BMS

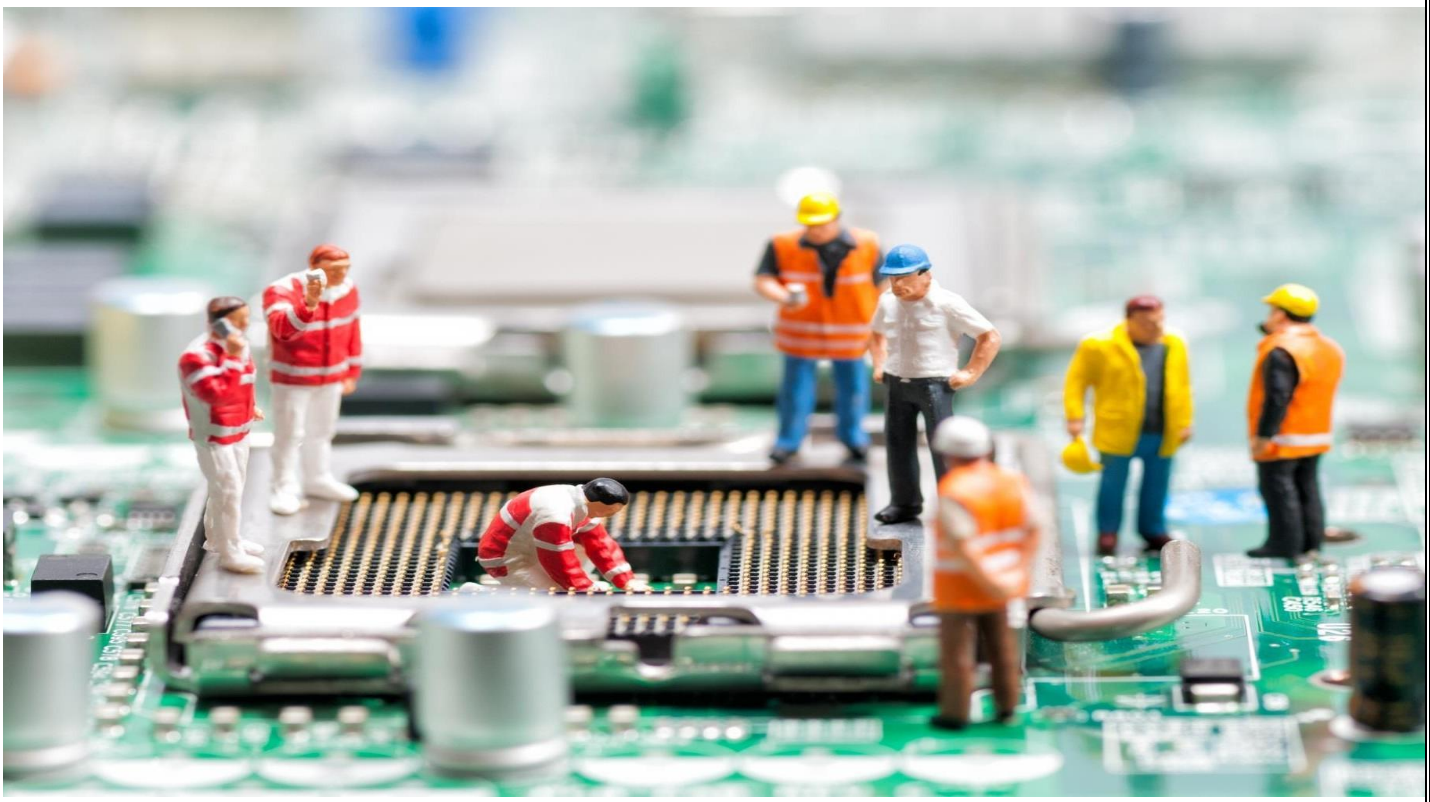
INSTITUTE OF TECHNOLOGY AND MANAGEMENT

Avalahalli, Doddaballapur Main Road, Bengaluru – 560064

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

RESEARCH COMPENDIUM

YEAR: 2019-20



RESEARCH
COMPENDIUM

Compiled by,

Mrs. Asha G.H

Dr.Dankan Gowda V

Dr. Ambika R
HoD, Dept. of ECE

About Institution

In view of the growing demand for technical education and with the goal of establishing a premier technical education on par with international standards, a new technical institution by `name 'BMS Institute of Technology and Management' was established in 2002. Currently, BMSIT & M offers seven UG, three PG programs and Ph.D. /M.Sc. (Engg.) in seven disciplines. BMSIT & M considers research to be of equal importance as academics for the betterment of an institution. Research culture has been embraced well by the faculty members and research scholars at BMSIT and M. In this report, we present an overview of the research activities of Information Science and Engineering, BMSIT & M.

Vision

“To emerge as one of the nation’s finest technical institutions of higher learning to develop engineering professionals who are technically competent, ethical and environment friendly for betterment of the society.”

Mission

“Accomplish stimulating learning environment through high quality academic instruction, innovation and industry – institute interface.”

About Department

The Department of Electronics and Communication Engineering started in 2002-03 with the vision: “To Provide Quality Education in Electronics, Communication and Allied Engineering fields to serve as Valuable Resource for Industry and Society”. The department is a “Research Center” recognized by VTU, has well experienced and qualified faculty members who inspire the students to face the competitive world. 16 faculties are pursuing their Doctoral Degrees. Consistent performance by the students in VTU examinations is a reflection of the efforts by all faculty members. The department is equipped with latest equipment and laboratory amenities to meet the global standards.

Vision

“Provide Quality Education in Electronics, Communication and Allied Engineering fields to Serve as Valuable Resource for Industry and Society”.

Mission

1. Impart Sound Theoretical Concepts and Practical Skills.
2. Promote Inter-Disciplinary Research.
3. Inculcate Professional Ethics.

Programme Educational Objectives

Graduates of the programme will:

PEO1: Work as professionals in the area of Electronics and allied engineering fields

PEO2: Pursue higher studies and involve in the interdisciplinary research work

PEO3: Exhibit ethics, professional skills and leadership qualities in their profession.

Programme Outcomes

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization for the solution of complex engineering problems.

2. **Problem analysis:** Identify, formulate, research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for public health and safety, and cultural, societal, and environmental considerations.

4. **Conduct investigations of complex problems:** The problems that cannot be solved by straightforward application of knowledge, theories and techniques applicable to the engineering discipline, that may not have a unique solution.

5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools, including prediction and modeling to complex engineering activities, with an understanding of the limitations.

6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with t h e society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. **Project management and finance**: Demonstrate knowledge and understanding of t h e engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. **Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Programme Specific Outcomes

Graduates will be able to:

PSO1: Exhibit competency in embedded system domain.

PSO2: Exhibit competency in RF and signal processing domain.

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

SLNO	Faculty	Title of the paper	Journal	Date	Indexing
1	Dr.JayadevaG S	Human Assisting Robot	International journal IJRECE Volume 7 Issue 2.	June 2019	Scopus indexed
2	Dr. C S Mala	Development of novel protocol for QoS improvement in WSN-RPEH	Inter International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3,	September 2019	Scopus Indexed
3	Prof. Saneesh Cleatus T	Feature Extraction and Classification of Visually Evoked Potentials Abstract: and its Application in Forensic Science	International Journal of Computer Sciences and Engineering. Volume 9 Issue No. 7	July 2019	Google Scholar
4	Dr. A Shobha Rani	Delay-Based Routing Mechanism for Load Balanced Routing in Wireless Mesh Networks	International Journal of Computer Science Engineering (IJCSE)	Vol.-7, Issue-5,31st May 2019	Scopus Indexed
5	Dr. A Shobha Rani	Scalable Local Route Repair-Hybrid Wireless Mesh Protocol (SLRR-HWMP) for IEEE 802.11s	International Journal of Computer Science Engineering (IJCSE) ISSN : 2319-7323 Vol. 8, pg 173-184	05 Sep-Oct 2019	Scopus Indexed
6	Dr. A Shobha Rani	Buffer Based Routing Mechanism for Load Balancing in Wireless Mesh Networks	International Journal of Computer Engineering and Technology (IJCET) Volume 10, Issue 1, pp. 1-10, Article ID: IJCET_10_01_001,	January-February 2019	Google Scholar
7	Dr.Vijayalakshmi G V	An RDAU-NET model for lesion segmentation in breast ultrasound images	International Journal(PLOS ONE)	23, Aug 2019	Scopus
8	Dr.Vijayalakshmi G V	A Multi-Sensor System for Silkworm Cocoon Gender Classification via Image Processing and Support Vector	International Journal(Sensors MDPI)	2019 June	Scopus and SCI

		Machine			
9	Dr.Vijayalakshmi G V	Bilingual text detection in natural scene images using invariant moments	International Journal(Journal of Intelligent & Fuzzy Systems)	22 November 2019	Scopus
10	Prof.Hamsavahini R.	Implementation of Automated Vehicle Identity Recognition System	International Journal of Engineering Science and Computing IJESC	June 209	Google Scholar
11	Prof.Hamsavahini R.	Multilingual speech to text Conversion using Matlab	IJSRD - International Journal for Scientific Research & Development	July 2019	Google Scholar
12	Prof. Rashmi N.	A new optimised interleaver design for high-dimensional data transmission in SCM-OFDM system	International Journal of Wireless and Mobile Computing (IJWMC)	Jan 2020	Scopus
13	Prof. Rashmi N.	Self-improved grey wolf optimization for estimating carrier frequency offset in SCM-OFDM systems	International Journal of Pervasive Computing and Communications. Vol. 18, No. 2, 202	Date: 02 January 2020	scopus and SCI
14	Prof. Rashmi N.	Performance Enhancement Of Scm-Ofdm System Using Carrier Frequency Offset Estimation And Correction	International Journal of Advanced Science and Technology. Vol.133 (2019), pp.11-18,	Date :Jan,2020	Scopus
15	Prof. Rashmi N.	Power efficiency enhancement using hybrid techniques for OFDM	International Journal on Emerging Technologies. 10(3): 260-267	2019	Google Scholar
16	Prof.Jagannath K. B.	Manifestation of intermediate phase in Cu doped Si-Te glasses	Elsiver-Journal of non crystalline solids	March 2020	Scopus indexed
17	Prof.Jagannath K. B.	Electrical switching and crystalline peak studies on Si20Te80−xSnx (1 ≤ x ≤ 7) chalcogenide bulk glasses	Journal of Non crystalline solids	May 2020	Scopus Indexed
18	Prof. Chandra Prabha.	Smart Real Time Manhole Monitoring	International Research Journal of Engineering	July 2019	Google Scholar

		System (2018-2019)	and technology, Volume 06, Issue 07		
19	Prof. Sabina R.	Effect of copper concentration on CTS thin films for solar cell absorber layer and photocatalysis application	Superlattices and Microstructures journal homepage: www.elsevier.com/locate/superlattices	May 2020	Scopus Indexed
20	Prof. Sabina R.	Temperature dependent growth of Cu2SnS3 thin films using ultrasonic spray pyrolysis for solar cell absorber layer and photocatalytic application	International, IOP science (Research material express)	21 August 2019	Scopus
21	Dr. Deepa N Reddy	A Statistical Approach to Spectrum Sensing Using Bayes Factor and p-Values	International Journal of Electrical Computer Engineering, Volume 9, No4, pp2910-2917	August 2019	Scopus
22	Dr.Shanthi Chandra	A Comparative Analysis of Performance of Several Wavelet Based ECG Data Compression Methodologies	International Journal Elsevier	01 Oct 2016	Scopus and SCIE
23	Dr. Dankan Gowda V.	Implementation of Automatic Air Traffic Control System: An Overview	GIS Science Journal Vol. 7, Issue 6	June 2020	Scopus Indexed
24	shashikala. J	Pothole Detection And Notification System	International Research Journal of Modernization in Engineering Technology and Science	July 2020	Google Scholar
25	Dr. Ambika R	Optimized Intelligent Traffic Light Control Using Image Processing	International Journal of Electrical Engineering and Technology	June 2020	Scopus
26	Shilpa Hiremath	Smart Moto Helmet	International Journal of Advanced research in basic engineering sciences	July 2019	Google Scholar

			and technology volume 5,issue 7		
27	Dr. Ambika R	Analysis Of Quantum Key Distribution In Cryptography And Its Applications	Journal of Seybold Report	July 2020	Scopus
28	Dr. Dankan Gowda V	Internet of Things: Internet Revolution, Impact, Technology Road Map And Features	Advances in Mathematics: Scientific Journal	July 2020	Scopus
29	Mamatha K R	Sparse Representation for Color Image Denoising	International Journal of Engineering and Computing	July 2019	Google Scholar
30	Dr.Anil kumar D	DRONE INTEGRATED WEATHER SENSORS FOR AGRICULTURE PURPOSE	International Journal of Electrical Engineering and Technology (IJEET)	July 2020	Scopus
31	Dr.Anil kumar D	Analysis Of Quantum Key Distribution In Cryptography And Its Applications	Journal of Seybold Report	July 2020	Scopus
32	Dr. Dankan Gowda V	Interpretation of One Dimensional Hilbert Transform in terms of Fourier Transform and its Applications	International Journal of Advanced Science and Technology	July 2020	Scopus
33	Dr. A Shobha Rani	Design and Implementation of Efficient Routing Algorithm for Wireless Sensor Networks	International Journal of Advanced Science and Technology	July 2020	Scopus
34	Dr. Dankan Gowda V	Design And Implementation Of Cryptcloud System For Securing Files In Cloud	Advances in Mathematics: Scientific Journal	July 2020	Scopus
35	Dr. Dankan Gowda V	FPGA Implementation of Low Power High Speed BTED Algorithm for 8 Bit Error Correction in Cryptography System	International Journal of Emerging Trends in Engineering Research	July 2020	Scopus

BMS INSTITUTE OF TECHNOLOGY AND MANAGEMENT

Avalahalli, Doddaballapur Main Road, Bengaluru - 560064

Publications(Conference) ACADEMIC YEAR - 2019-20

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Sl.No.	Faculty	Title of the paper	Conference	Date	Indexing
1	Dr. A Shobha Rani	Energy-Delay Based Route Request Scheme for Load Balanced Routing in Wireless Mesh Networks	4th International Conference on Communication and Electronics Systems	17th – 19th July 2019	Scopus Indexed, IEEE digital library
2	Prof.Mamatha K.R	Detection and analysis of plant leaf diseases using Convolutional neural Network	" International Conference on Recent innovative trends in Computer Science and applications",	25th-26th Oct 2019	Scopus Indexed,
3	Prof. Chandra Prabha.	Performance comparison of Conventional Neural Networks and Deep Learning Network of Cervical Cancer Diagnosis	International Conference	Nov 2019	Scopus Indexed,
4	Prof.Laxmisagar.H.S	A Survey on Automated Detection of Breast Cancer based	Proceedings of 2nd International Conference on innovative mechanisms for innovative applications	March, 2020	IEEE digital library
5	Prof.Laxmisagar.H.S	A survey on Automated methods used for WBC classification	2nd National Level Conference on Recent Trends in Engineering, Technology and Management Sciences (RTETMS-2020)" organized by Dr.K.V.Subba Reddy Institute of Technology, Kurnool, Andhra Pradesh.	26th June,2020.	IEEE digital library
6	Rashmi N	Automated Computer Vision based Weed Removal bot	sixth international conference on electronics ,	July 2-4,2020	Scopus Indexed, IEEE digital library

			computing and communication Technologies ,organised by IEEE Bangalore section.		
7	Rashmi N	Efforts to overcome the existing impairments in implementing Helicopter Satellite Communication System (HSCS)	third international conference on emerging trends inscience and technologies in engineering systems,ICETSE-2020, Organised bySJCIT, Bangalore	July19th 2020	Scopus Indexed, IEEE digital library

Human Assisting Robot

Anirban Sanyal[1], Dr. Jayadeva G.S[2], Arvind Srikanth[3], Sonal Pandey[4], Vibha K.V[5]
[2]Professor in Telecommunication Department, BMSIT&M
[1345]Students of VIII Semester, Telecommunication Department, BMSIT&M

Abstract- Various works are carried out on Human - following robots, for their immense potential to carry out mundane tasks like load carrying and monitoring of target individual through interaction. The recent advancements in vision and sensor technologies have helped in creating more user-friendly robots that are able to coexist with humans by leveraging the sensors for human detection, human movement estimation. But most of these sensors are suitable only for Line of Sight Objects. In case of loss of sight of the target, most of them fail to re-acquire their target. This work proposes a novel method to develop a human following robot using Bluetooth, GPS and Inertial Measurement Unit (IMU) on smartphones which can work under high interference environment and can reacquire the target if the target position is lost.

Keywords- Robotics, Arduino, helping mankind.

I. INTRODUCTION

Recent developments in the robotics world have developed robots and are user friendly, intelligent, and most importantly affordable. With these benefits of robotics, it is no wonder that they have found to be used in every field. The benefits of robots have increased their flexibility with being capable of performing a variety of tasks and applications. Robots also allow for increased production and profit margin because they can complete tasks faster. Robots have the ability to work around the clock since they do not require vacations, sick days, or breaks.

Robots are preferred over men doing the job specifically in hazardous environments, handling toxic substances, working with heavy loads and wherever repetitive tasks are involved.

This work on human assisting robot uses Bluetooth to have communication between the robot and the tag held by the user, here after called target. In this work mobile is used as a tag. Human assisting robots are generally used in hospitals, restaurants and airports, etc [1] to carry luggage. However most of these robots fail to re-acquire their target, if line of sight is lost [2].

This problem can be eliminated using GPS and GPS has the advantage of tracking the robot even its position data is lost. To enable the tracking system on robot, GPS is used along with the GY273 compass which helps in giving appropriate commands to the motor drivers.

The primary goal of our work was to design and fabricate a robot that not only tracks the target but also moves towards it with the help of GPS and compass.

The disadvantage of using the GPS is that it does not work indoor [3].

In this work all the processing is carried out by the Arduino UNO microcontroller and L298N motor drivers are used to drive the motors.

II. PROPOSED SYSTEM

The circuit connection of the proposed system is shown in Fig. (1). It mainly consist of Arduino UNO microcontroller, HMC5883L compass, NEO6m GPS, HC06 Bluetooth module, L298N motor driver, 12 V DC motors and batteries. The three challenges that are faced while assembling above components and making human assisting robot are (i) power consumption for driving the motor and other electronic systems, (ii) control system used to control the robot and (iii) tracking the robot.

To drive the motor 12V battery is used another 9V battery is used to power other electronic modules. For controlling functions Arduino UNO microcontroller which is based on ATmega328P is used. This acts as the brain of entire system. It receives the information from different modules such as compass, GPS etc, and checks information and identifies the location of target. Then it sends the instruction to motor driver for making movements. Then continued tracking is done using the data of GPS Neo 6m and magnetometer GY273. The position is determined by the GPS Neo 6m. Magnetometer GY273 is used to measure the direction of movement of target called heading.

A. Bluetooth connection establishment

The Bluetooth module used here is HC 06. To enable the Bluetooth password is 1234 or 0000. The Bluetooth has 6 pins such as State, Rx, Tx, enable, VCC and GND. The VCC of Bluetooth is connected to 5V of Arduino and GND is connected to GND of Arduino. Rx pin of Bluetooth is connected to the transmitter pin of Arduino and Tx is pin of Bluetooth is connected to the receiver pin of Arduino. The Bluetooth is basically establishes communication between the robot and the human.

B. Compass connection establishment

The compass used here is a magnetometer GY273. Two Way Interface (TWI) connections is established in order to interface it with the Arduino. It contains 5 pins VCC, GND, SCL, SDA and DRDY. VCC and GND connection are same as that of the Bluetooth. Serial Clock (SCL) is connected to the A5 analog pin of Arduino UNO and serial Data (SDA) is connected to the A4 analog pin of the Arduino. The magnetometer is used to determine the heading. The Fig. (2) below shows the connection diagram of the compass.

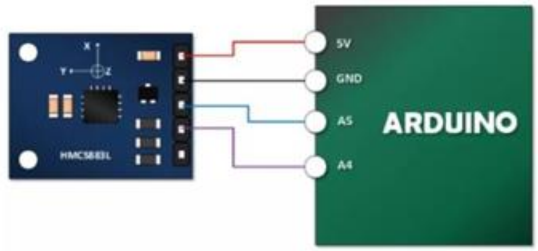


Fig.2: Connection of compass and Arduino UNO [4]

C. GPS connection establishment

This work uses GPS Neo 6m, UART connection is established in order to communicate with the Arduino board. The GPS has got 4 pins VCC and GND is connected as in Fig. (3) and the Rx is connected to the transmitter declared pin the Arduino board and Tx is connected to the receiver declared pin the Arduino. The GPS is used to get the location of the robot.

D. Motor Driver connection

The motor driver used in this project is L298d. the driver has got 4 input pins and 2 enable pins and output ports. The ground connection should be made common with Arduino and the driver. The four input pins are IN1, IN2, IN3 and IN4. These pins are connected to the digital pins of the Arduino. These pins are used to change the direction of the motor. IN1 and IN2 and used for motor1 and IN3 and IN4 are used for motor 2. The two enable pins are used to for speed control. ENA and ENB are the two enable pins for motor1 and motor2 respectively. The enable pins are connected to the PWM pins of the Arduino. The motors are connected to Arduino through motor drivers as the Arduino is not able to drive the required current. 12v Dc motors are used in this work. The connection diagram between L298n motor driver and Arduino is shown in Fig. (4)

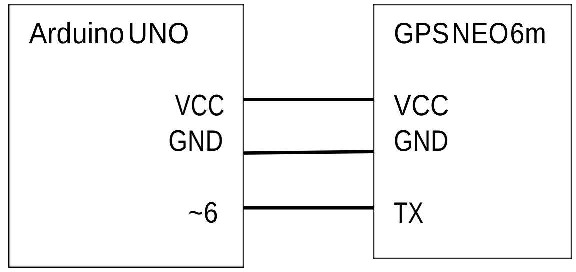


Fig.3: Connection of GPS

Here, we followed systematic approach which means starting from testing of each component to integrating it with other to make it function efficiently. Here each and every sensors and modules communicate among themselves to function efficiently. The above way of functioning is called decentralized approach. In decentralized approach there is no governing body. It is just the modules that work according to the algorithm.

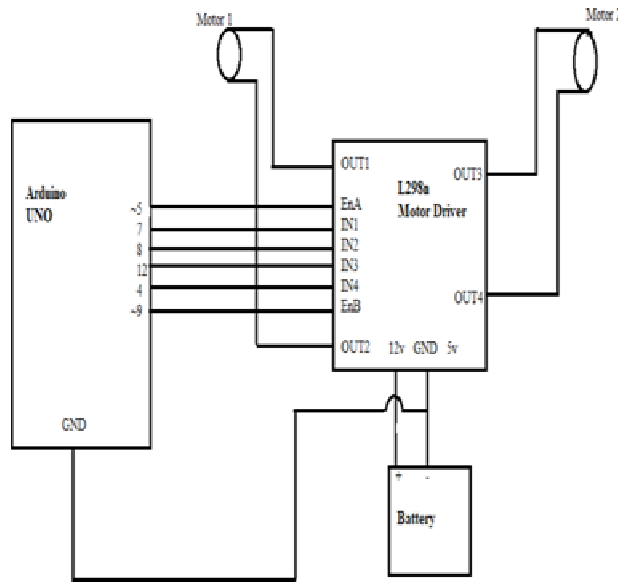


Fig.4: Connection of L298D with Arduino

E. Software Description

The system makes use of Arduino UNO that acts as a central control and processing unit along with Bluetooth to establish a connection between robot and target. The GPS along with compass is used to track or navigate the direction of the tag. Blynk platform is used to control our hardware via cloud using Bluetooth. After, extracting the information about the target, the tracking the target is done by GPS and Compass signals. In, the present prototype forward, backward, left and right turns motions are used to track target. As soon as the Arduino UNO unit receives the direction commands it executes commands provided in the program. Change in direction happens significantly whenever the target takes turn with respect to the hardware module.

III. METHODOLOGY

In this work there is no governing body, i.e it does not require any human intervention for command or direction. This means all the work is done by the modules interacting among themselves as per the algorithm. Hence we can say that the Autonomous connection is established.

Here Arduino is the brain of the entire system. It collects information from GPS, compass and Blynk platform via Bluetooth and give commands to the motor driver for controlling the robot. Everything starts with the connection with the Bluetooth. As the Bluetooth is connected to the smart phone it means there is a connection established between the target and the robot.

For control and navigation GPS Neo 6m is used along with magnetometer GY273. The Arduino gets the position of the robot and its direction of movement from these modules. For robot movement, DC motors are used. The H bridge motor driver is used to control the speed and the direction of the DC motor. PWM signal is used to control the speed and input pins are used to control the direction.

Whenever there is a change in the direction by the human the Arduino compares the position with and check the new direction with the help of magnetometer. Now it gives information to the motor driver through PWM pulse for changing the direction and follow the master. The algorithm of moving forward and turning is as shown in Fig. (5)

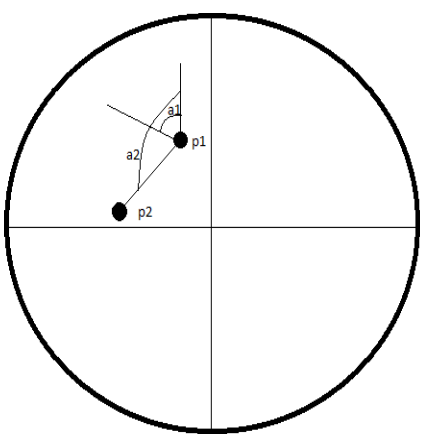


Fig.5: position determination

- Firstly, by using compass the direction is known where it's pointing (p1). So, in relation to the north pole it is called heading represented as a1.
- Secondly, position of android device can be estimated (p2). So, we can find the distance between the two points by using the common formula as:

  $D = \square(x2-x1)^2-(y2-y1)^2$
- Angle between the point p2 and the pole is known as bearing. It can be represented as a2.
- The turn angle can be estimated as : t=a2-a1 or t=bearing-heading
- Turn angle is calculated in order to move towards a particular point or direction.

The flow chart shown in Fig (6) shows the entire functioning of the robot.

IV. RESULTS

As expected the robot functions autonomously, without the intervention of the target / human. The algorithm written into the Arduino helps the robot to follow target by acquiring signals from sensors. Our robot was able to track the human (tag) independent of the direction (Left, Right, and straight) in which he was moving.

V. CONCLUSION

The paper has presented a Human-following robot which is being used for their immense potential to carry out mundane tasks like carrying load and for monitoring of an individual through interaction. However the present prototype is tested only with its movement on different direction and tracking the human who held the tag. In this work mobile is used as a tag. This system works efficiently in outdoor; however it is inefficient in indoors due to poor GPS signal. This work can be further improved by (i) adding obstacle detection sensors, (ii) scaling up the project for carrying weight and (iii) design wheels for moving in terrain surfaces. By adding the above features we can develop this work into product for public / commercial use.

VI. REFERENCES

[1]. V. Alvarez-Santos, X. M. Pardo, R. Iglesias, A. Canedo-Rodriguez, and C. V. Regueiro, " Feature analysis for human recognition and discrimination: Application to a person – following behavior in a mobile robot," Rob.

[2]. Auton. Syst., Vol. 60, no. 8, pp. 1021-1036,2012

[3]. BondaVennaPradeep,E.S.Rahul,Rao R Bhavani, "Follow me robot using bluetooth based position estimation" .1-6,2017

[4]. MHA Hamid, AH Adom, NA, Rahim MHF Rahiman, "Navigation of mobile robot using GPS".1-6, 5th international colloquium on signal processing & its applications(CSPA) 2009.

[5]. http://henrysbench.capnfatz.com

[6]. MD Sarmad Hassan, Mafaz Ali, Ali Fahim, "Design and Development of Human Following Robot" 1- 8,Student research paper conference Vol-2,No-15,july 2015.

[7]. Nicola Bellotto and H.Husing, "People Tracking And Identification With Mobile Robot"1-6,proceedings of the 2007 IEEE,International conference on Mechatronics and Automation,2007.

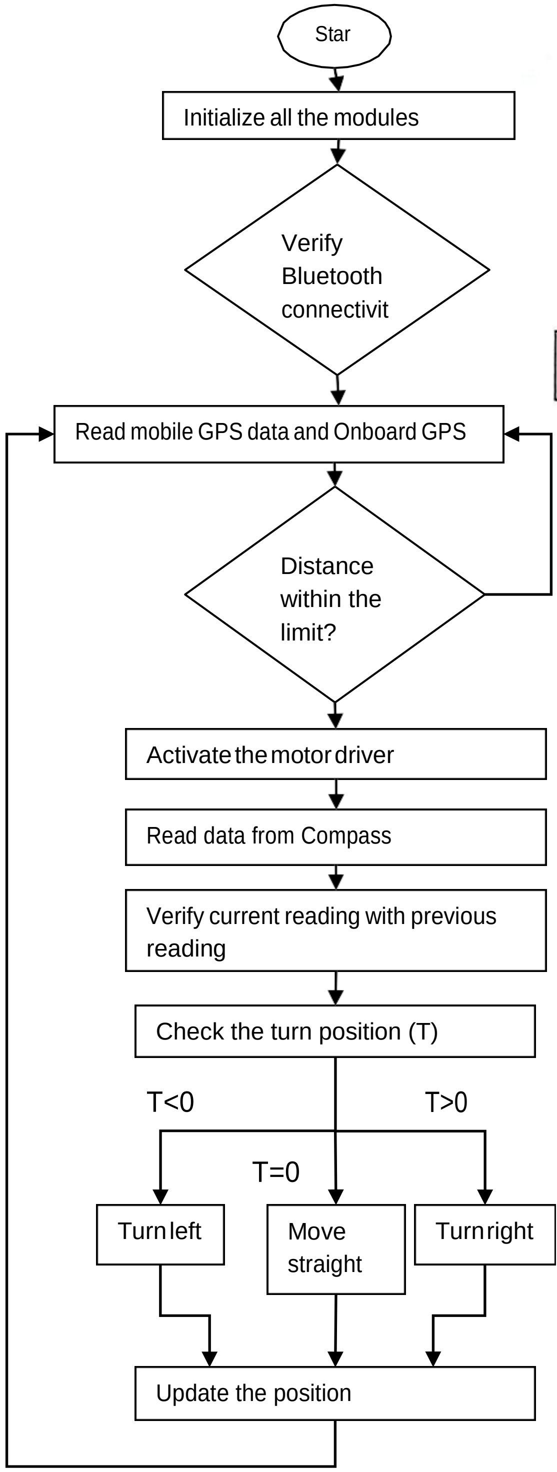


Fig.6: Flow chart of the system

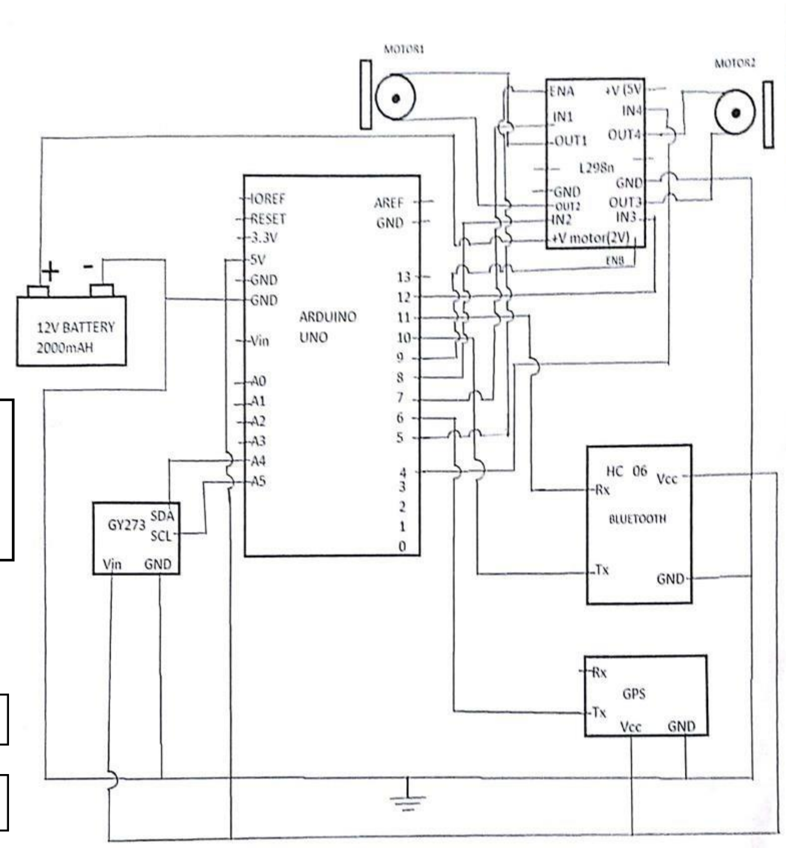


Fig.7: Circuit diagram of the system

Development of a Novel Protocol for Improvement of Qos Inwireless Sensor Networks: P-Rpeh

Sowmyashree M S, C S Mala

Abstract*: *The Wireless Sensor Networks (WSN) plays an important function in wireless communication because of its top notch utility and benefits. Wireless Sensor Networks have thousands or hundreds of potential nodes, which are minute computer like, and areable of measuring the physical characteristics of the neighboring environment location and thentransmit the gathered information using wireless radio links. This paper proposes an efficient scheme for resource allocation used for the transmission of data using priority in the nodes. The protocol, Medium Access Control (MAC) has been devisedfor evaluating the performance. The parameters considered are bandwidth, energy consumption, delay, throughput andpacket delivery ratio. A new MAC protocol,"P-RPEH"(Priority that is based on the Residual Power and Energy Harvesting rate), using priority in nodes with maximumenergyis proposed. The proposed method is compared with the PRIN (Priority in Node) and PRIMA (Priority based MAC)existing protocols. The results obtained from the new protocol proves that, the newproposed protocoloutperforms the existingmethods in terms of bandwidth, delay, energy, throughput and the ratio of packets delivered.

***Index Terms*: *Wireless Sensor Networks*, QOS, Medium Access Control, Priority.**

I. INTRODUCTION

The Wireless Sensor Networks (WSNs) has enormous number of sensors which are distributedin a random manner. These small/tiny sensor nodes are made up of a sensor unit, a processor, a power unit and communication capacities. These effective nodes makecommunication with each and every other nodeand are successfulinsharing and transmitting the sensed parameters such as humidity,temperature, motion detection etc. The communication reaches the collectingstation at once or through different intermediate nodes. The Quality of Service (QOS) which is important in the Wireless Sensor Networks is dependent on many parameters. Some of the parameters considered in proposed work to increase the efficiency are as follows:

i. Bandwidth
ii. Energy Consumption
iii. Delay
iv. Throughput
v. Packet Delivery Ratio

Bandwidth and Energy efficient routing protocols are required for obtaining the QOS.Further, there is a requirement for the standardized protocolin order to obtainthe information which is required by the sensors and are normallyavailable in the equipment of networking [1]. Prioritizing the nodes help in increasing the network performance which in-turn improves bandwidth andenergyefficiency.

Another parameter of importance is delay. In the process of collection of information, huge data statistics requirementshas to be gathered and accumulatedintime.Accumulatingrequired amount of data resultsina delay which may becomelonger. In this process, because of theinherent quality of the hyperlink used for the wireless transmission, thepackets which holds theinformation may get lost. This leads to a greater challenge in the transmission of real-time data, within therequired time frame.Some industrial applications allow delay-tolerance withinaspecifiedtime constraint [2].

If the packets are misplaced, the data has to be retransmitted again, which leads to a protracted postponement. In some situations such as monitoring, a long put off will delay the timing of obtaining data which causes extra losses. It is also crucial that a device makes a decision before the records are complete. This results in lack of information in some portions of the data.So care must be taken in such a manner that less delay is obtained. [3].

The other important evaluation parameters include the throughput and packet delivery ratio. There are numerous researches going on for the standardization of these factors to improve the QOS.

