	0.92	0.91
talk.politics.misc	0.85	0.87	0.86	0.95	0.97	Wheat	0.98	0.88	0.93	0.92
Average	0.92	0.92	0.89	0.89	0.93	Average	0.95	0.89	0.92	0.90
fivefold validation—Avg	0.91	0.92	0.89	0.89	0.92	fivefold validation—Avg	0.94	0.89	0.91	0.90
tenfold validation—Avg	0.90	0.90	0.88	0.89	0.91	tenfold validation—Avg	0.93	0.89	0.89	0.90

References

1. S. Arora, Y. Li, Y. Liang, T. Ma, A. Risteski, A latent variable model approach to pmi-based word embeddings. Trans. Assoc. Comput. Linguist. **4**, 385–399 (2016)
2. K. Hermann, P. Blunsom, The role of syntax in vector space models of compositional semantics, in *Proceedings of ACL 2013* (2013)
3. T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in *NIPS - Advances in Neural Information Processing Systems*, vol. 26 (2013b), pp. 3111–3119
4. G. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors. CoRR, abs/1207.0580 (2012)
5. N. Kalchbrenner, E. Grefenstette, P. Blunsom, A convolutional neural network for modelling sentences, in *Proceedings of ACL 2014* (2014)
6. A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet classification with deep convolutional neural networks, in *Proceedings of NIPS 2012* (2012)
7. A.X. Chang, C.D. Manning, TOKENSREGEX: Defining cascaded regular expressions over tokens. Computer Science Department, Stanford University, Stanford, CA, 94305
8. M. Pagliardini, P. Gupta, M. Jaggi, Unsupervised learning of sentence embeddings using compositional n-gram features. arXiv preprint arXiv:1703.02507 (2017)
9. Q. Le, T. Mikolov, Distributed represenations of sentences and documents, in *Proceedings of ICML 2014* (2014)
10. Y. Shen, X. He, J. Gao, L. Deng, G. Mesnil, Learning semantic representations using convolution neural networks for web search, in *Proceedings of WWW 2014* (2014)
11. D. Tang, B. Qin, F. Wei, L. Dong, T. Liu et al., A joint segmentation and classification framework for sentence level sentiment classification. IEEE/ACM Trans. Audio Speech Lang. Process. **23**(11), 1750–1761 (2015)
12. Y. Chen, Convolutional neural network for sentence classification. Master's thesis, University of Waterloo (2015)
13. J. Pennington, R. Socher, C. Manning, Glove: Global vectors for word representation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2014), pp. 1532–1543
14. T. Kenter, A. Borisov, M. de Rijke, Siamese cbow: Optimizing word embeddings for sentence representations. arXiv preprint arXiv:1606.04640 (2016)

Comparative Study Between MobilNet Face-Mask Detector and YOLOv3 Face-Mask Detector

V. Aadithya, S. Balakumar, M. Bavishprasath, M. Raghul, and P. Malathi

Abstract During this Covid-19, face masks are used to avoid cross-contamination as part of an infection protection strategy. Wearing a face mask can help avoid infection by preventing individuals from coming into contact with pathogens. When someone coughs, speaks, or sneezes, there is a chance that the infection will spread into the air and affect those nearby. So to prevent the rate of spreading, face masks are highly mandatory. Tracking every individual manually is an expensive task; therefore, we save a lot of time, cost and effort by automating this process. This proposed automation can be done using Artificial Neural Networks. YOLOv3 and MobileNetv2 are popular architectures used in different object detection applications. Hence, paper compares the above architectures by their performance, accuracy from outputs of both under different scenarios.

Keywords YOLOv3 · MobileNetv2 · Face mask · Artificial Neural Network · Image processing

1 Introduction

In December 2019, the first occurrence of the novel coronavirus SARS-CoV-2 was detected in the wholesale market in Wuhan, China. It was later found to be an infectious disease affecting the respiratory system. Face masks can be used to cover the nose and mouth as a precautionary measure, as they block most of the airborne particles. During the influenza pandemic, it has been demonstrated that wearing masks would minimise the risk of the disease spreading, as it will protect you and others from you. When things return to normal, and the risk of the disease spreading is still there, if each one of them wears a face mask, everyone in public places will be safer. We may use surveillance cameras mounted in public places to promote the security of individuals going to public places [1]. This camera may be used

V. Aadithya · S. Balakumar · M. Bavishprasath · M. Raghul · P. Malathi (✉)
Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: p_malathy@cb.amrita.edu

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_61

to distinguish individuals who do not wear masks. The increasing era of machine learning and image processing is advancing the world's technology. This can also be used to distinguish people not wearing masks rapidly, in this case, quickly. It is a challenge to recognise [2] those without masks in a large crowd of people on the street. Still, a computer can make it easy by identifying people without masks and then the security personnel can easily take proper action [3]. This can be applicable to any future pandemics or epidemics. We first need to collect relevant sample data to train the model, then pre-process and then extract the characteristics and define the region of interest to classify and obtain the classification parameters to perform. Sample data can be gathered from open sources on the internet. The concept of data augmentation has been applied to use the same image towards training the model for various conditions, probably by rotating the same image and still identifying it to add an adjustment to the image. We can also construct our own data set by applying a mask to a traditional image without a mask due to advancements in technology. There are complex new neural networks. Deep learning has recently been widely used for object detection tasks and faster and more accurate detections. We need a lightweight architecture that is simple to process even on embedded platforms that do not have a high-performance GPU like in the framework. Architectures such as MobileNet are thus being built.

In order to prevent cross-contamination, face masks are used as part of an infection control strategy. So to avoid the rate of spreading, we automate the process of finding the people not wearing masks. We will compare the performance of the two famous ways of classifying images algorithms, the Mobinet face-mask detector and the YOLOv3 face-mask detector, to find out which performs better for the given problem.

The most common deep learning-based object detection methods are regional proposal-based approaches and those based on a single-pipeline technique. The earlier system used to implement convolution neural networks in object detection is the region-based convolutional neural network (R-CNN) technique [4, 5]. The selection search method is used to generate region suggestions from the input photos, and a convolution neural network is used to extract features from the region proposals. For training the support vector machine, the extracted features are used. Quick R-CNN and Faster R-CNN were also proposed to reappear based on the R-CNN process [6]. But the process is time-consuming, and lately, many new faster models are there; one such model is YOLO.

YOLOv3 (You Only Look Once) is similar to faster R-CNN points that differentiate YOLO from Faster R-CNN in that it simultaneously produces classification and bounding box regression. In YOLO, a single CNN predicts numerous bounding boxes as well as class probabilities for specific boxes. It also increases detection efficiency by training on complete images. This model offers some advantages when compared to other object detection models. It uses a darknet-53 architecture trained on Imagenet stacked with 53 more layers, giving us a fully convolutional neural network with 106 layers granting YOLO an increased accuracy than its previous versions and making it a slower one. Since the new network combines the YOLOv2 (Darknet-19) network and the residual network, it has several short cut connections.

It has 53 convolutional layers, so they call it Darknet-53. We can use an SSD (single-shot detector) for a better speed as it will compute its feature map in just a single run on the input image. SSD can be implemented with mobilenetv2 for making the process faster. Mobilenetv2 [7] is also a CNN, but it was specially designed to perform well even on mobile devices. It is based on an inverted residual structure between the bottleneck layers, where the residual connections are located. The intermediate expansion layer uses lightweight depthwise convolutions to filter features as a source of nonlinearity. A complete convolution layer with 32 filters and 19 residual bottleneck layers make up the entire MobileNetV2 architecture.

2 Related Works

In [8] (Zhongyuan Wang, 2020), the author provides details about face masks. There are three types of masked face datasets, Masked Face Detection Dataset (MFDD), Simulated Masked Face Recognition Dataset (SMFRD) and Real-world Masked Face Recognition Dataset (RMFRD). Modern advanced face recognition algorithms are mainly based on deep learning technology based on large face samples. According to the requirements of the masked facial recognition system, this article contains some intuitive data sets. This paper presents three types of masked face datasets to address masked face recognition tasks: Masked Face Detection Dataset (MFDD), Real-world Masked Face Recognition Dataset (RMFRD) and Simulated Masked Face Recognition Dataset (SMFRD). In [9] (Javid et al., 2020), authors provide that how covid-19 transit between humans and how face masks will prevent it. In [10] (Wu et al., 2019), the author proposed helmet detection implemented using YoloV3. From CCTV footage, vehicles are classified using OpenCV and Python to obtain images and detect helmets. In [11] (Ming, 2019), for face identification [12], the author proposed an improved methodology that merged the Local Binary Pattern (LBP) method with sophisticated image processing techniques such as Contrast Adjustment, Bilateral Filter, Histogram Equalisation and Image Blending. The challenges concerning an enhanced face recognition algorithm and its use in an attendance management system are discussed in this research. Although much progress has been achieved in face detection and identification, there are still particular challenges in reaching the level of human accuracy. This article proposes a novel method that solves some difficulties that diminish accuracy by combining local binary pattern (LBP) algorithms with sophisticated image processing techniques (such as contrast adjustment, double-sided filtering, histogram flattening and picture fusion). The LBP code may be improved by using face recognition. The overall accuracy of the facial recognition system improves as a result. In [13] (Zhao and Li, 2020), the author discusses the YOLOV3 based object detection algorithm. YOLOv3 (You Only Look Once) is one of the most commonly used methods for detecting objects based only on deep learning. The initial width and height of the anticipated bounding box are calculated using the *k*-means clustering algorithm. The initial cluster's centre determines the computed breadth and height, and processing massive data sets takes a long time.

A novel clustering approach was created to estimate the range and beginning height of clusters to tackle these challenges—the bounding box as intended. According to the final results, the new approach has a faster convergence rate and may be used to initialise the width and height of the anticipated bounding box. Therefore, you can choose more representative values for the initial width and height from the final bounding box.

3 Methods

To detect unmasked faces in public places using an image processing based dynamic face-mask detecting system.

3.1 Mobilenetv2 Implementation

(See Fig. 1).

The process of training the mask detector is done by dividing it into two different modules and respective sub-modules. In training, we will load the mask detector from the disc and train the model for this data set using Keras and TensorFlow to serialise the mask detector to the disc. In the implementation, we'll load the mask detector after training it, repeat the face detection, and categorise each face as a mask or a non-mask. The data should be born-again into an exact format to help in additional economic training and better detection. The subsequent sections would offer the operation of the code used for training our network. ImagePaths are grabbed within the dataset. Then, assigning data with labels, process them through imagePath, and load image preprocessing. Changing the size to 224 $\times$ 224 pixels, converting to an array format, and scaling the pixel intensity in the input image to the range $[-1, 1]$ are amongst the preprocessing procedures. To the data and label lists, add the preprocessed image and its corresponding label. To ensure our data is in NumPy array format, one-hot encodes our category labels, which means that our data will be in the following form. Using scikit-learn, the training data is 80%, and the test data is 20%. During the training process, we will apply the mutation to our image in real-time for generalisation. Data augmentation refers to the process of determining the random rotation, zoom, shear, shift and flip parameters. We use the augmented object at training time. Load pre-trained ImageNet weights to MobileNet, exploit off the top of the network. We were then constructing a brand new, fully connected head and appending it to the bottom in situ of the recent head. Phase transition the base layers of the network. The weights of those base layers won't be updated during the backpropagation method, whereas the highest layer we tend to light will be tuned. After training completion, we'll evaluate the ensuing model on the check set. Here, we would create predictions on the test set, grabbing the best likelihood category

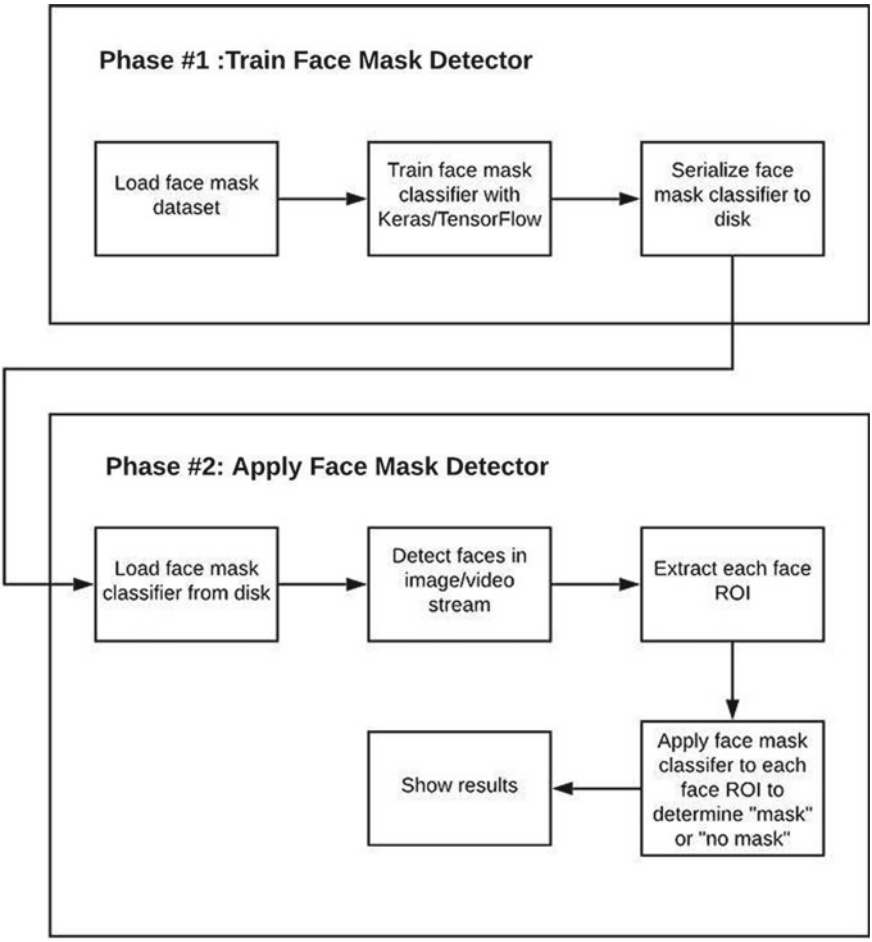


Fig. 1 Face-mask detection architecture using mobilenetv2

label indices. We plot the accuracy and losses together to check for overfitting. Our face-mask classification model is serialised to disc.

3.2 YOLOV3 Implementation

Here, we will discuss the working of the YOLO algorithm and how the algorithm detects the object in the image. In our project, we have an input image of 416×416 (Fig. 2).

Fig. 2 Face-mask detection architecture using yolov3

3.2.1 Detection Process

YOLO v3 detects three scales generated by carefully downsampling the input image's scale by thirty two, sixteen and eight, respectively.

The terrible first detection will be created by the 82nd layer. The eighty one primary layers will down-sample the image within the network in such the most straightforward way that once the idea reaches the 81st layer, it'll have a stride of 32. So, when our input size of the image is 416 × 416, the resultant output of the feature map will be thirteen × thirteen. And one detection will be done by mistreatment of the one × one kernel, which can provide us—the detection kernel of thirteen × thirteen × twenty one.

After that the feature map from the 79th layer will be trained with a few convolutional layers before being upsampled by 2X to dimensions twenty six × twenty six. From the previous layer 61, the feature map is concatenated. The merged feature maps are now processed through a few one x one convolutional layers again. The 94th layer then performs the second detection, producing a detection feature map of twenty six × twenty six × twenty one.

The 91st layer's feature map is concatenated with a feature map from the 36th layer using the same procedure as before. As before, a few one × one convolutional layers will follow to mix the knowledge from the previous 36th layer. We will create the ultimate third discovered at the 106th layer, which can offer us a map size of 52 × 52 × 21.

The responsibility of the thirteen × thirteen layers is going to be to detect the more prominent objects. In contrast, the fifty two × fifty two layer will be accountable for sleuthing more minor things, whereas the twenty-six × twenty-six layer can discover medium-sized objects.

3.2.2 Thresholding

The algorithm rule of YOLOV3 generates 10,647 boxes, most of which are irrelevant/redundant. Therefore, we must filter out unnecessary fields and discard them. This should be done using a confidence threshold and only maintain those cells that may have a significant confidence threshold. This step eliminates the detection of abnormal objects.

3.2.3 Non-max Suppression

After thresholding, we will try to find multiple boxes for each detected object.

However, we only need a box. Using Non-max suppression, the bounding box is calculated. Non-max suppression uses an idea called "intersection over union" or IoU. It takes two input boxes. As the name suggests, it can quantify the intersection and union of two boxes. After assigning IoU, the non-maximum suppression works as follows: Repeat till no boxes to process:

Select the box with the best chance of detection.

Take away all the boxes with a high IoU with the selected box.

Mark the chosen box as "processed".

This type of filtering ensures that there is only one bounding box on each recognised object.

3.2.4 Bounding Box Labelling

In the process of non-max suppression, we have neglected the class label. We will check if any bounding boxes have class labels as MASK, whilst merging to assign the class label.

If YES: The final merged bounding box will be false-positive as MASK.

Otherwise: It shows that all the bounding boxes which are merged, are assigned with the NO_MASK label. This results in the final merged bounding box being labelled as.

NO_MASK.

4 Results

(See Fig. 3, Table 1).

5 Conclusion

From the sample outputs that we have shown above, it is apparent that the YOLO implementation is preferred in cases of large crowds due to its good object detection base to identify multiple faces on the same image (when compared to mobile networks) and correctly finding the NO Mask classification when the face is hidden by hand. As MobileNetv2 is a lightweight architecture, it can easily be integrated with IoT devices to perform detection without compromising accuracy, and YOLO can be used in the case of heavily loaded scenarios.

Figure (1a)

Figure (1b)

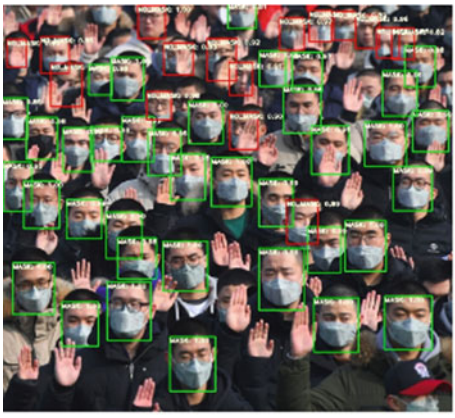


Figure (2a)

Figure (2b)

Fig. 3 Outputs from both mobilenet and yolo implementations under 2 sample scenarios

Table 1 Comparisons made from the output

Mobilenetv2	YOLOv3
Figure (1a) shows the result obtained from mobilenetv2 implementation, it was only able to find few people, and it couldn't identify most of them accurately. Then regarding the accuracy, it is highly less accurate, and the false-positive cases are easily possible. (Fig. (1b) wrongly identified as masked). The wrong predictions are where there is an obstruction in front of the mask	Figure (2a) shows the output obtained from the yolov3 implementation. It was able to find nearly all the people. Then regarding the accuracy, it is highly accurate, and the wrong cases are comparatively reduced. The reduced false-positive rate has proved this approach a better one. (Fig. (2b) Correctly identified as no Mask)
Figure (1b) is obtained by mobilenetv2 architecture. Wrongly predicted hand as a mask	Figure (2b) is obtained by YOLOv3 implementation. Correctly predicted No mask

References

1. R. Manjusha, L. Parameswaran, Design of an image skeletonization based algorithm for overcrowd detection in smart building, in *Lecture Notes in Computational Vision and Biomechanics*, vol. 30 (Springer Netherlands, 2019), pp. 615–629
2. P. Malathi, D. Bharathi, R.E. Vinodhini, Illumination invariant face recognition using fisher linear discriminant algorithm (FLDA). Int. J. Control Theory Appl. **9**(10), 4201–4210 (2016)
3. W. Lan, Y.W. Jianwu Dang, S. Wang, Pedestrian detection based on yolo network model. IEEE **121**, 435–447 (2018)
4. G. Yolcu, I. Oztel, S. Kazan, C. Oz, K. Palaniappan, T.E. Lever, F. Bunyak, Facial expression recognition for monitoring neurological disorders based on convolutional neural network. Multimedia Tools Appl **78**, 31581–31603 (2019)
5. R. Ravi, S. Yadhukrishna, R. prithviraj, A face expression recognition using cnn lbp. pp. 684–689 (2020)
6. T. Keshari, S. Palaniswamy, Emotion recognition using feature-level fusion of facial expressions and body gestures. pp. 1184–1189 (2019)
7. K.K.R. Sanj Kumar, G. Subramani, T. Senthil Kumar, L. Parameswaran, A mobile-based framework for detecting objects using SSD-MobileNet in indoor environment, in *Intelligence in Big Data Technologies–-Beyond the Hype*, vol. 1167 (Springer Singapore, Singapore, 2021), pp. 65–76
8. Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, H. Chen, Masked face recognition dataset and application. (6), 443–461 (2020)
9. B. Javid, M.P. Weekes, N.J. Matheson, Covid-19: should the public wear face masks?. BMJ, 369 (2020)
10. F. Wu, G. Jin, M. Gao, Z. HE, Y. Yang, Helmet detection based on im-proved YOLO v3 deep model, in *2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC),* pp. 363–368
11. S.M.B.F. Ming, An improved face recognition algorithm and its application in the attendance management system. (Elsevier BV, 2019)
12. T. Senthil Kumar, S.B. Sai Venkateshwaran, Evaluation of video analytics for face detection and recognition. Int J Appl Eng Res **10**, 24003–24016 (2015)
13. L. Zhao, S. Li, Object detection algorithm based on improved yolov3. Electronics **9**, 537 (2020)

Applying the RPBK22 Technique for Secure the Generalized Data

K. Shantha Shalini, S. Leelavathy, Kishore Pani, M. P. Dinakar, R. Guruprassath, and Sankarganesh

Abstract Nowadays, data is extremely important in the people's world. Recently, people have generated more data using internet media. As a result, this internet data lacks adequate protection and hackers may simply access it. This research work utilizes the RBPK22 method to overcome the difficulty. There are seven stages to implement this method. To find the mystery prime key S; To find the X_1 and X_2 values from prime numbers; To find the $\overline{X}_1$ and $\overline{X}_2$ values; To find the standard deviation values with the help of Step 2 and Step 3; To exchange a and b esteems from left in a matrix; To locate the T-test values and pair them that numbers from left to right; To exchange the odd and even equation a and b in a given lattice. The RPBK22 strategy gives extraordinary security, while appearing differently with ChaCha technique.

Keywords RPBK22 · Encryption · Decryption · Prime · T-test · ChaCha

1 Introduction

The enormous amount of information time has been purchased with sufficient freedoms for logical progress in enhancing medical care, monetary development, upgrading the educational framework, and other forms of entertainment. The examination of large amounts of data must go through several stages in order to obtain some significant qualities, which include few stages such as information reconciliation, information collection, data cleaning, data extraction, inquiry preparation, information demonstrating, and translation. Each level has various challenges such as heterogeneity, practicality, complexity, security, and protection. Because of its massive frameworks such as large volume, speed, and diversity, one of the major difficulties in analyzing huge information is security and protection. Even though

K. Shantha Shalini (✉) · S. Leelavathy · K. Pani · M. P. Dinakar · R. Guruprassath
Deparment of CSE, Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation, Ariyanur, India

Sankarganesh
Department of Electronics and Communication, PSR Engineering College, Sivakasi, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_62

there are mostly four attributes of huge information security: Foundation and structure security; Information protection; Information guideline; Vital and responsive security. To overcome these challenges, this research work applies the RBJ25 technique. The extra revolution XOR for ChaCha is considered as deficiency assault [1]. This creator has utilized a new hash idea for key speculating and ending conditions [2]. Creator has presented the bricklayer assault for establishing an investigation on ChaCha [3]. They focused on the security for Double-A [4]. They made a new plan for implementing a secure, quick and adaptable calculation [5]. SRB18 strategy is used to provide security to information [6]. SRB21 Phase 1 and SRB21 Phase 2 strategy are used to leverage security to the desired information [7, 8]. CBB21 [9], CBB22 [10], CBB20 [12], and RJB25 techniques are used to leverage security to information [13]. Various research works have been carried out to manage and secure data in health, bio statistics, network security, information security, and IOT environments [14, 15], [16–26].

2 Methods

This section discusses about the encryption and decryption methods. The proposed RPBK22 method has 7 phases. 1. To find the mystery prime key S. 2. To find the X_1 and X_2 values from prime numbers. 3. To find the $\overline{X}_1$ and $\overline{X}_2$ values. 4. To find the standard deviation values with the help of Step 2 and Step 3. 5. To trade a and b esteems from left in a matrix. 6. To locate the T-test values and pair them that numbers from left to right. 7. To exchange the odd and even equation a and b in a given lattice.

2.1 Encryption

- RPBK22 method is Table 1 for encryption.
- Input of the matrix (IP).

$$\text{IP} = \begin{bmatrix} 401/4 & 402/4 & 403/4 & 404/4 & 405/4 \\ 406/4 & 407/4 & 408/4 & 409/4 & 410/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

- Prime Numbers—1, 3, 5, 7, 11, 13, 17, 19, 23
- $X_1 = 3, 5, 7, 11$; $X_2 = 13, 17, 19, 23$
- **Using Eq. (2) and (9) and** Table 2**.**
- $\overline{X}_1 = \sum X_1/N_1$;

Table 1 Encryption algorithm

Steps	Encryption Procedure
I	To find a prime numbers in given matrix IP
II	Equally separate the two parts of prime numbers and apply those values in Eqs. (1), (2), and (3)
III	T-Test Formula $= (\overline{X}_1 - \overline{X}_2) / \sqrt{((S_1^2/N_1) + (S_2^2/N_2))}$ (1)
IV	To find the values for Eq. (1) with the help of Eqs. (2) and (3)
V	$\overline{X}_1 = \sum X_1/N_1 \quad \overline{X}_2 = \sum X_2/N_2$ (2)
VI	$S_1 = \sqrt{\sum (X_1 - \overline{X}_1)^2/(N_1 - 1)} S_2 = \sqrt{\sum (X_2 - \overline{X}_2)^2/(N_2 - 1)}$ (3)
VII	To pair the *T*-test value from left to right
VIII	To find the *n* value with the help of '*K*' and *K* is secret key
IX	If n is odd number where $n = 2k + 1$ $a^n + b^n = (a + b)\left(a^{n-1} - a^{n-2}b + a^{n-3}b^2 \cdots + b^{n-2}a - b^{n-1}\right)$ **(4)**
X	To apply the *n* value in Eq. (4)
XI	To merge the step *x* values from left to right
XII	Step *xi* values will be apply in given matrix
XIII	To find the *n* value with the help of '*K*'
XIV	If *n* is even number where $n = 2k$ $a^n + b^n = (a + b)\left(a^{n-1} - a^{n-2}b + a^{n-3}b^2 \cdots - b^{n-2}a + b^{n-1}\right)$ **(5)**
XV	To apply the *n* value in Eq. (5)
XVI	To merge the step *xv* values from left to right