The QOS challenges differ in different WSNs with respect to resource constraints such as node deployment, real time traffic, scalability, data redundancy, topology changes, and contention window size so on.

The paper is arranged into 5 sections: in which Section 1 introduces the parameters considered for QOS. Section 2 discusses on the prior art in MAC protocols. Section 3throws light on the proposed method *P-RPEH* protocol.Theresults of the proposed methodology are discussed in section 4.The summary/conclusion of the present work and future enhancements are discussed in section 5 and 6.

II. RELATED WORK

From the survey of literature which ambitions to explore on the different types of MAC protocols that are used in WSNs to achieve better bandwidth and energy efficiency with respect to priority based allocation for QOS. Time Division Multiple Access (TDMA) and Carrier Sense Multiple Access (CSMA) are the different types of MAC protocols that are usually used for transmission ofdata through the wireless networks.

PRIN based MAC protocol is one of the existing methods to improve the QOS. The PRIN uses static priority in source as well asin the intermediate nodes.AnandaKumarSubbramanya et al have assigned the highest priority to that node which maybeonly one hop away from sink node which aims to achieve a very good QOS parameters[4].

Revised Manuscript Received on September 25, 2019.
Sowmyashree M S, BMS Institute of Technology & Management
C S Mala, BMS Institute of Technology & Management

The amount of energy that is consumed is reducedby assigning differentpriority to all the incoming packets.
This hasincreasedthethroughput to some extent only.

Another methodisPRIMA, which is based on two phases. In the phase one, clusters are formed and in the phase two the channel is accessed. The clustering technique is used in MACto handle the issues related to network scalability in a better fashion.Jalel Ben-Othman et al [5], have used PRIMA protocol where in the channel access composes of the hybrid method where both TDMA and CSMA are used. TDMA controls the data and CSMA communicates with control messages. Thishas minimized the packet collision which the author claims to have reduced the energy consumption.

AQ-MAC is another Asynchronous MAC protocolwhich is usedfor obtaining the QOS inthe sensor networks, the efficiency of energyis obtainedby organizingthe incoming data into numerous static priorities.Fafoutis, Xetal[9], have discussed in detail the AQ-MAC protocol. The authors were able to moderately improve energy efficiency.

Bandwidth plays an important role in achieving good QOS. Bandwidth reduces congestion by assigning priority to the nodes. This is done using cluster formation. In the cluster formation process, the highest priority is assigned to theheadof clusters.The members of thecluster and the cluster head nodes communicate with single hop step, while the cluster head and the sink communicates with multiple hops. The data is sent using TDMA time slots inorder to utilize the bandwidth efficiently. [6]

D Mathavanet al have worked on the density of traffic. They have obtained a better bandwidth performance by creating adefinedpath tree. The path was based on the density of traffic. They have also separated the highest priority from the lowest priority in the tree. The author’s claim that by doing so congestion can be avoided [7].

To identify the class traffic,SSridevietal [10], have usedEWMAmethod.Inthis method, traffic classifier is inserted into the appropriate queue depending on the traffic. Then Priority is assigned to each queue. Depending on the priority the packets are scheduled. High and low priority is assigned depending on the source and transit traffic. The authors claim that by doing so, more bandwidth has been used.

Timeliness and also reliability has a greater impact on the performance of QOS. To obtain the timeliness of the data the delay during the transmission over the wireless network should be less which also yields efficient transmission [11].

III. PROPOSED METHOD

WSNs have wide applications which include homeautomation, defense, health, environment and many other areas. To work in all these complex fields a good augment support system is required. To fulfill the WSNs capacity, QOS requirements must be enhanced. This can be achievedbysetting priorities in the existing MAC protocols.

In the present work P-RPEH has been developed to enhance the QOS. In this method, the minimum and maximum nodes taken are 25 and 125 and are presented in the figures 1 and 2 respectively. The proposed methodology is presented in figure 3.All the mobile nodes deployed are in random fashion in the real time environment.The routing protocol used for implementation is Ad-Hoc On-Demand Distance Vector (AODV). The other protocols which are generally used for routing have energy constrains and flooding, which is overcome by using AODV. AODV establishes routes to destination on demand.

Then the priority to the nodes is assigned depending on the maximum Residual Power and Energy Harvesting rate.If in case the residual power is large and the rate of energy harvesting is small, then the priority is not assigned to the node. Or with small residual power and large energy harvesting rate, the priority is still not assigned. The priority is assignedonly when both residual power and energy-harvesting rate is maximum.
The other nodes which lack these criteria go to the sleep state. For transmission of data through wireless network TDMA is used .

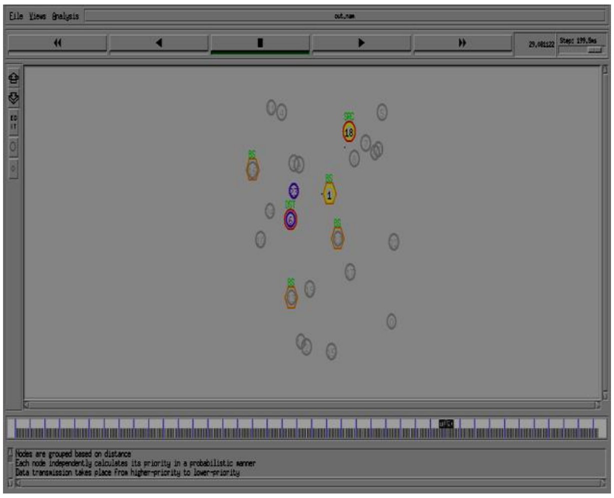


Fig 1: WSN Structure with minimum 25 nodes

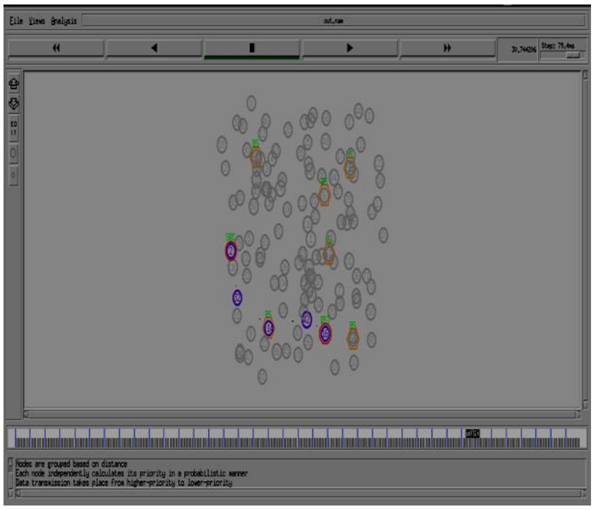


Fig 2: WSN Structure with maximum 125 nodes

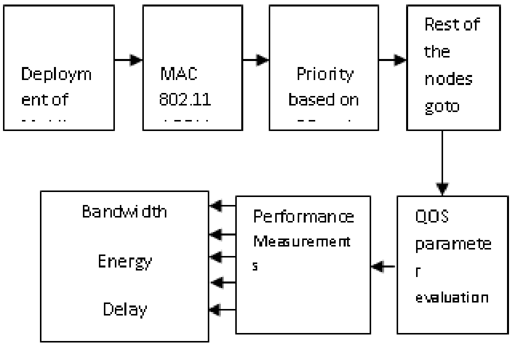


Figure 3: Proposed Methodology

The QOS performance parameters which are considered for the evaluation is listed below:

3.1 Bandwidth :

Bandwidth is calculated based on the RREQ (Route Request)Packets. These packets arethen sent to the destination node. These RREQ packets by the total amount of simulation time in the network, is expressed in equation (1) which denotes allocation of bandwidth.The source node at the initial stage checks theavailability of bandwidth. The RREQ packet is sent by the source if there exist a minimumof one time slot that is free at that instant of time, then the source sends a RREQ packet. If there isresponse within a fixed time period, then thenode which is the source node again sends the RREQ packet. If in case the response packet (RREP) is received, the path is set up for transmission. Once the path is set,the source node is responsible for allocating the time slots for starting thetransmission. During the process of path set up if there are three consecutive failures in finding thepath between source and destination, the source node aborts itself in sending the RREQ packets.

$$Bandwidth = \frac{B}{N} \quad \ldots\ldots.. (1)$$

Where Bdenotes the RREQ Packets sent to the destination nodein kbpsand$N$ is the total simulation time in seconds.

3.2 Energy Consumption:

If there are more number of nodes,the energy consumption is also more. A node drops some amount of energy for every packet transmitted orreceived, which is presented in equation(2)

$$Energy = \frac{E}{N} \ldots\ldots\ldots (2)$$

Where E is the amount of energy drop in joulesforthe transmission of each packet and N is the total simulation time in seconds.

3.3 Delay:

Delay is computed by the formula presented in equation 3, delay is nothing but the time that is taken by each packet hop from the sender to the receiver node.

$$Delay = \sum_1^n \quad Time\ spent\ on\ each\ Hop \ldots (3)$$

n = number of hops.

The unit for delay measurement is in seconds.

3.4 Throughput:

Throughput is computedas the number of total packets that are received by the destination node tothe total value of network time, aspresented in equation (4).

$$Throughput = \frac{T}{N} \ldots\ldots\ldots\ldots (4)$$

Where T is total number of received packets at the destination node. N is the total simulation time in seconds. It is measured in kilo bytes per second.

3.5 Packet Delivery Ratio:

The packet delivery ratio is calculated based on the number of packets that are actually sent from the source node to the number of packets that are received by the destination node. Equation (5) presents the same.

$$\text{Packet Delivery Ratio} = \frac{S}{W} \times 100 \ldots. (5)$$

S: The number of packets which are sent by the source node
W: Thenumber of packets that are received by the destination node.

It is obtained in terms of percentage.

The simulation procedure used for the proposed scheme:

Begin

1. Spread out all the nodes randomly in the real time environment.
2. Apply AODV routing Algorithm
3. Then apply the concept of priority to nodes.The priority is assigned based on P-RPEH method.
4. The remaining nodes goes to sleep state.
5. Evaluate the QOS performance parameters.
6. Generate the graphs respectively.

End

Following the above procedure, the QOS graphs were generated. The parameters used for simulation are indicated in the Table 1

Routing algorithm	AODV
Priority based on	Residual Power +Energy Harvesting Rate
Simulator used	NS2
Simulation start time	0.000000000 sec
Simulation End time	50.00000000 sec
Random Mobility Model	Random Way point
Propagation Model	Two-ray Ground
Traffic Model	Constant bit Rate (CBR)
Number of mobile nodes (Fixed Nodes)	25, 50, 75, 100 and 125

Channel Model	Omni Antenna
Packet Size	512
Network Interface types	Wireless ad-hoc
MAC Type	MAC/802_11, TDMA
Performance parameters	Bandwidth by RREQ packets, Delay, Energy Consumption and Throughput

Table 1

IV. SIMULATION RESULTS

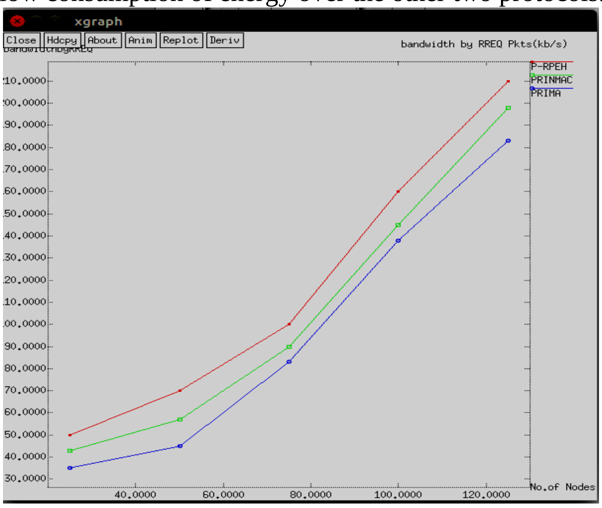
Here in section four the simulation results are discussed. The PRIN MAC and PRIMA Protocols were compared with P-RPEH, the present work.

The parametersconsidered for comparison are:

a. Bandwidth versus Number of Nodes
b. Energy versus Number of Nodes
c. Delay versus Number of Nodes
d. Throughput versus Number of Nodes
e. Packet delivery Ratio versus Number of Nodes.

4.1 .Bandwidth v/s Number of Nodes:
The allocation of resource such asbandwidth for the nodes which are of different sizes in the real time environment for various protocols wastaken into consideration. They were compared with the proposed new protocol. The comparison of graph is presented in the figure 4. It is clear from the graph that by varying the size of the nodes the size of bandwidth hasalso increased. The point to note here is that our present work gives a significant rise in bandwidth over the other two protocols.

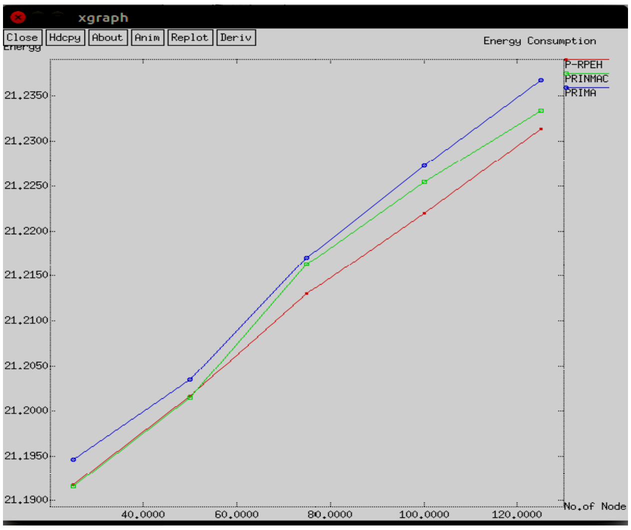
4.2 . Energy v/s Number of Nodes:
The consumption of energy likewise for various protocols was taken into consideration, and compared with PPREH, the proposed protocol. The graph with respect to energy v/s Nodes for different protocols isshown in figure 5. It is evident from the graph that asthere is increase in the size of the nodesthere is a decrease in the energy consumption.It also clearly indicates that the P-REPH protocol gives a significant low consumption of energy over the other two protocols.



No of Nodes

Figure 4: Number of Nodes *vs*Bandwidth

Bandwidth



No of Nodes

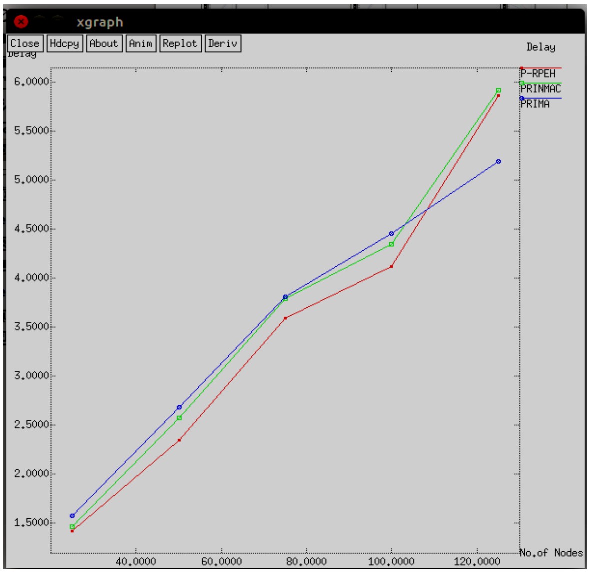
Figure 5: Number of Nodes *vs*Energy

4.3Delay v/s Number of Nodes:
The delay versusthe node Number for different protocols is presented in figure 6. It can be observed that as the number of nodes are increasing,delay decreases in general. Until 110 nodes the delay decreases greatly for P-RPEH protocol. There after there is a marginal increase in delay compared to PRIMA, but it is still better than PRINMAC. This result is further discussed in the bar graph at a later point in this paper.

4.4Throughput v/s Number of Nodes:
The next parameter throughput may also be considered as one of the important parameters to measure QOS. The graph for the nodes of different sizes in the real time environment for various protocols is present in figure 7. It is evident from the graph that as the node size increases the throughput also increases.The inference we draw from the graph is that there is a significant increase in throughput from the newalgorithm when compared with the other two existing protocols.

4.5 Packet Delivery Ratio v/s Number of Nodes:
The results of packet delivery ratio v/s the number of nodes is presented in figure 8. When the results of Packet delivery ratio of P-RPEH is compared with the PRINMAC and PRIMA protocols.The graph illustrates that asand when the size of the nodeincreases, the packet delivery ratioalso increases.Up tonode 100 there showsan increase in the packet delivery ratio for the protocol that is newly proposed, but after that the level of increase for the algorithm that we have developed is almost same when compared with the other two protocols.

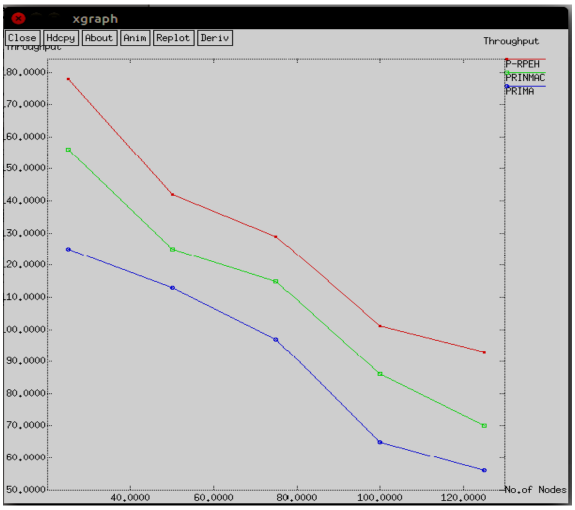
Delay



No of Nodes

Figure 6: Number of Nodes *vs*Delay

Throughput



No of Nodes

Figure 7: Number of Nodes *vs*Throughput

Packet Delivery Ratio

No of Nodes

Figure 8: Number of Nodes *vs*Packet Delivery Ratio.

The QOS parameters for all approaches are compared. After the comparison it is observed that the QOS parameter of our work yieldssignificantly better results than the existing ones which are PRINMAC and PRIMA protocols. The QOS parameters of interest of the newly developed protocol compared with the existing ones are presented in bar graph from figure 9 to figure 13. It is clearly evident from the figures that ournew proposed protocol outperforms better in the QOS parameter terms such as Bandwidth, Energy, delay, throughput and the packet delivery ratio in our prioritized method.

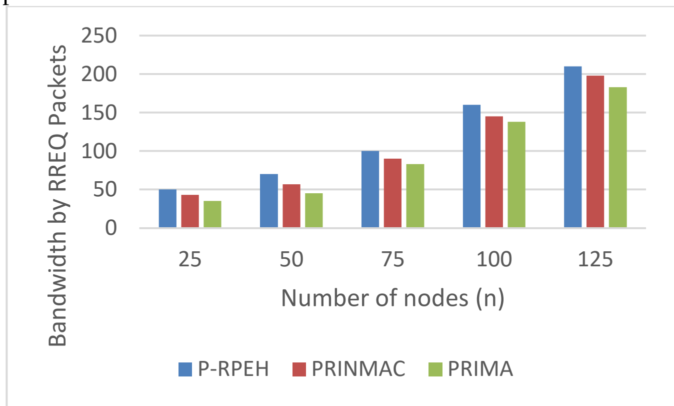


Figure 9: Comparison of *Bandwidth*

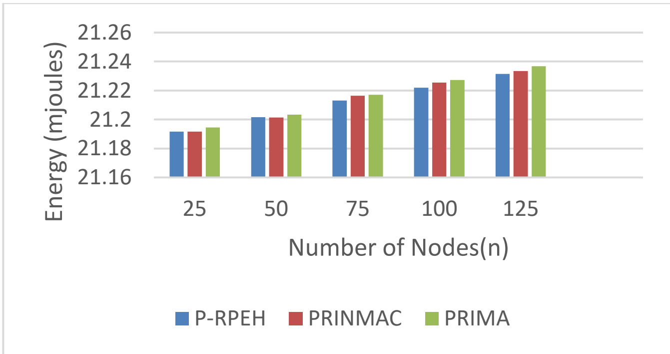


Figure 10: Comparison of *Energy*

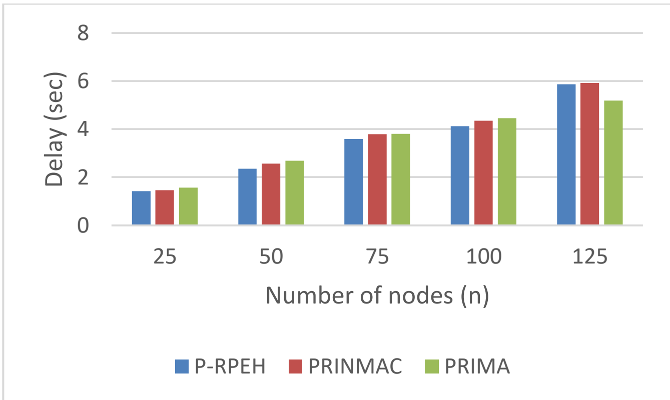


Figure 11: Comparison of *Delay*

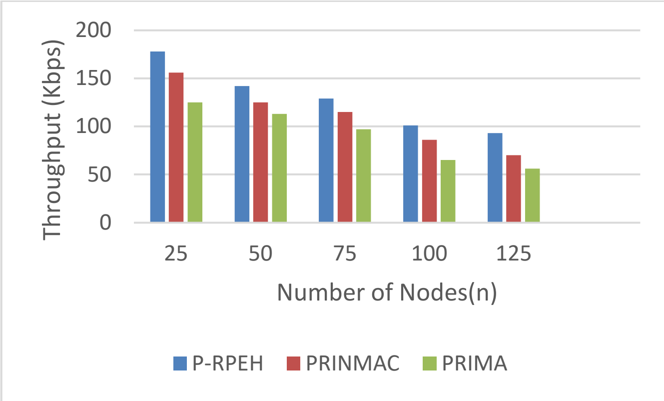


Figure 12: Comparison of *Throughput*

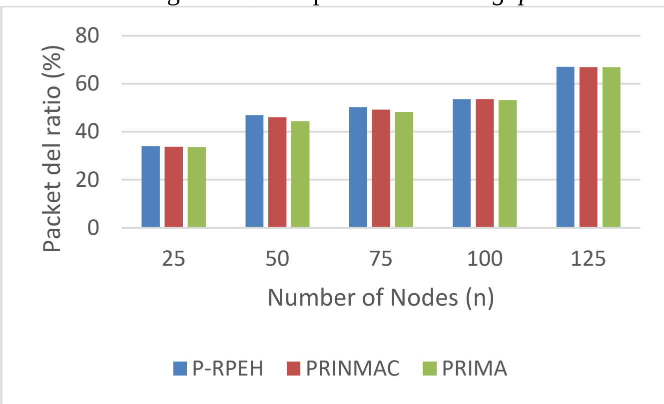


Figure 13: Comparison of *Packet Delivery Ratio*

The improvement in Performance of P-RPEH with the existing PRINMAC is presented in Table 3. Similarly,P-RPEH with that of PRIMAis presented in table 4.

No of Nodes	Parameters	P-RPEH	PRINMAC	Percentage improvement (P-RPEH with PRINMAC)
25		50	43	16.28
50		70	57	22.81
75		100	90	11.11
100		160	145	10.34
125	Bandwidth	210	198	6.06
25		21.19174	21.19457	0.01
50		21.20165	21.20345	0.01
75		21.21302	21.21631	0.02
100		21.222	21.22541	0.02
125	Energy	21.23134	21.23334	0.01
25		1.4127	1.4567	3.11
50		2.3456	2.5678	9.47
75		3.589	3.789	5.57
100		4.117	4.3452	5.54
125	Delay	5.8687	5.9234	0.93
25		178	156	14.10
50		142	125	13.60
75		129	115	12.17
100		101	86	17.44
125	Throughput	93	70	32.86
25		33.987	33.756	0.68
50		46.986	45.978	2.19
75	Packet	50.234	49.156	2.19
100	delivery	53.589	53.578	0.02
125	Ratio	66.998	66.912	0.13

Table 2

No of Nodes	Parameters	P-RPEH	PRIMA	Percentage improvement (P-RPEH with PRIMA
25		50	35	42.86
50	Bandwidth	70	45	55.56
75		100	83	20.48
100		160	138	15.94
125		210	183	14.75
25		21.19174	21.19456	0.01
50		21.20165	21.20343	0.01
75		21.21302	21.21703	0.02
100		21.222	21.22723	0.02
125	Energy	21.23134	21.23678	0.03
25		1.4127	1.567	10.92
50		2.3456	2.6789	14.21
75		3.589	3.8077	6.09
100		4.117	4.4567	8.25
125	Delay	5.8687	5.189	-11.58
25		178	125	42.40
50		142	113	25.66
75		129	97	32.99
100		101	65	55.38
125	Throughput	93	56	66.07
25		33.987	33.654	0.99
50		46.986	44.345	5.96
75	Packet	50.234	48.234	4.15
100	delivery	53.589	53.234	0.67
125	Ratio	66.998	66.845	0.23

Table 3

V. CONCLUSION

In this paper, we presented the P-RPEH, which is aQOS MAC protocol which improves the bandwidth as well as energy efficiency using priorityfor the wireless sensor networks. P-RPEH has combined the benefits of the residual power and energy harvesting rate,which aims to achieve significant amount of resource savings. To guarantee the Quality of Service, priority approach is used.

To conserve resources, P-RPEH protocol enables the nodes with maximum energy to transmit the data and lower priority nodes to go to sleep state. This avoids wasting of resources and uses TDMA schedule to transmit the data for communication process.

A great amount of work was carried out on P-RPEH Protocol. Its performance was analyzed through Simulation. The QOS parameters were compared with PRIMA and PRINMAC protocols. From the results obtained we conclude that the proposed work with a new protocol outperforms the other existing ones in terms of bandwidth, energy, delay, through-put and packet delivery ratio.

FUTUREENHANCEMENTS

The number of nodes considered in the present work is from 25 to 125. The work can be enhanced to greater number of nodes.The delay in the system can also be further decreased.

REFERENCES

1. SourayaYahiaoui, Mawloud Omar, AbdelmadjidBouabdallah, Enrico Natalizio, YacineChallal, "An energy efficient and QoS aware routing protocol for wireless sensorand actuator networks"International Journal of Electronics andCommunications (AEÜ) 83, Elsevier, 2017, pp-193-203.
2. Wenbo Zhang, Yue Liu, Guangjie Han, YongxinFeng,AndYuntao Zhao, "An Energy Efficient and QoS Aware Routing Algorithm Based on Data Classification for Industrial Wireless Sensor Networks",Special Section On Collaboration For Internet Of Things, IEEE, volume 6, 2018,pp-46495-46504.

3. X. Ju*et al.*, ``An energy conserving and transmission radius adaptive scheme to optimize performance of energy harvesting sensor networks," *Sensors*, vol. 18, no. 9, p. 2885, 2018.
4. Ananda Kumar Subramanian,,IlangoParamasivam, "PRIN: A Priority-Based Energy Efficient MAC Protocolfor Wireless Sensor Networks Varying the Sample Inter-Arrival Time", Wireless Pers Communication, Springer, 2016, pp-863-881.
5. Jalel Ben-Othman, Lynda Mokdad, Bashir Yahya, "An Energy Efficient Priority-based QoS MAC Protocol for Wireless Sensor Networks", IEEE ICC 2011 proceedings, 2011, pp-1-6.
6. Lokesh B Bhajantri, Siddalingesh S Hiremath, "Bandwidth Allocation and Scheduling in Distriduted sensor Networks", Internet of things and Cloud Computing Science Publishing Group, 2017, pp-30-37.
7. D Mathavan, R Karunakaran, J Kapelan, "Bandwidth Allocation in Application oriented networks based on Prioritization", SSRG International Journal of Mobile Computing & Applications, volume 4, Issue 1, 2017, pp-25-30.
8. Zarei, Z., Safavi, S. M., &Abbasi, A, "A MAC protocol for provisioning QoS and energyefficiency in WSNs. International Journal of Computer Theory and Engineering, 2012, pp89–493.
9. Fafoutis, X., Di Mauro, A., Vithanage, M. D., &Dragoni, N, Receiver-initiated medium access control protocols for wireless sensor networks. Computer Networks, 2015, pp55–74.
10. S Sridevi, M Usha, G Pauline Amertha Lithurin,"Priority Based Congestion control for heterogeneous traffic in multipath wireless sensor networks," International conference on computer communication and informatics,IEEE, january 2012, pg 1-5.
11. Y. Li, C. S. Chen, Y.-Q. Song, Z.Wang, and Y. Sun, `Enhancing real-time delivery in wireless sensor networks with two-hop information," *IEEE Trans. Ind. Informat.*, vol. 5, no. 2, 2009, pp 113-122.
12. Kakarla J, Majhi B, Battula R"A delay and energy aware coordination mechanism for WSAN",InternationalJournalof Communication System, 2016(October).
13. M. Liu, S. Xu, and S. Sun, "An agent-assisted QoS-based routing algorithm for wireless sensor networks,Journal of Networking Computer Application., vol. 35, no. 1, 2012, pp. 29 to 39.
14. M. Faheem and V. C. Gungor, " Energy efficient and QoS-aware routing protocol for wireless sensor network-based smart grid applications in the context of industry 4.0,Application of soft computing, vol. 68, no. 7, 2018, pp 910 to 922.
15. M. A. Razzaque, M. H. U. Ahmed, C. S. Hong, and S. Lee, "QoS-aware distributed adaptive cooperative routing in wireless sensor networks,Ad hoc Networks, vol. 19, Aug. 2014, pp. 28 to 42.
16. D.-R. Chen, " An energy-efficient QoS routing for wireless sensor networks using self-stabilizing algorithm, Ad hoc networks , vol. 26, Feb. 2016, pp. 240 to 255.
17. Andreou, P. G., Zeinalipour-Yazti, D., Samaras, G. S., &Chrysanthis, P. K, " A network-aware framework for energy-efficient data acquisition in wireless sensor networks", Journal of Network and Computer Applications, 46, 2014, pp227 to 240.
18. Doudou, M., Djenouri, D., Badache, N., &Bouabdallah, "A Synchronous contention-based MAC protocols for delay-sensitive wireless sensor networks: A review and taxonomy", Journal of Network and Computer Applications, 38, 2014, pp172 to 184.
19. Naderi, M. Y., Nintanavongsa, P., &Chowdhury, K. R. " RF-MAC: A medium access control protocol for re-chargeable sensor networks powered by wireless energy harvesting", IEEE Transactionson Wireless Communications, 13(99), 2014, pp3926 to 937.
20. Andrea M, Sabina R, "Priority Based Bandwidth Allocation in WSN", in proceedings of the 9th EAI International Conference on Performance Evaluation Methodologies and Tools, 2015, pp 137-144.
21. Tarek A, Rachid B Mourad O, "Bandwidth Assignment in a cluster based Wireless Sensor Network", in proceedings of the world Congress on Engineering, Vol. 2, 2013, pp 1-6.
22. Mary Cherian, T R Gopalkrishnan Nair, "Priority based bandwidth allocation in wireless sensor networks,"International journal of computer networks and communications, Volume 6, Issue 6, november 2014, pp 119-128.
23. C T Ee, and R Bajcsy, "Congestion Control and Fairness for many to one network routing in sensors," ACM SenSys, November 2004, pg 148-161.
24. SubhaKanta Swain and Pradipta Kumar Nanda," Priority Based Fairness Rate Control in Wireless Sensor Networks", IEEE WiSPNET 2017 conference,Pg no 2206-2213.
25. Shayesteh Tabatabaei, Mohammed Rahim Omrani." proposingamethodforcontrollingcongestion in wireless sensor networks using comparative fuzzy logic", Springer Wireless Personal Communications, June 2018, Volume 100, Issue 4, pp 1459–1476 .
26. Yitian Zhang , Kun Yang , Xiaofeng Lu, "Energy Efficient Prioritized Bandwidth Allocation Algorithm supporting RF Energy Charging " IEEE 2018 14th International Wireless Communication and Mobile Computing conference (IWCMC), pp no 1213-1218.
27. Yongbo Cheng, Shiliang Xiao, Jianpo Liu, FengGuo, Ronghua Qin, Baoqing Li, Xiaobing Yuan," An approximate bandwidth allocation algorithm for tradeoff between fairness and throughput in WSN", Springer - Wireless Networks, Volume 24, Issue 6, August 2018, pp 2165–2177
28. P. Mohanty, M. R. Kabat, "Energy efficient reliable multi-path data transmission in WSN for healthcare application", International JournalWireless. Network, 2016, vol. 23, no. 2, pp. 162-172.
29. S. Iniya Shree , M. Karthiga , C. Mariyammal, "Improving congestion control in WSN by multipath routing with priority based scheduling", IEEE- International Conference on Inventive Systems and Control (ICISC), 2017, pp 1-6.
30. D. Wu, J. He, H. Wang, C. Wang, and R. Wang, "A hierarchical packet forwarding mechanism for energy harvesting wireless sensor networks", IEEE Communication Magazine., vol. 53, Issue 8, Aug. 2015, pp. 92 to 98.

AUTHORS PROFILE

Mrs.Sowmyashree M S completed B.E (ECE), from Oxford college of engineering, Bangalore, M.Tech (Electronics and communication) from MVIT Bengaluru, and Pursuing Ph.D (WSN) from Visvesvaraya University, Belagavi. she has 12 years of teaching, 4 years of research experience. At Present, working as Asst.Professor, ETE Dept.,BMSIT, Bengaluru. Member for professional bodies, ISTE & IETE. Organized and attended many Workshops, FDPs and STTPs. Guided UG Projects . Published 11 Technical papers in National \ International journals \ Conference.

.

Dr C S Mala,is currently working as professor in ECE department. She has 32+ years of Teaching, Administration and Industry experience. She has also worked as special officer in VTU, headed the department of ECE in various colleges. She was also member, Board of Examiner ECE\TE in 2018-19. Also served as senior manager in in MANWISH Infotech Ltd. AS a researcher she has published papers in various National and International journals. She is also reviewer and session chair for National and International Conference.

ISSN 2321 3361 © 2019 IJESC

Research Article Volume 9 Issue No. 7

Feature Extraction and Classification of Visually Evoked Potentials and its Application in Forensic Science

Saneesh Cleatus.T[1], Mamatha. K. R[2], Hamsavahini. R[3]
Department of Electronics and Communication Engineering
BMS Institute of Technology and Management, Bangalore, India

Abstract:
In forensic science, there are various criminal analysis techniques used such as polygraph, narco-analysis test, brain fingerprinting. In any criminal act, the brain is always there, planning, executing and recording the crime. The fundamental difference between a perpetrator and a falsely accused innocent person is that, the former, having committed the crime, has the details of the crime stored in the brain whereas the latter who is innocent will not have the details in the brain. Various techniques like P300, P300-MERMER which are classes of EEG are used in interrogating a criminal mind. An analysis of Event Related Potentials(ERP) such as Visually Evoked Potential(VEP) within EEG signal is mainly limited to clinical diagnostics. In our work we developed an algorithm to extract and classify the features of visually evoked potentials (VEP). This technique proposed can be applied in the field of forensic science.