Table 2 X_1 and X_2 values

X_1	$(X_1 - \overline{X}_1)$	$(X_1 - \overline{X}_1)^2$	X_2	$(X_2 - \overline{X}_2)$	$(X_2 - \overline{X}_2)^2$
3	−3.5	12.25	13	−5	25
5	−1.5	2.25	17	−1	1
7	0.5	0.25	19	1	1
11	4.5	20.25	23	5	25
	$\sum (X_1 - \overline{X}_1)^2$	35		$\sum (X_2 - \overline{X}_2)^2$	52

- $\overline{X}_1 = (3 + 5 + 7 + 11)/4$; $\overline{X}_1 = 26/4$; $\overline{X}_1 = 6.5$
- $\overline{X}_2 = \sum X_2/N_2$;
- $\overline{X}_2 = (13 + 17 + 19 + 23)/4$ $\overline{X}_2 = 18$

- **Using Eqs. (3) and (8)**
- $S_1 = \sqrt{\sum (X_1 - \overline{X}_1)^2/(N_1 - 1)} \quad S_2 = \sqrt{\sum (X_2 - \overline{X}_2)^2/(N_2 - 1)}$
- $S_1 = \sqrt{(35/(4 - 1))} \; S_2 = \sqrt{(52/(4 - 1))}$
- $S_1 = \sqrt{(35/3)} \; S_2 = \sqrt{(52/(3))}$
- $S_1 = 11.66 S_2 = 17.33$

- **Using Eqs. (1) and (10)**
- T-Test Formula $= (\overline{X}_1 - \overline{X}_2) / \sqrt{((S_1^2/N_1) + (S_2^2/N_2))}$
- T-Test Formula $= (6.5 - 18) / \sqrt{((11.66^2/4) + (17.33^2/4))}$
- T-Test Formula $= 11.5 / \sqrt{((135.95/4) + (300.32/4))}$
- T-Test Formula $= 11.5 / \sqrt{((135.95 + 300.32)/4))}$
- T-Test Formula $= 11.5 / \sqrt{(436.27/4)}$
- T-Test Formula $= 11.5 / \sqrt{109.06}$
- "Pair the T-test value from left to right (1,1) (5,1) (0,9) (0,6) and swap it those numbers."

Step "1: (1,1)"

$$\text{TTE} = \begin{bmatrix} 401/4 & 402/4 & 403/4 & 404/4 & 405/4 \\ 406/4 & 407/4 & 408/4 & 409/4 & 410/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

where TTE is T-Test Encryption.

Step "2: (5,1)"

$$\text{TTE} = \begin{bmatrix} 401/4 & 406/4 & 403/4 & 404/4 & 405/4 \\ 402/4 & 407/4 & 408/4 & 409/4 & 410/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

Step "3: (0,9)"

$$\text{TTE} = \begin{bmatrix} 410/4 & 406/4 & 403/4 & 404/4 & 405/4 \\ 402/4 & 407/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

Step "4: (0,6)"

$$\text{TTE} = \begin{bmatrix} 407/4 & 406/4 & 403/4 & 404/4 & 405/4 \\ 402/4 & 410/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

- To find the n value; $a = 2$, $b = 3$, $k = 2$, $n = 5$, ***n* is odd number**
- **Using Eqs. (4) and (7)**
  - $a^2 + b^3 = (2 + 3)(2^{5-1} - 2^{5-2}3 + 2^{5-3}3^2 - 2^{5-4}3^3 - 2^{5-5}3^4 + 3^{5-4}2 - 3^{5-3}2 + 3^{5-2}2 - 3^{5-1})$
  - $a^2 + b^3 = (5)\ (2^4–2^33 + 2^29–2^127\text{-}2^081 + 3^12–3^22 + 3^32–3^4)$
  - $a^2 + b^3 = (5)\ (16–24 + 36–54\text{-}81 + 6–18 + 54–81)$
  - $a^2 + b^3 = (5)\ (-146)$
  - $a^2 + b^3 = (5,1), (4,6)$

Step "5: (5,1)"

$$\text{TEB} = \begin{bmatrix} 407/4 & 402/4 & 403/4 & 404/4 & 405/4 \\ 406/4 & 410/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

Step "6: (4, 6)"

$$\text{TEB} = \begin{bmatrix} 407/4 & 402/4 & 403/4 & 404/4 & 410/4 \\ 406/4 & 405/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

- To find the n value: $a = 2$, $b = 3$, $k = 2$, $n = 4$
- ***n* is even number**

Using Eq. (5) and (6)

- $a^2 + b^3 = (2 + 3)(2^{4-1} - 2^{4-2}3 + 2^{4-3}3^2 - 2^{4-4}3^3 + 3^{4-3}2 - 3^{4-2}2 + 3^{4-1})$
- $a^2 + b^3 = (5)\ (2^3–2^23 + 2^19–2^027 + 3^12–3^22 + 3^3)$
- $a^2 + b^3 = (5)\ (16–12 + 18–54 + 6–18 + 27)$
- $a^2 + b^3 = (5)\ (-17)$
- $a^2 + b^3 = (5,1), (7,0)$

Step "7: (5, 1)"

$$\text{TEB} = \begin{bmatrix} 407/4 & 406/4 & 403/4 & 404/4 & 410/4 \\ 402/4 & 405/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

Step "8: (7, 0)"

$$\text{TEB} = \begin{bmatrix} 408/4 & 406/4 & 403/4 & 404/4 & 410/4 \\ 402/4 & 405/4 & 407/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

2.2 *Decryption*

- RPBK22 method is Table 3 for decryption.
- Using Eq. (6)
- "Pair of numbers are (0,7), (1,5)".

Step "1: (0, 7)"

Table 3 Decryption algorithm

Steps	Decryption procedure
I	To find the n value with the help of 'K'
II	If n is even number where $n = 2\,k$ $a^n + b^n = (a+b)\left(a^{n-1} - a^{n-2}b + a^{n-3}b^2 \cdots - b^{n-2}a + b^{n-1}\right)$ (6)
III	To apply the n value in Eq. (6)
IV	To merge the step III values from right and apply
V	To find the n value with the help of 'K' and K is secret key
VI	If n is odd number where $n = 2\,\mathrm{k} + 1$ $a^n + b^n = (a+b)\left(a^{n-1} - a^{n-2}b + a^{n-3}b^2 \cdots + b^{n-2}a - b^{n-1}\right)$ (7)
VII	To apply the n value in Eq. (7)
VIII	To merge the step VII values from right and apply
IX	To track down an indivisible number in framework TTD
X	Similarly independent the two pieces of indivisible numbers
XI	$S_1 = \sqrt{\sum (X_1 - \overline{X}_1)^2/(N_1 - 1)} S_2 = \sqrt{\sum (X_2 - \overline{X}_2)^2/(N_2 - 1)}$ (8)
XII	$\overline{X}_1 = \sum X_1/N_1 \qquad \overline{X}_2 = \sum X_2/N_2$(9)
XIII	T-Test Formula = $(\overline{X}_1 - \overline{X}_2) / \sqrt{((S_1^2/N_1) + (S_2^2/N_2))}$(10)
XIV	To consolidate the qualities step XIII from right to left
XV	Step XIV values apply in the matrix

$$\mathrm{DB} = \begin{bmatrix} 407/4 & 406/4 & 403/4 & 404/4 & 410/4 \\ 402/4 & 405/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

Step "2: (1, 5)"

$$\mathrm{DB} = \begin{bmatrix} 407/4 & 402/4 & 403/4 & 404/4 & 410/4 \\ 406/4 & 405/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

• **Using Eq. (7)**; "Pair of numbers (6, 4), (1, 5)".
Step "3: (6, 4)"

$$\mathrm{DA} = \begin{bmatrix} 407/4 & 402/4 & 403/4 & 404/4 & 405/4 \\ 406/4 & 410/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

Step "4: (1, 5)"

$$\mathrm{DA} = \begin{bmatrix} 407/4 & 406/4 & 403/4 & 404/4 & 405/4 \\ 402/4 & 410/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

• Pair the T-test value from right to left (6,0), (9,0), (1,5), and (1,1) and swap it those numbers.

$$\mathrm{TTD} = \begin{bmatrix} 407/4 & 406/4 & 403/4 & 404/4 & 405/4 \\ 402/4 & 410/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

where TTD is T-Test Decryption.
Step "5: (6,0)"

$$\text{TTD} = \begin{bmatrix} 410/4 & 406/4 & 403/4 & 404/4 & 405/4 \\ 402/4 & 407/4 & 408/4 & 409/4 & 401/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

Step "6: (9,0)"

$$\text{TTD} = \begin{bmatrix} 401/4 & 406/4 & 403/4 & 404/4 & 405/4 \\ 402/4 & 407/4 & 408/4 & 409/4 & 410/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

Step "7: (1,5)"

$$\text{TTD} = \begin{bmatrix} 401/4 & 402/4 & 403/4 & 404/4 & 405/4 \\ 406/4 & 407/4 & 408/4 & 409/4 & 410/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

Step "8: (1,1)"

$$\text{TTD} = \begin{bmatrix} 401/4 & 402/4 & 403/4 & 404/4 & 405/4 \\ 406/4 & 407/4 & 408/4 & 409/4 & 410/4 \\ 411/4 & 412/4 & 413/4 & 414/4 & 415/4 \\ 416/4 & 417/4 & 418/4 & 419/4 & 420/4 \\ 421/4 & 422/4 & 423/4 & 424/4 & 425/4 \end{bmatrix}$$

3 Result and Discussion

The proposed algorithm RPBK22 encryption performance compared with ChaCha. ChaCha concept moves all diagonal values to the 1st column.

The three by three matrix has 24 bytes of fie size; the six by six matrix has 76 bytes of file size; the ten by ten matrix has 312 bytes of file size; the fifteen by fifteen matrix has 812 bytes of file size; the twenty by twenty matrix has 1531 bytes of file size; and the forty by forty matrix has 6580 bytes of file size as shown in Table 4.

From Fig. 1, the RPBK22 technique and has compared the encryption speed in seconds. The encryption performance of the speed 1.99 (s), 2.33 (s), 2.95 (s), 3.43

Table 4 RPBK21 encryption performance

File size	ChaCha	RPBK22
24	1.69	1.99
76	1.29	2.33
312	2.73	2.95
822	2.64	3.43
1531	3.4	3.79
6580	2.27	4.1

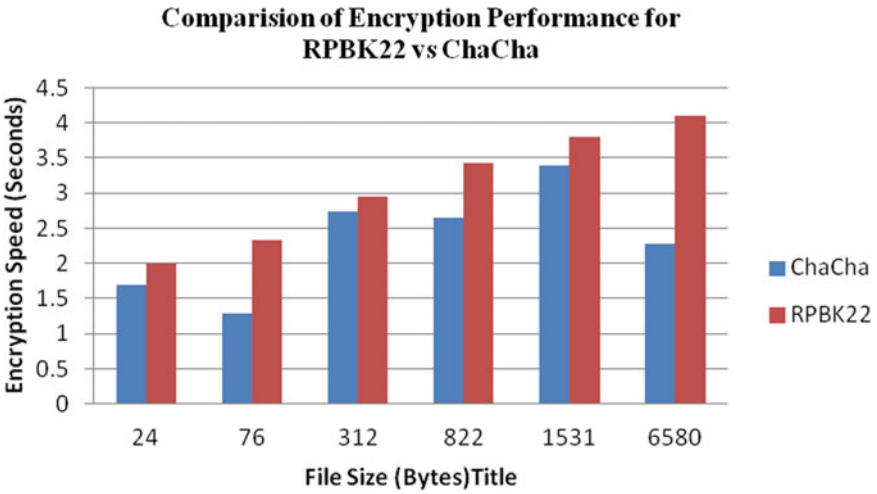


Fig. 1 Encryption performance

(s), 3.79 (s) and 4.1 (s) for the RPBK22 and 1.69 (s), 1.29 (s), 2.73 (s), 2.64 (s), 3.4 (s), and 2.27 (s) for the ChaCha. The RPBK22 gives more protection of the data; when compared to existing techniques.

4 Conclusion

The data globe is the current global situation, where the huge amount of data are generated via internet media; it is unrestricted; it does not have exceptional security. The proposed research work has successfully addressed these challenges by implementing the ChaCha method. This method effectively hacks data from software developers. The proposed RPBK21 strategy has successfully implemented 7 phases by finding the mystery prime key S; X_1, X_2, $\overline{X}_1$ and $\overline{X}_2$ values from prime numbers. Also, this research work has detected the standard deviation values. Further, it has traded a and b esteems from left in a matrix. Also, it has couple the integers from

left to right to locate the T-test values, and it successfully exchanges the undividable and non-negative whole integers in a given lattice by successfully swapping the M pair numbers. The RPBK21 strategy gives extraordinary security, while appearing differently about ChaCha and RBJ25 techniques. Finally, it adds the prime variables tasks for establishing information security.