I. INTRODUCTION

Forensic science is the scientific method of gathering and examining information about the past which is then used in a court of law. It is used to for the scientific assessment of DNA, blood samples, bones and so on. Forensic plays an important role in criminal investigations. There are various other scientific methods such as psychological test, narco analysis test, polygraph test, brain Mapping etc. EEG is a promising tool in forensic science since this is a non-invasive technique and it does not induce any drugs in human body. More over the EEG signals are directly taken from human scalp and are electrical variations in various parts of human brain and it is also a function of persons thought and emotions. Hence this is one of the best method to retrieve the kind of information stored in the brain. Novel approaches in the detection and processing of EEG signals and its analysis is very important since these techniques are very promising not only in the field of forensic analysis, but in various other clinical diagnosis also. Till date, various techniques like P300, P300-MERMER which are event related potentials a class of EEG signals are used to interrogate a criminal mind. In this paper, an analysis of a class of event related potentials that is visually evoked potential is proposed that can be applied in the field of forensic science. Visually Evoked Potentials(VEP) within EEG signals are mainly used in clinical diagnostics. VEP is exploited in the field of forensics in the present project. A visual evoked potential is an evoked potential caused by a visual stimulus, such as an alternating checkerboard pattern on a computer screen. Responses are recorded from electrodes that are placed on the back of your head and are observed as a reading on an electroencephalogram (EEG). These responses usually originate from the occipital cortex, the area of the brain involved in receiving and interpreting visual signals. The EEG (electroencephalography) based system determines whether specific information is stored in a person's memory or not. Information of a crime registered in the brain is revealed by a distinctive pattern in the VEP (Visual Evoked Potentials) extracted from EEG. We are presently demonstrating this by the variation in the EEG pattern for a particular class of images. These variations may be observed for features extracted from persons EEG signals and a level of abnormality in the feature values can be utilised for further analysis. There are number of methods used in the extraction of features from EEG signals. Asish Panat et all, [1] uses wavelet transform to extract the features related to the mood/emotions of a person. A real time data of 256 samples per second is were used in the analysis. Four level decomposition of wavelet transform was implemented for feature extraction. The Statistical Parameters considered were mean, standard deviation, variance, skewness, entropy, power, and RMS value. It is observed that there is a significant change in the values calculated for RMS value, Power and Entropy of the signal when calculated for the EEG of the person when he is in ‘Angry’ mood and when he is in ‘Sad’ mood. Then values of features Power and RMS value of the data hold the higher values in case of the Angry emotion as compared to the Sad emotion, whereas, the values of feature Entropy show a lower value in the Angry emotion as compared to the sad emotion. Murugappan [2] et all, used the combination of surface Laplacian (SL) filtering, time-frequency analysis of wavelet transforms (WT) and linear classifiers are used to classify discrete emotions (happy, surprise, fear, disgust, and neutral). The validation of statistical features was performed using 5-fold cross validation. In this work, K nearest neighbour (KNN) outperformed linear discriminant analysis (LDA) by offering a maximum average classification rate of 83.04% on 62 channels and 79.17% on 24 channels, respectively.

Methodology.
We have analysed Visually Evoked Potential (VEP) within the Electroencephalographic (EEG) signals by extracting its features and classifying them. We have extracted the following features of Visually Evoked Potential (VEP) signals.

- Variance
- Standard Deviation
- Simple Square Integral

- RMS
- Waveform Length
- Entropy

EEG response after viewing different pictures (visual stimuli) was stored in three classes i.e. stimulus1 (S1), stimulus2- match (S2-match) and stimulus2- no match (S2-no-match). The EEG signals which were obtained after viewing repeated pictures (S2 repeated as S1) were categorized as “match” group and the EEG signals obtained after viewing non- repeated pictures (S2 is completely different from S1) were categorized as “no-match”. A random EEG signal corresponding to S1 is taken and classified into either match or no-match using KNN classification technique. For this particular classification technique, the above mentioned features are extracted using wavelet transform. These features are extracted using wavelet transform with the help of six-level wavelet decomposition. With the help of these extracted features, the EEG signals were classified into “match” group and “no-match” group using K-Nearest Neighbour (KNN) classification technique. The bock diagram of the proposed model is as shown in figure 1. It consists of testing and training signals. The testing data contains 10 EEG signals (S1 stimuli) and training data has 20 EEG signals (S2 stimuli with both ‘match’ and ‘no-match’). The wavelet decomposition and feature extraction are implemented on both the testing and training signals. The features of testing signals are compared with the features of training signals and based on close resemblance of features with the training signals, the testing signals are either classified into either ‘match’ or ‘no-match’ group using KNN classification technique.

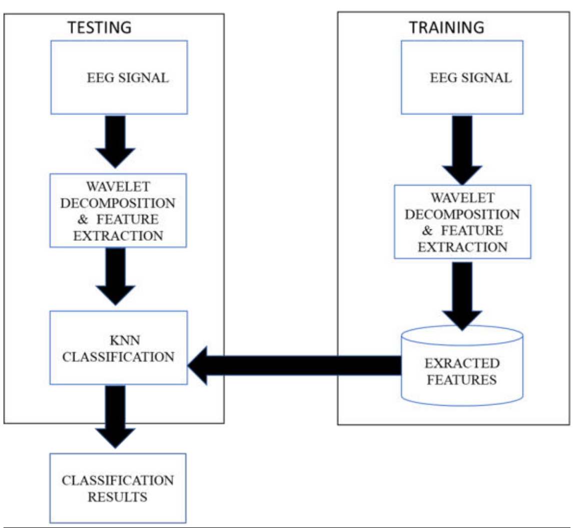


Figure.1. Block Diagram of the Model

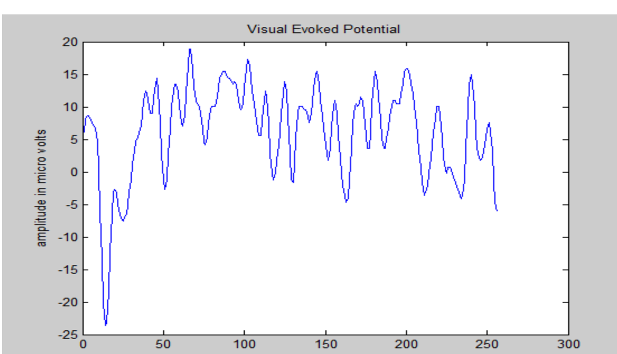


Figure.2. Visual Evoked Potential signal (input signal)

Wavelet Decomposition

In wavelet decomposition, the EEG signal is divided into a number of segments depending upon the decomposition levels used and the input to the wavelet transform is shown in above figure 2. The input shown in figure 2 is a visual evoked potential signal in which the potential is evoked by showing a visual stimulus. In our case the visual stimulus is the image which falls under match and no-match category. We have used a six-level wavelet decomposition technique. This results to six decomposed segments. The mother wavelet used is Daubechies. During wavelet decomposition, down sampling (decimation) is used after every level of decomposition and for this reason, the number of samples keeps on reducing with the increase in the number of decomposition levels as shown in figure 3.

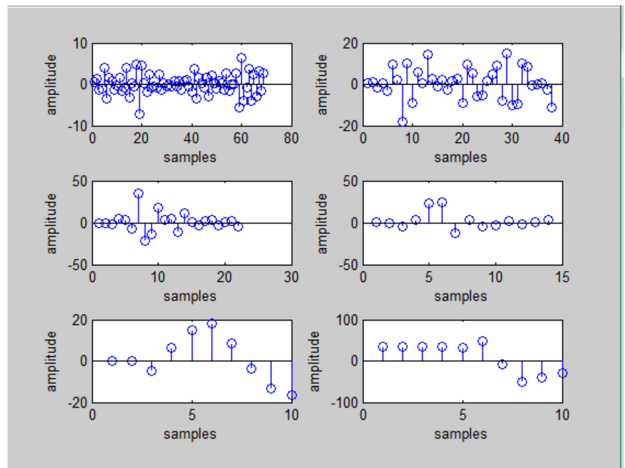


Figure.3. Decomposed Signal

Feature Extraction

Six statistical parameters like Variance, Standard Deviation, Simple Square Integral, RMS, Waveform Length and Entropy are calculated. The features are calculated on each component of the decomposed signal individually[7]. Figure 4 shows the plot of extracted features of an EEG signal which consists of six plots for six features separately. The first one is the plot for standard deviation in which x-axis shows six decomposed signal level and y-axis shows the magnitude values of standard deviation. In a similar way, variance, simple square integral (energy), entropy, waveform length and RMS value are plotted in figure 4.

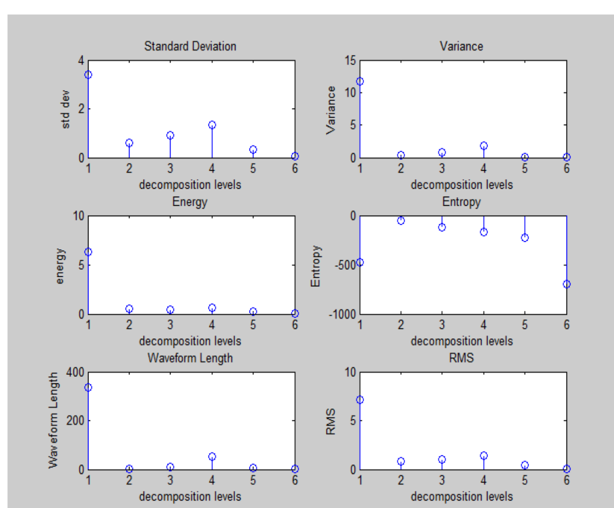


Figure.4. Extracted features of an EEG signal.

Classification
Using the extracted features, the EEG signals are classified into either 'match' or 'no-match' group. First, the Euclidian distance between the testing signal and training signal are calculated, that is, the Euclidian distance is calculated between the signal corresponding to S1 stimulus and S2 stimulus (match/no-match). It is given by:

$$\sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$$

If the testing signal lies nearer to 'match' training signal, then the testing signal is classified into 'match' group else it is classified into 'no-match' group. Figure 5 shows the K -Nearest Neighbour plot for a single testing signal with only one feature being considered i.e. variance.

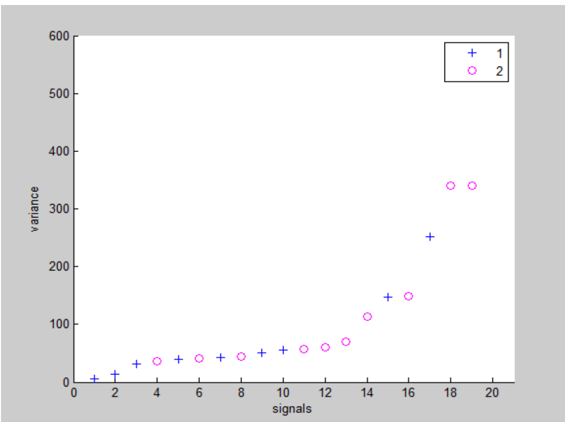


Figure.5. KNN Plot of a single testing signal

In figure 5, the Euclidian distances of the variance between a single testing signal and each of the 20 training signals are calculated. If the distance between the testing signal and the 'match' training signal is least i.e. if the distance is closer to zero, then the testing is classified into 'match' group otherwise the signal is classified into 'no-match' group.

Results
In order to group a testing signal into 'match' or 'no-match' using K -Nearest Neighbour classification technique, the value of K was varied from 3 to 11. The best results were obtained for K=7 where 9 out of 10 testing signals were classified into 'match' group thereby giving a classification accuracy of 90%. Blue plus indicates match class and pink circle indicates no match class.

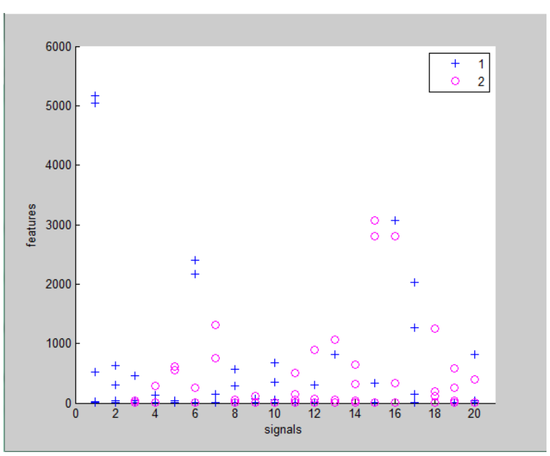


Figure.6. KNN plot of all the ten testing signals.

The Euclidian distances of all the features between 10 testing signal and each of the 20 training signals are calculated. If the distance between the testing signal and the 'match' training signal is least i.e. if the distance is closer to zero, then the testing signal is classified into 'match' group else the testing signal is classified into 'no-match' group. In figure 6, nine out of ten test signals comes under match class with 90% efficiency.

II. CONCLUSION

In this work, a person's response to the visual objects is studied. It states that if a person is subjected to different types of visual inputs, different variations in EEG were obtained which facilitated in classifying the signals into different classes. The extracted features using wavelet decompositions and further classification shows 90% efficiency. We hereby propose that, in criminal investigations we can use this method by analysing the familiarity of the person with the crime scene. If the person is familiar with the crime scene, there is 90% chance that, his/her VEP signal falls under the category of match signal.

III. SCOPE OF FUTURE WORK

Till date the forensic analysis using EEG signals is concentrated to standard studies of P300, P300-MERMER Polygraph etc. The present project proposes to record deviations in EEG signals and use the same to decipher the sanctity of an event occurred in past. This idea, after further studies may be of helpful in forensic detection of mental thoughts.

Acknowledgement
We are thankful, to the department of electronics and communication engineering, BMSIT for providing facility to simulate our algorithm using MATLAB. We are thankful to the students Shruti Singh, Sridevi N Wagle, Surabhi Mishra and Ravi Kumar S for patiently evaluating the performances of these algorithms and for bringing out several ideas during the execution of this work.

IV. REFERENCES

[1].Asish Panat, Anita Patil, and Gayatri Deshmukh."Feature Extraction of EEG Signals in Different Emotional States." Proceedings of 8th IRF International Conference, 04th May-2014, Pune, India, ISBN: 978-93-84209-12-4

[2].Murugappan, Murugappan, Ramachandran Nagarajan, and Sazali Yaacob. "Combining spatial filtering and wavelet transform for classifying human emotions using EEG Signals." Journal of Medical and Biological Engineering31.1 (2011): 45-51.

[3]. Vigneshwari, C., et al. "Analysis of finger movements using EEG signal. "International Journal of Emerging Technology and Advanced Engineering 3.1 (2013): 583-588.

[4]. Snodgrass, Joan G., and Mary Vanderwart. "A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity." Journal of experimental psychology: Human learning and memory6.2 (1980): 174.

[5]. Palaniappan, Ramaswamy, Paramesran Raveendran, and Shogo Nishida. "Multi-channel noise reduced visual evoked potential analysis." 123.10 (2003): 1721-1727.

[6]. Sergyán, Szabolcs. "Color histogram features based image classification in content-based image retrieval systems." Applied Machine Intelligence and Informatics, 2008. SAMI 2008. 6th International Symposium on. IEEE, 2008.

[7]. Asha, H. V., Sandhya Paul, and H. D. Kavana. "Fetal electrocardiogram extraction using Robust Independent Component Analysis." 2014 International Conference on Communication and Signal Processing. IEEE, 2014.

IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Delay-Based Routing Mechanism for Load Balanced Routing in Wireless Mesh Networks

Keerthi D S[1*], Shobha Rani A[2], Basavaraju T G[3]

[1]Department of Electronics& Communication Engineering, Malnad College of Engineering, Hassan, India
[2]Department of Electronics and Communication Engineering, BMSIT, Bengaluru, India
[3] Department of Computer Science & Engineering, Govt. SKSJTI, K R Circle, Bengaluru, India

[*]Corresponding Author: keerthikumards@gmail.com, Tel.: +91-9611138618

DOI: https://doi.org/10.26438/ijcse/v7i5.501506 | Available online at: www.ijcseonline.org

Accepted: 17/May/2019, Published: 31/May/2019

***Abstract*—**The tremendous growth in usage of internet technology has led to the development of various wireless networks. One of the wireless networks that have gathered lot of attention is Wireless Mesh Network (WMN). WMN is being preferred because of its numerous advantages such as better coverage area, communication with other networks, low energy consumption, cost effectiveness, increased network capacity and is compatible with all IEEE 802.11 standards. However there are several challenges degrades the network performance. Normally WMNs adopts shortest path algorithm for route establishment and most of the existing algorithms are not fully accounting the factors that impacts network performance, which in turn introduces unbalanced load distribution issues in the WMN. The design of a novel delay-based link quality metric is given in this paper which utilizes the real-time statistics from the wireless driver to consider the wireless contention, congestion, and channel loss. The proposed delay metric is additive in nature and introduce less routing overhead compared to existing mechanisms. Simulation results illustrate that the proposed DBL-AODV protocol significantly enhances the performance of the network by reducing the routing overhead and the routes having high delay are avoided from the process of packet forwarding compared to the standard protocol.

***Keywords*—**Delay, Load balancing, AODV,WMN

I. Introduction

WMN is a multi-hop wireless network comprising of Mesh Gateway (MG) connected to wired systems. Mesh Routers (MR) forwarding the traffic to Internet Gateway and Mesh Clients (MC) are the end users. Mesh nodes are tiny radio transmitters which operate similar to that of a wireless router. These nodes deliver the packets as per the destination even when the nodes are not directly connected. Thus mesh network is reliable even when one of the nodes is not effective. WMN [1] has advantage over Mobile ad hoc network (MANET) by providing high broadband internet access in which clusters of stationery mesh routers with each one supporting multiple radios. Some wire connected MRs behaves as internet gateways (IGWs) to other MRs and provides internet connection.
Features of WMN: Wireless mesh network is a multi-hop wireless network. This provides connectivity even to the users in the non-line of sight conditions and hence increases the throughput. In this way they provide reliable communication than the cellular networks. WMNs exploit IEEE 802.11 radio technology and are easily integrated with the other wireless networks such as wireless sensor, cellular, wireless fidelity (Wi-Fi), WiMedia, worldwide interoperability for microwave access (WiMAX) etc. through the mesh routers which are equipped with multiple interfaces.

WMNs have highly adaptable network architecture, easier deployment and configuration, robust and reliable service coverage, lower operation cost and require less installation time. They are self-configuring; the network spontaneously includes a new node into the existing structure without the involvement of a network administrator. They are self-healing; the network automatically identifies the most secure and the fastest path to transmit the data, even if the nodes fail. As more and more nodes are installed in the network, the network becomes wider and faster. WMN can also be used for wide variety of applications such as transport systems, wireless sensor network, surveillance system, broadband home networking, building automation, health and medical system, community and neighbourhood networking.

Some of the basic routing metrics in wireless mesh networks are: Hop count is the basic routing metric which determines the path with minimum number of nodes between the source and destination. It is the simplest routing metric and is commonly used with AODV, DSR, DSDV routing protocols. However it doesn't take into consideration the transmission rate and packet loss.

Expected Transmission Count (ETX) is the number of transmissions required to successfully transmit a unicast packet on a link. The ETX of a path is the total summation of ETX of each link along that path. The ETX of a link is measured by

$$ETX=1/((D_f \times D_r)) \quad (1)$$

where, D_f is the forward packet delivery ratio. D_r is the reverse packet delivery ratio. This routing metric considers only the packet drop ratio and do not consider transmission rate.

ETT is defined as the time required to successfully transmit a packet at the MAC layer. It is an extension of ETX and considers average size of the packet and bandwidth of the link

$$ETT=ETX \times S/B \quad (2)$$

where, S is the average packet size and B is the bandwidth. This routing metric increases the network throughput and doesn't consider the traffic-load and interference of a link.

WCETT routing metric is an extension form of ETT. The WCETT metric of a path say 'p' is defined by

$$WCETT(p) = (1-\propto)\sum_{link\in p} ETT_l + \propto \times max_{1\le c\le k} X_c \quad (3)$$

where, $\propto$ is the tunable parameter ranging between 0 and 1.X_c is the total summation of the links that are on channel 'c', 'k' describes different channels used in transmission path. WCETT has all the advantages of ETT except isotonicity. This routing metric considers intra-flow interference and doesn't take into account inter-flow interference. Metric of interference and channel switching (MIC) considers intra-flow interference, inter-flow interference and also supports load balancing. Since MIC is an improvement over WCETT, it considers the isotonicity of the path.

The MIC is a combination of two components, IRU and CSC.The MIC for a path p is defined by

$$MIC(p) = \frac{1}{(N \times \min(ETT)_{link})} * \sum_{link\ l\in p} IRU_1 + \sum_{node\ i\in p} CSC_i \quad (4)$$

where, N is the total number of nodes in the network.

Since wireless mesh networks are a type of packet switching networks, the routing protocol directs the packets from their source to their final destination through intermediate nodes. Several routing protocols exist for these types of networks [1]. Each protocol adopts a routing strategy. Based on these routing strategies, the mesh routing protocols are characterized into Proactive routing, Reactive routing and Hybrid routing protocols.

Low Capacity, management of the traffic and gateway nodes and end to end fairness problem are the major issues in the employment of WMNs. In order to reduce the existing challenges in WMNs load balancing plays a vital role. The entire paper is organized as below. The related work is discussed in Section 2. The suggested novel routing scheme is defined in section 3. The simulation results, performance comparison and analysis are presented in section 4. Lastly we have drawn inference in section 5.

II. RELATED WORK

In [2], authors have described a load balancing Multipath protocol MM-AOMDV. This protocol takes the advantages of multipath routing protocol in order to decrease the Route Discovery overhead. MM-AOMDV determines less congested routing paths bearing higher probability of transmission using metrics like load on the channel, remnant energy at node and channel access contentions. Thus, MM-AODV sets off route maintenance procedure less number of times compared to AOMDV.

The authors [3] have proposed a novel traffic predicting multipath routing algorithm called MRATP which comprises three important tasks such as multipath routing, congested discovery and load balancing. The results obtained show that MRATP performs effectively when compared with other algorithms due to its high scalability, better adaptability, desirable success ratio, robustness and lower end to end delay.

A prediction based adaptive load balancing (PALB) mechanism [4] has been proposed for MANETs. The PALB mechanism operates on wavelet analysis. The prediction of network traffic permits minimum congestion in traffic by adaptive distribution of traffic load among multiple disjoint paths thereby reducing packet dropping probability and end-to-end packet delay which in turn balances the network's energy consumption.

Adriana Hava et.al [5] has presented a novel load balancing algorithm that enhances QoS measures of videos in Wireless Mesh Networks. This novel mechanism achieves load balancing across the nodes by constantly evaluating the performance at WMN nodes and forwarding traffic to less congested nodes in the WMN. It employs a novel hybrid hierarchical architecture which is an amalgamation of centralized and distributed approach. The simulations performed depict that the proposed solution achieves better results over the classic approach.

The authors Huaiyu Wen et.al [6] has proposed a cross layer routing protocol for load balancing. The optimal path is discovered by using packet delivery rate and node load as performance parameters and by considering default hop count metric and cross-layer data of MAC layer. The main drawback of the proposed protocol is that it fails to achieve a high throughput while keeping a minimized end to end delay.

The mechanism in [7] employs routing and MAC scheduling to provide rigorous time delays for delay-sensitive traffic. The routing is implemented in MAC layer of the IEEE 802.16 mesh network using its specific features. The next hop node is determined by every node of the network on its own by utilizing pseudo-random election algorithm in coordinated distributed scheduling of the standard IEEE 802.16. Thus the algorithm renders itself scalable by working in a distributed manner and the possible path loop is avoided by developing a loop cancelling mechanism.

Interference Delay Aware (IDA) Routing Metric [8] is the integration of Cumulative inter-flow Load Interference (CLI), Channel Switching Load Cost (CSLC) and Estimated Transmission Delay (ETD).

CLI considers the interflow interference caused by the interfering nodes and is defined as

CLI (P) = $\sum_{link^c_{x,y\epsilon p}} CLI^c_{x,y} = \sum_{link^c_{x,y\epsilon p}} \sum_{j\epsilon | N^c_x\ U\ N^c_y|} LT^c_j$ (5)

where, $link^c_{x,y\epsilon p}$ =Link between Node x and y over a channel ‘C’ for path ‘P’

U N^c_y = union of interfering nodes $x\ and y$ respectively within two hops during transmission.

LT^c_j = Transmission time required by the Node j to carry traffic load through channel ‘C’ and is defined by

$$LT^c_j = \frac{TL^c_j}{B} \quad (6)$$

where, TL^c_j =Queue length

S=Packet size

B=Bandwidth

Channel Switching Load Cost (CSLC) considers the intra flow interference within two hops. It is given by

$$\text{CSLC(P)} = \sum_{node\ j\epsilon p} LC_j \quad (7)$$

LC_j =Intra-flow interference factor of node j.

In our work we have proposed a modified AODV protocol called Delay Based Load Balancing Ad-hoc on Demand Protocol (DBL-AODV), which considers response time of all nodes along the path as a routing metric for route discovery.

III. Delay Based Load Balancing AODV Protocol

The DBL-AODV protocol performs the following modifications to the existing protocol during routing:

- Route selection is based on the hop count and response time
- Whenever response time increases, packets are routed via alternate paths
- Based on the response time RREQ packets are considered or discarded

Delay Based Load Balancing AODV Protocol (DBL-AODV) is a modification of the standard AODV protocol. It uses response time as routing metric to opt the best path to the destination. The proposed method finds end-to-end delay for all the available paths to the destination during route discovery. The source node broadcasts RREQ towards the destination node. If route request message reached in the neighbour node or rely node, this node has to find the delay of the route request and it updates in routing table and checks its sequence number to find the freshness of route request. Sequence number is updated in its own routing table by rely nodes. The destination node finds the overall delay and it to be updated in route reply message. The path with higher delay indicates that it is heavily loaded and the path with least delay indicates the path is not congested. Once the source node gets a route reply, it selects the best route having minimum delay and avoiding packet transmission via congested route thus increasing the overall performance of the WMN. The flow chart of the DBL-AODV is as shown in Figure 2.

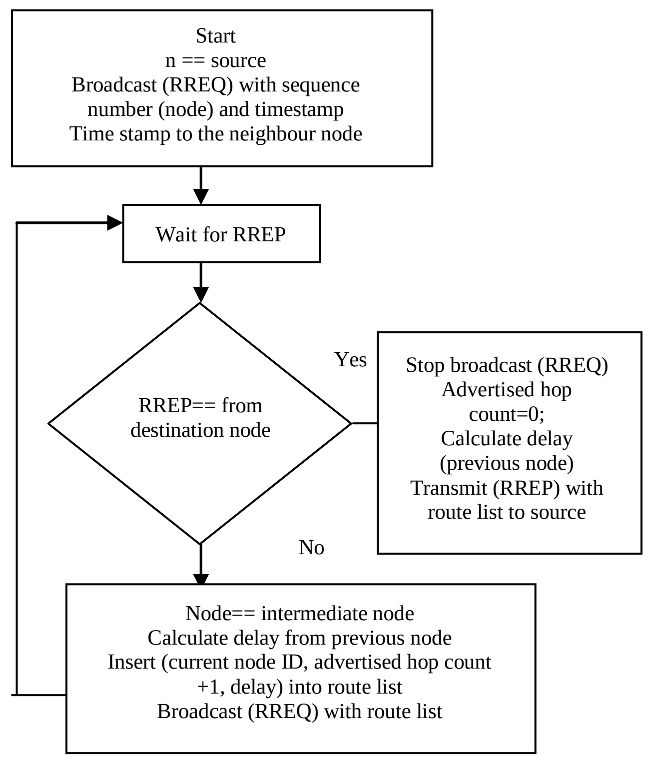


Figure 2. Flow Chart of DBL-AODV Protocol

The delay at every node in WMNs is an aggregation of different delays like input queuing delay, transmission delay, propagation delay, retransmission delay, processing delay and output queuing delay. The network delay analysis involves splitting of packet delay into a series of node delays where delay at each node is defined as the difference in interval of packet arrival at a node and the next node. Thus, the equation below determines that the nodal delay can be decomposed into its component delays as shown.

$$\tau_{nodal} = \tau_{proc} + \tau_{queue} + \tau_{trans} + \tau_{prop} \quad (8)$$

The processing delay τ_{proc} at each node is explained as the period of time exploited by a node for detecting errors, searching a feasible link for further transmission based on destination address and time for interpreting packet header while a packet is being processed. Although the processing appears complicated, this kind of delay is normally insignificant in contrast to other time delays. The time utilized for placing a packet into the communication media is the transmission delay τ_{trans} and the below equation represents the computation of the same.

$$\tau_{trans} = \frac{L}{R} \quad (9)$$

where L represents packet length in bits and R gives the rate of transmission in bits per unit time. The time duration expended by a packet in a queue for allowing the transmission of other packets moving across the same link is called queuing delay τ_{queue}. The queuing delay τ_{queue} and the transmission delay τ_{trans} are relatedby the following equation.

$$\tau_{queue} = \tau_{trans} * L_{queue} \quad (10)$$

where L_{queue} denotes the average length of queue and it relies on load factor.Load factor is described as the ratio of attempted link transmission to the maximum rate of link transmission. For a load factor of less than 0.5, the typical mean queue length will be less than 1. The queue length increases when the load factor exceeds 1. The transmission delay is the significant constituent of the node delay for a rarely loaded link. However, queue delay prevails for a densely loaded link.

The propagation delay τ_{prop} is defined as the duration that a signal utilizes for propagating from one node to the other via communication medium. It is derived by the following equation.

$$\tau_{prop} = \frac{d}{s} \quad (11)$$

$$\tau_n = \sum_{i \in n}(2D_i + E_i) + \tau_{WRj} + \tau_p \quad (12)$$

where, τ_n represents response time of the RREP packet n
i denotes a node on the path n
D_i represents internal delay in node i precluding the extra delay added.
τ_{WRj} is the delay in generating RREP packet in response to RREQ packet.
τ_p represents the propagation delay.
Internal Delay is given by

$$N_i = \tau_R + \tau_{Proc} + \tau_{Fi} + \tau_p \quad (13)$$

where τ_R represents RREQ/ RREP packet receiving delay
τ_{Proc} is the processing time of a packet
τ_{Fi} stands for the packet forwarding time in node i.

IV. Performance Evaluation

We conducted a number of simulations to evaluate the effectiveness of the proposed algorithm over standard AODV protocol. The performance of the DBL-AODV and standard AODV are compared in terms of throughput, total number of packets received, end to end delay and jitter.

The simulation environment consists of 25, 36, 49, 64 and 81 node densities which are placed randomly in a 1500m x 1500m area. The traffic is generated from three sources and traffic type chosen is Constant Bit Rate with each packet having size of 512 bytes. The simulations are run over 300 seconds.

The other parameters used for performance analysis are listed in Table 1.

Table 1. Mesh Network parameters in simulation

Parameter	Value
Protocols	AODV, DBL-AODV
No. of Nodes	25, 36, 49, 64, 81
Radio type	802.11b Radio
MAC Protocol	802.11s
Antenna Model	Omni-directional
Path Loss Model	Two ray propagation
Traffic Type	CBR
No. of CBR	3

In the simulation, the performance of DBL-AODV is evaluated and compared with Ad hoc On-demand Distance Vector routing protocol (AODV).From Figure 4 it is observed that AODV gives lesser throughput since packets are forwarded through the shortest path irrespective of the load along the path whereas the DBL-AODV considers the response time as a parameter for route selection. Based on the response time proposed protocol changes the route discovery process. In DBL-AODV packets are forwarded through the path having lesser response time and lower hop count among the all available paths to the destination. Hence proposed protocol gives higher throughput as compared to the AODV protocol. From the Figure 3, it is also observed that when node densities increase, the performance of the AODV decreases but DBL-AODV yields higher throughput. For lower node densities, the performance of the proposed protocol is increased by 5% but for higher node densities, the performance of the DBL-AODV is increased by 15% compared to AODV protocol which is even true for total number of packets received. From Figure 5, it is seen that end to end delay for DBL-AODV is lesser than the AODV protocol since it avoids the packet transfer through the highly congested path. For the higher node densities, delay is very much less than the AODV protocol. This is even same for the Jitter.

Table 2. Throughput Comparison

Node Density	AODV	DBL-AODV
25	296406	301138
36	187938	218924
49	125380	286159
64	86255.1	159105
81	111987	138974

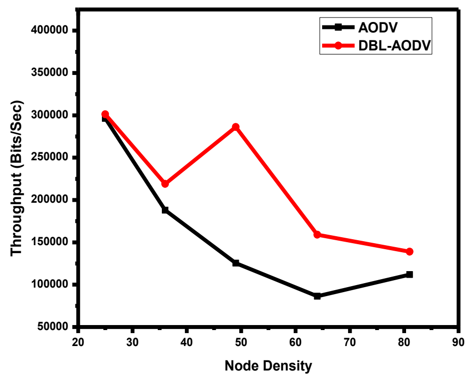


Figure 3. Variation of throughput v/s node densities

Table 3. Total Messages Received

Node Density	AODV	DBL-AODV
25	21707	21788.7
36	13702	15707
49	9180.33	17725.5
64	6314	11410
81	8197	10000.3

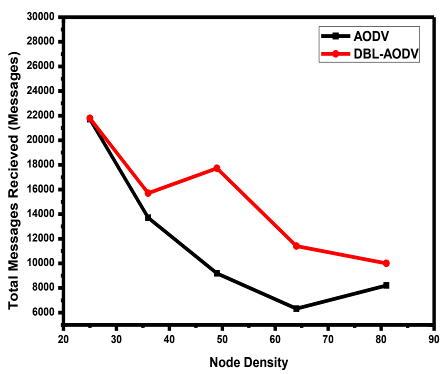


Figure 4. Variation of total messages received v/s node densities

Table 4. End to End Delay Comparison

Node Density	AODV	DBL-AODV
25	3.8457	2.43911
36	3.48773	2.98289
49	1.50114	1.45233
64	0.73018	0.6242
81	2.70725	1.0865

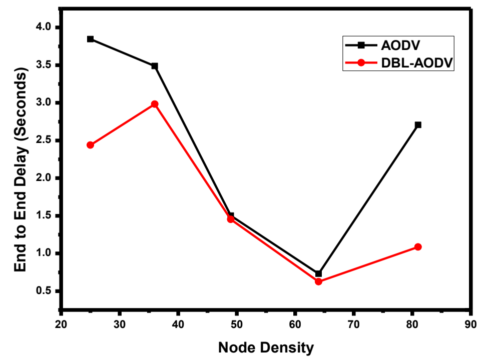


Figure 5.Variation of End to End Delay v/s node densities

Table 5. Jitter

Node Density	AODV	DBL-AODV
25	0.01224	0.01107
36	0.02336	0.01284
49	0.03944	0.00509
64	0.067	0.05347
81	0.05494	0.00713

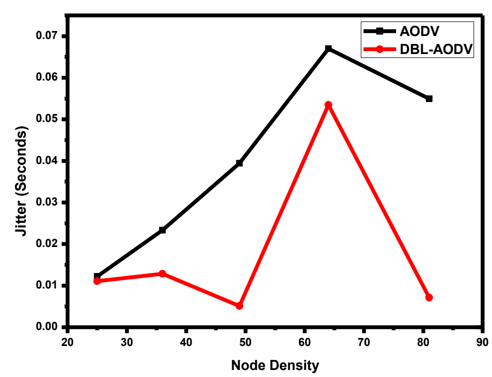


Figure 6.Variation of Jitter v/s node densities

V. CONCLUSION AND FUTURE SCOPE

In this work, DBL-AODV protocol is proposed which balances the load in the network by considering delay as the routing metric while performing the route discovery process. The DBL-AODV routing protocol adopts the best route with possible minimum delay and hop count, thereby overcoming the limitations of the existing multipath routing protocol. From the results it is clearly evident that DBL-AODV outperforms AODV in terms of throughput, total number of packets received, delay and jitter. The overall performance of the wireless mesh network is improved by around 5% for lower node density and around 15% for higher node density as compared to the standard AODV protocol.

ACKNOWLEDGMENT

This work is supported by the Technical Education Quality Improvement Program Phase III with the World Bank assistance. The authors would like to thank the State Project Facilitation Unit, Directorate of Technical Education, Government of Karnataka, National Project Implementation Unit, MHRD, New Delhi for their support and encouragement.

.REFERENCES

[1] F. Akyildiz, X. Wang, and W. Wang “Wireless mesh networks: a survey”, in Computer Networks Journal, vol.47, 2005, pp.445-487.

[2] Sujatha P. Terdal, Dr V.D Mytri, Dr. A Damodaram “Multiple Metrics based Load Balancing Routing Protocol for Mobile Ad Hoc Networks” First Asian Himalayas International Conference on Internet 3-5 Nov. 2009

[3] Shouyi YIN, Xiaokang LIN “Adaptive Load Balancing in Mobile Ad hoc Networks”, IEEE Communications Society /WCNC 2005, pp.1982-1987, 2005.