References

1. C. Bagath Basha, S. Rajaprakash, Enhancing the security using SRB18 method of embedding computing. Microprocess. Microsyst. (2020)
2. C. Bagath Basha, S. Rajaprakash, Securing Twitter data using phase I methodology. Int. J. Sci. Technol. Res. 1952–1955 (2019)
3. C. Bagath Basha, S. Rajaprakash, Applying the SRB21 phase II methodology for securing Twitter analyzed data, in *AIP Conference Proceedings of the International Conference on Mechanical Electronics and Computer Engineering* (2020)
4. C. Bagath Basha, S. Rajaprakash, Applying the CBB21 phase 2 method for securing twitter analyzed data. Adv. Math. Sci. J. 1085–1091(2020)
5. C. Bagath Basha, S. Rajaprakash, V.V.A. Harish, M.S. Krishna, K. Prabhas, Securing Twitter analysed data using CBB22 algorithm. Adv. Math. Sci. J. 1093–1100 (2020)
6. U.M. Prakash, K. Kottursamy, K. Cengiz, U. Kose, B.T. Hung, 4x-expert systems for early prediction of osteoporosis using multi-model algorithms. Measurement, **180**, 109543 (2021)
7. A. Saranya, K. Kottilingam, A survey on bone fracture identification techniques using quantitative and learning based algorithms, in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)* (IEEE, 2021), pp. 241–248
8. R.D. Rajagopal, S. Murugan, K. Kottursamy, V. Raju, Cluster based effective prediction approach for improving the curable rate of lymphatic filariasis affected patients. Clust. Comput. **22**(1), 197–205 (2019)
9. K. Cengiz, R. Sharma, K. Kottursamy, K.K. Singh, T. Topac, B. Ozyurt, Recent emerging technologies for intelligent learning and analytics in big data, in *Multimedia Technologies in the Internet of Things Environment* (Springer, Singapore, 2021), pp. 69–81
10. R. Mehra, K. Pachpor, K. Kottilingam, A. Saranya, An initiative to prevent japanese encephalitis using genetic algorithm and artificial neural network, in *2020 International Conference on Computational Intelligence (ICCI)* (IEEE, 2020), pp. 142–148
11. A. Shanthini, G. Manogaran, G. Vadivu, K. Kottilingam, P. Nithyakani, C. Fancy, Threshold segmentation based multi-layer analysis for detecting diabetic retinopathy using convolution neural network. J. Ambient Intell. Hum. Comput. 1–15 (2021)
12. K.P. Vijayakumar, K.P.M. Kumar, K. Kottilingam, T. Karthick, P. Vijayakumar, P. Ganeshkumar, An adaptive neuro-fuzzy logic based jamming detection system in WSN. Soft. Comput. **23**(8), 2655–2667 (2019)
13. R. Arul, G. Raja, K. Kottursamy, P. Sathiyanarayanan, S. Venkatraman, User path prediction based key caching and authentication mechanism for broadband wireless networks. Wirel. Pers. Commun. **94**(4), 2645–2664 (2017)
14. G. Raja, K. Kottursamy, A. Theetharappan, K. Cengiz, A. Ganapathisubramaniyan, R. Kharel, K. Yu, Dynamic polygon generation for flexible pattern formation in large-scale uav swarm networks, in *2020 IEEE Globecom Workshops (GC Wkshps)* (IEEE, 2020), pp. 1–6
15. V. Ponnusamy, K. Kottursamy, T. Karthick, M.B. Mukeshkrishnan, D. Malathi, T.A. Ahanger, Primary user emulation attack mitigation using neural network. Comput. Electr. Eng. **88**, 106849 (2020)

SRSIoT: Toward a Specialized Approach of SRS for Smart IoT Applications

Darshan Pradeep Pandit and Chowdary Ch Smitha

Abstract Every year, there is a significant increase in IoT devices, which are monitored and controlled by smart IoT applications. The developed applications are utilized by a wide range of users, who may or may not be familiar with IoT devices. In order to achieve a successful application design model, the article presents a software engineering methodology for IoT devices that include people and their needs in a smart environment. Clear user requirements aid in the identification and analysis of IoT devices for various user requirements based on the varied smart application. The appropriate requirement analysis establishes a solid foundation for creating IoT applications for smart environments. This article presents an overview of IoT applications and investigates why the existing Software requirement specification (SRS) technique is insufficient to capture all elements of an IoT application. Hence, this research work proposes that the SRS toward IoT application should be different from the conventional Software product.

Keywords Smart environment · Software engineering · IoT · Stakeholder · Requirement

1 Introduction

The Internet of Things (IoT) has grown popular and has attracted a lot of attention from researchers since it is an intriguing issue that contributes a smart environment to human existence. The cluster of infrastructures that connects linked devices enables data management, data mining, and access to the data generated by those devices [1]. The Internet of Things (IoT) is a global infrastructure network that connects physical and virtual objects by utilizing data collection and transmission mechanisms.

D. P. Pandit (✉) · C. C. Smitha
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

C. C. Smitha
e-mail: smitha@kluniversity.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_63

IoT provides sensor and connection capabilities with device unique identification for independent cooperative services and application development. The Internet of Things network distinguishes network connectivity, sensitive data capture, data transmission, and interoperability. Users in smart IoT environment interact with each other through various types of connections such as RFID, Wi-Fi, Bluetooth, and ZigBee. GSM, GPRS, 3G, and LTE technologies are responsible for forming wide area IoT network [2, 3]. The smart applications provide a platform for interaction between the connectivity devices and sensors from the IoT infrastructure. With the development of IoT infrastructure, there should be a software engineering approach. But nowadays, users are adopting the available IoT application and applying it in a smart environment, which may contain unnecessary function and sensors embedded into it. This makes to consume battery life and increase memory as there is a lot of unnecessary information generation, which in return waste the users time and money. In order to avoid wastage, user requirements are thoroughly identified and analyzed to develop IoT solutions. This solution should be designed and developed with respect to software engineering standards, which helps the device and sensors to be fault tolerant and reusable for a smart environment. Before implementing any device and sensors in IoT infrastructure, there is a requirement to identify the need of that device. Once the user requirements are acquired, the design will be easy and it will also save development time and cost of the application. It also eliminates unnecessary functions, which are simply embedded into the applications. These unnecessary functions may consume more memory and power, which may slow down the application. Hence, the application should be designed based on the user perspective. This paper deals with software engineering approach, which provides the detailed view of the stakeholders as per the applications of IoT and varying user requirement and stakeholders according to the IoT application. There is a need to develop Software engineering approach to identify and analyze users and their requirements in order to design IoT application.

2 Software Requirement Specifications

SRS enables us to define the many features of a smart application as well as how the application is anticipated to behave. It outlines the processes required to create a product that meets all user and business requirements. Sample Software Requirement Specifications (SRS) for smart agricultural and healthcare IoT environments are provided in the tables.

3 Software Requirement Specification for IoT Applications (SRSIoT)

SRS plays a vital role in the development of an IoT application. By monitoring SRS, the developer can monitor the IoT application. The SRS serves as a template for software developers, which is useful throughout the design process. We suggest, based on Tables 1, 2, and 3 for SRSIoT, which must be solely added in traditional SRS and collected when designing and developing IoT applications.

The requirements gathering maps the users and developers view to design application for a smart environment. The following Fig. 1 depicts systematic requirement analysis model for IoT application.

Table 1 SRS for usecases in IoT enabled smart agriculture

Sr. no	Conventional SRS template	Description
1.1	Purpose	To propose a precise, site-specific, and sustainable agriculture for smart farming with new applications and opportunities [4]
1.4	Product scope	To provide a comprehensive monitoring system for agriculture includes its design and architectural implementation with focus on smart farming
2.1	Product perspective	The system mainly aims at in-situ evaluation of the leaf area index (LAI) which acts as an important crop parameter
2.2	Product functions	1. To provide long-term continuous crop monitoring system 2. An Apache server is responsible for generating queries and providing graphical user interface (GUI) - Web-based
2.3	User classes and characteristics	1. WSN WLAN PLNM Internet 2. UMTS/LTE 3. MQTT 4. Farm 5. FMIS (farm management information system) 6. IoT Gateway 7. User
2.4	Operating Environment	1. Based on Raspberry Pi 3, Each and every mote is connected via USB to a Linux system 2. Messaging protocol MQTT absolutely used at the IoT layer only 3. Via SPI or I2C bus platforms and its GPIOs, external sensors can be connected [5]

(continued)

Table 1 (continued)

Sr. no	Conventional SRS template	Description
2.5	Design and implementation constraints	1. For monitoring incoming data of connected clusters of sensor and rebooting the affected sensors is performed by Raspberry Pis 2. This Iot context utilizes cost-efficient, fully equipped, small and single-board computer [6]
2.7	Assumptions and dependencies	1. Environmental Challenges a. harsh weather conditions b. condensation water c. dry seasons 2. Wildlife-related Challenges a. animals and wildlife [7]
3.1	User interfaces	1. Web-based graphical user interface and database queries by apache server 2. In GUI provides the information regarding current status of sensor
3.2	Hardware interfaces	1. TelosB1-based platform a. Ground-level Clustered sensors b. Corresponding sensor 2. Raspberry Pis (RPi). Via PLMN 3. Open-source COTS mote used for commercial purpose a. Seamlessly integrating temperature and soil moisture sensors as a agricultural sensors 4. IEEE 802.15.4 low-cost wireless network [8] 5. To amplify communication range and improve radio performance external antennas are used [9]
3.3	Software interfaces	1. TinyOS2 2. Make use of time stamps and sequence numbers (SNs) with the help of Universal asynchronous receiver transmitter (UART) and the radio interface 3. 802.15.4 radio chip is useful in providing signal strength and link quality information [4]
3.4	Communications Interfaces	1. Via public land mobile networks (PLMN) communication, internet connectivity is established realized by LTE 2. MQTT communication provides data transport
4.1	System feature 1	The paper aims at providing data persistence, analytics, and visualization for accumulating decision by farmers

(continued)

Table 1 (continued)

Sr. no	Conventional SRS template	Description
4.1.1	Description and priority	This paper discuss about LAI profiles including fine-grained spatio-temporal resolution for long-term continuous crop monitoring system. It is a widely used important function which helps in providing vital performance conditions of the plants [10]
4.1.2	Stimulus/Response sequences	Protocols regarding timing synchronization are not required. Due to continuously reachable ground sensors by their cluster heads and in order to receive events of ground packets, they adjust their reference sampling
4.1.3	Functional requirements	Base station collects the information gathered by sensors and transmitted to a farm management information system (FMIS)
5.1	Performance requirements	1. In monitoring WSN, TelosB1-based platform (8 MHz TI MSP430 MCU, 10 kB RAM), act as basic sensor [11] 2. A central ALIX.6F2 router provides integration of Computers with UMTS connectivity
5.2	Safety requirements	1. As occasional software crashes and freezing is observed, mechanical timers were installed as an operational safety for turn off all devices during the night phases on regular basis 2. UPS 3. optical filter and diffuser accessory
5.3	Security requirements	In this paper, the central gateway develop a fixed SSH reverse tunnel to the Internet server that allows remote access to Pis and sensors that could manually be reprogrammed and modified to unexpected challenges [12]
5.4	Software quality attributes	1. Sensor redundancy 2. Return on investment (RoI)
5.5	Business rules	1. A simple star-topology is used within each cluster to form Sensor motes into clusters 2. Cluster heads are connected via WLAN [13] 3. No demand for routing protocols

Table 2 SRS for usecases in IoT enabled smart healthcare

Sr. no	Conventional SRS template	Description
1.1	Purpose	To monitor heart rate particularly when a person fall into weakness [14]
1.4	Product scope	To provide implementation of the heart rate and temperature monitoring system with focus on the senior citizens [15]
2.1	Product perspective	To determine heart beat rate for each minute and sends a short message (SMS) to administration warning to the cell phone
2.2	Product functions	1. Healthcare tracking system [16] 2. Trace and monitor patients
2.3	User classes and characteristics	1. Pulse sensor 2. Temperature sensor 3. GSM module 4. microcontroller 5. Server [17]
2.4	Operating environment	ANN system is inferred to signify the situation of heart rate and the present heart rate
2.5	Design and implementation constraints	Once the connections are established, a finger is kept on pulse rate sensor and temperature sensor to examine physiological parameters
2.7	Assumptions and dependencies	The light on pulse rate sensor provides the working condition, if the light is off then check for the faults
3.1	User interfaces	The system makes use of GUI for representing the results of patient's obtained from their pulse rate
3.2	Hardware interfaces	1. heart beat sensor 2. temperature sensor 3. power supply unit 4. Transmitter and receiver 5. zigbee module 6. lcd [18]
3.3	Software interfaces	1. Hostinger.in helps in creating URL 2. PHP language 3. HeartRateFree application to test beats per minute (bpm) a. When Resting or Tired or While at Work place b. After Waking up c. After Exercising d. Before Sleep
3.4	Communications interfaces	Concerned people would get SOS SMS through GSM module
4.1	System feature 1	1. Heartbeat monitoring 2. detecting abnormalities

(continued)

Table 2 (continued)

Sr. no	Conventional SRS template	Description
4.1.1	Description and priority	In this system displays the results on the screen when user place his finger on the pulse sensor and grip the temperature sensor with his fingers [19]
4.1.2	Stimulus/response sequences	The system generates prior alarm when the pulse seems to be dropping
4.1.3	Functional requirements	GSM module will send SMS to doctors, family members or any other concerned person, if any abnormalities are detected in the system
5.1	Performance requirements	1. Algorithms related to Sensor fusion permit to monitor and sense core temperature and dehydration level 2. New sophisticated equipped models helps in the measuring of heart rate activities and variability while breathing which helps in examining parameters relating to a person's fitness [20]
5.2	Safety requirements	1. Timely updates 2. generation of alarms
5.3	Security requirements	The person is authorized by any given time through login details [21]
5.4	Software quality attributes	1. The system consists of Arduino Uno board helps in providing IDE (Integrated development environment) with a cross platform application written in the language Java 2. Hardware and a part of software with standard functions are included in the board
5.5	Business rules	By amalgamating hardware and software reading can be sent to the server continuously [22]