[4] Li Zhi-yuan, WANG Ru-chuan, Bi Jun-lei “A Multipath Routing Algorithm Based on Traffic Prediction in Wireless Mesh Networks”, Fifth International Conference on Natural Computation 2009, pp. 115-119 (2009)

[5] Adriana Hava, Gabriel- MiroMuntean, YacineGhamri-Doudane and John Murphy “A New Load Balancing Mechanism for Improved Video Delivery Over Wireless Mesh Networks”, IEEE 14th International Conference on High Performance Switching and Routing (HPSR), pp. 136-141, 2013

[6] Huaiyu Wen, Guang Chun Luo “Load Balance Routing Protocol in Wireless Mesh Network based on Cross-layer knowledge”, International Conference on Computational and Information Sciences 2013, pp. 1352-1355, 2013

[7] Yajun Li, Yuhang Yang and Chengyu Cao “A Novel Routing Algorithm in Distributed IEEE 802.16 Mesh Networks”, IEEE Communications Letters, Vol. 13, No. 10, October 2009 pp. 761-763.

[8] Anh-Ngoc Le, Dong-Won Kum, You-Ze Cho “Load-aware routing protocol for multi-radio Wireless Mesh Networks” International Conference on Communications and Electronics (ICCE), 2008, pp.138-143.

[9] EkramHossain , Kin K. Leung “Wireless Mesh Networks: Architectures and Protocols”, Springer, 2008.

[10] QimingTian “A New Interference-Delay Aware Routing Metric for Multi-Interface Wireless Mesh Networks” 6th International Conference on Wireless Communication Networking and Mobile Computing (WiCOM), 2010, pp.1-5.

[11] Iqbal M, Xinheng Wang, Wertheim, D Xu Zhou “Load balanced multiple gateway support in wireless mesh networks for broadband services” Wireless and Optical Communication Conference, 2009, pp.1-5.

[12] Dong-Won Kum, Won-KyeongSeo, Jae-In Choi, You-Ze Cho “NAV-based Load Aware Routing protocol in Wireless Mesh Networks”, Information and Communication Technology Convergence(ICTC),2010,pp.218-219.

[13] Gálvez, J.J, Ruiz, P.M., Skarmeta A.F.G. “A distributed algorithm for gateway load-balancing in Wireless Mesh Networks” Wireless Days (WD), 2008, pp.1 - 5.

[14] Xian Xiao-dong, Fu Yun-qing, Wang Song-jian, Liang Shan “A routing protocol based on combinative metric for wireless mesh network” International Conference on Networking, Sensing and Control (ICNSC), 2009, pp.803-806.

[15] V. Prasad, VS. Sunsan, "*Multi path dynamic routing for data integrity and delay Minimization differentiated services in wireless sensor network*", International Journal of Scientific Research in Computer Science and Engineering, Vol.4, Issue.4, pp.20-23, 2016

[16] A. Sebastian, S. Sivagurunathan, "*A Survey on Load Balancing Schemes in RPL based Internet of Things*", International Journal of Scientific Research in Network Security and Communication, Vol.6, Issue.3, pp.43-49, 2018

Authors Profile

Keerthi D S. received his B.E. and M.Tech degrees from Visvesvaraya Technological University (VTU), Belagavi. He is pursuing Ph.D. under the guidance of Dr. T. G. Basavaraju at the Department of Computer Science and Engineering Research Centre, GSKSJ Technological Institute affiliated to VTU, Belagavi. He is currently working as an Assistant Professor at Malnad College of Engineering, Hassan, Karnataka. His research interest is in the field of Wireless Mesh Networks.

Shobha Rani A. received her B. E Electronics and Communication Engineering) degree from Bangalore University in the year 2000, M. Tech (Digital Electronics and Communication) Degree from Visvesvaraya Technological University (VTU), Belgaum in the year 2005, and Ph. D (Networking) degree from Kuvempu University, Shimoga in the year, 2014. She is currently working as a Associate Professor in the Department of ECE at BMS Institute of Technology and Management, Bengaluru, India. She has 13 technical articles in good reputed journals and conferences. Her research interests include designing and implementing efficient protocols for wireless networks such as Ad-hoc, Sensor and Mesh networks.

T. G. Basavaraju has obtained his B.E degree in Computer Science and Engineering from Kuvempu University, Shimoga, Karnataka, M.E. degree in Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, Karnataka and Ph.D. degree from Jadavpur University, Kolkata. Currently, he is working as Professor and Head, Department of Computer Science and Engineering, Government SKSJ Technological Institute, Bangalore. He has more than 20 years of teaching and industrial experience. He has published several papers in national and international journals. He has also authored and co-authored several text books in the area of Computer Networking. His research interests are Wireless Mobile Ad hoc Networks, Wireless Sensor Networks and Wireless Mesh Networks.

Scalable Local Route Repair-Hybrid Wireless Mesh Protocol (SLRR-HWMP) for IEEE 802.11s

Chethan K C
Department of Computer Science and Engineering
Government Engineering College, Hassan, Karnataka, India
hanichethan@gmail.com

Shobha rani A
Department of Electronics and Communication
B M S Institute of Technology and Management, Bangalore
shobharani@bmsit.ac.in

Dr. T G Basavaraju
Department of Computer Science and Engineering
Government SKSJTI, Bangalore, India
Tg.braju@gmail.com

Abstract — Mesh Wireless Networks provide encouraging approach to provide network connectivity to the users at low cost. Wireless Mesh Networks became an important research area nowadays due to many of its advantages. Wireless Mesh Network uses Hybrid Wireless Mesh Routing Protocol (HWMP) protocol as a default path selection protocol and Air Time Link Metric (ALM) as default metric. HWMP is a MAC layer protocol used to create a path from source to destination. Even though HWMP is a default path selection protocol for 802.11s, many challenges remain unresolved, especially in routing. The results of many researches reveal that Control overhead is the main cause for not achieving scalability in all types of networks. In this paper, we have proposed a localized route repair mechanism named Scalable Local Route Repair-Hybrid Wireless Mesh Protocol (SLRR-HWMP) for Wireless Mesh Network. The proposed routing mechanism shows better performance in terms of throughput and end to end delay, especially in a large network; and hence provides scalability. This paper also shows the comparison of SLRR-HWMP and HWMP in terms of Throughput and end to end delay. Qualnet simulation tool is used for evaluating the proposed mechanism. The results shown by evaluation reveal that proposed routing technique shows better results compared to HWMP protocol.

Keywords - WMN- Wireless Mesh Networks, HWMP- Hybrid Wireless Mesh Protocol, Scalable, Local Route Repair.

I. INTRODUCTION

Mesh Networks are getting popular because of flexible infrastructure for providing Internet. WMN acts as a backbone network, so that different networks can be integrated and form a Mesh network. WMN uses existing infrastructure rather than creating new infrastructure. So WMNs are very helpful especially in rural areas where there is no enough infrastructures for forming a network. WMN combines the features of Different networks such as WLAN, Ad Hoc, mobile networks and other types of networks. WMN shows many advantages compared to other types of networks. WMN have characteristics of self-reliance, self-ruling and self-serving. So, managing of WMN is also easier. WMN with the above-mentioned advantages becomes an interesting topic for the researchers. WMN is an extension of an existing Wireless Local area network designed to provide scalability.

A. Architecture of wireless Mesh Networks.

Architecture of WMN consists of Mesh Points (MP), Mesh Access Points (MAP), Mesh Portal Points (MPP) and Stations (STA)[1][3] as shown in figure 1.

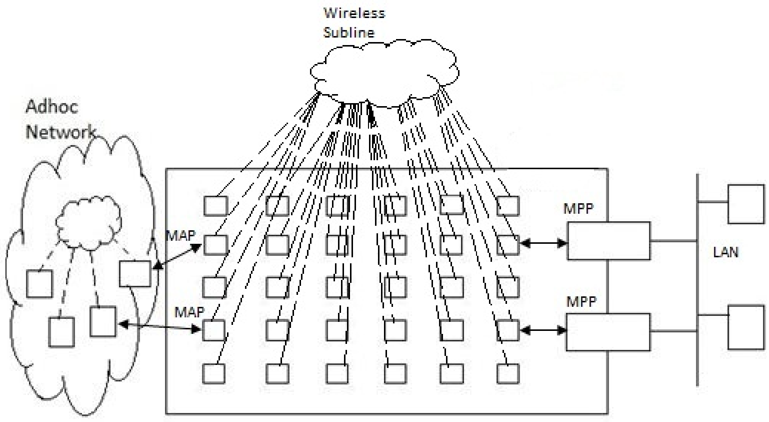


Fig.1

Mesh Points/Mesh Routers (MP) acts as backbone routers for forwarding the traffic from source to destination. Most of the functionalities are implemented in Mesh Points. Mesh points act as a router forwarding a packet from source to destination. Mesh Points are provided with multiple interfaces to accept the requests from various types of Wired/Wireless stations, Mesh Access Points (MAP) and Mesh Portal Points (MPP). Mesh Access Points (MAP) are stations situated in between mixed type of stations and Mesh Routers. MAP is responsible for providing interface to 802.11 stations and Mesh Points. Mesh Portal Points (MPP) act as an intermediate node between Mesh Points and 802.3 networks. MPPs serve as gateways for non mesh network. Mesh Stations are devices, which take part in exchanging of messages to/from wireless mesh networks. Mesh stations are equipped with NIC card and protocol stack and can connect to the mesh routers. Mesh Stations can communicate with mesh network through Mesh Routers, MAP or Mesh Portal Point.

The rest of the paper is organized as shown below: Sector II gives the working of HWMP Protocol, Division III gives the related work, Section IV sheds light on proposed routing mechanism, Algorithm and Mathematical model for proposed mechanism, Section V gives a brief overview of Routing metric used by HWMP protocol, and Section VI discusses Simulation Parameters, Results and discussions.

II. HYBRID WIRELESS MESH PROTOCOL

A. Route Discovery Process

The default route selection protocol for Wireless Mesh Networks is HWMP [2]. Path selection is nothing but finding a way to reach a given destination. The source that needs to find a route to the destination uses route discovery process. The source sends a Path Request (PREQ) frames to all their neighbors. PREQ frame is attached with Source MAC address, Destination MAC address, Path discovery ID, Time to Live (TTL), Hop count to originator, Metric and Target count and other fields [1]. If any of the neighboring nodes is a target address, it replies with a Path Reply (PREP) Frame else forwards the PREQ frames to their neighbors. This operation continues till it out reach the destination. While forwarding the PREQ frames to the next nodes, each node decrements the TTL, modifies hop count and metric to determine hop count and metric to the originator. Each station also discards PREQ frame, if the same frame reaches the nodes again. Each station uses the sequence numbers to discard the redundant frames. Once the PREQ frame reaches the destination, the destination learns the better hop count to originator. Then the destination sends PREP Frame to the originator through nearest route calculated; based on least hop count or metric. The PREP frame consists of Target MAC address, Hop count to the Target, Metric, TTL and other fields. The originator uses this path to forward the data frames. The Process of route discovery process is shown in figure 2.

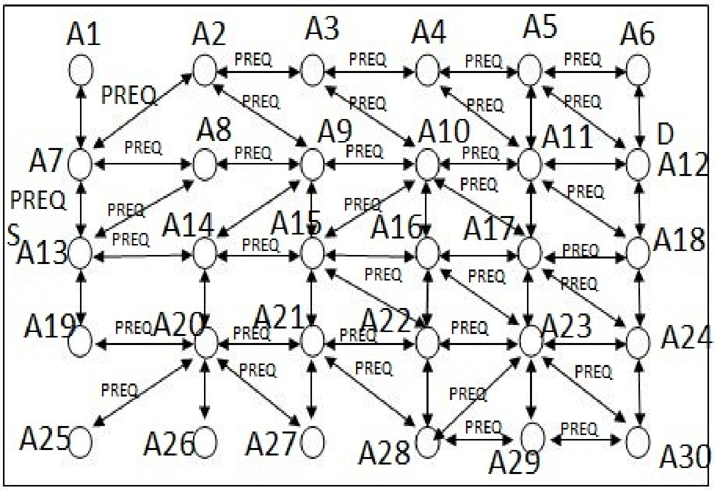


Fig.2: PREQ packets from source node.

Initially Source node A13 sends PREQ request to their neighboring nodes. These neighbor nodes forward the PREQ frames until it reaches destination node A12. When node A12 receives PREQ frames,it finds the shortest route towards source and PREP will be sent back towards A13.The PREP sent by the destination node is shown in fig 3. The route formed by this process is A13-A8-A9-A10-A11-A12.

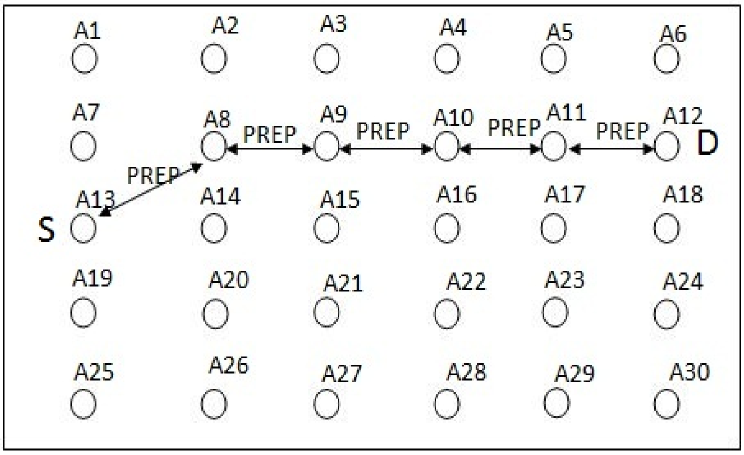


Fig.3: PREP packets from destination node

B. Route Maintenance in HWMP

When error occurs in nodes or links due to occurrence of faults; sends RERR frame to precursor nodes. Precursor nodes maintain the route information related to path. After receiving the Route Error (ERR)/ Path Error (PERR) frames each node checks and invalidates the path, which is affected by errors; and forward RERR frames towards source. The source also invalidates the existing path towards destination and reinitiates path discovery process as shown in figure 2.

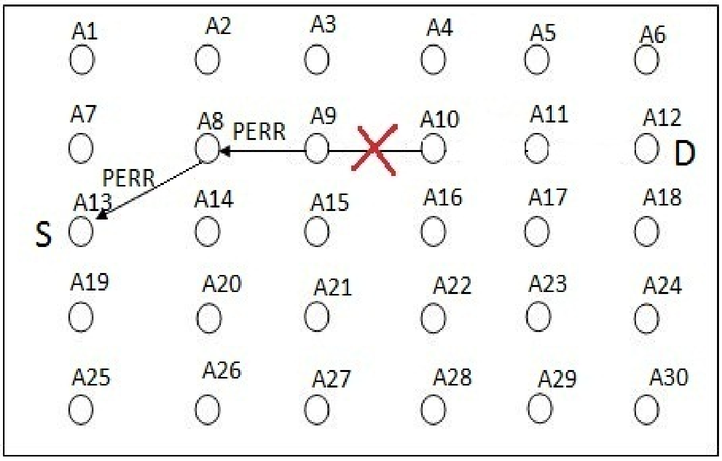


Fig.4: RERR from immediate neighbor node

As shown in fig 4, when the link between the nodes A9 and A10 is broken due to faults; usually it broadcasts the RERR frames to their neighbors. RERR frame are forwarded until they reach source node. Even though RERR contains minimum fields, it causes more overhead in a large network.

Here considering this problem; solution is worked out to avoid too much overhead. When node receives RERR frames, it unicast RERR frames to source; instead broadcasting. As shown in figure 3, the origin node begins route discovery process, when it receives RERR frames from node A8.

III. RELATED WORK

Some Researchers have proposed self-healing mechanisms for Wireless Mesh Networks.

S Menaka and M K Jayanthi [4] shed light on construction of stable route for Mobile Ad hoc network in the event of a link/node failure due to the movement of nodes. They proposed a new idea of finding a stable route using a cross layered routing metric. They stated that once the link/node is failed, the connectivity would take time to resume therefore steadiness of way is important in routing. Hence the authors considered stability factor and Received Signal Strength of a link as a metric. Here the researcher has found the principle that prevention is better than cure i.e. even before the connectivity is disconnected, routing scheme gracefully selects an alternate route. The researchers have considered Dynamic Source Routing (DSR) protocol to design a new approach.

Michel Bahr [5] describes his proposed routing for IEEE 802.11s mesh network. IEEE 802.11s standard supports multi hop routing. The Bahr revealed new mesh data frame format for 802.11s standard. He also defined the format of frames and working of HWMP. The routing in Mesh Network is different from routing in Ad hoc network. In WMN, the nodes/routers/Mesh Points are static in nature and in Ad hoc network the stations are mobile in nature. WMN acts as a backbone for Wireless LAN Access points. Ad hoc network uses different routing protocols such as Ad hoc on demand Distance Vector Routing (AODV) protocol and Dynamic Source Routing (DSR). Since the architecture of Mesh wireless network is different from Ad hoc Network, the protocols designed for Mobile Ad Hoc network may not be suitable for WMN. At present, Wireless Mesh Network using HWMP as a standard routing protocol, which uses MAC address for finding a shortest path; whereas AODV and DSR uses IP addresses for finding a shortest path.

Kishwer Abdul Khaliq . [8] Proposed a congestion avoidance technique. Here modifications are done to HWMP protocol for IEEE 802.11s for avoiding the congestion. This proposed mechanism offers localized rerouting, which minimizes overhead. This mechanism works similar to HWMP protocol but checks for threshold queue level. The proposed mechanism considers Queue level to detect congestion. Queue level is set to some threshold value. When queue level is below a threshold rate, it onwards the PREQ packet to another nodes. When Queue size reaches threshold value the neighbor node sends Congestion Control Notification Frame (CCNF) to its neighbors. These neighbors send PREQ (Path Request) frame to discover a new route to destination. Once PREQ frames reaches destination, sends PREP frame to source to form an alternate Route. Khaliq also shows that his routing mechanism has better performance compared to HWMP protocol, but there is a chance of more overhead due to sending of CCNF frames, so it may degrade the scalability features.

IV. PROPOSED ROUTING MECHANISM

In the proposed routing mechanism, when the node receives RERR frame from the node which is affected by the fault, normally it broadcast RERR frame to their intermediate nodes. All the neighbor-nodes which receive RERR frame, they forward RERR frame until they reach the Source. Then source reinitiates route discovery process to find the way from starting node to goal node.

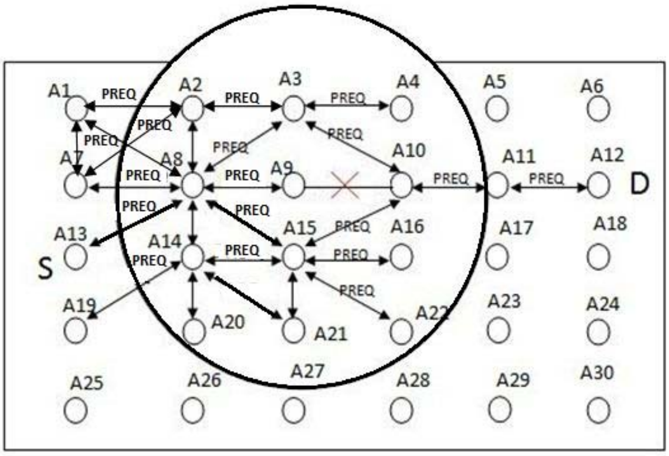


Fig.5: PREQ packets after applying LRR-HWMP.

In this proposed mechanism, the node which receives RERR frame is responsible for finding a new route towards destination instead of source node. This particular node acts as a origin node and goal node is kept the same. Once it receives RERR frame; it triggers path inventing process by sending PREQ frames to their neighbor nodes. Here instead of sending the PREQ packets to entire network, it limits the journey of PREQ packets; the PREQ packets are broadcasted only to few number of hops (2 hops) which are assigned initially as shown in figure 5.

A. Methodology

In case of failure, only the node which has the path towards the source receives RERR frame i.e. the RERR frame is unicast towards the source node. This node is responsible for finding the alternate path towards destination. So make the node which received the RERR as a source (temporary source node) and next node to error node towards destination as temporary destination node. The temporary source node is in charge of discovering a path to temporary destination node and hence broadcast the PREQ frames with limited hops to their neighbors. The PREQ frames are forwarded until it reaches temporary destination node. The temporary destination node sends a PREP frames to the temporary source node along with minimum hops or better metric. This proposed mechanism avoids forming a path twice from source node to temporary source node and temporary destination node to destination node; since path is already available from source node to temporary source node and temporary source node to temporary destination node i.e. A13- A8 and A11-A12. So in this paper efforts are made to repair the route locally.

B. Algorithm

- Let Ns is the sender, N_d is the destination, Ni is the intermediate nodes where i=1,2,3...
- N_{iu} (A8) is the immediate upstream intermediate node which receives Error; N_{ie} (A9) is the Error node which sends error.
- N_{id} (A10) is the immediate downstream node to error node.
- N_{idn}(A11) is the next immediate downstream node from error node

If path between N_{ie} (A9) and Nid (A10) is broken

Nie (A9): Send (RERR frame)

N_{iu} (A8): Receive (RERR frame)

Number of hops = 2

Temporary source = Niu (A8)

Temporary Destination= Nid (A10)

Niu (A8): Send (PREQ Frames)
Nid (A10): Receive (PREQ frames)
A10 sends PREP Frame to A8 via Different path
PREP Frame consists of shortest route based on Metric

- Locally route is Formed and unicast the route to A8 and A10
- Join the local route to the main route so that it can again form route A13-A8-A9-A10-A11-A12.

C. Mathematical Model

a) Existing Routing Mechanism

Normally Total Time taken for processing a Frame along the network from source to destination by HWMP protocol in the first iteration is given by.

$$\text{Total Time} = \sum_{i=1}^{n} t_{mi} + t_{pi} \quad \ldots (1)$$

Where n is the total nodes, t_{mi} is the time taken to calculate metric value and t_{pi} is the time taken to calculate remaining part of the frame. The above equation can be written corresponding to the figure 2 is given by

$$\text{Total Time} = \sum_{i=1}^{A8} t_{mi+} t_{pi} + \sum_{i=A8}^{A11} t_{mi+} t_{pi} + \sum_{i=A11}^{A12} t_{mi+} t_{pi} \quad \ldots (2)$$

Where A8, A11, A12 are the intermediate nodes from source towards destination. t_{mi} and t_{pi} are time taken to calculate the metric and remaining frame respectively.

Total Time taken for processing a Frame along the network from source to destination by HWMP protocol in the first iteration is same as equation 2. When the Link is broken between two nodes it reinitiates route discovery process than also the time taken by these nodes is same as in Equation 2.

b) Proposed Routing Mechanism

According to Proposed routing mechanism Total Time taken for processing a Frame along the network from source to destination by HWMP protocol in the first iteration is same as equation 2. But when the link is broken between two nodes, locally route will be repaired. So the Time taken for HWMP protocol in the second iteration is given by

$$\text{Total Time} = \sum_{i=A8}^{A11} t_{mi} + t_{pi} \quad \ldots (3)$$

V. ROUTING METRIC

Routing metrics plays an important role in judging the cost or efficiency of a route. In other words the design of routing metrics helps to identify the capable path to the destination in a network . Basic metrics are designed based on hop count, are not efficient for selecting the efficient path and some other parameters are also need to be considered. The design of a routing metric is a very challenging research problem because it helps to capture various parameters related to route stability, Asymmetry of wireless links, choosing optimum path and to avoid loops in routing, basic load interference and Quality of Service parameters. In this default routing metric called Air Time Link Metric (ALM) is considered as default routing metric for calculating a path from source to destination. The ALM metric is calculated by using following equation.

$$\mathbf{Ca= [O+Bt/r]*(1/(1-e_f))}$$

Where O and Bt are constants, r is the rate of transmission, ef is the frame error rate and Bt is the size of a frame.

VI. SIMULATION PARAMETERS

Parameters	Values
Number of nodes	16, 20, 24, 28, 32
Terrain	1500*1500
MAC layer	802.11s
Protocol	HWMP, SLRR-HWMP
Simulation Time	300 sec
Application	Constant Bit Rate
Packet Size	512 bytes
Packet Rate	2mbps
Transmission Range	150m
Topology	Random

Simulation results and Discussion

We have used Qualnet tool for measuring the performance of HWMP routing protocol. Qualnet is Proprietary tool which provides very good support for implementing a new routing mechanism. The experiment has been carried out in windows Operating system environment. Constant Bit Rate (CBR) application is used to create traffic in network. Random topology is used here to place the nodes in a network.

Simulations have been done by considering the below parameters:

Packet Delivery Ratio (PDR): Ratio of the total packets Received by destination to the Number of packets Sent by starting node.

End to End Packet Delay (E to E Delay): is defined as the Time used by the data packets to reach destination.

Control Overhead (CO): Specifies the total control packets used by routing protocol to route a packet from origin node to goal node.

- *Scenario A*

In scenario A, we have considered number of nodes and simulated for different parameters by considering number of nodes. Here the HWMP performance is evaluated by considering the normal execution of HWMP protocol and by considering broken route from onset node and finishing node. The same scenarios are considered and evaluated by considering our proposed mechanism named as SLRR-HWMP protocol. The results are calculated for Packet Delivery ratio, End to End delay and Control Overhead.

The figure 6 indicates the throughput when the nodes use HWMP protocol in the normal mode i.e. HWMP (N), and by introducing fault i.e. HWMP (D) and our proposed mechanism called SLRR-HWMP. Here we can observe that as the node number increases, more packets will be dropped; in turn reduces throughput. This situation is worse in the case of introducing the fault. This graph shows slight improvement for the proposed approach i.e SLRR-HWMP over HWMP (D).

The figure 7 indicates the Average End to End delay, when the nodes use HWMP protocol in the normal mode i.e. HWMP (N), and by introducing fault i.e. HWMP (D) and our proposed mechanism called SLRR-HWMP. Here we can observe that that as the node number increases, the average End to End delay increases when introducing a fault i.e. in HWMP (D). Because of path failure and time to repair will be more compared to HWMP (N).But in the case of SLRR-HWMP protocol, the average End to End delays decreases with respect to HWMP (D).

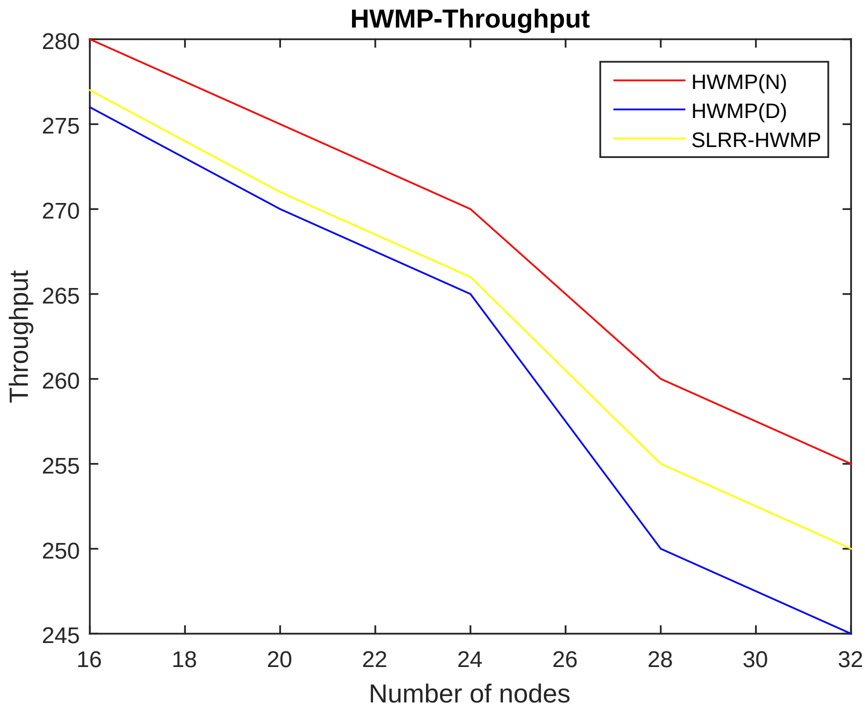


Fig.6: HWMP-Throughput

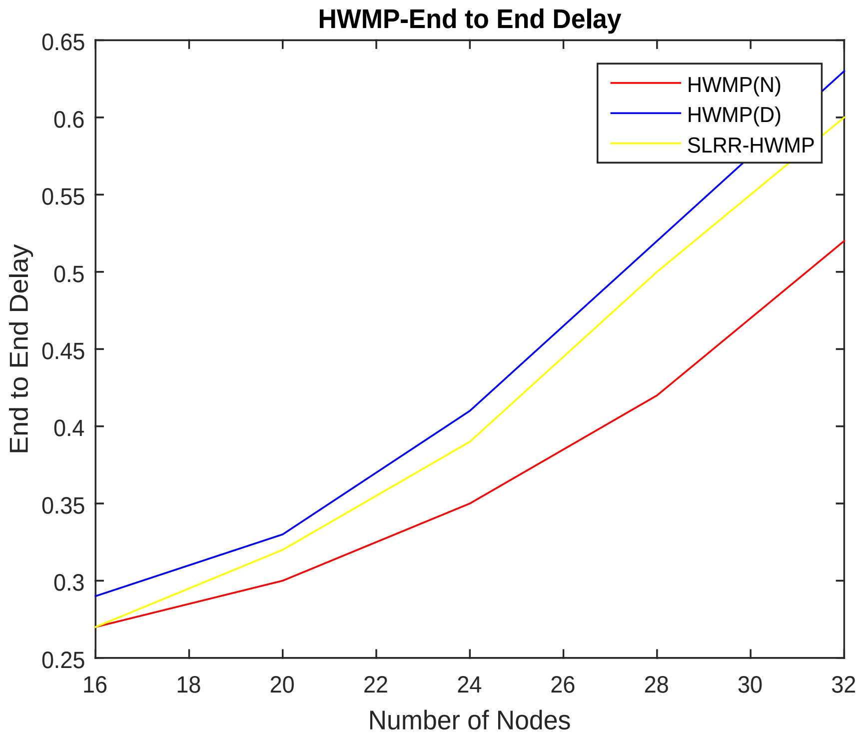


Fig.7: HWMP-End to End delay

The figure 8 indicates the Control Overhead, when the nodes use HWMP protocol in the normal mode i.e HWMP (N), and by introducing fault i.e HWMP (D) and our proposed mechanism called SLRR-HWMP.

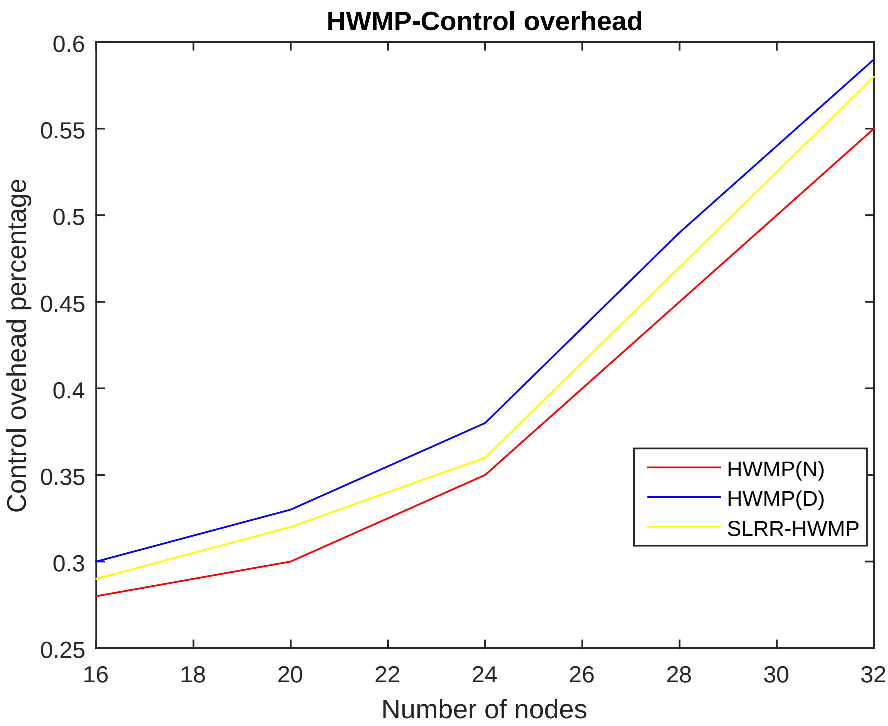


Fig.8: HWMP-Control overhead

The figure 8 also shows that as the node number increases, the control overhead increases when introducing a fault i.e. in HWMP (D). Because of path failure and control packets for discovering a route will be more compared to HWMP (N). But in the case of SLRR-HWMP protocol, the control overhead decreases when compared to HWMP (D).

- *Scenario B*

In scenario B, we have considered multiple inputs and multiple outputs and evaluated for different parameters. Here the HWMP efficiency is evaluated by considering the normal execution of HWMP protocol and by considering broken route from source node and destination node. The same scenarios are considered and evaluated by considering our proposed mechanism named as SLRR-HWMP protocol. The results are calculated for Packet Delivery ratio, End to End delay and Control Overhead.

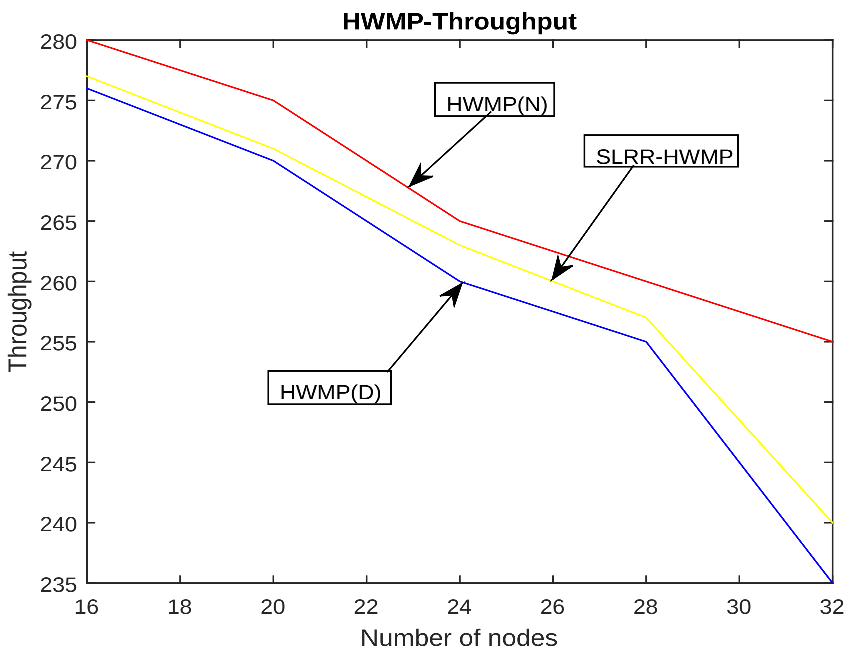


Fig.9: HWMP-Throughput (Multiple inputs and multiple outputs).

The figure 9 unveils the throughput when the nodes use HWMP protocol in the normal mode i.e. HWMP (N), and by introducing fault i.e HWMP (D) and our proposed mechanism called SLRR-HWMP in the case of multiple input and multiple output. Here we can observe the sudden packet loss rate, as the node number increases; which in turn reduces throughput. This situation is worse in the case of introducing the fault. This graph shows slight improvement in the throughput for the proposed approach i.e. SLRR-HWMP over HWMP (D).

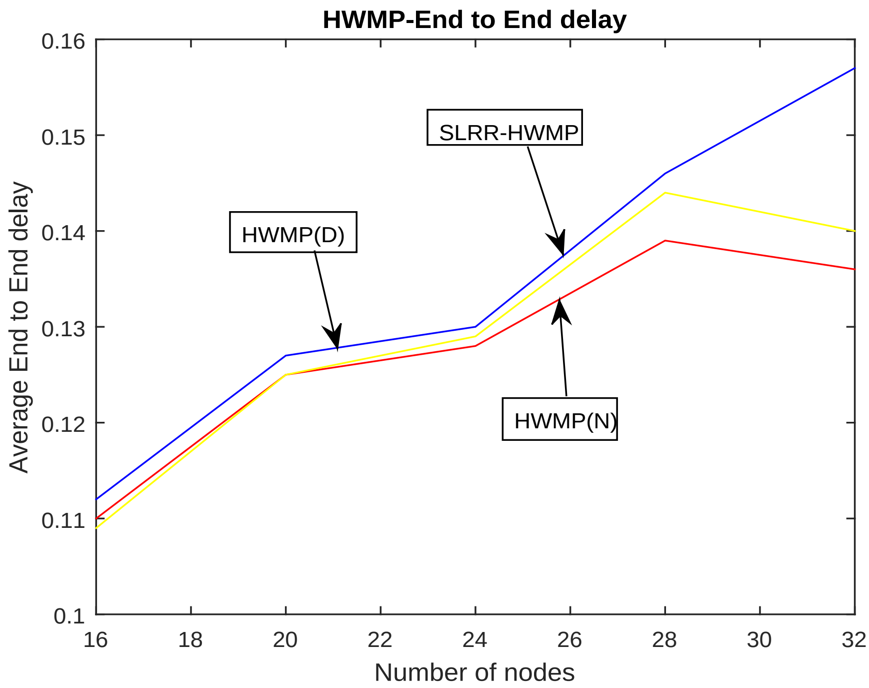


Fig.10: HWMP-End to End Delay (Multiple inputs and multiple outputs).

The figure 10 discloses the Average End to End delay, when the nodes use HWMP protocol in the normal mode i.e. HWMP (N), and by introducing fault i.e. HWMP (D) and our proposed mechanism called SLRR-HWMP. Here we can observe that that as the node number increases, the average End to End delay increases when introducing a fault i.e. in HWMP (D). Because of path failure and time to repair will be more compared to HWMP (N). But in the case of SLRR-HWMP protocol, the average End to End delay decreases compared to HWMP (D).

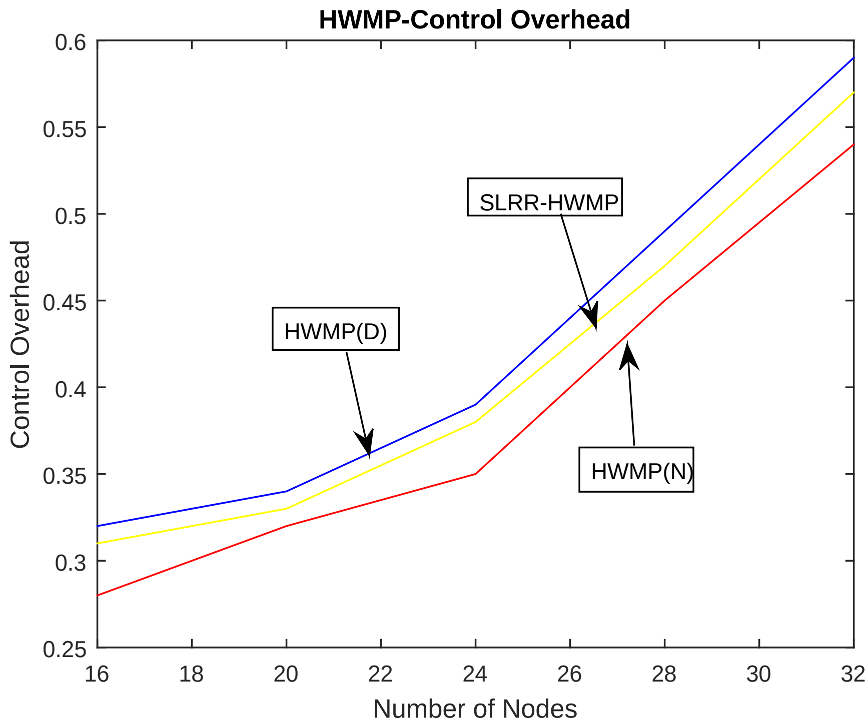


Fig.11: HWMP-Control overhead (Multiple inputs and multiple outputs).

The figure 11 represents that, as the number of nodes increases, the control overhead increases. This situation worsens when introducing a fault i.e. in HWMP (D). Because of path failure and control packets for discovering a route will be more compared to HWMP (N). But in the case of SLRR-HWMP protocol, the control overhead decreases when compared to HWMP (D).

The results reveal that, after applying proposed routing mechanism (SLRR-HWMP), results are improved slightly in terms of Packet Delivery ratio, End to End delay and control overhead since we have considered only a few numbers of nodes as shown in graph 3.

In the proposed routing mechanism, the default values of initial RREQ, RREQ TTL Increment, RREQ TTL threshold and mesh diameter values are limited to Minimum values. So when any node receive RERR packet from the neighbor node, it initiates path discovery process by calling Expanding Ring Search algorithm (with Mesh Diameter=2) and hence it sends RREQ packets only to their two hop neighbors and route is formed locally instead starting from the source node. This mechanism slightly reduces number of control packets and hence reduces overhead.

Since SLRR-HWMP limits the number of PREQ packets used for route formation; LRR-HWMP repairs the route locally. LRR-HWMP protocol shows better accomplishment in terms of Packet Delivery Ratio, Routing Overhead and Average E to E Delay. Especially for a large network; this proposed routing mechanism is a boon. Since it minimizes number of PREQ packets.

CONCLUSION

Wireless Mesh Networks getting popular because of flexible architecture for providing Internet. WMN is an extension of an existing Wireless Local area network designed to provide scalability. Even though HWMP protocol is the default routing protocol for 802.11s (Wireless Mesh Network), there are many challenges related to routing and remain unsolved. This is an attempt to design and evaluate a new routing mechanism called Local Route Repair mechanism for HWMP protocol called SLRR-HWMP. Fresh endeavor of research in this paper presents local route repairing technique; it considers only the neighboring nodes to repair a route locally and hence minimizes overhead, Time and cost. Under this mechanism route is repaired locally and avoids unnecessary control overhead and thereby increasing the scalability. The results also reveal better performance compared to existing protocol. It is evident from our newly attempted work of research output that SLRR-HWMP routing mechanism will be a boon for a large network.

REFERENCES

[1] Akyildiz, I F. Wang, and Wang W. “Wireless Mesh Networks: a survey”, Computer Networks 47 (2005) pp.445-487,jan 2005.
[2] Ling He,Jun Huang, Feng Yang, ―A Noval Hybrid Wireless Routing Protocol for WMN’s‖, International Conference on Electronics and Information Engineering,2010.
[3] Riduan M Abid, Taha BenBrahim, Saad Biaz, “IEEE 802.11s Wireless Mesh Networks for Last-mile Internet Access: An Open Source Real-World indoor Testbed Implementation, Technical report #CSSE10-02.
[4] S Menaka and M K Jayanthi, “Adaptive and self healing Routing for Mobile Adhoc Networks using cross Layer Design”, Journal of Computing and Information Technology-CIT 22,2014,4,227-236.
[5] Michael Bahr, “Proposed Routing for IEEE 802.11s WLAN Mesh Networks”.
[6] Md Asri Bin Ngadi, Saquib Ali, Abdul Hanan Abdullah and Rashid Hafeez Khokhar, ―A Taxonomy of Cross Layer routing metrics for Wireless Mesh neteorks‖, EURASIP journal on Wireless Communication and Networking,2012.
[7] Pranjal Pandey, SS Sathish, Joy Kuri,Haresh Dagale, “Design and Implementation of IEEE 802.11s mesh nodes with Enhanced features”
[8] Rosy, shwani Kush , “Self-Healing in Wireless Routing Using Neighbour nodes”, IJACT.
[9] Kishwar Abdul Khaliq, Muhammad Sajjad Akbar, Amir Qayyum, Ehsan Elahib, Ameer Zaheer,” Congestion Avoidance Hybrid Wireless Mesh Protocol (CA-HWMP) for IEEE 802.11s”, 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014), Procedia Computer Science 32 (2014) 229 – 236.
[10] Joseph D. Camp and Edward W Knightly, “The IEEE 802.11s Extended Service Set Mesh Networking Standard.
[11] Incheon Park, Jinguk Kim, Ida Pu,”Blocking Expanding Ring Search Algorithm for efficient Energy Consumption in Mobile Ad Hoc Networks”
[12] Shwetha Mishra, Jyoti Singh, Arti Dixit, Shiva Prakash, “Modified Expanding Ring Search Algorithm for Ad Hoc Networks”, International Journal of Computer Science and Information Technologies (IJCSIT),Vol 3(3),2012,4067-4070.
[13] Purvi N Ramanuj, Hiteishi M Diwanji, “Enhanced Expanding Ring Search Algorithm for AODV”, Special Isuue of International Journal of Computer Applications (0975-8887) on Wireless Communication and Mobile Networks,no.13,Jan 2012.

International Journal of Computer Engineering & Technology (IJCET)
Volume 10, Issue 1, January-February 2019, pp. 1-10, Article ID: IJCET_10_01_001
Available online at
http://www.iaeme.com/ijcet/issues.asp?JType=IJCET&VType=10&IType=1
Journal Impact Factor (2016): 9.3590(Calculated by GISI) www.jifactor.com
ISSN Print: 0976-6367 and ISSN Online: 0976–6375

BUFFER BASED ROUTING MECHANISM FOR LOAD BALANCING IN WIRELESS MESH NETWORKS

Keerthi D S, Shobha Rani A and T G Basavaraju
Department of Electronics and Communication Engineering, MCE, Hassan, India,
Department of Electronics and Communication Engineering, BMSIT, Bengaluru, India
Department of Computer Science and Engineering, Govt. SKSJTI, Bengaluru, India

ABSTRACT

In recent years it is witnessed that the Wireless Mesh Networks (WMNs) are becoming the most promising technology as they offer low cost broadband wireless connectivity, larger coverage area, high flexibility and easy deployment. WMNs are an extension of existing wireless networks. WMN is an emerging technology; however, there are certain challenges that still exist in the network such as scalability, load balancing, mobility, power management etc. Here we have proposed a novel routing protocol which considers buffer occupancy of intermediate nodes for route selection. Simulation results convey that the proposed protocol outstandingly enhances the performance of the network by balancing the traffic load among less congested nodes compared to the standard protocol.

Keywords: Buffer, Load balancing, WMN.

Cite this Article: Keerthi D S, Shobha Rani A and T G Basavaraju, Buffer Based Routing Mechanism for Load Balancing in Wireless Mesh Networks. International Journal of Computer Engineering and Technology, 10(1), 2019, pp. 1-10.
http://www.iaeme.com/IJCET/issues.asp?JType=IJCET&VType=10&IType=1

1. INTRODUCTION

WMNs are one of the most favorable wireless technologies which find several applications, e.g., broadband home networking, enterprise networking, community and neighborhood networks, building automation etc. They are an extension of wireless ad-hoc networks and they aim at delivering a cost effective and robust, high bandwidth network over a particular coverage area. WMNs [1] are dynamically self-healing, self-organized and self-configured as the nodes in the network automatically establish and maintain mesh connectivity among themselves which in turn provide benefits such as effortless network maintenance, low up-front costs and dynamic and reliable service coverage. A more sophisticated routing protocol along with an efficient routing metric can be used to enhance the performance of the WMNs. WMNs employ a number of routing protocols where every protocol maintains its own routing strategy [2, 3].

These protocols are categorized as proactive, reactive and hybrid based on the strategies adopted by them.

Load Balancing is a method of evenly distributing the traffic across two or more nodes for effectively mediating communication and achieving redundancy even if there is any link failure. Load balancing can be achieved by increased throughput, minimal overheads and optimal utilization of resources. Whenever the routes are exploited repeatedly over a longer duration, the packet concentration elevates in intermediate nodes. This creates bottlenecks and thereby degrades the network performance due to congestion and also results in longer delays. The caching technique used in most of the on-demand routing protocols causes additional mass of load on certain nodes. In WMN, congestion is the major cause for the uneven distribution of traffic that causes underperformance of the network. Congestion is a condition where multiple packets are present in network exceeding existing resource available and thereby deteriorating the performance of the network. These packets completely fill up the buffer space and possibly overflow, manifesting in excess end to end delay, packet drop and low network throughput and degraded bandwidth.

The organization of the paper is as follows. The related work is discussed in Section 2. The proposed routing scheme BBL-AODV is completely described in section 3. The simulation results, performance comparison and analysis are presented in section 4. The inference of the paper is in section 5.

2. RELATED WORK

Many protocols exist for load balancing in wireless networks. Most of them have considered parameters like link occupancy, queue size, delay etc. as a metric for route selection. In [4], the authors have proposed a buffer aware routing protocol which takes buffer space of neighbouring nodes into account while taking routing decision. Even though the protocol performs well it introduces more delay while performing route discovery.

In [5], the authors present a congestion control approach by advertising the buffer space to the surrounding neighbour nodes. This approach is independent of the routing protocol but it needs additional mechanisms when inappropriate traffic is generated in the network. Congestion avoidance mechanism [6] deals with high load and fast mobility, which selects the best path based on minimum value of buffer queue occupancy instead of minimum hop count as in standard AODV.

Kruti N Kapadia et al. [7] have put forth an airtime congestion aware metric for a multi-radio WMN. It uses airtime link metric for estimating RTT for delay calculation and channel utilization. With the combination of both parameters, load is uniformly distributed over less congested nodes. Raffaele Bruno et al. [8] have developed an algorithm based on queuing model for detection of congested paths in heterogeneous mesh networks. The algorithm is implemented at the nodes with the objective of forwarding traffic to less congested gateways based on their queue occupancy.

Dhurandher et al [9] have presented a novel protocol for distributed admission control and interference aware admission control for WMN with the aim of providing QoS guarantee for multimedia applications. The protocol calculates suitable admission control ratio that ensures loss rate within the tolerable limit. In addition to this it also computes end to end delay to exercise flow request. ECARP [10] is proposed for heavy traffic loads in MANETs and this protocol considers the number of packets in node buffer to regulate congestion among nodes. CARM [11] prefers less congested, high throughput links for traffic by employing weighted delay, retransmission count, and queuing delay in buffer.

In our work we have proposed a modified AODV protocol called Buffer Based Load Balancing Adhoc on Demand Protocol (BBL-AODV), which considers node buffer occupancy as well as average buffer occupancy of all neighbouring nodes as a routing metric during route discovery.

3. BUFFER BASED LOAD BALANCING AODV PROTOCOL

Buffer Based Load Balancing AODV Protocol (BBL-AODV) is an enhancement over the standard AODV protocol. It uses a path metric which allows source node to select the best path to the destination based on buffer occupancy of intermediate nodes. The standard AODV protocol discovers the route by broadcasting/flooding RREQ message and establishes a path when RREP message is received. The path selection is based on minimum hop count and sequence number irrespective of packet handling capacity. In Buffer Based Load Balancing (BBL) Protocol, each node maintains information about its node buffer occupancy in its routing table. Each node calculates its average buffer occupancy *avg_buff_occ* based on node buffer occupancy *node_buff_occ* of all neighboring nodes and itself. The node is assigned with a threshold value *thr* for optimum buffer utilization.This threshold value defines buffer space that a node requires to forward received packets. Whenever a node receives RREQ packet, it compares its current node buffer occupancy with threshold *thr*. If the current buffer occupancy is less than threshold *(node_buff_occ< thr*), the node broadcasts RREQ packets or else it discards. Therefore, load is distributed along the less congested path and this enhances the performance of the wireless mesh network.

The average buffer occupancy is calculated as follows:

$$\text{avg\_buff\_occ} = \frac{\text{node\_buff\_occ} + \sum_{j=1}^{n} \text{node\_buff\_occ}}{n+1} \qquad (1)$$

where ‘n’ represents total neighboring nodes.

The node_buff_occ represents the buffer occupancy of the node which is given by the following equation.

$$node\_buff\_occ = \frac{\sum_{i=1}^{b} buff_{i}\_pkt}{b} \qquad (2)$$

where, $buff_i$ _ pkt represent total number of packets in every buffer present.

‘b’ refers to the number of buffers in each node.

The BBL-AODV Algorithm for Route Selection is described as follows.

Step 1: Node has data to send

Step 2: Check the active nodes for transmission // active refers to a node already participating in

data transfer

Step 3: If (node= =active)

then go to step 4

else go to step 7

end if

Step 4: Calculate *node_buff_occ* of each intermediate node.

Step 5: If node_buff_occ<thr,

then forward the packet

else rebroadcast RREQ

end if

Step 6: Broadcast RREQ and wait for RREP

Step 7: Receive RREP go to step 4

4. PERFORMANCE EVALUATION

This section discusses parameters and performance metrics chosen for simulation of the proposed BBL-AODV protocol with standard AODV protocol. The performance of the proposed routing mechanism is evaluated by carrying out simulations using the licensed network simulator Qualnet v6.1.

4.1. Simulation Parameters

We have done substantial number of simulations for studying the performance and effectiveness of the suggested BBL-AODV against the standard AODV protocol in a wireless mesh network. The simulation is run for 300 seconds on a grid topology of simulation area 1500 x 1500 m^2at lower and higher data rates for different Node densities. The nodes are placed in the simulation area in a uniform fashion with 150 m as inter-node distance. The traffic source selected is Constant Bit Rate traffic with each packet size being set to 512 bytes. The Table 1 below describes the simulation parameters employed for performance analysis.

Table 1 Simulation Parameters

Parameter	Value
Protocols	AODV, BBL-AODV
No. of Nodes	25, 36, 49, 64, 81,121
Radio type	802.11b Radio
MAC Protocol	802.11s
Antenna Model	Omni-directional
Path Loss Model	Two ray propagation
Traffic Type	CBR
No. of CBR	3

4.2. Performance Metrics

The effectiveness of the proposed and the standard protocol is compared using the performance metrics like throughput, total number of packets received, end to end delay and jitter.

- *Throughput:* The rate at which message is delivered successfully over a communication channel and is measured in bits per second.
- *Total number of packets received:* It is the number of packets received at the destination.
- *End-to-end delay:* It is the time taken by a packet to traverse from source to destination.
- *Jitter:* It is measuring time difference in packet inter-arrival time.

4.3. Simulation Results

In this paper, the results are presented for different node densities at low and high data rates for both BBL-AODV and standard AODV protocol.

4.3.1. Low Data rate

In this scenario packets are forwarded through three CBR connections. Each source generates a low traffic of 1 packet/sec and is routed to the destination for a period of 300 sec. The above procedure is repeated for 36, 49, 64, 81 and 121 node densities respectively. After each simulation, the performance metrics such as throughput, delay, jitter and total messages received are recorded.

Figure.1-4 shows throughput, total packets received, end-to-end delay and jitter with varying node densities for BBL-AODV and AODV. It is observed that throughput remains almost same for BBL-AODV and standard AODV protocol. This is because the traffic generated in the network is within the buffer threshold limit for each node. So, both protocols use default routing strategy for path selection in the network. This is true even for total packets received, end-to-end delay and jitter respectively.

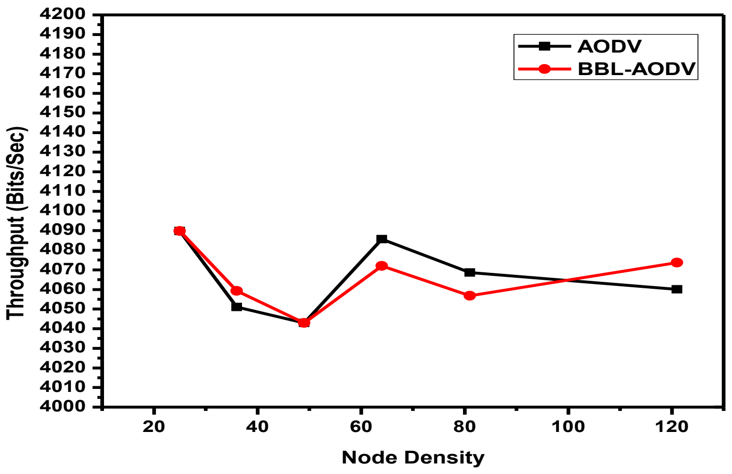


Figure 1 Variation of throughput v/s node densities

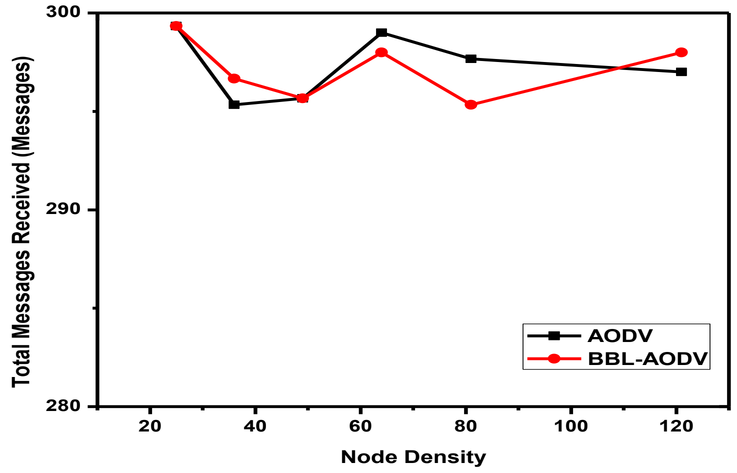


Figure 2 Variation of total messages received v/s node densities

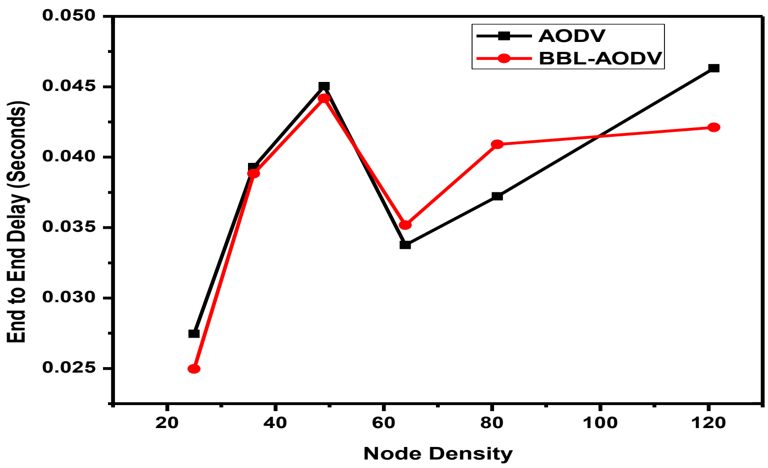


Figure 3 Variation of End to End Delay v/s node densities

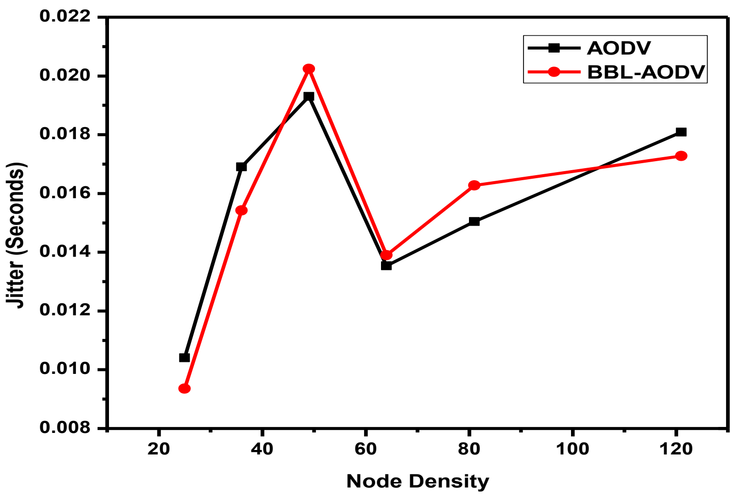


Figure 4 Variation of Jitter v/s node densities

4.3.2. High Data rate

In this section, the proposed protocol is compared with AODV protocol for a high data rate. The packets are forwarded through three CBR connections. Each source generates high traffic that is 100 packet/ sec and is routed to the destination for a period of 300 sec. The above procedure is repeated for 36, 49, 64, 81 and 121 node densities respectively. After each simulation the performance metrics such as throughput, delay, jitter and total messages received are recorded.

Figure 5 shows the variation of throughput with respect to the node densities for BBL-AODV and AODV protocol. It is observed that throughput for our proposed protocol is greater than AODV. This is due to the fact that when the load is very high, AODV failed to handle congestion thus resulting in degradation of the performance of the network. But the performance of the BBL-AODV is good since it uses the knowledge of buffer availability of each node during route discovery and avoids packet forwarding through the highly congested nodes.

Figure 6 depicts the total message received v/s node density. It is evident that the number of packets received at the destination is more in BBL-AODV as compared to AODV. This is due to the fact that the packet drop ratio is less in BBL-AODV as it avoids nodes having buffer occupancy greater than threshold which in turn increases the packet reception at the destination.

In Figure 7 it can be observed that the end-to-end delay in case of BBL-AODV is less than that of AODV. As traffic flow increases, AODV will experience higher delay due to congestion. Therefore, data packets take more time to reach the destination. BBL-AODV captures the link quality by monitoring buffer utilization. It minimizes congestion along the path by computing buffer occupancy based on the proposed metric a*vg_buff_occ* and *node_buff_occ* thus nodes are not over flooded. This results in optimum utilization of all network resources thereby allowing data packets to take less time to reach destination. Hence end to end delay of BBL-AODV is lower than that of AODV. Since delay variation in BBL-AODV is less than AODV, it is suitable for real time applications. Hence BBL-AODV performs better than AODV protocol.

Figure 8 shows the jitter variation for BBL-AODV and AODV. It is observed that BBL_AODV experiences lesser jitter for increase in the node density and traffic rate when compared to AODV. Hence the proposed protocol outperforms the AODV protocol in terms of throughput, total number of packets received, end to end delay and jitter.

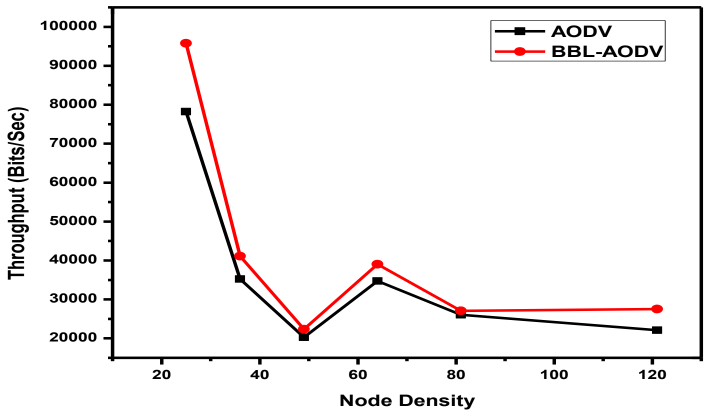


Figure 5 Variation of throughput v/s node densities

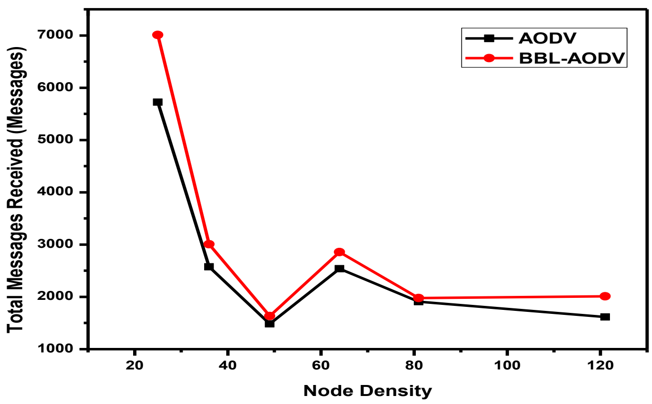


Figure.6. Variation of total messages received v/s node densities

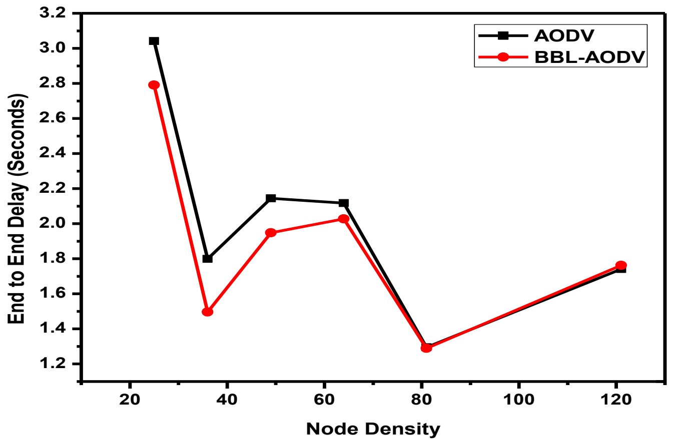


Figure 7 Variation of End to End Delay v/s node densities

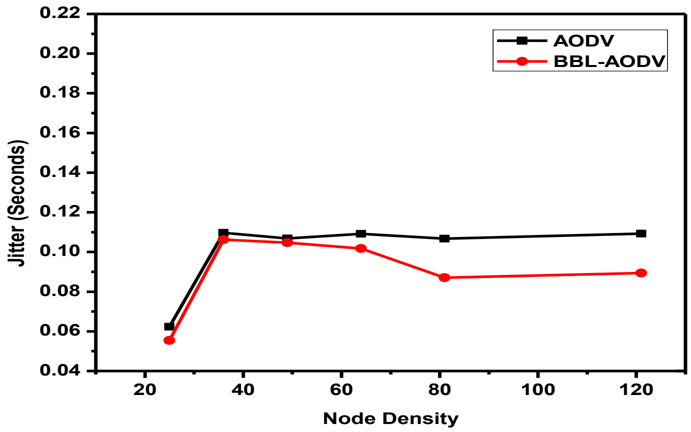


Figure 8 Variation of Jitter v/s node densities

5. CONCLUSION

In this work, we have proposed BBL-AODV; an effective buffer-based routing protocol for wireless mesh network. It captures link quality by monitoring buffer utilization. It minimizes congestion along the path by computing buffer occupancy based on the proposed routing metric. Simulations are done to analyze the proposed protocol's performance for different node densities with varying data rate. The metrics used to estimate the performance of the proposed protocol and standard protocol are throughput, delay, jitter and total messages received. From the results, it is evident that the proposed protocol outperforms the standard AODV protocol scheme with reference to the predefined performance metrics and hence results in optimum utilization of all network resources and increases the overall performance of the wireless mesh network.

ACKNOWLEDGEMENT

This work is supported by the Technical Education Quality Improvement Program Phase II with the World Bank assistance. The authors would like to thank State Project Facilitation Unit, Directorate of Technical Education, Government of Karnataka, National Project Implementation Unit, MHRD, New Delhi for their support and encouragement.

REFERENCES

[1] F. Akyildiz, X. Wang, and W. Wang, Wireless mesh networks: a survey, Computer Networks Journal, vol.47, 2005, pp.445-487.

[2] Eslami, M, Karimi, O, Khodadadi T, A survey on wireless mesh networks: Architecture, specifications and challenges" Control and System Graduate Research Colloquium (ICSGRC), 2014, pp. 219 – 222.

[3] M. L. Sichitiu, Wireless mesh networks: Opportunities and challenge, Proc. of the Wireless World Congress, (Palo Alto, Ca), May 2005.

[4] Kamal Mistry, Sanjay Srivastava, R. B. Lenin, Buffer aware routing in interplanetary Ad Hoc Network, First International Communication Systems and Networks and Workshops, 2009, pp. 1- 10.

[5] Jani Lakkakorpi, Mikko Pitkänen and Jörg Ott, using buffer space advertisements to avoid congestion in mobile opportunistic DTNs" Proceedings of the 9th IFIP TC 6 international conference on Wired/wireless internet communications, 2011, pp. 386-397.

[6] Maroufi B.E., Rahmani M.D., Rziza M. "Congestion Avoidance in AODV Routing Protocol Using Buffer Queue Occupancy and Hop Count", Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015, pp.193-200.

[7] Kruti N. Kapadia. Dayanand D. Ambawade "Congestion aware load balancing for multi-radio Wireless Mesh Network" International Conference on Communication, Information & Computing Technology (ICCICT) 2015.

[8] Raffaele Bruno, Marco Conti, Antonio Pinizzotto "A Queuing Modeling Approach for Load-Aware Route Selection in Heterogeneous Mesh Networks", IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks & Workshops 2009.

[9] Sanjay Kumar Dhurandher, Isaac Woungang, Mohammad S. Obaidat "A Distributed Adaptive Admission Control Scheme for Multimedia Wireless Mesh Networks" IEEE Systems Journal Volume: 9, Issue: 2, June 2015, pp. 595 – 604.

[10] T. G. Basavaraju, S. K. Sarkar, C. Puttamadappa, and M. A. Gautham, "ECARP: an efficient congestion adaptive routing protocol for mobile Ad hoc networks," in Proceedings of the 6th International Conference on ITS Telecommunications Proceedings, pp. 715–718, Chengdu, China, June 2006.

[11] X. Chen, H.M. Jones, and A. D. S. Jayalath, "Congestion-aware routing protocol for mobile ad hoc networks," in Proceedings of IEEE Conference of Vehicular Technology, pp. 21–25, October 2007.

[12] Keerthi D S, T.G. Basavaraju "Load Balancing Routing Mechanisms For Wireless Mesh Networks: A Survey", in Fourth International Conference on "Emerging Research In Computing, Information, Communication And Applications (ERCICA)", 2016. (Published in Springer proceedings)

[13] Sooyeol Yang,Youngmi Baek,Junhyung Kim,Keuchul Cho and Kijun Han "A routing metric for load balance in wireless mesh networks", 11th Intrrnational Conference on Advanced Communication Technology (ICACT) Volume: 03,2009,pp.1560-1565.

[14] Ekram Hossain, Kin K. Leung "Wireless Mesh Networks: Architectures and Protocols", Springer, 2008.

[15] Peipei Zhao, Jiansheng Qian "The Study of Wireless Mesh Network Routing Algorithm Based on Interference Model", Fourth International Conference on Genetic and Evolutionary Computing (ICGEC), 2010, pp.98-101.

[16] Benyamina D, Hafid A. and Gendreau M "Wireless Mesh Networks Design-A Survey" Communications Surveys &Tutorials, IEEE, Volume: 14, Issue: 2, 2012, pp. 299 - 310.

[17] Guerin J, Portmann M and Pirzada A "Routing metrics for multi-radio wireless mesh networks" Telecommunication Networks and Applications Conference (ATNAC), Australasian, 2000, pp.343-348.

[18] Bari S.M.S, Anwar F and Masud M.H. "Performance study of hybrid Wireless Mesh Protocol (HWMP) for IEEE 802.11s WLAN mesh networks" International Conference on Computer and Communication Engineering (ICCCE) ,2012, pp. 712 - 716.

BIOGRAPHY

Keerthi D S. received his B.E. and M.Tech degrees from Visvesvaraya Technological University (VTU), Belagavi. He is pursuing Ph.D. under the guidance of Dr. T. G. Basavaraju at the Department of Computer Science and Engineering Research Centre, GSKSJ Technological Institute affiliated to VTU, Belagavi. He is currently working as an Assistant Professor at Malnad College of Engineering, Hassan, Karnataka. His research interest is in the field of Wireless Mesh Networks.

Shobha Rani A. received her B. E Electronics and Communication Engineering) degree from Bangalore University in the year 2000, M. Tech (Digital Electronics and Communication) Degree from Visvesvaraya Technological University (VTU), Belgaum in the year 2005, and Ph. D (Networking) degree from Kuvempu University, Shimoga in the year, 2014. She is currently working as a Associate Professor in the Department of ECE at BMS Institute of Technology and Management, Bengaluru, India. She has 13 technical articles in good reputed journals and conferences. Her research interests include designing and implementing efficient protocols for wireless networks such as Ad-hoc, Sensor and Mesh networks.

T. G. Basavaraju has obtained his B.E degree in Computer Science and Engineering from Kuvempu University, Shimoga, Karnataka, M.E. degree in Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, Karnataka and Ph.D. degree from Jadavpur University, Kolkata. Currently, he is working as Professor and Head, Department of Computer Science and Engineering, Government SKSJ Technological Institute, Bangalore. He has more than 20 years of teaching and industrial experience. He has published several papers in national and international journals. He has also authored and co-authored several text books in the area of Computer Networking. His research interests are Wireless Mobile Ad hoc Networks, Wireless Sensor Networks and Wireless Mesh Networks.