Table 3 A template for SRSIoT

Sr. no	SRS IoT	Description
6	Sensor/mote	< Specify the sensor/mote name, total number of sensors, Type of sensor, version and manufacturer >
6.1	Purpose	< Define the different task to be performed by sensor as per the need >

(continued)

Table 3 (continued)

Sr. no	SRS IoT	Description
6.2	Scope	< Define the limitations and desired result from sensor >
6.3	Cost	< Cost of sensor > < Installation cost > < Computational Cost >
6.4	Energy consumption	< Battery lifetime consideration for sensors > < Enabling power saving mode for sensors when it is in idle state > < Selecting need based communication system and protocols > < Avoid redundant utilization of sensors >
6.5	Licensing of design schematics	< Licensing for the circuit design schematics >
6.6	Availability	< Availability of sensor in the current market > < Fixing of Sensor (repairable) >
6.7	Scalability	< sensor system should be scalable enough to handle updated processes > < Sensor system should be selected with a focus toward scalable applications >
6.8	Reliability	< Sensor works or functions properly as per its stated intended specification without any failure > < Sensor should be trustworthy enough for providing accurate information >
6.9	Maintainability	< Rectify the detected error or defects > < Replace or fix faulty sensors > < Determine the service interval for the applications in Smart environment > < Operating lifetime of a sensor >
6.10	Reconfigurability	< Sensor system should be reconfigurable in terms of hardware and software upgrade >
6.11	Interoperability	< Sensor system should be compatible with heterogeneous environment and support various operations >
6.12	Portability	< Sensor system should operate at different platform > < Sensor system should be able to transfer data and application from one device to other >
6.13	Safety	< Identify the risk associated with the sensors >
6.14	Time and space complexity	< Sensor system should select based on response time, acquisition time and processing time > < Sensor system should select based on accuracy required and memory constraints >

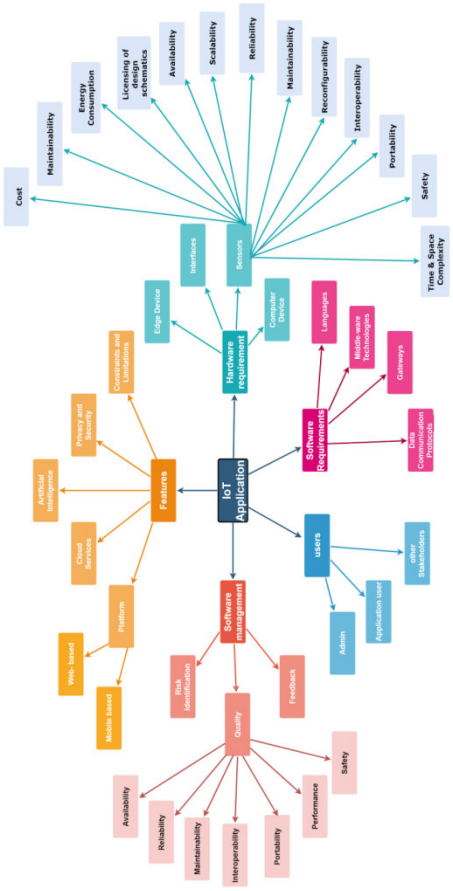


Fig. 1 Software requirement specification for IoT applications (SRSIoT)

4 Conclusion

This paper argues that the existing SRS method is not adequate to capture all aspects of IoT application to full-circle stages of software engineering. In order to support this argument, a research study has been performed on IoT applications, which enable the need of software engineering approach. Furthermore, the SRS for IoT applications for our proposed argument that SRS for IoT applications should be different from the traditional Software product. While designing and developing an application in a smart environment, SRS for IoT must keep the criteria listed in Table 3 in consideration.

References

1. B. Dorsemaine, J.-P. Gaulier, J.-P. Wary, N. Kheir, P. Urien, Internet of things: a definition and taxonomy. in *2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies*, January (IEEE, Cambridge, UK, 2016). https://doi.org/10.1109/NGMAST.2015.71
2. Z.K.A. Mohammed, E.S.A. Ahmed, Internet of things applications, challenges and related future technologies. in *2016 12th International Conference on Innovations in Information Technology* (IIT), January, (IEEE, Al-Ain, United Arab Emirates, 2017). https://doi.org/10.1109/INNOVATIONS.2016.7880054
3. H. Bach, W. Mauser, Sustainable agriculture and smart farming, earth observation open science and ınnovation. ISSI Scientific Report Series 15, January (2018). https://doi.org/10.1007/978-3-319-65633-5_12
4. P. Salami, H. Ahmadi, Review of farm management ınformation systems (FMIS). Earth Observ. Open Sci. Innov. New York Sci. J. **3**, 87–95 (2010)
5. S. Heble, A. Kumar, K.V.V Durga Prasad, S. Samirana, P. Rajalakshmi, U. B. Desai, A low power IoT network for smart agriculture. in *2018 IEEE 4th World Forum on Internet of Things* (WF-IoT), May (IEEE, Singapore, 2018). https://doi.org/10.1109/WF-IoT.2018.8355152
6. K.A. Patil, N.R. Kale, A model for smart agriculture using IoT. in *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication* (ICGTSPICC), June, (IEEE, Jalgaon, India, 2017). https://doi.org/10.1109/ICGTSPICC.2016.7955360
7. J. Shenoy, Y. Pingle, IOT in agriculture. in *2016 3rd International Conference on Computing for Sustainable Global Development* (INDIACom), October (IEEE, New Delhi, India, 2016)
8. S. Navulur, A.S.C.S. Sastry, M.N. Giri Prasad, Agricultural management through wireless sensors and internet of thing. Int. J. Electri. Comput. Eng. (2017). https://doi.org/10.11591/ijece.v7i6.pp3492-3499
9. B.L.V.S. Aditya, K. Subrahmanyam, Embedded linux based agricultural field monitoring and automation using IoT. Int. J. Eng. Technol.(UAE) (2018)
10. I.A. Lakhiar, G. Jianmin, T.N. Syed, F. Ali Chandio, N. Ali Buttar, W. Ahmed Qureshi, Monitoring and control systems in agriculture using ıntelligent sensor techniques: a review of the aeroponic system. (2018). https://doi.org/10.1155/2018/8672769
11. P. Lashitha Vishnu Priya, N. Sai Harshith, N.V.K. Ramesh, Smart agriculture monitoring system using IoT. Int. J. Eng. Technol.(UAE) (2018)
12. J. Bauer, N. Aschenbruck, Design and ımplementation of an agricultural monitoring system for smart farming. in *2018 IoT Vertical and Topical Summit on Agriculture—Tuscany* (IOT Tuscany), June, (IEEE, Tuscany, Italy, 2018). https://doi.org/10.1109/IOT-TUSCANY.2018.8373022

13. Innovative data transmission and routing strategies of wireless body area sensor networks: taxonomy, review, issues and constraints of state of the art contributions in recent literature. Int. J. Eng. Technol.(UAE) **7**, 696–704 (2018). https://doi.org/10.14419/ijet.v7i2.7.10925
14. N. Gupta, H. Saeed, S. Jha, M. Chahande, S. Pandey, Study and implementation of IOT based smart healthcare system. in *2017 International Conference on Trends in Electronics and Informatics* (ICEI), February, (IEEE, Tirunelveli, India, 2018). https://doi.org/10.1109/ICOEI.2017.8300718
15. A.M. Wajih, S. Tanin, A. Sami, A heartbeat and temperature measuring system for remote health monitoring using wireless body area network. Int. J. Bio-Sci. Bio-Technol. (2016). https://doi.org/10.14257/ijbsbt.2016.8.1.16
16. B. Alberto, W. Klaas, Advances in physical activity monitoring and lifestyle interventions in obesity: a review. Int. J. Obesity (2011). https://doi.org/10.1038/ijo.2011.99
17. S.M. Riazul Islam, D. Kwak, Md. Humaun kabir, M. Hossain, and Kyung-sup Kwak, The ınternet of things for health care: a comprehensive survey. IEEE Access (2015). https://doi.org/10.1109/ACCESS.2015.2437951
18. M. Ahmed, D. Mohammed, F. Ahmed, H. Najmuldeen, O. Shokhan, Classification of ınternet of things (IoT) models in healthcare systems. (2017)
19. M. Thangaraj, P.P. Ponmalar, S. Anuradha, Internet of things (IOT) enabled smart autonomous hospital management system—a real world health care use case with the technology drivers. in *2015 IEEE International Conference on Computational Intelligence and Computing Research* (ICCIC), March, (IEEE, Madurai, India, 2016). https://doi.org/10.1109/ICCIC.2015.7435678
20. R.K. Kodali, G. Swamy, B. Lakshmi, An implementation of IoT for healthcare. in *2015 IEEE Recent Advances in Intelligent Computational Systems* (RAICS), June, (IEEE, Trivandrum, India, 2016). https://doi.org/10.1109/RAICS.2015.7488451
21. M. Sai Prasanthi, V.B. Katragadda, H. Perumalla, B. Sowmya, Hybrid approach for securing the IoT devices. Int. J. Innov. Technol. Explor. Eng. **8**(4), 147–151 (2019)
22. M. Praveena, M. Kameswara Rao, Survey on big data analytics in healthcare domain. Int. J. Eng. Technol. (UAE) **7**, 919–925 (2018). https://doi.org/10.14419/ijet.v7i2.7.11097
23. M.N. Salman, P. Trinatha Rao, M.Z. Ur Rahman, Adaptive noise cancellers for cardiac signal enhancement for IOT based health care systems. J. Theoret. Appl. Info. Technol. (2017)
24. V. PremaLatha, S. Karimunnisa, U. Harita, Augmented review on internet of things for health care. J. Adv. Res. Dynam. Control Syst. (2017)
25. C.N.S. Kumar, M.A. Hussain, A review on employee's health monitoring system using IOT. Int. J. Eng. Technol.(UAE) (2018)
26. Web link. https://www.getkisi.com/blog/internet-of-things-communication-protocols
27. Web link. https://medium.com/omarelgabrys-blog/software-engineering-introduction-part-1-b79238ec97ee

Image Classification Based on Convolutional Neural Network

P. Lakshmi Prassanna, S. Sandeep, Kantha Rao, T. Sasidhar, D. Ragava Lavanya, G. Deepthi, N. Vijaya SriLakshmi, P. Mounika, and U. Govardhani

Abstract The main goal of this research work is to use Neural Networks to recognize and categorize images. As an input, a fixed dataset will be selected. Further, this research work will consider two distinct types of creatures and determine the animal species. The difficulty of recognizing images of animals is used to solve the CAPTCHA challenge. It is simple and easy, but data shows that cats and dogs are particularly difficult to distinguish automatically. CNNs are the simple neural networks with at least one layer that use convolution rather than generic environment multiplication. We shall visualize and distinguish a code made up of 0's and 1's that is taken- cat as 0 and Dog as 1. This research work aims to train the neural network and recognize the animals. As a result, it demonstrates that such animal finding may be completed by a convolution neural network.

Keywords Image classification · Data resources · Testing data · Pooling layers · Convolutional neural networks

1 Introduction

- The process of extracting information from an image is referred to as image classification. Thematic maps are commonly generated by using the image classification raster. There are two types of classification: supervised and unsupervised, which depends on the relationship between the analysts and thus, the machine during classification. Within the Multivariate toolset, there are diferent methods for performing both supervised and unsupervised classification. Since classification may be a multi-step procedure, the image classification toolbar was designed to give an interactive environment for performing classifications. The toolbar not only helps in the workflow of unsupervised and supervised classification but it also provides features for processing the input data, generating training samples

P. L. Prassanna (✉) · S. Sandeep · K. Rao · T. Sasidhar · D. R. Lavanya · G. Deepthi · N. V. SriLakshmi · P. Mounika · U. Govardhani
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_64

and signature files, and evaluating the standard of the training samples and signature files. CNN is a deep learning algorithm that can obtain an input image, attach weight and bias value to a variety of aspects/substance within the image and distinguish one from the other.

- The main aim of this paper is to implement neural networks for image recognition. Neural Networks are used to recognize the animal species. In this experiment, we will take a dataset and split it into a number of epochs, after which we will quantify the proportion of cats and dogs depending on those epochs. Previous papers have used authentication mechanisms such as captcha to collect the images. However, we take the images of animals with datasets that do not have security signs and measure the precision within a certain number of animals.

1.1 Supervised Classification

To classify a picture, supervised classification uses the spectral signatures obtained from training samples. The training samples can be quickly generated to reflect the groups you want to extract using the ımage classification toolbar. Users can also be able to quickly generate a signature file from the training samples, which the multivariate classification tools can use to identify the data.

1.2 Unsupervised Classification

During a multiband image, unsupervised classification detects spectral classes (or clusters) without the analyst's interference. The ımage classification toolbar promotes unsupervised classification by providing access to clustering methods with the ability to study cluster norm and classification tools.

2 Literature Survey

There are theoretically innumerable groups under which a single image can be categorized [1]. Manually inspecting and classifying images are a time-consuming operation, which becomes almost difficult when dealing with a large number of images, say 10,000 or maybe 100,000 [2].