Journal of Intelligent & Fuzzy Systems 37 (2019) 6773–6784
DOI:10.3233/JIFS-190339
IOS Press

Bilingual text detection in natural scene images using invariant moments

Karan Maheshwari[a,1], Alex Noel Joseph Raj[b,1,*], Vijayalakshmi G.V. Mahesh[c,1], Zhemin Zhuang[b], Elizabeth Rufus[a], Palaiahnakote Shivakumara[d] and Ganesh R. Naik[e]
[a] *Vellore Institute of Technology, Vellore, Tamil Nadu, India*
[b] *Department of Electronic Engineering, Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Shantou University, Shantou, China*
[c] *Department of Electronics and Communication Engineering, BMS Institute of Technology and Management, Bangalore, Karnataka, India*
[d] *Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lampur, Malaysia*
[e] *MARCS Institute, Western Sydney University, Sydney, Australia*

Abstract. In today's world, there have been lots of unique optical character recognition systems. One drawback of these systems is that they cannot work effectively on natural scene images where the text is not only subject to different orientations, lightning, and background but can be of multiple scripts as well. The paper, proposes a state of the art algorithm to detect texts of different dialects and orientations in an image. The whole text detection pipeline is divided into two parts. First, extraction of probable text regions in an image is performed based on a combination of statistical filters, which results in a high recall. These regions are then fed to an Artificial Neural Networks (ANN) based classifier which classifies whether the proposed regions are text or non-text, which increases the overall precision. The validity of the algorithm is verified on the most challenging bilingual text detection dataset MSRA-TD500 and a promising F1 score of 0.67 is reported.

Keywords: Text detection, entropy and variance filters, invariant moments, artificial neural networks, bilingual text detector

1. Introduction

Text or more formally written communication is one of the two ways of communication humans use. From the cave paintings which ages back to the era of Neanderthals to the sign posts available on streets now, humans have always relied on a written way of communication to pass on information. It is the only way with which our species have been recording its work and passing it generation by generation. The current scenario is that the text is not just restricted to being written on artifacts or wall paintings, but when one sees around one can observe that there is a piece of text written on almost all man-made objects whether it is a product description, advertisement or some information. Today there is a demand for systems which can robustly detect texts of all kinds in an image, which includes variation in not just lightning, style, size but also can cover different dialects. These systems will then be able to aid the process of searching by the content in an image or video in search engines or end to end text recognition and machine translation systems.

Text understanding problem in images and videos can be divided into the following sub problems: (i) detection, (ii) localization, (iii) tracking, (iv) extraction and enhancement, and (v) recognition (OCR) [1].

[1] Authors contributed equally.

*Corresponding author. Alex Noel Joseph Raj, Department of Electronic Engineering, Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Shantou University, Shantou, China. E-mail: jalexnoel@stu.edu.cn.

The paper concentrates mainly on text detection and localization in natural images. Although lots of work [2–4] has been done in the field of text detection, but text detection and localization in natural images is challenging because text can be of different fonts, colors, and orientations; subjected to uneven lighting conditions and distortions and finally natural images have a diverse range of objects and patterns which are similar to text [5].

Text detection and localization approaches can be grouped into (a) texture based classifier techniques [6–9]:- identify text regions by considering them as unique textures that are distinguishable from background; (b) connected component techniques [10–20]; exploit the text regularity by extracting features such as edges, strokes, color etc. (c) Sliding windows based techniques, [21–23] which use windows to locate text lines or regions and later employ machine learning for classification and (d) hybrid techniques [5, 24]. A detailed review of these techniques is presented in section (2).

Here a region based text proposal approach is proposed that combines the use of statistical filters to detect thetext proposal regions through background elimination and later employ translation and rotation invariant shape features (Zernike Moments) to classify the regions as text or non-text. The combination provides fewer text proposal regions per image which are very low compared to other approaches like [11, 14] which gives around 100–1000 regions per image. Further, unlike other methods which focus on text detection in the uni-lingual dataset, the proposed method is testedon a very challenging dataset released by [10] containing both English and Chinese text. The paper is organized as follows: Section 2 provides a detailed review of the text detection methods. The overview of the proposed algorithm is presented in section 3. The details of the experiments and the results are discussed in section 4. Finally, section 5 and 6 contains limitations and conclusion respectively.

2. Literature survey

Numerous research have been done in the area of text detection pertaining to real-world scenes in images and videos. Elaborate surveys can be found in [1–3]. Most of these works focus on the horizontal text. Since the inception of multi-oriented datasets like MSRA-TD500, the community has started paying attention to the real-life scenarios of text detection. A recent survey which explains about the different state of the art algorithms, challenges, and future directions can be found in [2]. Although most of the existing works use the ICDAR ('03 '05 '11 '13) dataset, the proposedwork focuson MSRA-TD500dataset, since ICDAR datasets only contain horizontal English text lines.

The texture-based techniques consider the text as a unique texture which is different from the background [6–9] employ features based on texture properties like regional intensities, wavelet coefficients, and filter outputs to localize the text regions. These methods are computationally costly and are subject to rotational and scale variations. These techniques are limited to texts in a horizontal orientation and fail to work with vivid backgrounds and patterns like brick structure, glass or leaves behind the text because the texture properties of these are similar to text.

The component-based techniques [10–20] are generally a cascade of classifiers with increasing strength. The first level of weak classification is the extraction of probable regions of text by proposing candidates based on some sort of clustering or grouping. The second level of classification is to screen out the non-text candidates based on manually designed threshold-based rules or feature-based learning classifiers. These methods are invariant to rotation and scale change but have a drawback of generating a high number of false positives. Lately, these techniques have become the mainstream approach in text detection, but very few works have been done on texts of arbitrary orientations [10, 11].

Sliding window-based techniques, [21–23] use sliding template (window) of different scales for a coarse identification of the text lines followed by feature extraction. These techniques distinguish text and non-text based on local and global geometric properties of text followed by machine learning-based filtering as done by the authors in [21, 22]. Researchers in [23] slide the window across the whole image and do a variant of K- Means Clustering on input vector obtained after ZCA whitening. As these methods scan the whole image, the performance is good, but there is a tradeoff with speed as these approaches suffer from severe computation complexity. Recently to overcome the computational load of sliding window technique, [11–14] employ different approaches based on region proposals as done in object detection (RCNN, Fast RCNN, Faster RCNN, Mask RCNN) to generate probable proposals of text region rather than text lines. This approach has a computational advantage over sliding window since the whole is not required for analysis. Further in [11], the authors use concept of edge boxes with a region-based network and generate 100–1000 text proposals per image, which are then classified using a cascade of weak and strong classifiers and then sent for recognition, whereas [14] uses Faster Region based Convolution Neural Networks on detecting text regions in an image.

Hybrid techniques [5, 24] are those who employ the advantages of both texture-based and component-based techniques. These methods are computationally less expensive than texture-based techniques and generate less false positives (candidates) compared to component-based techniques. Similar to other algorithms these two algorithms can only detect horizontal text. This work employs a hybrid approach whereboth texture-based and component-based techniques are used to identify text regions and report a very promising F1 score.

3. Overview of the proposed algorithm

Two aspects motivate our algorithm: (a) based on the recent success of symmetry window techniques for text localization [21] where a symmetrical template is employed only to localize probable text lines, here the probable text regions are identified instead of the text line and classify them using ANNs. This will drastically reduce the computational load as symmetrical template window is not required to run across the whole image and (b) to increase the visibility and readability, the text is normally portrayed on distinguishable constant gray level regions [6]. Hence efficient detection and elimination of background regions would enable faster text detection. Our text detection pipeline can be divided into:

1. Reduction of Area of Text Analysis
2. Generation of probable text proposal regions
3. Feature extraction and classification of probable text proposals

3.1. Reduction of area of text analysis

Text detection in natural scenes is a challenge since the images contain a wide variety of information. Hence grouping similar regions and removal of unneeded content would greatly reduce the area of text analysis. The approach is similar to the texture based technique discussed in [6, 21, 22], but the novelty lies in the usage of statistical features instead of texture features in detecting the dissimilar regions and usage of shape features for accurately classifying the text and non-text regions. Here a combination of statistical filters is proposedto accentuate the dissimilar neighbors and determine the text proposal regions.

Entropy is a measure of randomness and variance is a measure of variability [25]. Since text regions exhibit certain regularity and are distinct from other objects, these statistical features can be employed to highlight the edges of the dissimilar regions. The intent is to calculate the entropy and variance of each pixel and employ a dynamically calculated threshold value to determine if the pixels belong to text or non-text proposal regions.

Here a sliding window technique is employed to compute the entropy and the variance of the centre pixel of an nxn window of the input image $I_{Original}$. The window is moved along the horizontal and vertical directions with a fixed step size to obtain the entropy image I_E and the variance image Iv using Equations (1 and 2). Here considering the computational efficiency, the window size and step length were fixed as 31×31 and 1 respectively.

$$I_E = \sum_i \sum_j p(i, j) \log(p(i, j)) \qquad (1)$$

$$I_V = \sum_i \sum_j (i - \mu)^2 p(i, j) \qquad (2)$$

where,

p(i,j) is the probability of occurrence of an intensity level in an image.

μ is the mean of the image.

From the outputs shown in Fig. 1(a) itcan be verified that the I_E and I_v images can be employed to emphasize dissimilar neighbors by brightening edges pixels. Since text regions exhibit spatial symmetry and geometric regularity they have higher entropy and variance values when compared to homogeneous background regions [25]. Hence a simple threshold mechanism can be used to eliminate background pixels from the text pixels, but since statistical features vary for images, the threshold value cannot be preset. Therefore, based on the analysisit wasfound that pixels with entropy and variance above the median value can be considered as text candidates, but this approach would fail if the text is less distinguishable from the background or the background as it has a geometric pattern similar to text. To overcome the problem and to effectively identify the text proposal regions, a two-step process is adopted wherein atfirst the pixels that are above the median value of

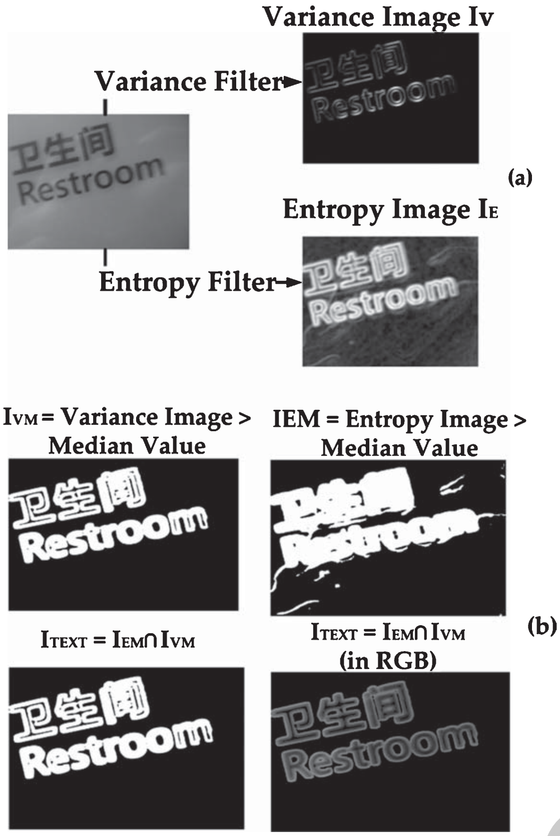


Fig. 1. (a) Illustration of outputs of Entropy and Variance filters. (b) Illustration of extraction of common pixels based on Equations (3–5) with a sample image taken from MSRA-TD500 dataset.

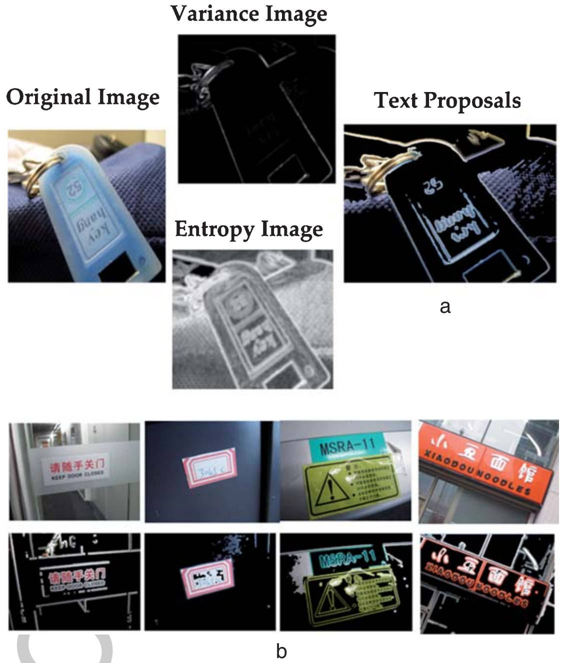


Fig. 2. Text proposal obtained from the original image after background removal; (a) Text proposals generated through the combined use of statistical filers; (b) Sample Image from MSRA-TD500 data set.

both I_E and I_v images are determined and later the common pixels are foundamong the identified ones. Equations (3–5) explain the process mathematically and Fig. 1(a) illustrate the procedure of identifying the text proposal regions.

$$I_{EM} > median(I_E) \tag{3}$$

$$I_{VM} > median(I_V) \tag{4}$$

$$I_{TEXT} = I_{EM} \cap I_{VM} \tag{5}$$

Here I_{EM} and I_{VM} represent the images obtained by considering pixels that are greater than the median value of entropy and variance respectively and I_{TEXT} represents the image containing the probable text regions obtained from common regions of I_{EM} and I_{VM}.

It can be seen from the Fig. 2(a) and (b) the procedure eliminates the background regions and highlights the probable text regions, thus reducing the area of text analysis significantly since only these regions will be considered for further analysis. The experimental analysis of this step on MSRA-TD500 data set presented a recall close to 0.96, and the quantitative analysis of the reduction in the search area is presented in section (4).

3.2. *Generation of probable text proposal regions*

The regions identified in I_{TEXT} as probable text regions are extracted as individual sub-regions by considering an 8×8 connectivity. The use of statistical filters in locating the probable text regions reduces the text search area significantly, and on average 17 text proposals per image, are produced which are comparatively lower than a number of proposals generated in other approaches like [11, 14] which gives around 100–2000 regions per image. The features extracted from text proposal can be used to train a classifier to classify the regions as text as non-text, but it is observed in a few cases that the text proposals are of abnormal size since background also becomes a part of the extracted sub-regions as shown in Fig. 3. Extracting features from these abnormal sub-regions

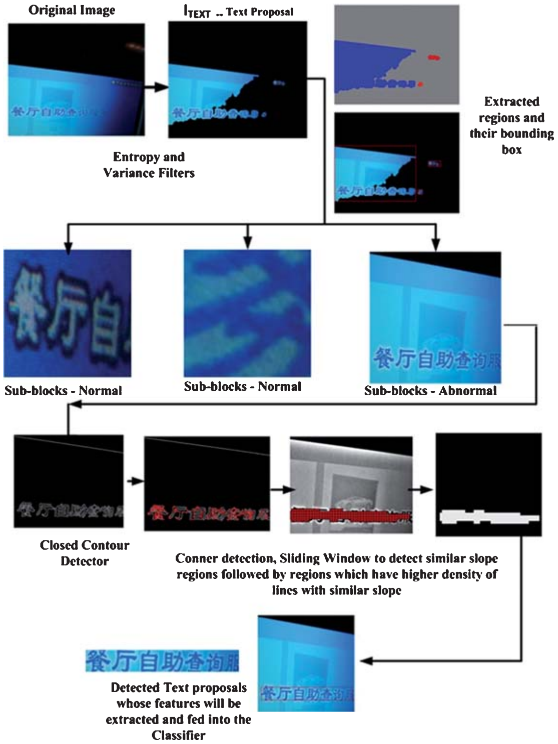


Fig. 3. Illustration of Pruning algorithm with a sample image from MSRA-TD500 data set.

will directly result in poor classification. Hence to improve the accuracy, the abnormal sub-regions are fed to a stronger proposal generator algorithm which prune these abnormal regions into further sub-regions by accurately localizing the text regions.

The abnormal text regions are pruned based on the assumption that (a) characters within the text regions have closed contours [27] and (b) the characters within the text regions have similar slopes due to their regularity and geometric symmetry [6].

The sub-regions that are greater than 60% of the original image (in terms of area) are considered as abnormal and are fed to this stronger proposal generator algorithm. Since the characters within the text regions have closed contours, herethe algorithm explained in [27] is employed to identify the regions which have closed contours (edges). Later the corners of these text proposal regions are found by the technique given in [28] and determine the slope between the successive corner points in different orientations. Later the corners points with similar slope are identified and connected. Post this a cascade of a sliding window of sizes 128×128 to 256×256 with an increment of 32 pixels,is used to detect the regions that provide a higher density of lines with a similar slope. Here it is believed that the text line has the same slope throughout. These regions are the new text proposals which are extracted out of previously generated abnormal text regions from (5). The pruning algorithms serve two purposes (a) localizes the text regions within the abnormal text regions and (b) further sub-divides the abnormal text regions them into smaller sub-regions. The whole process can be visualized in Fig. 3.

3.3. Feature extraction and classification of probable text proposals

3.3.1. Feature extraction using Zernike Moments

The text proposal regions, obtained through the combination of statistical filters contain text, and non-text information isefficiently described using image descriptors for successful classification. A good choice of features enables the proposed system to provide better accuracy. Since text regions exhibit spatial symmetry and geometric regularity, shape characteristics [26] can be employed to differentiate between text and non-text regions. The requirements of a good shape descriptor are (a) invariance to change in rotation, scale and translation, (b) provide features with low redundancy and large discrimination ability and (c) present hierarchical representation, i.e., furnish coarse (global) to finer (local) details. Literature suggests that moment-based descriptors satisfy the above requirements. Zernike moments (ZM) [29] due to their orthogonality, have lowest feature redundancy, and also it is observed that their hierarchical nature allows the lower order moments provide global information, and higher-order moments provide local information respectively [30, 31]. Further, it is computationally less expensive to compute the higher order moments. Therefore, the shape features are extracted using Zernike descriptors and the classifier is trainedto classify the sub-regions as text and non-text.

Let $S(x, y)$ be the sub-regions extracted from $\mathrm{I_{TEXT}}$.

$\mathrm{I_{TEXT}} = \{S_1(x, y),\ S_2(x, y), \ldots\ldots S_n(x, y)\}$
where n be the number of sub-regions per image.

To compute ZM, each sub-block $S(x, y)$ is re-scaled to $\mathrm{N} \times \mathrm{N}$ dimension and projected on to set of complex Zernike polynomials as given in Equation (6)

$$\mathrm{A_{nm}} = \frac{\mathrm{m}+1}{\pi} \int_{\mathrm{x}} \int_{\mathrm{y}} \mathrm{S(x, y)} \mathrm{V^{*}_{nm}}(\rho, \theta) \mathrm{dx\, dy} \quad (6)$$

where $x^2 + y^2 \leq 1$

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho)\exp(jm\,\theta) \quad (7)$$

where
$V_{nm}(x, y)$ is the Zernike Polynomial
n is the order of the polynomial,
m is the repetition factor such that $|m| \leq n$ and $n - |m|$ is even,
ρ is the length of the vector from the origin to the pixel located at spatial location (x,y) and is given by $\rho = \sqrt{x^2 + y^2}$,
θ is the angle of the vector from the origin to the pixel located at spatial location (x,y) from the x-axis in counter clockwise direction and,
$R_{nm}(\rho)$ is the radial polynomial defined as,

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|} (-1)^s \frac{(2n+1-s)!}{s!(n-|m|-s)!(n+|m|+1-s)!} \rho^{n-s} \quad (8)$$

The ZM obtained from the above equation is a complex quantity given by,

$$A_{nm} = R_{ZM} + jI_{ZM} \quad (9)$$

$$|A_{nm}| = \sqrt{R_{ZM}^2 + I_{ZM}^2} \quad (10)$$

Here $|A_{n,m}|$ represent shape descriptor obtained from Zernike moments for the order n and repetition factor m. To obtain the complete shape information a feature vector is formedby concatenating the moments obtained by varying n and m. For the proposed work, 72 different ZM are computed for each $S(x, y)$ by varying n from 0 to 15 and concatenate them to form a feature vector (FV) as shown in Equation (10). The FV is later used for training the classifier.

$$FV = \{|A_{00}|, |A_{01}|, |A_{20}|, |A_{22}|, |A_{31}|, |A_{33}|, \ldots, |A_{15,15}|\} \quad (11)$$

The feature vectors so obtained have strong class discrimination ability and hence can differentiate the text and non-text sub-regions effectively. The FV obtained for text images, and non-text images are visualized in Fig. 4(a) and (b).

From the figures, it can be noticed that the ZM magnitudes for both text (red) and non-text (black) regions have negligible likeliness between them. This motivates the classifier for easier distinction between the text and non-text regions. Further, from Fig. 4(c), it can be observed that there are eight prominent

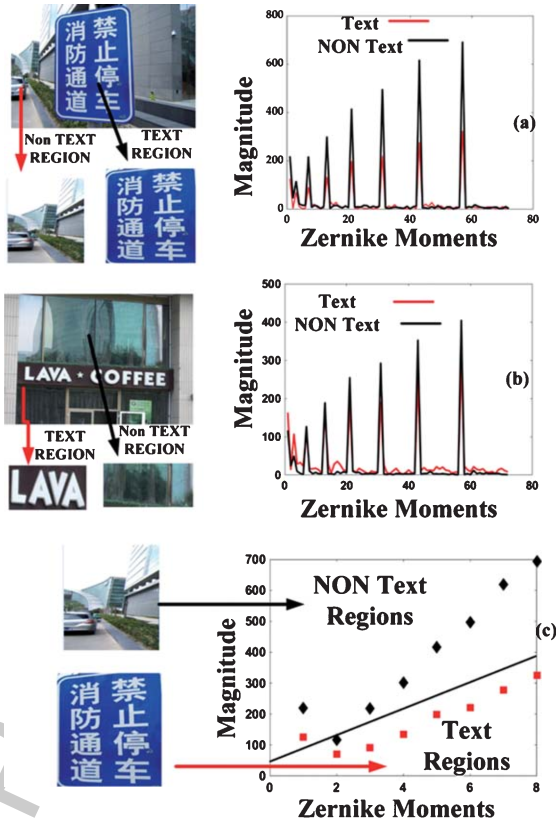


Fig. 4. (a) and (b) Illustration of Zernike magnitudes between Text (Red) and Non-text region (Black); (c) Feature space showing the significant Zernike features (Red) for Text and Non-Text regions (Black).

peaks (for both text and non-text regions) at positions 1, 3, 7, 13, 21, 31, 43 and 57 on the x-axis. These peaks correspond to the magnitudes of Zernike moments: $|A_{00}|$, $|A_{20}|$, $|A_{40}|$, $|A_{60}|$, $|A_{80}|$, $|A_{10,0}|$, $|A_{12,0}|$ and $|A_{14,0}|$ respectively that are significant and hold larger descriptive information. The significant features in the feature space for Fig. 4(a) and (b) are depicted in Fig. 4(c).

3.3.2. Classification using ANN

The features extracted from all the text and non-text sub-regions form the input that is fed to the artificial neural network (ANN) for classification. The proposed method was evaluated considering the MSRA-TD500 dataset with 500 images of which 300 are training images and 200 test images. The evaluation comprises of training and testing process. ANN classifier is deployed here that is capable of learning by example and process the information in parallel. ANN has layered architecture that includes: the input, hidden layers, and output layer. Each layer has neurons with its associated activation function that process the data presented to it. The ANN maps various inputs to output by finding a non-linear relationship between them. The mapping is governed by the adjustable parameters of the network (weights and biases). These parameters are adjusted during the training process by providing the training dataset to the network. The training devises an optimal learned model that is saved and assessed in the testing process.

For all the training images, text and non-text sub regions were identified and extracted from the 300 training images. For each region of text and non-text ZM feature vector is extracted with feature vector size of [72, M] where M is the number of samples in the dataset. These feature vectors are labeled with two classes C1 and C2 that represents text and non-text respectively. With this, a complete training set is achieved as shown

$$\mathrm{X_r} = [\mathrm{T_r}, \mathrm{y}] \tag{12}$$

where

$\mathrm{T_r}$ = [Zernike feature vectors for all text and non-text sub regions of training images] and, y = [C1, C2].

This training datais fed to a simple 3-layer ANN as displayed in Fig. 5. The network's input is a feature vector of length 72. The network has two hidden layers of 36 and 18 nodes respectively and an output layer with 1 node. The number of learnable parameters is 2628 ($72 \times 36 + 36$) in the1st layer, 666 ($36 \times 18 + 18$) in the 2nd layer and 19 ($18 \times 1 + 1$) in the final layer. Thus a total number of 3313 learnable parameters or weights of the network are to be modified during training process. The network now takes in $\mathrm{X_r}$ as the input for learning. The hidden layers find non-linear patterns in the data and pass them to the output layer, which predicts the probability with which a particular input belongs to a class, in this case (C_1 is text and C_2 is non-text).

Neurons of each layer use sigmoid non-linearity which can be defined as, $\mathrm{y} = \frac{1}{1+\mathrm{e}^{-\mathrm{z}}}$ and, $\mathrm{z} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i w_{ij}} + \mathrm{b_j}$, where $\mathrm{x_i}$ belongs to the input feature vector, $\mathrm{b_j}$ represents the bias term and $\mathrm{w_{ij}}$ represents the corresponding weight from node i to node j. The weights are updated using gradient descent, which can be stated as, $\Delta \mathrm{w_{ij}} = -\eta \frac{\partial \mathrm{E_i}}{\partial \mathrm{w_{ij}}}$, in which the learning rate η is set to 0.001 and a batch size of 32 is fixed. Thetotal number of epochs for training is set to 5000, with a minimum error based early stopping.The loss function considered for training is the mean squared error (MSE) [32, 33] as shown,

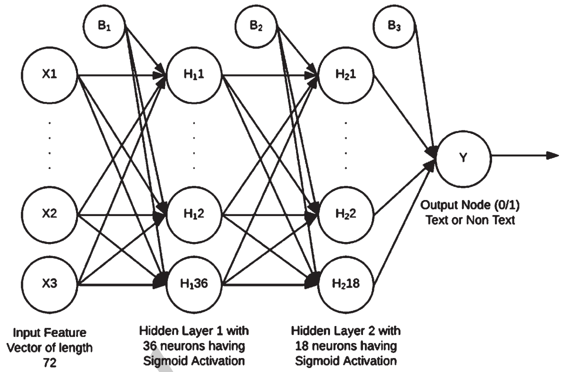


Fig. 5. The neural network structure.

$$MSE = \frac{1}{m} \sum\nolimits_{i=1}^{M} (y^i - t^i)^2 \tag{13}$$

where

y^i is the network output of the network and,

t^i is the desired output from the network.

MSE is used as the loss function since when used with sigmoid activation function in a shallow network(as incorporated in the proposed work), gives similar or better results when compared with cross-entropy loss function [34]. Also, it has been observed in many works that cross entropy loss function works better with softmax as an activation function when compared with sigmoid,thus acombination of MSE and sigmoid for a shallow network gives optimum results. Further, with a better initialization, MSE outperformscategorical cross-entropy based networks in terms of less error achieved [33, 34]. The training process is presentedin Fig. 6. From the plot, it can be observed that training is stopped at 3665th epoch when a minimum error of 0.001 is achieved. The network required 04 hours and 15 minutes of training time. The network was trained on a NVIDIA Quadro P4000 GPU.

As mentioned earlier that MSRA-TD500 has a set of 300 train and 200 test images. In thisapproach training set widens up and all the proposals generated from those 300 images become the training setwith 5133 proposals.The features of these proposals are extracted and train the ANN. A group of volunteers labeled the whole training dataset.

The training process is followed by testing, where, the proposals generated from 200 test images becomes the test dataset. During testing the learned

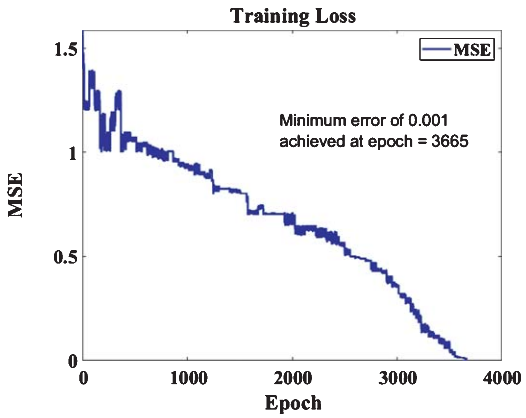


Fig. 6. Graph depicting training loss per epoch.

ANN model is evaluated for its accuracy by providing the test set $X_t = [T_t]$ obtained from 200 testing images (3592 proposals) to it.Finally, for the regions that are classified as text (C1), a bounding box was plotted around those regions on the actual image, and that becomes the final output of thealgorithm. The flowchart which explains the entire flow ofthe algorithm is shown in Fig. 7. Further the performance of the model is assessed using the confusion matrix or contingency table.

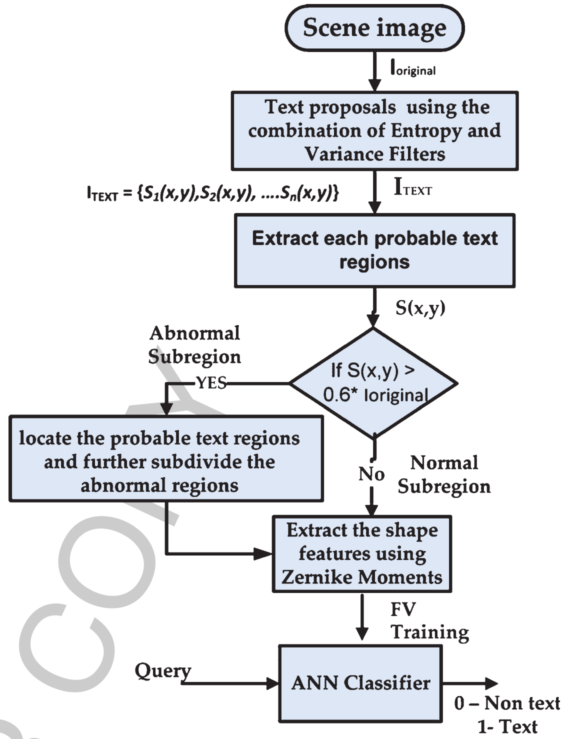


Fig. 7. The flow diagram of the proposed algorithm.

4. Results and discussions

The functionality of the proposed algorithm was testedon one of the most challenging text detection dataset proposed by a team of researchers from Microsoft Research Asia [10]. This dataset is referred to as MSRA-TD500 an openly released public dataset that contains 500 natural scene images covering more than one dialect, English and Chinese for text detection. These images were carefully split by the creators themselves in the ratio of 60:40 for training and testing. Thus the dataset has300 train and 200 test images. Not only the dataset is known for its toughness and variety but also it serves as a benchmarking dataset for all new algorithms.

To illustrate the efficacy of our algorithm, three different experiments were performed; (1) Reduction in Area of Text Analysis; (2) Computation of performance metrics of the ANN classifier and, Precision, Recall, and F1 score of the proposed algorithm and (3) Comparison of the proposed algorithm with other text detection techniques.

4.1. Reduction in area of text analysis

The combined use of statistical filters followed by the two-step threshold detection procedure explained in section (3.1) eliminates the homogeneous background regions and improves the computation efficiency by reducing the search area of text regions. The resultant I_{TEXT} corresponds to the image containing the probable text as shown in Fig. 1. Accordingly, after the filtering process, it can be noted from the figure that the background is eliminated to highlight the probable text regions thus reducing the Area of Text Analysis. The total reduction in the area is quantified through Equation (14)

$$Reduction\ in\ Area\ of\ Text\ Analysis = \sum_i^N \frac{Non\ Zero\ pixels\ in\ I_{Text}}{Total\ Pixel\ in\ I_{Orignal}} \tag{14}$$

Here N represents the total number of images considered. For MSRA-TD500 dataset where N = 500, the reduction achieved through the combined use of statistical filters was 72.85%.

4.2. Performance measures

On average, the algorithm was able to locate 17 sub-regions regions per image, and later these regions were resized to 256×256 and then fed into the Zernike feature based neural network for training and classification. The training set had 300 images which generated 4493 sub regions, out of which 4102 sub regions were normal, and 391 were of abnormal size and therefore required to be pruned as per the procedure explained in section (3.2). The pruning algorithm further generated 1031 probable regions of text from those abnormally sized probable text regions. Thus, the training set generated 5133 regions that included 3977 non-text regions and 1156 text regions which were manually labeled as C1 (text) and C2 (non-text) respectively. The position of the sub-region is extracted and stored separately and is later used to plot the bounding box if the sub region is classified as positive. Also, coordinates of the sub-blocks (x, y, width, and height) with respect to the original image are extracted for future reference. Two different performance measures have been presented (a) Classifier Accuracy and (b) Evaluation with respect to the ground truth.

4.2.1. Classifier Accuracy

For each sub-block, a 72-dimensional Zernike based shape feature vector was computed which were then labeled and later used to train the ANN. The ANN was trained with 5000 epoch limits with a learning rate of 0.001. During training, the trainable parameters of the network are modified to produce the optimal model. A confusion matrix [35] is framed from the result of the training process to compute the training accuracy of the model which is displayed in Table 1.

From the matrix, the training accuracy is computed using,

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FN}+\text{FP}} \tag{15}$$

which was found to be 98.81%.

Here, True Positive (TP): Region of text detected as text,

False Positive (FP): Region of non-text detected as text,

False Negative (FN): Region of text detected as non-text,

True Negative (TN): Region of non-text detected as non-text,

4.2.2. Evaluation with respect to the ground truth

To illustrate how good, the proposed method localizes the text within a scene image, thetext localization outputs are compared with the ground truth results. Accordingly, a bounding box is plotted around all the positive classifications (TP and FP) and then it is compared with ground truth. To have a realistic comparison with other text detection techniques,an evaluation procedure detailed in [10] was followedto compute the final F1 score:

Positive Predictions (TP):-

i. If the prediction of the model overlaps with 50% of the pre-defined ground truth and the angle between them is less than $\pi/8$, then they are treated as valid. This protocol used by [10] allows us to have a fair comparison
ii. Maximal suppression is adopted, i.e., if the proposed method predicts two bounding boxes, and if the bigger box contains a smaller box covering the text region then the bigger box is suppressed (refer. Fig. 8)

Based on the assumptions made, the bounding boxes are plotted around the positive classifications of both training and testing sets. From Tables 1 and 2 it is seen that the training process with 300 images (5133 sub-images) produced positive classifications of 1215 whereas the testing with 200 images (3592

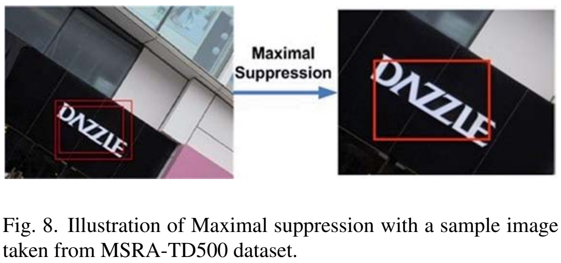


Fig. 8. Illustration of Maximal suppression with a sample image taken from MSRA-TD500 dataset.

Table 1
Confusion matrix from training ANN

	Text region	Non-Text region	Accuracy
Text region	1155 (TP)	1(FN)	98.81%
Non-Text region	60 (FP)	3917(TN)	

Table 2
Confusion matrix from testing ANN

	Text region	Non-Text region	Accuracy
Text region	655 (TP)	42(FN)	95.9%
Non-Text region	103 (FP)	2792(TN)	

Table 3
Performance Evaluation with respect to the ground truth text bounding box

	Precision	Recall	F1score
Training Set	0.77	0.64	0.70
Testing Set	0.76	0.59	0.67

Fig. 9. Text detected on sample images from the MSRA-TD500 dataset using the proposed algorithm.

sub-images) produced 758 positive classifications respectively. These bounding box predictions were then compared with ground truth label, and the performance measures such as Precision, Recall, and F1 score [35] are calculated based on Equations (16–18), and results are summarized in Table 3. Further, text detected on few sample images from the dataset are shown in Fig. 9.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{17}$$

$$\text{F1 Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{18}$$

4.3. *Performance comparison with other text detection techniques*

The performance of the proposed method is compared with the [7, 10, 17, 36–38], and their F1 scores are reported in Table 4. The techniques illustrated in Table 4 with * were collectively obtained from [10], the authors have reported having tested these techniques on MSRA-TD500 dataset, and the technique with ** was re-evaluated by us on the dataset.

Table 4
Performance evaluation with other text detection techniques

Authors	Precision	Recall	F1 score
Chen et. al. [7]*	0.05	0.05	0.05
C. Yao et. al. [10]*	0.63	0.63	0.60
Yin et al. [17]*	0.71	0.61	0.65
Proposed	**0.76**	**0.59**	**0.67**
Ephstein et. al. [36]**	0.25	0.25	0.25
Kang et al. [37]*	0.71	0.62	0.66
Chen, Huizhong, et al. [38]**	0.24	0.48	0.32

These results logically compare the performance of the proposed algorithm with that of the state-the-of-art algorithms. These performances have been directly referred from the high-quality research articles. With a high F1 score of 0.67, the proposed method provides a better classification of text and non-text regions when compared to other algorithms tested on the same bilingual dataset.

From Table 4, it can be inferred that the proposed method has a better Precision and F1 score compared to other state-of-art techniques. This illustrates that the proposed method can be effectively employed to detect text regions in natural scenes which have diverse objects in different orientations, brightness, and texture. To further show the efficacy of the proposed method, images comparing the proposed algorithm's performance with [36, 37] are presented in Fig. (10)

5. Limitations of the proposed method

The proposed algorithm gives very promising results on a very challenging dataset and effectively localizes texts of different dialects at arbitrary orientations within natural images. Still, the performance is far from perfect and the algorithm is observed to failrepetitively on two cases. 1)When the text is on the reflective surface as observed in Fig. 11(a) and 11(b). 2) When the text has less contrast with respect to the background and background has a strong recurring pattern similar to the text as shown in Fig. 11(c) and 11(d). The failures areattributed to thepoor accuracy in locating the text proposal regions using the statistical filters rather than the classifier.

a b c d

Fig. 10. Comparison with ground truth and other techniques on MSRA-TD500 dataset. (a) Ground truth (b) Text detection using [38], (c) Text detection using [36] and (d) Proposed method.

Fig. 11. Failure of text detection (a) and (b) on a reflective surface; (c) and (d) on a recurring pattern with sample images from from MSRA-TD500 dataset.

6. Conclusion and future works

The work presents a unique state of the art pipeline for text detection in natural images. After scrutinizing theperformance on test images, it can be stated that the hybrid approach is better than existing techniques regarding performance and speed due to the following reasons. 1) Very high recall (0.96) of the output of Equation (5) ensures that texts of different dialects are captured. The interplay of variance and entropy filters makes sure that vivid texts are captured. There are very few False Negatives (Text Detected as Non-Text) when the performance on Eq. (5) was evaluated. 2) The novel method employs rotational and translational invariant moments as an input feature vector to the classifier ensures that the algorithm detects texts of all orientation. The next tasks to be carried can be categorized as follows 1) to improve the performance of the algorithm by keeping the limitations, mentioned in Section 5, in mind; and (2) once detection is perfected the problem of text recognition of the detected text will be approached.

Acknowledgments

The project is supported by the Research Start-up Fund Subsidized Project of Shantou University, ChinaGrant No: NTF17016. The National Science Foundation of China, Grant. No: 61471228 and the Key Project of Guangdong Province Science and Technology Plan, Grant. No: 2015B020233018.

References

[1] K. Jung, K.I. Kim and A.K. Jain, Text information extraction in images and video: A survey, *Pattern recognition* **37**(5) (2004), 977–997.

[2] Y. Zhu, C. Yao and X. Bai, Scene text detection and recognition: Recent advances and future trends, *Frontiers of Computer Science* **10**(1) (2016), 19–36.

[3] Q. Ye and D. Doermann, Text detection and recognition in imagery: A survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37**(7) (2015), 1480–1500.

[4] J. Liang, D. Doermann and H. Li, Camera-based analysis of text and documents: A survey, *International Journal of Document Analysis and Recognition (IJDAR)* **7**(2-3) (2005), 84–104.

[5] Y. Zheng, Q. Li, J. Liu, H. Liu, G. Li and S. Zhang, A cascaded method for text detection in natural scene images, *Neurocomputing* **238** (2017), 307–315.

[6] S.M. Hanif and L. Prevost, Text detection and localization in complex scene images using constrained adaboost algorithm. In *Document Analysis and Recognition, 2009 ICDAR'09 10th International Conference on*, IEEE, 2009, pp. 1–5.

[7] X. Chen and A.L. Yuille, Detecting and reading text in natural scenes. In *Computer Vision and Pattern Recognition, 2004 CVPR 2004 Proceedings of the 2004 IEEE Computer Society Conference on*, IEEE, Vol. 2, 2004, pp. II–II.

[8] K. Wang, B. Babenko and S. Belongie, End-to-end scene text recognition. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, IEEE, 2011, pp. 1457–1464.

[9] L. Neumann and J. Matas, Real-time scene text localization and recognition. In *Computer Vision and Pattern Recogni-*

tion (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 3538–3545.

[10] C. Yao, X. Bai, W. Liu, Y. Ma and Z. Tu, Detecting texts of arbitrary orientations in natural images. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2012, pp. 1083–1090.

[11] M. Jaderberg, K. Simonyan, A. Vedaldi and A. Zisserman, Reading text in the wild with convolutional neural networks, *International Journal of Computer Vision* **116**(1) (2016), 1–20.

[12] M. Jaderberg, A. Vedaldi and A. Zisserman, Deep features for text spotting. In *European Conference on Computer Vision*, Springer, Cham, 2014, pp. 512–528.

[13] L. Gomez and D. Karatzas, A fast hierarchical method for multi-script and arbitrary oriented scene text extraction, *International Journal on Document Analysis and Recognition (IJDAR)* **19**(4) (2016), 335–349.

[14] Y. Nagaoka, T. Miyazaki, Y. Sugaya and S. Omachi, Text Detection by Faster R-CNN with Multiple Region Proposal Networks. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, IEEE, 2017, pp. 15–20.

[15] W. Huang, Z. Lin, J. Yang and J. Wang, Text localization in natural images using stroke feature transform and text covariance descriptors. In *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1241–1248.

[16] J. Matas, O. Chum, M. Urban and T. Pajdla, Robust wide-baseline stereo from maximally stable extremal regions, *Image and Vision Computing* **22**(10) (2004), 761–767.

[17] X.C. Yin, X. Yin, K. Huang and H.W. Hao, Robust text detection in natural scene images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**(5) (2014), 970–983.

[18] W. Huang, Y. Qiao and X. Tang, Robust scene text detection with convolution neural network induced mser trees. In *European Conference on Computer Vision*, Springer, Cham, 2014, pp. 497–511.

[19] L. Gomez and D. Karatzas, Object proposals for text extraction in the wild. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*, IEEE, 2015, pp. 206–210.

[20] C.L. Zitnick and P. Dollár, Edge boxes: Locating object proposals from edges. In *European Conference on Computer Vision*, Springer, Cham, 2014, pp. 391–405.

[21] Z. Zhang, W. Shen, C. Yao and X. Bai, Symmetry-based text line detection in natural scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2558–2567.

[22] X. Zhang, Z. Zhang, C. Zhang and X. Bai, Symmetry-based object proposal for text detection. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*, IEEE, 2016, pp. 709–714.

[23] T. Wang, D.J. Wu, A. Coates and A.Y. Ng, End-to-end text recognition with convolutional neural networks. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, IEEE, 2012, pp. 3304–3308.

[24] Y.F. Pan, X. Hou and C.L. Liu, A hybrid approach to detect and localize texts in natural scene images, *IEEE Transactions on Image Processing* **20**(3) (2011), 800.

[25] H.R. Sheikh and A.C. Bovik, Image information and visual quality. In *Acoustics, Speech, and Signal Processing, 2004 Proceedings (ICASSP'04) IEEE International Conference on*, IEEE, Vol. 3, 2004, pp. iii–709.

[26] Z. Liu and S. Sarkar, Robust outdoor text detection using text intensity and shape features. In *Pattern Recognition, 2008. ICPR 2008 19th International Conference on*, IEEE, 2008, pp. 1–4.

[27] C. Harris and M. Stephens, A combined corner and edge detector. In *Alvey Vision Conference*, Vol. 15, No. 50, 1988, pp. 10–5244.

[28] G. Azzopardi and N. Petkov, A CORF computational model of a simple cell that relies on LGN input outperforms the Gabor function model, *Biological cybernetics* **106**(3) (2012), 177–189.

[29] A. Tahmasbi, F. Saki and S.B. Shokouhi, Classification of benign and malignant masses based on Zernike moments, *Computers in Biology and Medicine* **41**(8) (2011), 726–735.

[30] V.G. Mahesh and A.N.J. Raj, Invariant face recognition using Zernike moments combined with feed forward neural network, *International Journal of Biometrics* **7**(3) (2015), 286–307.

[31] A.N.J. Raj and V.G. Mahesh, Zernike-Moments-Based Shape Descriptors for Pattern Recognition and Classification Applications. In *Advanced Image Processing Techniques and Applications*, IGI Global, 2017, pp. 90–120.

[32] A. Papoulis and S.U. Pillai, Probability, random variables, and stochastic processes, *Tata McGraw-Hill Education* (2002).

[33] G.P. Zhang, Neural networks for classification: A survey, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **30**(4) (2000), 451–462.

[34] P. Golik, P. Doetsch and H. Ney, Cross-entropy vs. squared error training: A theoretical and experimental comparison. In *Interspeech*, Vol. 13, 2013, pp. 1756–1760.

[35] D.T. Larose and C.D. Larose, Discovering knowledge in data: An introduction to data mining, John Wiley & Sons, 2014.

[36] B. Epshtein, E. Ofek and Y. Wexler, Detecting text in natural scenes with stroke width transform. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* IEEE, 2010, pp. 2963–2970.

[37] L. Kang, Y. Li and D. Doermann, Orientation robust text line detection in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 4034–4041.

[38] H. Chen, S.S. Tsai, G. Schroth, D.M. Chen, R. Grzeszczuk and B. Girod, Robust text detection in natural images with edge-enhanced maximally stable extremal regions. In *Image Processing (ICIP), 2011 18th IEEE International Conference on IEEE*, 2011, pp. 2609–2612.

RESEARCH ARTICLE

An RDAU-NET model for lesion segmentation in breast ultrasound images

Zhemin Zhuang[1], **Nan Li**[1], **Alex Noel Joseph Raj**[1]*, **Vijayalakshmi G. V. Mahesh**[2], **Shunmin Qiu**[3]

1 Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Department of Electronic Engineering, Shantou University, Shantou, Guangdong, China, **2** Department of Electronics and Communication Engineering, BMS Institute of Technology and Management, Bengaluru, Karnataka, India, **3** Imaging Department, First Hospital of Medical College of Shantou University, Shantou, Guangdong, China

* jalexnoel@stu.edu.cn

Citation: Zhuang Z, Li N, Joseph Raj AN, Mahesh VGV, Qiu S (2019) An RDAU-NET model for lesion segmentation in breast ultrasound images. PLoS ONE 14(8): e0221535. https://doi.org/10.1371/journal.pone.0221535

Editor: Varenyam Achal, Guangdong Technion Israel Institute of Technology, CHINA

Received: May 19, 2019

Accepted: August 8, 2019

Published: August 23, 2019

Data Availability Statement: Training and validation dataset: http://ultrasoundcases.info/category.aspx?cat=67. Testing dataset:(1) Imaging Department of the First Affiliated Hospital of Shantou University. https://github.com/xbhlk/STU-Hospital.git, (2)Breast Ultrasound Lesions Dataset (Dataset B) https://doi.org/10.1109/JBHI.2017.2731873.

Funding: This work was supported by Key Project of Guangdong Province Science & Technology Plan (No. 2015B020233018, Zhenmin Zhuang); The National Natural Science Foundation of China

Abstract

Breast cancer is a common gynecological disease that poses a great threat to women health due to its high malignant rate. Breast cancer screening tests are used to find any warning signs or symptoms for early detection and currently, Ultrasound screening is the preferred method for breast cancer diagnosis. The localization and segmentation of the lesions in breast ultrasound (BUS) images are helpful for clinical diagnosis of the disease. In this paper, an RDAU-NET (Residual-Dilated-Attention-Gate-UNet) model is proposed and employed to segment the tumors in BUS images. The model is based on the conventional U-Net, but the plain neural units are replaced with residual units to enhance the edge information and overcome the network performance degradation problem associated with deep networks. To increase the receptive field and acquire more characteristic information, dilated convolutions were used to process the feature maps obtained from the encoder stages. The traditional cropping and copying between the encoder-decoder pipelines were replaced by the Attention Gate modules which enhanced the learning capabilities through suppression of background information. The model, when tested with BUS images with benign and malignant tumor presented excellent segmentation results as compared to other Deep Networks. A variety of quantitative indicators including Accuracy, Dice coefficient, AUC(Area-Under-Curve), Precision, Sensitivity, Specificity, Recall, F1score and M-IOU (Mean-Intersection-Over-Union) provided performances above 80%. The experimental results illustrate that the proposed RDAU-NET model can accurately segment breast lesions when compared to other deep learning models and thus has a good prospect for clinical diagnosis.

Introduction

Breast cancer, next to skin cancer is a disease which seriously endangers women health [1, 2]. With the development of modern medicine, if breast cancer is diagnosed early, the survival rate of patients is greatly improved. The diagnosis of a breast tumor can be divided into

(No. 61471228, Zhenmin Zhuang); and The Research Start up Fund Subsidized Project of Shantou University, China: NTF17016 (Alex Noel Joseph Raj).

Competing interests: The authors have declared that no competing interests exist.

invasive diagnosis and non-invasive diagnosis. Invasive diagnosis, which mainly refers to biopsies, causes physical damage to the tissues, whereas non-invasive diagnosis refers to the examination of the breast lesion area, using either X-ray, MRI(Magnetic Resonance Imaging) or Ultrasound (US) imaging examination. Among various examinations, using US images, due to its low radiation, low cost and real-time output capabilities have become a preferred choice for breast tumor diagnosis.

Image segmentation in BUS images refers to extracting the region of interest (lesion) from the normal tissue region. Fig 1 presents few BUS images with both benign and malignant tumors and it is understood that morphology of the tumor varies significantly from the surrounding tissues. This attribute forms the basis for localization and segmentation of tumors using various Machine Learning and Deep Learning techniques. The quality of the segmentation directly affects the accuracy and reliability of the diagnosis results. Due to the nature of the acquisition process, US images are affected by noise and other image artifacts' that greatly increase the difficulty of the segmentation process. Horsch et al. [3] proposed an algorithm for BUS lesion segmentation, where the images were initially pre-processed for noise removal using a median filter. Later the processed images were intensity inverted, multiplied with Gaussian constraint function and thresholded to provide potential lesion boundaries by suppressing distant pixels. Finally, Average radial derivative function aids as a utility function to maximize the actual lesion margins. Although the threshold method is fast, parameters such as the center, height, and width were required to be provided manually for better segmentation results. Xu and Nishimura [4] proposed an algorithm for BUS segmentation using Fuzzy C-Mean (FCM) clustering which required prior initialization of a number of clusters and the noise tolerance level. These initializations were not generalized and depended on the experience, thereby affecting the overall segmentation result. Gomez et al. [5] proposed a method similar to [3] for breast ultrasound lesions segmentation. Here CLAHE and Anisotropic diffusion filter were successively used to enhance the contrast and reduce the speckle noise associated with BUS images. Then, the watershed transformation algorithm was used for finding potential lesion boundaries which were further refined by the Average radial derivative function to determine the final contour of the lesion. An overlap ratio of about 86% was reported by the authors. Daoud et al. [6] introduced a semi-automatic active contour model, which required users to provide an initialization (circular contour) within the tumor. Later, statistical parameters calculated based on the envelope signal-to-noise ratio was iteratively used to move the coordinates of the initial contour towards the tumor boundary. However, the segmentation outputs of the model largely depend on the initial contour. When the initial contour is not well positioned, the ideal segmentation outputs were not achieved. virmani et al. [7] studied the application of despeckling filtering algorithm in BUS image segmentation. The study included (a) finding the optimum number of despeckle filters from an ensemble of 42 filters and (b) evaluation of the segmentation outputs of the benign and malignant tumor. The first set of experiments provided 6 optimum filters that retained the edges and features of the image. Measures such as Beta metrics and Image Quality score were used to assess the performance of the filters. Next, the speckle-removed BUS images were segmented by the edge-based active contour model proposed by Chan and Vese [8]. The performance of the segmentation algorithm was quantitatively evaluated using Jaccard index and qualitatively by the radiologists. It was stated that the DPAD(detail preserve anisotropy diffusion) filter was able to obtain clinically acceptable images. The proposed method was tested on 104 ultrasound tumor images (43 benign and 61 malignant) and an average Jaccard index of 79.52% was reported. Daoud et al. [9] proposed a method based on super-pixels to segment the lesions in BUS images. To begin with, the BUS image was decomposed into coarse hyper-pixels to obtain the initial contour of the tumor and later the coarse pixels were refined to super-pixels to improve the final contour

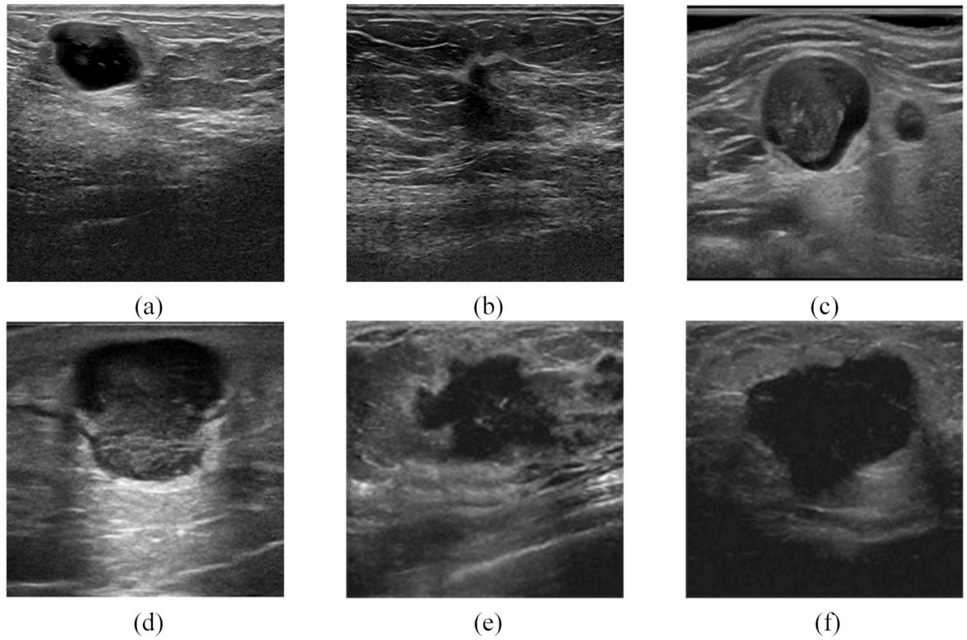


Fig 1. Benign and malignant tumors. (a) and (b) were obtained from Breast Ultrasound Lesions Dataset(Dataset B) [13]. (c) and (d) were acquired from Gelderse Vallei Hospital in Ede, the Netherlands [14]. (e) and (f) were obtained from the Imaging Department of the First Affiliated Hospital of Shantou University. It can be seen from these six figures that there are obvious differences between the tumor morphology and the surrounding tissues.

https://doi.org/10.1371/journal.pone.0221535.g001

of the segmented tumor. The two-stage pipeline provided segmentation results which were comparable to the ground truth. Panigrahi et al. [10] proposed a novel hybrid clustering technique comprising of Multi-scale Gaussian kernel induced Fuzzy C-means (MsGKFCM) and Multi-scale Vector Field Convolution (MsVFC) to segment the region of interest within the BUS images. Initially, the BUS images were preprocessed using speckle reducing anisotropic diffusion technique [11] and then clustered as probable lesion segments using MsGKFCM. Later cluster centers were presented as inputs to MsVFC to obtain the accurate lesion boundary. The technique was tested on 127 US images and various performance measures were used to evaluate the technique. Accordingly, the average values of Jaccard index and dice similarity scores were 93.1% and 93.3% respectively. Zhuang et al. [12] proposed a fractal based technique to segment US images. Here the images of the carotid artery were enhanced using fuzzy technique and later segmented using fractal length. It was reported that fractal length based segmentation presented more accurate segmentation results than the fractal dimensions. The technique presented high qualitative values of DSC, Precision, Recall and F1 score (0.9617, 0.9629, 0.9653 and 0.9641 respectively), together with a low value of APD (1.9316).

In recent years, with the continuous development of the convolutional neural networks (CNN), semantic segmentation algorithms employing deep learning architectures have become popular. These models combine both shallow and high-level features and thus provide accurate results when compared to traditional algorithms which mainly depend on shallow features. However, the application of deep learning in medical images is still in its infancy. Xu et al. [15] proposed a method for BUS image segmentation using CNN. Volumetric (3D) mammary US images were presented to CNN to segment the US images into four major tissues: skin, fibrous gland, mass, and adipose. The idea was to treat the segmentation as a classification problem where every pixel is associated with a class label. Therefore, a large number of BUS data samples are collected and the annotated images were trained on an 8 layered CNN model comprising of convolution, pooling, fully connected and the softmax layers. To provide the classification, the softmax layer was modified to output a probability distribution array with 4 elements, whose maximum value represented one of the four class labels. An F1 score of above 80% was reported. Lian et al. [16] proposed the Attention guided U-Net model based on U-Net architecture which incorporated attention masks for accurate iris segmentation. The use of attention masks enabled the Atten-UNet to localize on the iris region instead of the whole eye. The contracting path (encoder pipeline) of Atten-UNet presented the probable iris bounding box coordinates which were then used as a mask to focus more on the iris region thereby avoiding false segmentation outputs due to the background. The model was tested on UBIRIS.v2 and CASIA-IrisV4- Distance dataset. The mean error rates achieved were 0.76% and 0.38%, respectively. Xia and Kulis [17] proposed a fully unsupervised deep learning network referred to as W-Net model. The model concatenates two U-Nets for dense prediction and reconstruction of the segmentation outputs. Also, post-processing schemes, such as fully connected Conditional Random Field and Hierarchical segmentation were successively employed to provide accurate segmentation edges and merging of over-segmented regions respectively. The model was evaluated on the Berkeley Segmentation Database (BSDS300 and BSDS500). An overlap of 60% and 59% with respect to ground truth was reported for the two datasets. Tong et al. [18] proposed a U-Net model to segment the pulmonary nodules in CT images. Initially, the pulmonary parenchyma was obtained through binary segmentation followed by the use of morphological operators. Later the segmented lung parenchyma is divided into 64x64 cubes and introduced to a modified U-Net model comprising of residual modules instead of plain neural units. The new model provided an improvement in the training speed and also prevented over-fitting. The model presented better segmentation outputs when compared to other segmentation algorithms such as Level set [19] and Graph-cut [20] techniques.

Here we propose an RDAU-NET (Residual Dilated Attention Gate) architecture to segment the lesions in BUS images. Our contributions are as follows: (a) propose a model similar to [21] where residual units replace the plain neural units in the encoder-decoder structure of the U-Net structure to extract more features from the BUS image, (b) addition of dilated convolution model to the end of encoder pipeline to obtain semantic information from a large receptive field and (c) inclusion of Attention Gate(AG) system, in the skip connection part of the encoder-decoder section to suppress the irrelevant information and to effectively improve the sensitivity and prediction accuracy of the model. Figs 2 and 3 illustrates a few of the test images along with the segmentation results realized by the proposed model. The rest of the paper is organized as follows. Section 2 describes the RDAU-NET network structure, Section 3 explains the dataset and the augmentation technique adopted for training the RDAU-NET, Section 4 presents the experimental results followed by discussion and conclusion.

Methods

RDAU-NET network model structure

Our model is based on the U-Net architecture proposed by [22]. It has 6 residual units along the encoder pipeline which extracts the relevant features from the BUS images. Each residual unit includes a pooling operation and therefore presents a downsampled feature map at the end of the encoder pipeline. The smaller feature maps tend to reduce the accuracy of the semantic segmentation, and hence the outputs of the encoder pipeline are fed to a series of dilated convolution module with 3x3 convolution kernels and dilation ratios of 1, 2, 4, 8, 16, 32 respectively. The module outputs feature maps computed from the large receptive field which aid in improving the overall segmentation accuracy. Later, the features maps of dilated convolution module are summed and fed into a decoder pipeline consisting of 5 residual

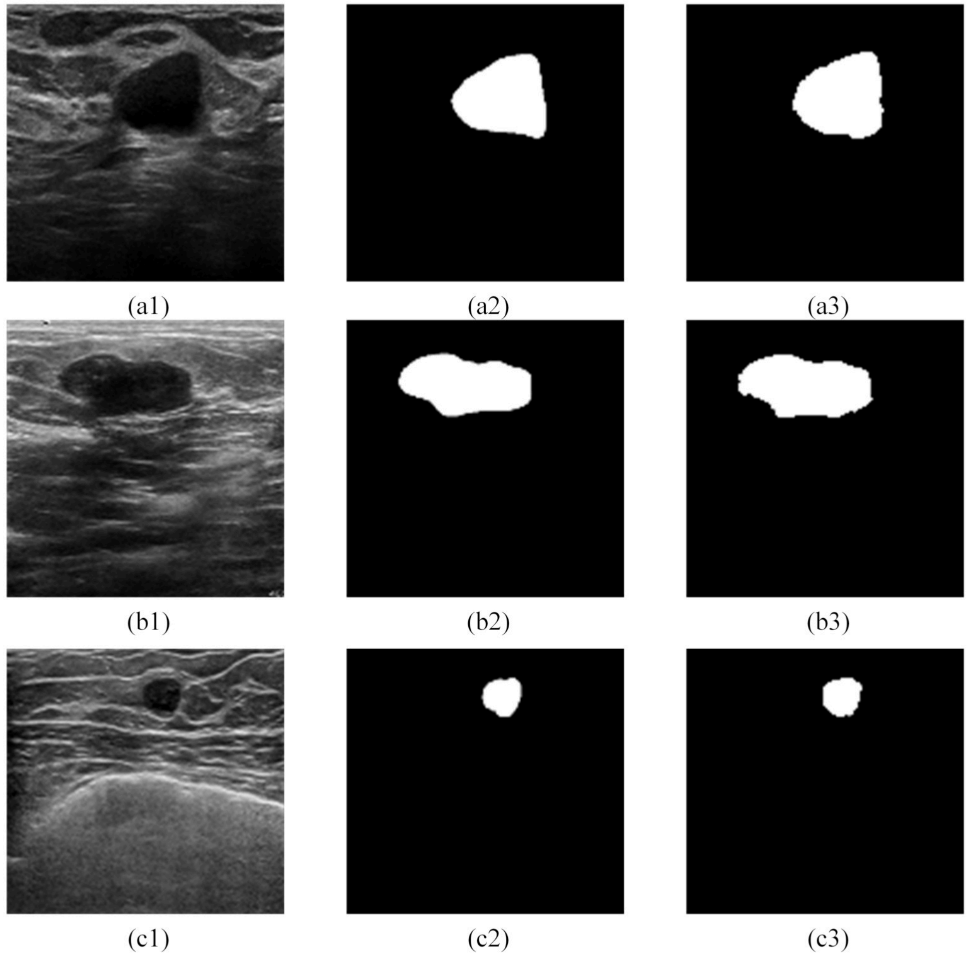


Fig 2. Ultrasound breast tumor segmentation based on the RDAU-NET model. Here (a1)—(c1) were obtained from Dataset B [13]. (a2), (b2), (c2) are gold standard and (a3), (b3), (c3) are the results of the RDAU-NET model segmentation.

https://doi.org/10.1371/journal.pone.0221535.g002

units. The decoders assist in upsampling the feature maps by concatenating the detailed feature outputs of the decoder with the corresponding high-level semantic information of the encoder. Normally traditional U-Net [22] use copying and cropping technique to facilitate the learning process, but we replace them with Attention Gate (AG) module which concentrates on learning the lesions rather than the unnecessary background. Further, the decoder pipeline restores the segmentation outputs to input image resolution and the final 1x1 convolution module presents the classification label of each pixel. Fig 4 illustrates the proposed RDAU-NET model and the following sections explain each module in detail.

Residual network

With the increase in the number of layers, the network will have better learning ability as it progresses. However, during training, as the network starts to converge, the accuracy gets saturated and network performance degrades rapidly due to the problem referred to as the "vanishing gradients". Therefore, we introduce residual units into the U-Net to avoid performance degradation during the training process. He et al. [23] proposed a residual learning correction

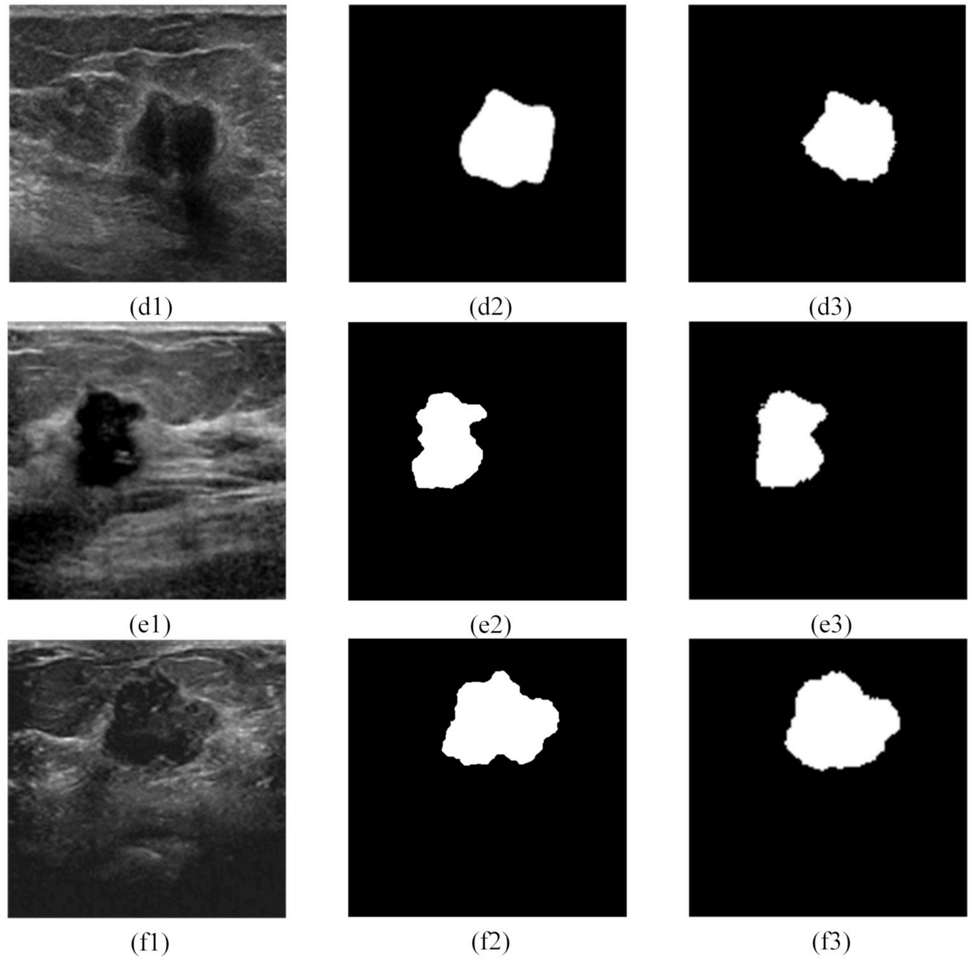


Fig 3. Ultrasound breast tumor segmentation based on the RDAU-NET model. Here(d1)was obtained from Dataset B [13] and (e1) and (f1) were acquired from the Imaging Department of the First Affiliated Hospital of Shantou University. Also(d2), (e2) and (f2) are gold standard and (d3), (e3), (f3) are the results of the RDAU-NET model segmentation.

https://doi.org/10.1371/journal.pone.0221535.g003

scheme to avoid performance degradation which is expressed in Eq (1)

$$y = F(x, \{W_i\}) + x \tag{1}$$

Here x and y are the input and output vectors of the residual block and W_i is the weight of the corresponding layer. The function $F(x, \{W_i\})$ is the residual function which when added to x proved easier to train and learn the features than learning directly from the input x. Also, Eq (1) solved the "vanishing gradient" problem associated with deep networks. By taking the partial derivative of y with respect to x(Eq (2)), we can understand that the partial derivative is always greater than 1, and thus the gradient does not disappear with the increase of the number of layers.

$$\frac{\partial y}{\partial x} = 1 + \frac{\partial F(x, \{W_i\})}{\partial x} \tag{2}$$

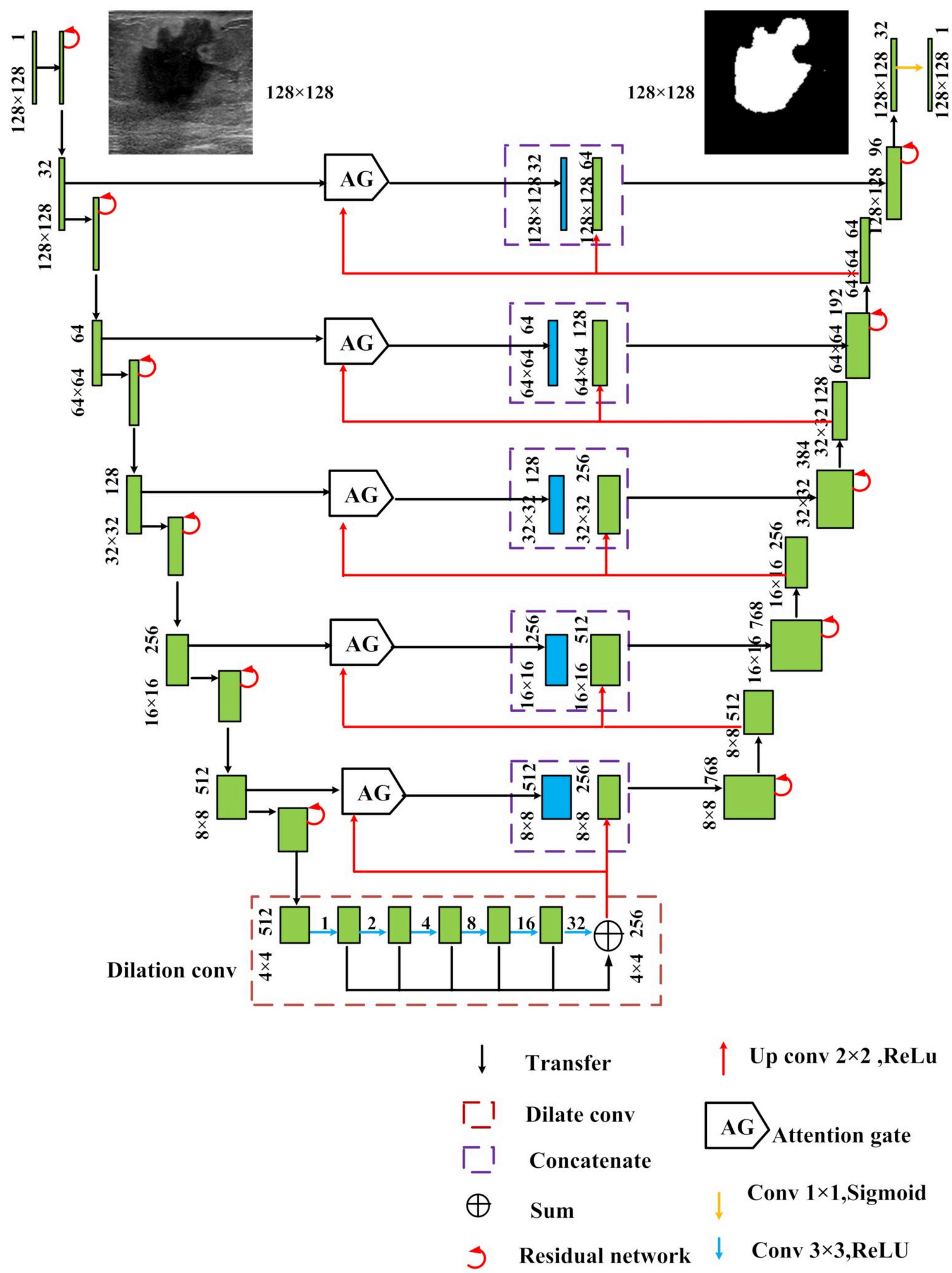


Fig 4. RDAU-NET model structure. The numbers above the boxes (green) indicate that the size of the input along with the number of channels. For example, 128x128 1 indicate the input resolution and the number of channels respectively. The blue box represents the outputs from Attention Gate module.