Before a decade, multiple issues with computer vision were saturating on their precision [3]. However, as deep learning methods became more commonly used, the precision of those problems has increased significantly [4]. One of the major issues was image classification, which is defined as predicting the image's category [5]. Cat and dog ımage classification is an example of how cat and dog images are categorized [6].

The main aim of this paper is to use state-of-the-art object detection techniques to acheive high precision [7]. For the image classification challenges, a convolution neural network was developed [8–10]. We would be able to review, protect, and manage habitats and ecosystems more effectively if we received accurate, comprehensive, and up-to-date information about wildlife presence and behavior across extensive geographic regions [11–13]. Currently, such data is often obtained manually at a high cost because it is sparsely and infrequently obtained [14–17]. We look at the opportunity to gather such data automatically with the potential to turn many areas in genetics, ecology, and zoology into big data science [18]. Many animals destroy crops or threaten humans in places such as airports or farming areas near parks by necessitating the implementation of a device that senses animal presence and sends an alert in the name of public safety [19, 20].

3 Existıng Work

A CNN or ConvNet is a form of deep, feed forward artificial neural network that has been used to successfully analyze the visual imagery. CNN contrasts any image pixel by pixel, and the parts it compares during image recognition are referred as attributes. CNN is trained by comparing the estimated attribute matches in two pictures that are taken in the same place. Any neuron in CNN is attached to a small group of neurons underneath it, allowing it to accommodate less weights and reduce the amount of required neurons.

4 Proposed Work

This project focuses on classifying the images of cats and dogs because mostly cats and dogs will be similar and it is not easy to differentiate cats and dogs. Previously, it has been done on Support Vector Machine (SVM) but presently it is implmented on Convolutional Neural Network (CNN). Therefore, ımage classification based on Convolutional Neural Network (CNN) was implemented on online platform. This project aims to collect the data and differentiate between cats and dogs.

5 Image Classificatıon Using CNN

The primary ojective of this research work is to use neural networks to recognize the images. Neural Networks are used to train the model and recognize which animal species they belong to. This experiment considers a dataset and splits it into a number of epochs, after which we will quantify the proportion of cats and dogs depending on those epochs. Previous papers have taken the images with security like captcha.

However, we are capturing images of animals with datasets that do not include security indicators and calculating the accuracy within a certain proportion of animals. The *n*th layer of a Convolutional Neural Network is made up of the convolutional layer and pooling layer. Depending on the complexity of the images, the number of such layers may be expanded even more to capture even more low-level data, albeit at the expense of more computing resources. The picture grouping of Dogs versus Cats has persisted for some time. The Dogs versus Cats challenge in Kaggle is attempting to decipher the Captcha challenge, which entails the quick separation of images. It's simple for humans, but research shows that cats and dogs are difficult to distinguish. To address this problem, many of them have worked on or are working on developing machine learning classifiers.

On the Asirra dataset, a classifier that backed color features leverage 56.9% accuracy. A SVM classifier with a mixture of color and texture features has achieved an accuracy of 82.7%. A pooling layer performs a standard down sampling procedure on the function charts, decreasing their in-plane dimensionality and thereby applying translation invariance to minor shifts and distortions, as well as reducing the number of learnable parameters. It is important to note that none of the pooling layers have learnable parameters, but filter duration, stride, and padding are hyper parameters in pooling operations, comparable to convolution processes.

5.1 Block Diagram

A convolution neural network is used to solve the problem and produce an improved output and outcome in our project, and this research work will only focus on a subset of these videos. The cumulative number of images in our dataset will be 10,000. Keras will use it to make templates. As an input, a pre-defined dataset is selected. Two different kinds of animals are examined to determine the animal species. Also, this research work attempts to complete the CAPTCHA challenge, which includes differentiating the dog and cat pictures. Convolutional networks are neural networks with at least one layer that uses convolution instead of general matrix multiplication (Fig. 1).

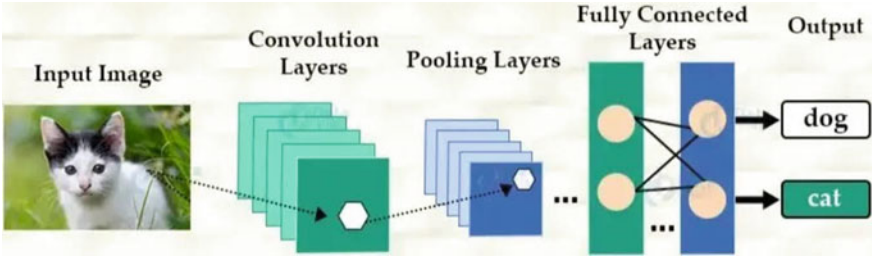


Fig. 1 Block diagram for image classification of dogs and cats

Fig. 2 Classification of cats and dogs

To begin, we take an input image of a dog or a cat, apply convolution layers to the image, then it categorizes the images as a cat or a dog. We will visualize and differentiate a code, which can be used in the form of 0's and 1's and classify them like cats as 0 and Dog as 1 Fig. 2).

Here, the images of cats are classified as 0, and images of dogs are classified as 1.

5.2 Technologies Employed

Python is an interpreted, general-purpose software application. With some of its influential need for significant applicator, Python's styling theory stresses writability. Its terminology concepts and entity methodology was developed to maintain the statement simple and logical programming for both small and large programs.

The Susceptible-Infectious-Recovered (SIR) model is a compartmentalized framework in which the quantity of contagious S, bacterial I, and rescued or expired (or immune) entities R are taken into consideration. The below are the potential deductions:

$$\begin{aligned} \mathrm{d}S(t)/\mathrm{d}t &= -(\beta/N)SI \\ \mathrm{d}I(t)/\mathrm{d}t &= ((\beta/N)S - \gamma)I \\ \mathrm{d}R(t)/\mathrm{d}t &= \gamma I \end{aligned}$$

It is noticed that the permanent N has never been the kingdom's community, but rather a populace made up of prone (S), afflicted (I), and restored (R) persons (R). In the new structure, the total populace $N = S + I + R$ is stable at all times as $\mathrm{d}S(t)/\,\mathrm{d}t + \mathrm{d}I(t)/\mathrm{d}t + \mathrm{d}R(t)/\,\mathrm{d}t = 0$.

6 Methodology

The picture grouping of Dogs versus Cats has been around for about some time. The Kaggle competition on Dogs vs. Cats has been attempted to solve the CAPTCHA challenge, which involves in quickly separating the images. Humans find it easy, but research shows that cats and dogs are more difficult to tell apart immediately.

To cope with this challenge, many of them have worked on or are working on developing machine learning methods. On the Asirra dataset, a classifier that backed color features reached 56.9% accuracy. A SVM classifier with a mixture of image representation obtained an accuracy of 82.7%.

Multiple factors images, which have been three or four hidden nodes are found in fully connected layers. A activation function has its own kernel, which is a kernel function. Feature selection maps have different kernels. Per function map in previous layers is convolutionalized by kernels.

A redistributing layer performs a standard down sampling procedure on the function charts, decreasing their in-plane dimensions and thereby applying linearity to minor movements and deviations, as well as reducing the number of training samples.

It's worth noting that none of the pooling layers have gray values, while filter duration, stride, and packing are functions in pooling operations, similar to compression techniques. This initiative, moreover, is divided into two main parts since the first and second stages are the most important in the installation of this method. The two stages of this project are explored in detail in this section to provide a better understanding of the creation and method of image classification.

This paper is divided into 4 stages:

Gathering Requirements

Eestablishment of Image classification

Convolution Layers
Pooling Layers

6.1 Gathering the Requirements

The first stage of this initiative centered on gathering all the criteria required to create this image identification. This project continues since a huge number of people have worked on Support Vector Machine (SVM). A dataset of cats and dogs is required to complete this project, and the dataset will be used to do so.

We'll see what happens when we put the data to the test. This is where their knowledge comes into existence. Following a detailed understanding of the system's nature, the emphasis could turn to designing and collecting all the datasets required in the project's corresponding phases.

6.2 Establishment of Image Classification

The emphasis will also be on designing this technique after all of the specifications have been gathered, as well as an idea for it. This project's method entails developing computer code that uses several services to distinguish between images of dogs and cats. This project can be completed using image recognition toolsets but we will be using the Python programming language to achieve our goals.

6.3 Convolution Layer

The convolution layer is the central component of a convolutional network, and it is responsible for the majority of computational function. It maintains the spatial relationship between pixels, allowing it to derive and learn features from them. It captures the original image's local dependencies. The image is represented as a matrix, and convolved feature map or activation map is obtained by moving the feature matrix over the image matrix, which is also a matrix. Changing the values in the filter matrix helps one to run a number of operations.

6.4 Pooling Layer

The dimensionality of the function map in this layer will be minimized to obtain smaller or shrinked maps, which reduce the parameters and computation. Pooling from the rectified and downsized function map may be max, average, or sum pooling.

7 Results

Here in this Fig. 3, we are printing the summary of the model, i.e., VGG16. There are so many layers or types in VGG16 like input Layer, Conv2D, Maxpooling 2D and so on.

Here in Fig. 4, we are testing and training the dataset of cats and dogs. A total of 1000 cats and dogs have been taught and verified. The total number of training pictures of cats and dogs is 2000, including 1000 validation images (Fig. 5).

```
[15] shivams_model.summary()

     Model: "vgg16"
     _________________________________________________________________
     Layer (type)                 Output Shape              Param #
     =================================================================
     input_1 (InputLayer)         [(None, 224, 224, 3)]     0
     _________________________________________________________________
     block1_conv1 (Conv2D)        (None, 224, 224, 64)      1792
     _________________________________________________________________
     block1_conv2 (Conv2D)        (None, 224, 224, 64)      36928
     _________________________________________________________________
     block1_pool (MaxPooling2D)   (None, 112, 112, 64)      0
     _________________________________________________________________
     block2_conv1 (Conv2D)        (None, 112, 112, 128)     73856
     _________________________________________________________________
     block2_conv2 (Conv2D)        (None, 112, 112, 128)     147584
     _________________________________________________________________
     block2_pool (MaxPooling2D)   (None, 56, 56, 128)       0
     _________________________________________________________________
     block3_conv1 (Conv2D)        (None, 56, 56, 256)       295168
     _________________________________________________________________
     block3_conv2 (Conv2D)        (None, 56, 56, 256)       590080
     _________________________________________________________________
     block3_conv3 (Conv2D)        (None, 56, 56, 256)       590080
     _________________________________________________________________
     block3_pool (MaxPooling2D)   (None, 28, 28, 256)       0
```

Fig. 3 Summary of the model

```
total training cat images: 1000
total training dog images: 1000
total validation cat images: 500
total validation dog images: 500
--
Total training images: 2000
Total validation images: 1000

[25] batch_size = 128
     epochs = 15
     IMG_HEIGHT = 150
                                                    ✓ 0s   co
```

Fig. 4 Testing and training

```
predictions = model.predict(img_array)
score = predictions[0]
print(
    "This image is %.2f percent cat and %.2f percent dog."
    % (100 * (1 - score), 100 * score)
)
```

```
This image is 99.04 percent cat and 0.96 percent dog.
```

Fig. 5 Here, we are calculating the accuracy of cats and dogs

8 Conclusion

The image classification based on Convolutional Neural Network (CNN) is used to classify various sorts of animals, however, in our project, we have used just cats and dogs as input and are conducting image classification on that dataset. On that dataset, we are testing and training. We are removing certain images, such as those that are blurry or duplicates of cats and dogs. On the final dataset, we will perform the accuracy of dogs and cats. We can identify only limited number of species or we can compare it with any of the two animals but not the dataset containing many animals. As long as the dataset is increasing, the project becomes complex to identify the animal species.

References

1. J. Elson, J. Douceur, J. Howell, J. Saul, Asirra: a CAPTCHA that exploits interest-aligned manual image categorization. in *Proceedings of ACM CCS* (2007), pp. 366–374
2. M. Ramprasath, M. Vijay Anand, S. Hariharan, Image classification using convolutional neural networks. Int. J. Pure Appl. Mathem **119**(17), 1307–1319 (2018)
3. O.M. Parkhi, A. Vedaldi, A. Zisserman, C.V. Jawahar, Cats and dogs. in *Computer Vision and Pattern Recognition* (CVPR), 2012 *IEEE Conference on* (IEEE, June), pp. 3498–3505
4. M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional neural networks (2013). arXiv preprint arXiv:1311.2901
5. B. Liu, Y. Liu, K. Zhou, Image classification for dogs and cats (2020)
6. P. Golle, Machine learning attacks against the Asirra CAPTCHA. in *Proceedings of the 15th ACM Conference on Computer and Communications Security* (ACM, 2008), pp. 535–542
7. L. Mansourian, M.T. Abdullah, L.N. Abdullah, A. Azman, Evaluating classification strategies in bag of SIFT feature method for animal recognition. Res. J. Appl. Sci. Eng. Technol. 1266–1272 (2015)
8. M.S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. Palmer, C. Packer, J. Clune1, Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. (2017). ArXiv , 1703.05830v5
9. P. Chattopadhyay, R. Vedantam, R.R. Selvaraju, D. Batra, D. Parikh, Counting everyday objects in everyday scenes. in *The IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 1604.03505v3 [cs.CV], 9 May (2017)

10. A. Krizhevsky, I. Sutskever, G. Hinton, Imagenet classification with deep convolutional neural networks. In Adv. Neural Info. Process. Syst. **25**, 1106–1114 (2012)
11. S. Shaikh, M. Jadhav, N. Nehe, U. Verma, Automatic animal detection and warning system. Int. J. Adv. Found. Res. Comput. **2**, 405–410 (2015)
12. V. Mitra, C. Wang, G. Edwards, Neural network for LIDAR detection of fish. in *Proceedings of the International Joint Conference on Neural Networks* (2003), pp. 1001–1006
13. S. Sharma, D. Shah, Real-time automatic obstacle detection and alert system for driver assistance on Indian roads. Indones. J. Electri. Eng. Comput. Sci. **1**, 635–646 (2016)
14. H. Ragheb, S. Velastin, P. Remagnino, T. Ellis, Human action recognition using robust power spectrum features. in *Proceedings of the 15th International Conference on Image Processing* (ICIPO'08), (San Diego, CA, 2008), pp. 753–756
15. H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. in *Proceedings of the 26th Annual International Conference on Machine Learning*, (ACM, 2009), pp. 609–616
16. F.-F. Li, J. Johnson, S. Yueng, Lecture 9: CNN architectures. (2017)
17. J. Sánchez, F. Perronnin, High -dimensional signature compression for large-scale image classification. in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, (IEEE, 2011) pp. 1665–1672
18. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks. (2015)
19. A. Krizhevsky, Learning multiple layers of features from tiny images. Master's thesis, Department of Computer Science, University of Toronto (2009)
20. P.V.V. Kıshore, P.V.V. Prasad, C.R. Prasad, R. Rahul, 4-Camera model for sign language recognition using elliptical fourier descriptors and ANN. in *International Conference on Signal Processing and Communication Engineering Systems—Proceedings of SPACES 2015, in Association with IEEE* (2015), pp. 34–38

A Novel Facial Emotion Recognition Scheme Based on Graph Mining

Jyoti S. Bedre and P. L. Prasanna

Abstract Recently, learning human feelings through analyzing facial expressions has been acquiring immense attention. Facial expressions play a decisive role in expressing the mental status and are vital in social interactions. Recognition and analyzation of facial expressions aids in understanding one's emotions in all situations even when verbal communication fails. This paper describes the fundamental phases involved in the emotion detection process. It explores the various face emotion detection schemes like machine learning, deep learning and others employed in literature and compares the diverse emotion recognition approaches in terms of FER techniques, datasets, emotions, number of emotions and publication years. This article then reviews the several graph-based techniques used for facial emotion recognition (FER) in existing studies and explains their benefits in the FER process. Finally, this work in this paper presents the vital findings discovered during the study for achieving further improvement in the FER process and developing better approaches for superior FER performance.

Keywords Facial expression · Facial emotion · Emotion recognition · Graph-based techniques

1 Introduction

The crucial steps involved in the FER are face acquisition, face expression extraction and emotion/expression detection, depicted in Fig. 1.

The face acquisition is the preliminary step where it may identify face in input images or detect face in initial frame and track the face in rest of the frames in case of image sequences. The expression extraction step involves extraction of facial

J. S. Bedre (✉)
Department of Computer Science and Engineering, K L University, Hydrabad, India
e-mail: jyoti.phd2020@klh.edu.in

P. L. Prasanna
Department of Computer Science and Engineering, KLEF, Vaddeswaram, Guntur 522502, India
e-mail: pprasanna@kluniversity.in

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_65

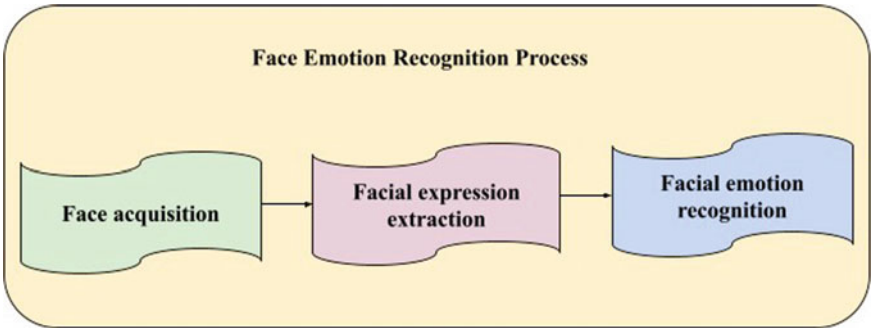


Fig. 1 General framework of FER

features. In this step, feature vectors are extracted through applying appearance-based or geometric-based schemes either to the entire face image or to the particular regions of the face image. Here, the classifier will be trained for distinguishing the facial emotional status based on recognized action units. Emotion detection is the final step in emotion analysis. In this step, the different facial emotions are classified or detected depending upon the features extracted in the previous step.

2 Facial Emotion Recognition Techniques

In this section, different techniques depicted in Fig. 2 such as machine learning (ML), deep learning (DL) and others employed for FER in existing works are investigated.

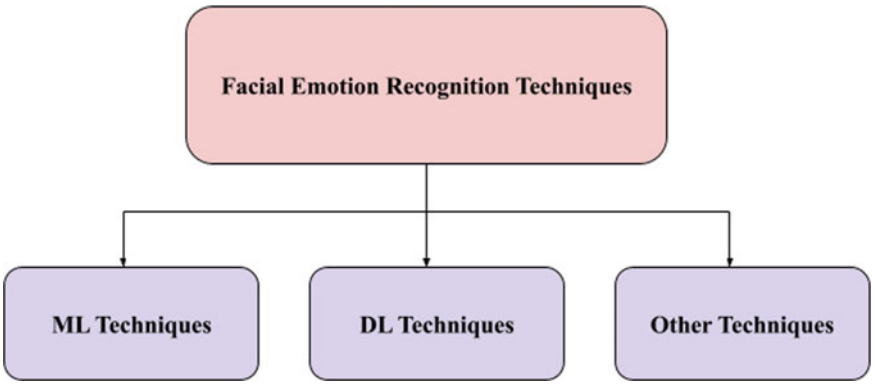


Fig. 2 FER techniques

2.1 ML-Based Techniques

Lopes et al. [1] employed an ML-based approach for FER in the elderly. Viola Jones with haar features was utilized for extracting the faces. Gabor filter was exploited for extorting the face characteristics. The extracted characteristics were categorized using multiclass support vector machine (M-SVM) technique. FER results reported that the aging negatively influenced the FER tasks. Li and Oussalah [2] recognized diverse facial emotions using ML approaches. It employed an adaptive AdaBoost method and haar transform for face detection and principal component analysis (PCA) with K-nearest neighbor (KNN) for emotion classification. The approaches employed exhibited 75% accuracy in FER. Abdul-Hadi and Waleed, [3] employed SVM method for FER. They effectively recognized the multi-emotions and achieved 92.88% accuracy. Munasinghe [4] exploited a random forest (RF) scheme for recognizing the face expressions. The faces were identified in every image through a histogram of oriented gradients and sliding window technique. A feature vector was estimated after determining the facial landmarks followed by emotion detection using the RF classifier. Drume and Jalal [5] employed a multi-level categorization method for recognizing face expressions from facial images. Experimental results reported that the proposed multilevel approach recognized facial expressions with 94% recognition accuracy. Ghimire et al. [6] proposed a ML approach for FER using salient geometric attributes. The facial points were initialized using graph matching method and tracked using Kanade-Lucas-Tomasi tracker. The geometric attributes were extracted from triangle, line and point comprising tracked outputs of facial points followed by emotion categorization using SVM. Krestinskaya and James [7] discussed an ML approach for identifying the face expressions. Pixel normalization was performed for eliminating the intensity offsets. After extracting facial attributes, emotion classification and detection was achieved through a KNN, RF and min–max classifier. Results illustrated an improvement in FER performance from 92.85 to 98.57%. Xiaoxi et al. [8] employed ML techniques for FER. The appearance and geometric attributes were extracted from face images, and emotions were categorized and recognized using SVM and deep Boltzmann machine (DBM). Comparison of FER performance of SVM and DBM revealed that SVM outperformed the DBM in expression classification.

2.2 DL-Based Techniques

Liu et al. [9] proposed a DL-based FER scheme for identifying distinct facial emotions. A low-ranked multimodal fusion approach was exploited for enhancing the efficacy of FER. Experiments revealed that the proposed DL-FER approach was capable of capturing both local and global facial data and significantly performed well than several competitive algorithms. Jain et al. [10] employed deep convolutional neural networks (DCNNs) for classifying facial emotions. The proposed

DCNN approach outperformed the state-of-the-art methods in FER. Chen et al. [11] exploited softmax regression with deep sparse autoencoder (SRDSAN) technique for FER. The high-level facial attributes were extracted using DSAN, and facial expressions were classified using SR method. Kumar et al. [12] exploited CNN approach for detecting face expressions. Viola Jones scheme was used for face detection prior to expression recognition. This approach provided >90% FER accuracy. Luh et al. [13] proposed a deep neural network (DNN) with YOLOv3 approach for determining human face expressions. The proposed YOLOv3-DNN approach gained good FER rate compared to other FER methods. Sang and Ha [14] presented a productive scheme for making the learnt deep attributes in dense convolutional networks (DenseNets) more discriminative, thereby exploiting them in recognizing face emotions more effectively. Lasri et al. [15] utilized CNNs for recognizing face emotions of students. This approach involved three stages namely face identification, normalization and expression recognition. The faces were identified through a Haar-like detector and were normalized. About 70% accuracy in FER was attained through this approach. Jain et al. [16] executed FER via a hybrid DL technique. The proposed FER system employed CNNs and recurrent neural networks (RNNs) for emotion classification and identification. The FER system's performance was examined under distinct circumstances. Furthermore, hyperparameters were utilized for system tuning. The fusion of CNN-RNN methods significantly boosted the overall detection result and FER performance. Yang et al. [17] employed a DL scheme for FER. A weighted mixture DNN (WMDNN) was exploited for automatically extracting features which were useful for FER activities. The dual channels of face images containing local binary pattern (LBP) and grayscale facial images were processed using WMDNN. The emotion-related attributes of grayscale images were extracted through refining a semi-VGG16 network whereas attributes of LBP images were extracted through a shallow CNN. The both channel outputs were combined in a weighted manner and a softmax classification was used for estimation of final recognition result. The proposed WMDNN-based dual-channel approach outperformed the FER techniques based on single-channel deep networks. Results confirmed that the employed DL methods namely AlexNet and GoogleNet displayed outstanding FER performances with enhanced recognition accuracy.

2.3 Others

Majumder et al. [18] performed FER using geometric features via Kohonen self-organizing map (KSOP) approach. It employed a 26-dimensional geometric face feature vector comprising eyebrow, lip and eye feature points as input for describing the facial variations because of the different expressions. Performance evaluation of KSOP with the traditional ML techniques revealed that the KSOP performed well in detecting basic six expressions while ensuring improvement in mean recognition rate and mean classification accuracy. Ko and Sim [19] presented a methodology for FER in image sequences. The active appearance (AA) frameworks and

dynamic Bayesian network (DBN) with Kalman filter were employed for modeling the temporal aspects of face emotions. The proposed AA-DBN approach showed a greater FER performance level through attaining >90% recognition accuracy. Mistry et al. [20] utilized a shape and appearance-based NN method for recognizing face emotions. The face real-time tracking and extraction of geometric and texture features of images was conducted using an active appearance framework. The extracted shape and texture features were utilized for detecting seven common face emotions and 18 face actions. This approach was capable of detecting facial emotions and actions from images up to 60° of pose changes. Nicolai and Choi [21] conducted FER using fuzzy logic technique. The face of the subject and its facial features along with the relevant detecting points were extracted. This approach attained 78.8% detection rate. (Siddiqi et al. [22]) employed a hierarchical detection scheme for FER. The local and global attributes were extracted using PCA and emotions were categorized using a hidden Markov model (HMM) and linear discriminant analysis (LDA). This approach attained 98.7% FER accuracy. Abdat et al. [23] used a bimodel mechanism for detecting emotions from physiological signals and facial expressions via a feature-level fusion. The features were extracted from physiological cues and face expressions. Feature transformation was done using PCA and best sub-features were selected using mutual information. Findings showed that the feature-level fusion with PCA provided good FER results.

2.4 Literature Survey

See Table 1.

2.5 Publication Year-Wise Analysis of Existing Works on FER

The various existing works on FER discussed here are analyzed based on publication years, FER datasets and FER techniques. The reviewed FER works in previous subsections published in the recent 11 years are depicted in Figure 3. Analyzing the research works on FER, it could be observed that more number of research works published in 2018 are reviewed in this survey, followed by 2017, 2019, 2015, 2013, 2012, 2014, 2010, 2016, 2011 and 2020.

Table 1 Existing studies on FER

Authors	FER techniques	Emotions recognized	Number of recognized emotions	Datasets	Publication year
Lopes et al.	M- SVM	Happiness,	7	Lifespan	2018
		Neutral,		database	
		Anger,			
		Sadness,			
		Disgust,			
		Fear,			
		Surprise			
Li and Oussalah	Adaptive AdaBoost, PCA, KNN	Happiness, neutral, surprise, disgust, sadness	5	Indian face database (IFD) and taiwanese face database	2010
Abdul-Hadi and Waleed	SVM	Smile, crying, no- smile, laughing	4	Image database	2020
Munasinghe	RF	Sadness, surprise, anger, contempt, happiness, disgust	6	CK+	2018
Drume and Jalal	PCA, SVM	Neutral, sad, angry, happy, surprise	5	–	2012
Ghimire et al.	SVM	Anger, happy, neutral, surprise, disgust, sad, fear	7	MMI, CK+ , Multimedia Understanding Group (MUG)	2017
Krestinskayaan d James	KNN, RF, Min- max classifier	Fear, sadness, anger, happiness, neutral,	7	JAFFE	2017
		surprise, disgust			
Xiaoxi et al.	DBM, SVM	–	–	FERC-2015	2017

(continued)

Table 1 (continued)

Authors	FER techniques	Emotions recognized	Number of recognized emotions	Datasets	Publication year
Jain et al.	DCNNs	Sad, happy, neutral, disgust, surprise, fear, angry	7	JAFFE, CK+	2019
Chen et al.	SRDS	Surprise,	7	JAFFE, CK+	2018
	AN	angry, disgust, sadness, fear, happiness, neutral			
Kumar et al.	CNNs	Fearful neutral, disgusted, surprised, happy, angry, sad	7	Facial expression recognition challenge (FERC)-2013, CK+	2017
Luh et al.	YOLO v3-DNNs	Anger, happy, neutral, disgust, sad, surprise, fear	7	JAFFE, RaFD, CK+	2019
Mohammadpour et al.	CNN	Contempt, happiness,	7	CK	2017
		anger, fear, sadness, surprise, disgust			
Sang and Ha	DenseNets	Happy, surprise, neutral, sad, fear, angry, disgust	7	FERC-2013	2018‘
Lasri et al.	CNN	Fear, happy, disgust, angry, neutral, sad, surprise	7	FERC-2013	2019
Jain et al.	CNN, RNN	Surprise, happy, fear, angry, sad, neutral, disgust	7	MMI, JAFFE	2018
Yang et al.	WMDNN	Surprise, happiness, fear, sadness, anger, disgust	6	CK+ ,Oulu-CASIA, JAFFE	2017

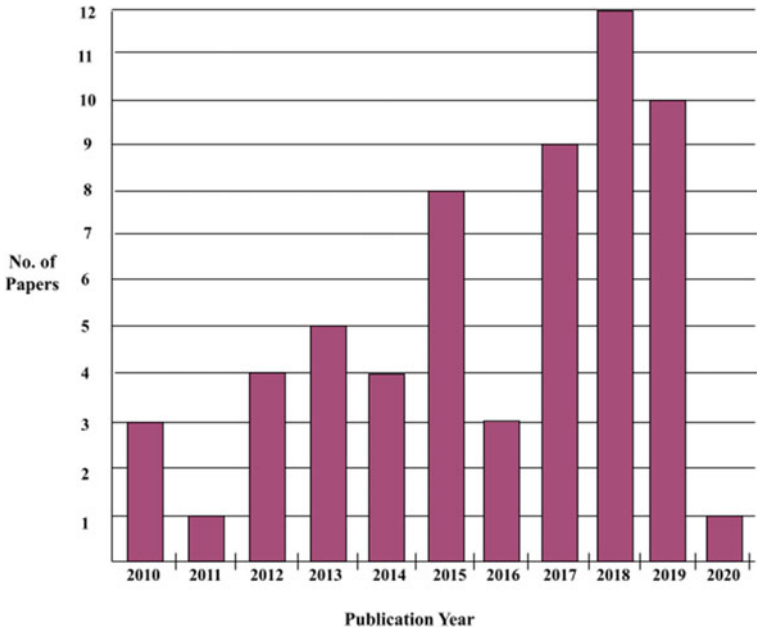


Fig. 3 Publication year-wise analysis of FER methods

3 Graph-Based Techniques for Facial Emotion Recognition

Apart from ML, DL and other techniques, recently graph-based techniques have gained extensive popularity in FER tasks. Graph-based techniques provide an effective and interesting approach for recognizing face emotions in order to learn and understand human behavior. The crucial benefit of graphs is that they provide a robust way for representing unstructured information. Moreover, they are increasingly significant in representing complex structures like images and therefore are very useful in the FER process. Some of the research studies exploiting graph mining techniques for FER are discussed in this section.

Hassan and Mohammed [24] proposed a FER method based on graph mining. It exploited graph theory and graph mining concepts for expression recognition. The face region was represented as a graph of nodes and edges. The customary sub-graphs in every emotional subject were determined using sub-graph mining technique (gSpan algorithm). It performed binary classification after encoding the selected subgraphs for classifying the expression of facial image. The proposed FER approach showed 90% accuracy in facial emotion detection. Jiang et al. [25] exploited graph-based approach for FER. Facial expression detection was accomplished via graph embedding from scrambled face images. This approach performed better even on chaotic patterns of disorganized faces and consistently attained greater accuracy on employed face expression datasets. Ngoc et al. [26] used a graph neural

network (GNN) technique for FER. In the proposed graph structure, nodes were defined using landmarks, whereas edges were built using Delaunay technique. The proposed GNNs helped in capturing emotional information via inherent characteristics of face. The proposed GNN approach achieved 50.65 and 98.47% accuracy with two distinct datasets. Hanmandlu et al. [27] employed a graph-based method for identifying facial expressions. This approach utilized fiducial points as facial attributes for distinguishing faces. This approach exhibited 96.67% recognition accuracy. Moreover, it was insensitive to illumination changes and pose changes on frontal images. Lopez et al. [28] exploited another graph-based scheme for FER. A fusion of local and global methods were used for constructing the graphs whose nodes encoded the facial attributes. Experimental outputs indicated that this approach outperformed the similar well-known FER methods.

4 Conclusion

In this paper, a general FER framework was provided along with major procedures involved in FER such as face image acquisition, expression extraction and expression detection. This paper investigated the different approaches exploited for FER. It explored ML, DL and other approaches for FER. Apart from these, it also explored graph-based approaches. A literature survey was conducted in terms of FER methods, datasets, publication years, emotions recognized and number of recognized emotions, and analyzed publication year-wise existing works on FER. The ML, DL and other FER methods showed good FER performances. The graph-based schemes provided more effective and innovative ways for recognizing face emotions with improved results for recognizing face emotions. Though graph-based methods showed improved performances than ML, DL and other FER schemes, many of them utilized the entire graph for detection purposes and generated numerous features which further increased the feature size. Therefore, better graph mining approaches are required for executing FER more effectively through reducing the extracted final feature size and providing a greater FER rate.

References

1. N. Lopes, A. Silva, S.R. Khanal, A. Reis, J. Barroso, V. Filipe, J. Sampaio, Facial emotion recognition in the elderly using a SVM classifier, in *2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW),* (IEEE, 2018), pp. 1–5
2. J. Li, M. Oussalah, Automatic face emotion recognition system, in *2010 IEEE 9th International Conference on Cyberntic Intelligent Systems,* (IEEE, 2010), pp. 1–6
3. M.H. Abdul-Hadi, J. Waleed, Human speech and facial emotion recognition technique using SVM, in *2020 International Conference on Computer Science and Software Engineering (CSASE),* (IEEE, 2020), pp. 191–196

4. M.I.N.P. Munasinghe, Facial expression recognition using facial landmarks and random forest classifier, in *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, (IEEE, 2018), pp. 423–427
5. D. Drume, A.S. Jalal, A multi-level classification approach forfacial emotion recognition, in *2012 IEEE International Conference on Computational Intelligence and Computing Research*, (IEEE, 2012), pp. 1–5
6. D. Ghimire, J. Lee, Z.N. Li, S. Jeong, Recognition of facial expressions based on salient geometric features and support vector machines. Multimedia Tools Appl. **76**(6), 7921–7946 (2017)
7. O. Krestinskaya, A.P. James, Facial emotion recognition usingmin-max similarity classifier, in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, 2017), pp. 752–758
8. M. Xiaoxi, L. Weisi, H. Dongyan, D. Minghui, H. Li, Facialemotion recognition, in *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*, (IEEE, 2017), pp. 77–81
9. Z. Liu, H. Suo, B. Yang, A novel facial expression recognition scheme based on deep neural networks, in *International Conference on Artificial Intelligence and Security*, (Springer, Cham, 2019), pp. 612–622
10. D.K. Jain, P. Shamsolmoali, P. Sehdev, Extended deep neural network forfacial emotion recognition. Pattern Recogn. Lett. **120**, 69–74 (2019)
11. L. Chen, M. Zhou, W. Su, M. Wu, J. She, K. Hirota, Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction. Inf. Sci. **428**, 49–61 (2018)
12. G.R. Kumar, R.K. Kumar, G. Sanyal, Facial emotion analysis usingdeep convolution neural network, in *2017 International Conference on Signal Processing and Communication (ICSPC)*, (IEEE, 2017), pp. 369–374
13. G.C. Luh, H.B. Wu, Y.T. Yong, Y.J. Lai, Y.H. Chen, Facialexpression based emotion recognition employing YOLOv3 deep neural networks, in *2019 International Conference on Machine Learning and Cybernetics (ICMLC)*, (IEEE, 2019), pp. 1–7
14. D.V. Sang, P.T. Ha, Discriminative deep feature learning for facialemotion recognition, in *2018 1st International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, (IEEE, 2018), pp. 1–6
15. I. Lasri, A.R. Solh, M. El Belkacemi, Facial emotion recognitionof students using convolutional neural network, in *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)*, (IEEE, 2019), pp. 1–6
16. N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, M. Zareapoor, Hybriddeep neural networks for face emotion recognition. Pattern Recogn. Lett. **115**, 101–106 (2018)
17. B. Yang, J. Cao, R. Ni, Y. Zhang, Facial expression recognition using weighted mixture deep neural network based on double-channel facial images. IEEE Access, **6**, 4630–4640 (2017)
18. A. Majumder, L. Behera, V.K. Subramanian, Emotion recognition fromgeometric facial features using self-organizing map. Pattern Recogn. **47**(3), 1282–1293 (2014)
19. K.E. Ko, K.B. Sim, Development of a facial emotion recognition method based on combining AAM with DBN, in *2010 International Conference on Cyberworlds*, (IEEE, 2010), pp. 87–91
20. K. Mistry, L. Zhang, S.C. Neoh, M. Jiang, A. Hossain, B. Lafon, Intelligent appearance and shape based facial emotion recognition for a humanoid robot, in *The 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014)*, (IEEE, 2014), pp. 1–8
21. A. Nicolai, A. Choi, Facial emotion recognition using fuzzy systems, in *2015 IEEE international conference on systems, man, and cybernetics*, (IEEE, 2015), pp. 2216–2221
22. M.H. Siddiqi, S. Lee, Y.K. Lee, A.M. Khan, P.T.H. Truc, Hierarchicalrecognition scheme for human facial expression recognition systems. Sensors **13**(12), 16682–16713 (2013)
23. F. Abdat, C. Maaoui, A. Pruski, Bimodal system for emotionrecognition from facial expressions and physiological signals using feature-level fusion, in *2011 UKSim 5th European Symposium on Computer Modeling and Simulation,* (IEEE. 2011), pp. 24–29

24. A.K. Hassan, S.N. Mohammed, A novel facial emotion recognition scheme based on graph mining. Defence Technol. **16**(5), 1062–1072 (2020)
25. R. Jiang, A.T. Ho, I. Cheheb, N. Al-Maadeed, S. Al-Maadeed, A. Bouridane, Emotion recognition from scrambled facial images via many graph embedding. Pattern Recogn. **67**, 245–251 (2017)
26. Q.T. Ngoc, S. Lee, B.C. Song, Facial landmark-based emotion recognition via directed graph neural network. Electronics **9**(5), 764. (2020)
27. M. Hanmandlu, D. Gupta, S. Vasikarla, Face recognition usingElastic bunch graph matching, in *2013 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, (IEEE, 2013), pp. 1–7
28. F.A.P. López, H.M. Mora, J.A.G. Selva, A neural network frameworkfor face recognition by elastic bunch graph matching. Recent Adv. Electroscience Comput. **75**, (2015)

Correction to: A Unique Interlinking Converter Control for Hybrid AC/DC Islanded Microgrids

M. Jayachandran, Gundala Srinivasa Rao, and Ch. Rami Reddy

Correction to:
Chapter "A Unique Interlinking Converter Control for Hybrid AC/DC Islanded Microgrids" in: P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93, https://doi.org/10.1007/978-981-16-6605-6_12

In the original version of the book, the following belated correction has been incorporated:

In chapter "A Unique Interlinking Converter Control for Hybrid AC/DC Islanded Microgrids", the affiliation of author "M. Jayachandran" has been changed from "Puducherry Technological University, Puducherry, India" to "Sri Manakula Vinayagar Engineering College, Puducherry, India"

The chapter and book have been updated with the changes.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-981-16-6605-6_12

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_66

Retraction Note to: Optimized Lower Part Constant-OR Adder for Multimedia Applications

Mahendra Vucha and A. L. Siridhara

Retraction Note to:
Chapter "Optimized Lower Part Constant-OR Adder for Multimedia Applications" in: P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_19

The authors have retracted this article because Figures 1, 2 and 3, Tables 1, 2, 3 and 4, as well as parts of the text, were duplicated from a previously published article by Ayad Dalloo, Ardalan Najafi and Alberto Garcia-Ortiz [1]. Both authors agree to this retraction.

[1] A. Dalloo, A. Najafi and A. Garcia-Ortiz, "Systematic Design of an Approximate Adder: The Optimized Lower Part Constant-OR Adder," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 26, no. 8, pp. 1595-1599, Aug. 2018, doi: 10.1109/TVLSI.2018.2822278

The retracted version of this chapter can be found at
https://doi.org/10.1007/978-981-16-6605-6_19

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*,
Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6_67

Author Index

P. Karrupusamy et al. (eds.), *Sustainable Communication Networks and Application*, Lecture Notes on Data Engineering and Communications Technologies 93,
https://doi.org/10.1007/978-981-16-6605-6