https://doi.org/10.1371/journal.pone.0221535.g004

Normally $F(x, \{W_i\})$ and x have different dimensions and hence a correction term W_s is added to the input to match the dimension as shown in Eq (3).

$$y = F(x, \{W_i\}) + W_s x \tag{3}$$

In the proposed RDAU-NET model, the inputs to the residual unit of encoder pipeline are effectively convolved with a standard 3x3 kernel and the skip connection with the 1x1 kernel

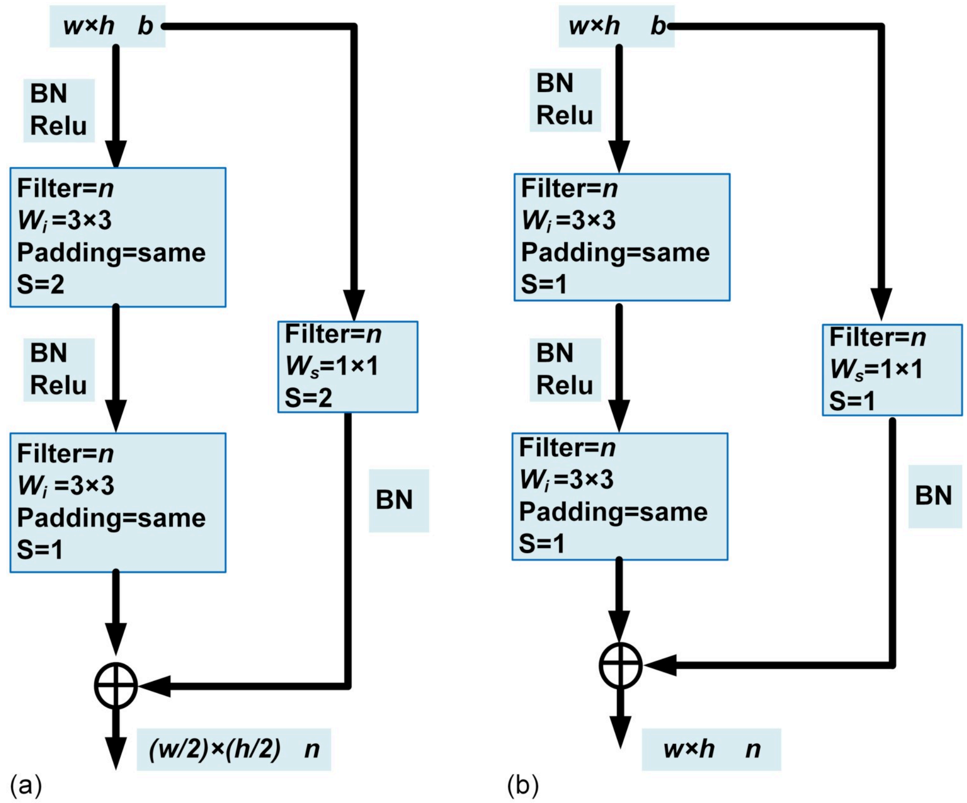


Fig 5. Residual units of encoder and decoder pipeline. (a) Residual units of encoder pipeline. Here *w*, *h*, and *b* represent the width, height, and channels of the input feature map, respectively. BN is batch normalization. Relu is an activation function and *n* is the number of filters. In the encoding process, the values of *n* are 64,128,256,512 and 512 for layers 2,3,4,5,6 respectively. (b)Residual units of decoder pipeline. Here values of *n* are 512, 256, 128, 64 and 32 for layers 5,4,3,2,1 respectively.

https://doi.org/10.1371/journal.pone.0221535.g005

(W_s) was used to match the dimensions of the residual function. A detailed structure of the Residual unit employed in the encoder pipeline is shown in Fig 5(a). Here w x h corresponds to the width and height of the input and b represents the number of channels. Further BN, Relu, and S represent the batch normalization, activation function and the stride length (pooling operation) respectively. Also, *n* denotes the number of filtering operations performed per layer. In our model *n* takes values such as 64,128,256,512 and 512 corresponding to layers 2 to 6 of the encoder pipeline. It should be noted that the S is fixed to 2 for all the layers except for the first residual unit where it is set to 1. The decoder pipeline consists of 5 residual units which emulate the residual units of encoder section with S = '1' to allow the input and output to have the same resolution. Fig 5(b) illustrates the residual unit of the decoder pipeline. The proposed RDAU-NET model avoids performance degradation issues and greatly reduces the difficulty involved in training a deep network. Further, it effectively improves the feature learning ability and is beneficial for the extraction of complex feature patterns of BUS images, thus improving the segmentation results significantly.

Dilation convolution

In CNN architectures, due to convolution and pooling operations, the network present feature maps with less spatial information that affects the overall segmentation accuracy. Since the encoder pipeline of the U-Net represents an FC-CNN (Fully Connected CNN), dilated

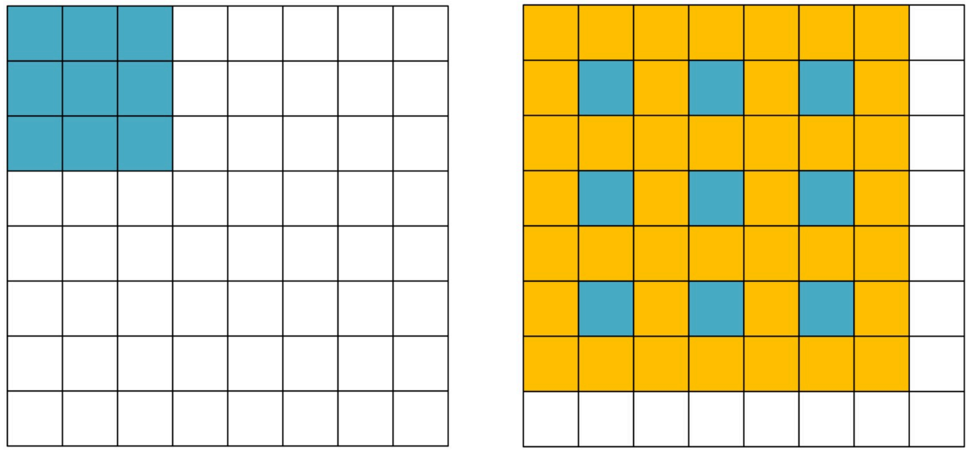


Fig 6. Illustration of receptive fields for r = 1 and r = 2. (a) and (b) illustrate the visual field of a 3x3 convolution kernel with $r = 1$ and $r = 2$ respectively. When $r = 2$, though the kernel parameters remain the same, the receptive field has increased to 7x7 (shown as the orange and blue parts in(b)) when compared to traditional convolution ($r = 1$, as shown in the blue part of (a)). Therefore the dilation process increases the size of the receptive field and compensates for the subsampling.

https://doi.org/10.1371/journal.pone.0221535.g006

convolution modules are often employed in U-Nets [24, 25] to improve the receptive field. Eq (4) illustrates the dilated convolution operation between the input image *f(x,y)* and kernel *g(i,j)*.

$$z(x, y) = \sigma\{\sum_{i,j} f(x + i \times r, y + j \times r) \times g(i, j) + \beta\} \tag{4}$$

Here σ is Relu function, β is a biased unit, and r represents the dilation parameter that controls the size of receptive fields. In general, the size of the receptive field can be expressed as:

$$N = \{(K_{size} + 1) \times (r - 1) + K_{size}\}^2 \tag{5}$$

Where K_{size} is the size of the convolution kernel, r is the dilation parameter, and N is the size of the receptive field, which is illustrated in Fig 6.

In the RDAU-NET model the feature maps of size 4x4 obtained at the end of the encoder pipeline are fed into a series of dilated convolution modules with $r = 1,2,4,8,16,32$ and $N = 3x3$, 7x7, 15x15, 31x31, 63x63 and 127x127 respectively and the outputs of the six convolutions are added, upsampled (by a factor of 2) and then fed into the decoder pipeline as shown in Fig 4. It should be noted that in the dilation convolution module, output feature maps have the same size as that of inputs but contain information from a wide range of receptive fields which greatly improves the feature learning ability of the network as illustrated in Fig 7.

Attention Gate (AG) module

Although dilated convolutions improve the feature learning ability of the network, still there are difficulties in reducing the false predictions of small objects that have large shape variations [26]. This is mainly due to the loss of spatial information in the features maps obtained at the end of the encoder pipeline. In order to improve the accuracy, the existing segmentation framework schemes [27–29] rely on the addition of object positioning models to simplify the task obtaining the spatial attributes. Oktay et al. [26] proposed the attention U-Net network, which integrated the AG module into the U-Net model to realize spatial localization and

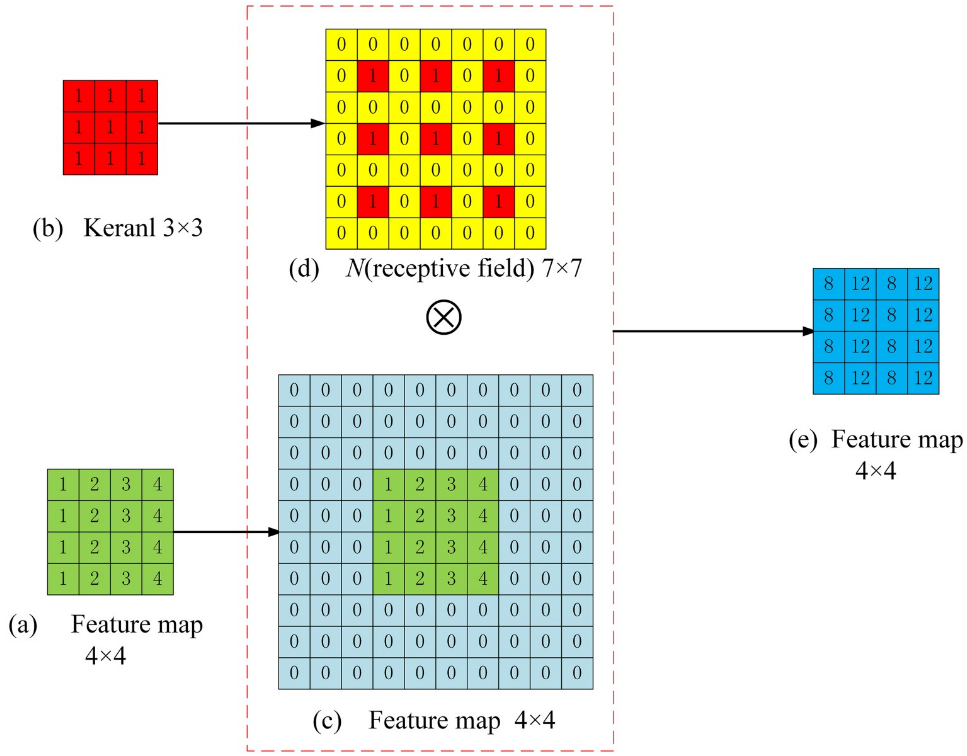


Fig 7. Illustration of the dilated convolution module. Here the dilation parameter r = 2, the stride size S = 1, the input feature map is 4x4, kernel size is 3x3 and receptive filed *N* is 7x7. After processing using dilated convolution, the size of the original feature map remains the same but the receptive fields increases while keeping the parameters of the model intact.

https://doi.org/10.1371/journal.pone.0221535.g007

subsequent segmentation. The AG module eliminated the need for training multiple models which required a large number of additional training parameters. In addition, compared to the positioning model used in multi-level U-Net network, the AG module gradually suppresses the feature response in the irrelevant background regions and strengthens the learning ability of foreground [30].

The AG model derives attention coefficients that aid in improving the segmentation accuracy. Here the coefficients are computed by combining "rich feature maps with low spatial information" obtained from the upsampled decoder layers with the high-level semantic outputs of the corresponding encoder layer. Once the gating coefficients are computed, they are element-wise multiplied with the encoder output to retain the significant activation [31]. The structure of the AG module is shown in Fig 8 and the attention coefficients are computed as

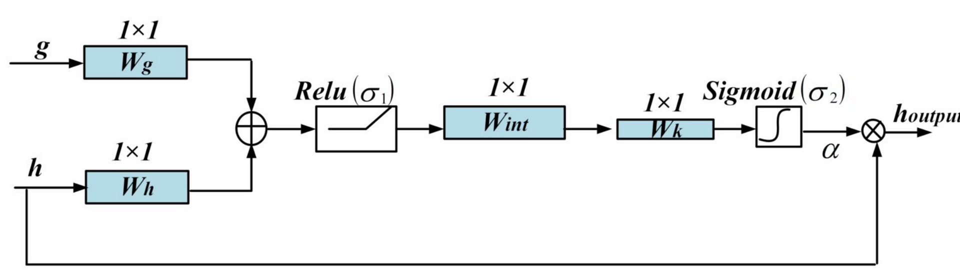


Fig 8. Schematic diagram of the Attention Gate (AG).

https://doi.org/10.1371/journal.pone.0221535.g008

based in Eqs (6) and (7)

$$\alpha = \sigma_2\{W_k[W_{int}(\sigma_1(W_h \times h + W_g \times g + b_{h,g})) + b_{int}] + b_k\} \quad (6)$$

$$\sigma_2(x) = \frac{1}{1 + exp(-x)} \quad (7)$$

Here $\alpha \in [0, 1]$ denote the computed attention coefficients, g and h represent the feature maps presented to the inputs of AG module from the decoder and encoder pipelines respectively and W_g, W_h, W_{int}, W_k indicate the convolution kernels. We choose the kernel size as 1x1 to reduce the number of training parameters and the computational complexity. Also σ_2 is sigmoid activation function which limits the range between 0 and 1 and σ_1 is the Relu function. Here sigmoid was chosen over softmax since it provided dense activations at the output [26]. AG module outputs the constructive features through elementwise multiplication of α with the corresponding encoder layer output as given by Eq (8).

$$h_{output} = \alpha \times h \quad (8)$$

The output of the AG module (h_{output}) filters out the irrelevant context information and aggravate the useful feature information that effectively improves the sensitivity and prediction accuracy of the model. Further, when compared to [31], since the h and g are of the same resolution, our AG module eliminates the need for the computationally intensive interpolation operation and thus operate faster with less memory requirement.

Materials

Data collection

This study considered a total of 1062 BUS images obtained from three different sources: (a) GelderseVallei Hospital in Ede, the Netherlands [14], (b) First Affiliated Hospital of Shantou University, Guangdong Province, China, and (c) BUS images obtained from Breast Ultrasound Lesions Dataset (Dataset B) [13]. The performance evaluation was based on cross-validation where the training set was used to train the proposed RDAU-NET model and the validation set was considered for fine-tuning the parameters. The optimized model was tested for segmentation performance and generalization ability using the samples of the testing set. For training and validation, we used the BUS images from [14]. The training and validation set contained 730 and 127 samples respectively. The test set consisted of 205 samples comprised of 163 BUS images obtained from Dataset B [13]and 42 BUS images provided by the Imaging Department of the First Affiliated Hospital of Shantou University. The BUS images obtained from Shantou First Affiliated Hospital were acquired using the GE Voluson E10 Ultrasound Diagnostic System(L11-5 50mm broadband linear array transducer, 7.5MHz frequency) The training, validation and the test images contained both malignant and benign BUS lesions. The RDAU-NET structure proposed in the work uses Keras (2.1.6) framework and calls Tensorflow (1.11.0). The entire model was executed using GPU TITAN XP with operating system Ubuntu version 14.04, CUDA version 9.0, cuDNN version of 7.1.2 and graphics card's memory of 12GB.

Data processing

To accomplish the task of cross-validation on the dataset, the training dataset had to be labeled. The BUS images from [14] were manually segmented and labeled(ground truth) by the

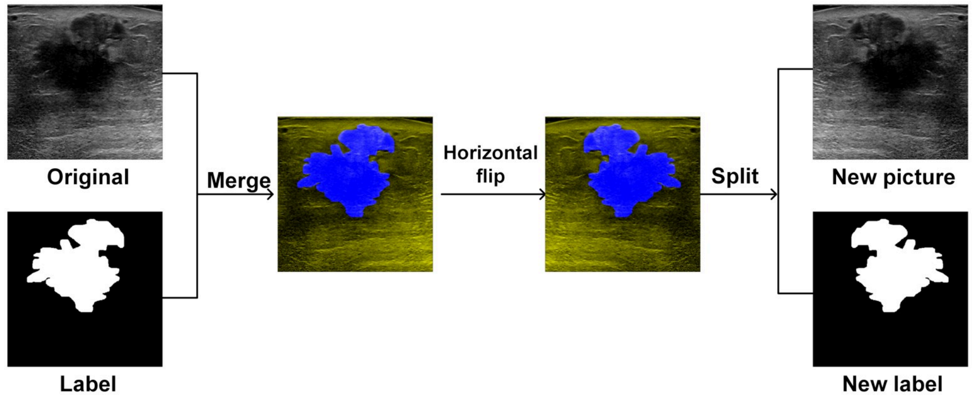


Fig 9. Data augmentation illustrating horizontal flipping to expand the training dataset.

https://doi.org/10.1371/journal.pone.0221535.g009

specialist with more than 7 years of experience at the First Affiliated Hospital of Shantou University. To achieve a good segmentation under the limited number of training samples, data augmentation [32] was performed to expand the training data set. Here we first merge the BUS images and their ground truths together and then perform four affine transformations (shift along the vertical axis, shift along the horizontal axis, shear transformation and flipping about the horizontal plane) to obtain a new transformed image. Later the transformed image and its new ground truth are separated and appended to the training set as additional training images. Thus 730 images of the training set were expanded to obtain 2919 images. The data augmentation process is pictorially illustrated in Fig 9.

Results and discussion

Two separate experiments were performed to illustrate the effectiveness of the proposed RDAU-NET model: (a) The best input image resolution that can provide a good qualitative and quantitative segmentation results when used with RDAU-NET model (b) Performance evaluation of the segmentation outputs of RDAU-NET model over FCN8s, FCN16s [33], U-Net [22], SegNet [34], Residual U-Net [35], Squeeze U-Net [36], Dilated U-Net [37], RAU-NET (Residual-Attention-UNet), DAU-NET(Dilated-Attention-UNet), and RDU-NET (Residual-dilated-UNet), The performance of the segmentation outputs was evaluated using the 9 evaluation indices: Accuracy(Acc), Precision(Pc), Recall, Dice coefficient(DC), Mean-Intersection-Over-Union(M-IOU), Area-Under-Curve(AUC), Sensitivity(Sen), Specificity (Sp) and F1score(F1). These performance indicators were computed as follows:

1. Dice coefficient [38]: It represents the degree of similarity between the segmented output of the proposed model and the gold standard. The higher the similarity between the tumor region and the gold standard, the greater is the Dice coefficient and better the segmentation result. The Dice coefficient is calculated as,

$$Dice\ \ Coefficient(DC) = \frac{2 \times (X \cap Y)}{(X + Y)} \tag{9}$$

   Also, the dice coefficient loss (Dice_loss) is the loss and is computed as follows.

$$Dice\ \ Coefficient\ \ Loss(Loss) = 1.0 - \frac{2 \times (X \cap Y)}{(X + Y)} \tag{10}$$

   where, X is the gold standard, which is the average result marked by two clinical experts, Y

Table 1. Definition of the abbreviations.

Category	Actual lesion	Actual non-lesion
Predicted lesion	True Positive(TP)	False Positive(FP)
Predicted non-lesion	False Negative(FN)	True Negative(TN)

https://doi.org/10.1371/journal.pone.0221535.t001

is the tumor area segmented by the model and $X \cap Y$ represents the area of overlap between the gold standard and the segmented output of the model.

2. Mean-Intersection-over-Union(M-IOU) [39]: is defined as the average ratio between the intersection and union of the gold standard and the segmented output of the model. It provides a measures coincidence between the gold standard and the segmented output of proposed the model. Higher the coincidence, greater is the M-IOU and better is the segmentation accuracy M-IOU is expressed as follows. Where N is the number of IOU.

$$IOU = \frac{X \cap Y}{X \cup Y} \tag{11}$$

$$M-IOU = \frac{\sum_{i=1}^{N} IOU_i}{N} \tag{12}$$

3. Performance indicators that are obtained from the confusion matrix: The Accuracy, Precision, Sensitivity, Specificity, and F1 score. These are associated with true positive (TP), true negative (TN), false positive (FP) and false negative (FN) of the confusion matrix. Here we have explained them in the Table 1. TP, FP, FN and TN are the numbers of pixels corresponding to the four categories and the formula of performance indicators is shown in Table 2.

4. The area under the curve (AUC): AUC is the area under the receiver operating characteristic (ROC) curve. It represents the degree or the measure of separability and indicates the capability of the model in distinguishing the classes. Higher the AUC better is the segmentation output and hence the model.

Qualitative and quantitative analysis of the RDAU-NET model for different input image resolutions

As a preliminary experiment, the segmentation task on BUS images was performed with 4 different network input sizes of 64x64, 96x96, 128x128 and 256x256 pixels. During the experiment, the number of training epochs was set to 300 and the batch size for 64x64, 96x96,

Table 2. The formula of performance measure.

Performance Measure	Formula	Description
Accuracy(Acc)	$\frac{TP+TN}{TP+FP+FN+TN}$	A ratio of the number of correctly predicated pixels to the total number of pixels in the image
Precision(Pc)	$\frac{TP}{TP+FP}$	A ratio of the number of correctly predicated lesion pixels to the total number of predicted lesion pixels
Recall/sensitivity(Sen)	$\frac{TP}{TP+FN}$	A ratio of the number of correctly predicated lesion pixels to the total number of actual lesion pixels
F1score(F1)	$2 \times \frac{Precision \times Sensitivity}{Precision+Sensitivity}$	A measure for the accuracy
Specificity(Sp)	$\frac{TN}{TN+FP}$	A ratio of the number of correctly predicated non-lesion pixels to the total number of actual non-lesion pixels

https://doi.org/10.1371/journal.pone.0221535.t002

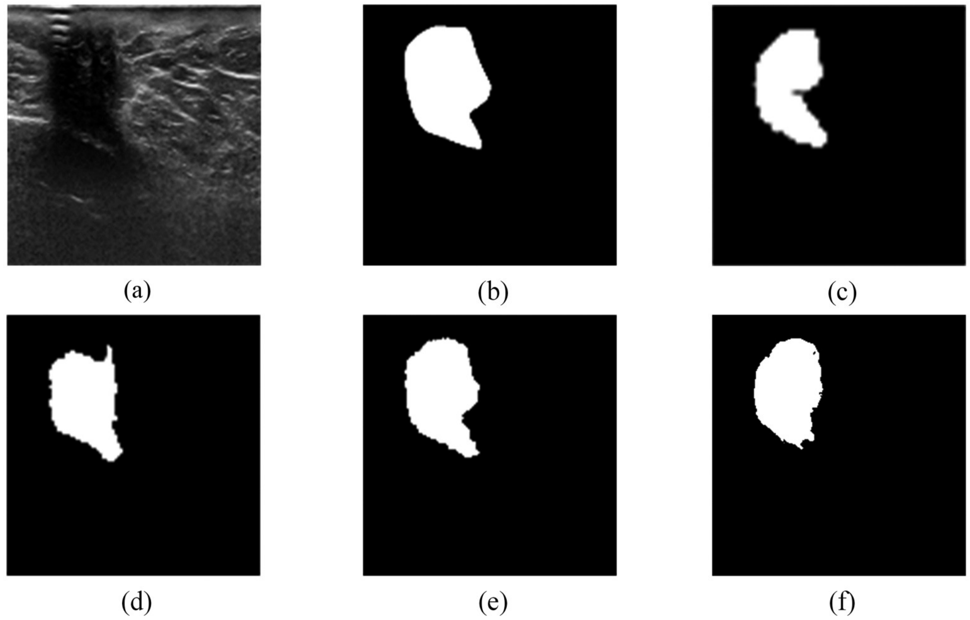


Fig 10. **Sample image from Dataset B [13].** (a) image of malignant invasive ductal carcinoma. (b) Gold standard. (c-f) are the results of segmentation for input sizes are 64x64, 96x96, 128x128, and 256x256 respectively.

https://doi.org/10.1371/journal.pone.0221535.g010

128x128 was selected as 32 while the batch size of 256x256 was 16. The batch sizes were mainly chosen to reduce the computational overhead and satisfy the memory requirements. The segmentation results of the experiment are shown in Fig 10 and Table 3 illustrates the performance metrics computed using Eqs (9) to (12) and Table 2. From Fig 10(a) to Fig 10(f), it can be seen intuitively that the best automatic segmentation results were obtained for the input image size of 128x128 pixels (Fig 10(e)). Also, the performance metrics (Table 3) emphasize that the maximum values are obtained for the input size of 128x128 pixels. In terms of computation time, though the inputs of size 64x64 pixels presented the least time, their segmentation results were not accurate. Therefore, the experiments were on focused on using 128x128 as the input image resolution for further evaluations and comparisons.

Performance evaluation of the segmentation outputs of the RDAU-NET with other models

Qualitative comparison with other models. For the qualitative performance comparison, the segmentation results of FCN8s, FCN16s, SegNet, U-Net, Residual U-Net, Squeeze U-Net, Dilated U-Net, RAU-NET, DAU-NET, RDU-NET, and RDAU-NET models are presented in Figs 11–13. In all these cases the input images that were tested were of size 128x128 and segmented outputs are of the same size.

Table 3. Quantitative evaluation of BUS images of different input sizes.

Input Size	Loss	Acc	DC	Sen	Sp	F1	Pc	M-IOU	AUC	Train(min)
64 × 64	0.2033	0.9758	0.7966	0.7921	0.9914	0.7968	0.8471	0.7863	0.9094	75
96 × 96	0.1713	0.9775	0.8286	0.8232	0.9920	0.8291	0.8669	0.8019	0.9186	105
128 × 128	0.1530	0.9791	0.8469	0.8319	0.9934	0.8478	0.8858	0.8067	0.9227	140
256 × 256	0.1664	0.9668	0.8335	0.8208	0.9935	0.8403	0.8807	0.7992	0.9147	435

https://doi.org/10.1371/journal.pone.0221535.t003

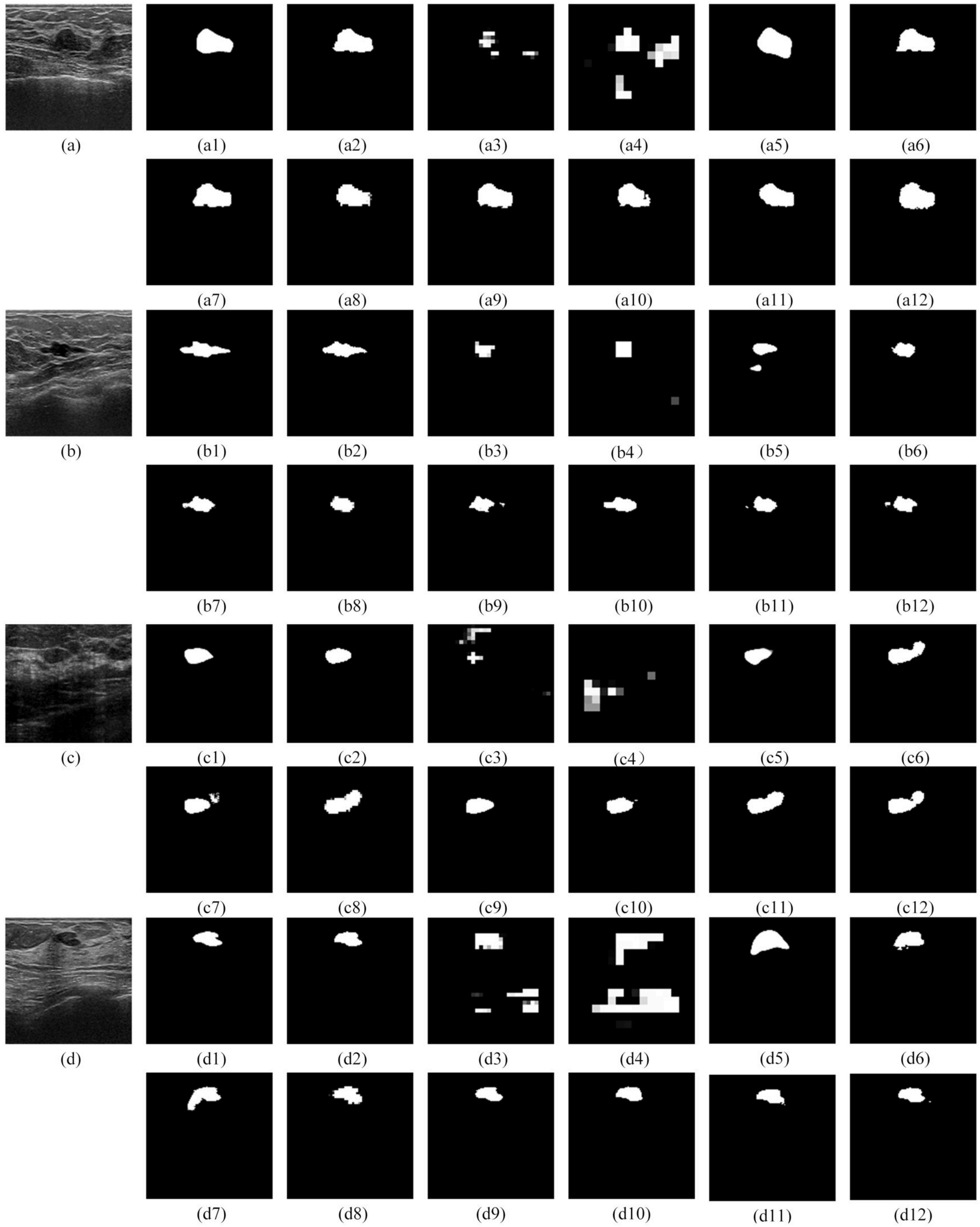


Fig 11. Segmentation outputs for the BUS images from the test dataset. The test dataset was obtained from Dataset B. Fig 11(a—d) illustrate the results for test images obtained from Dataset B. (a1), (b1), (c1), (d1) are the gold standard. (a2)—(a12), (b2)—(b12), (c2)—(c12), (d2)—(d12) are the segmentation results from RDAU-NET, FCN8s, FCN16s, SegNet, U-Net, Residual U-Net, Squeeze U-Net, Dilated U-Net, RAU-NET, DAU-NET, RDU-NET respectively.

https://doi.org/10.1371/journal.pone.0221535.g011

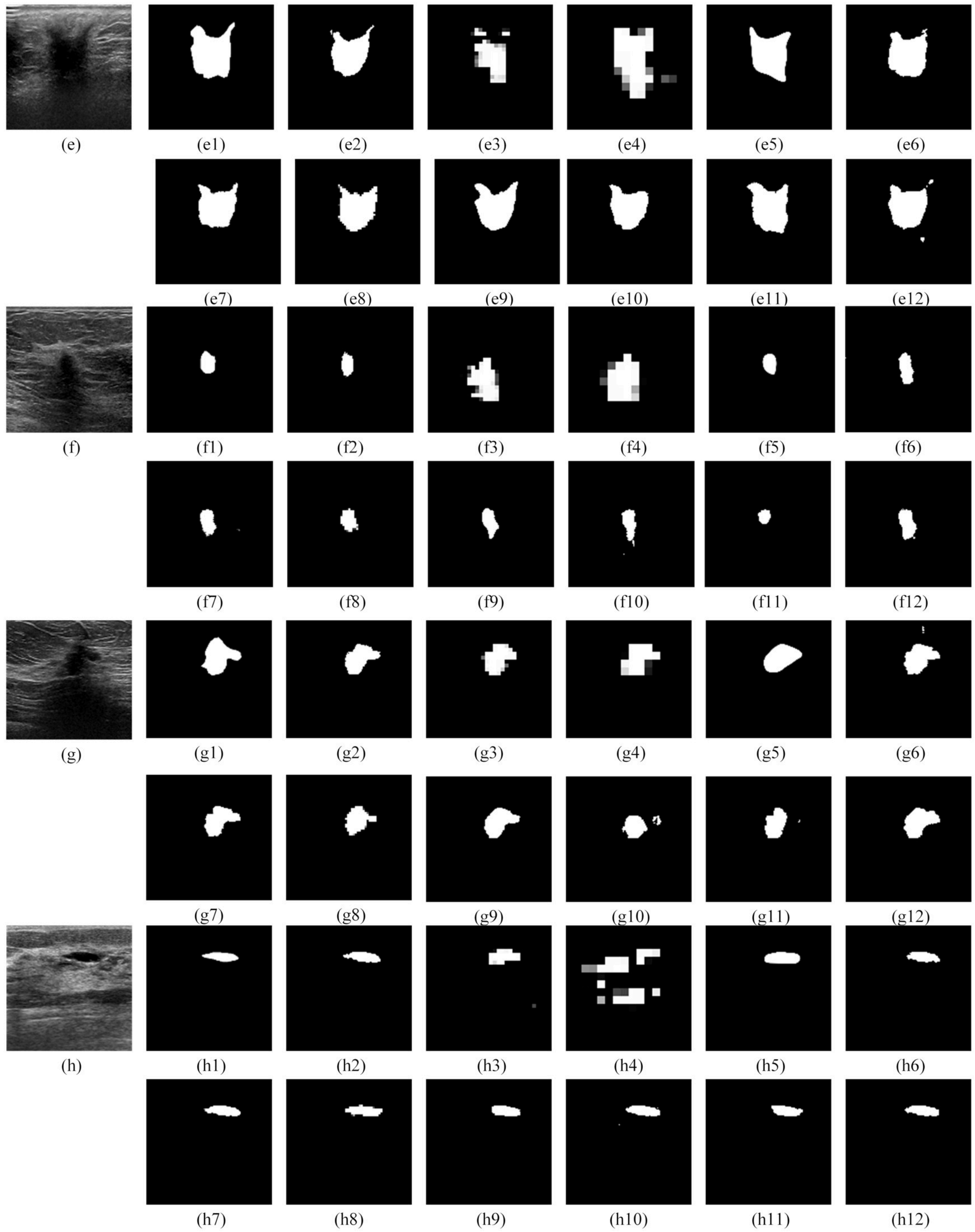


Fig 12. Segmentation outputs for the BUS images from the test dataset. The test dataset was obtained from Dataset B. Fig 12(e—h) illustrate the results for test images obtained from Dataset B. (e1), (f1), (g1), (h1) are the gold standard. (e2)—(e12), (f2)—(f12), (g2)—(g12), (h2)—(h12) are the segmentation results from RDAU-NET, FCN8s, FCN16s, SegNet, U-Net, Residual U-Net, Squeeze U-Net, Dilated U-Net, RAU-NET, DAU-NET, RDU-NET respectively.

https://doi.org/10.1371/journal.pone.0221535.g012

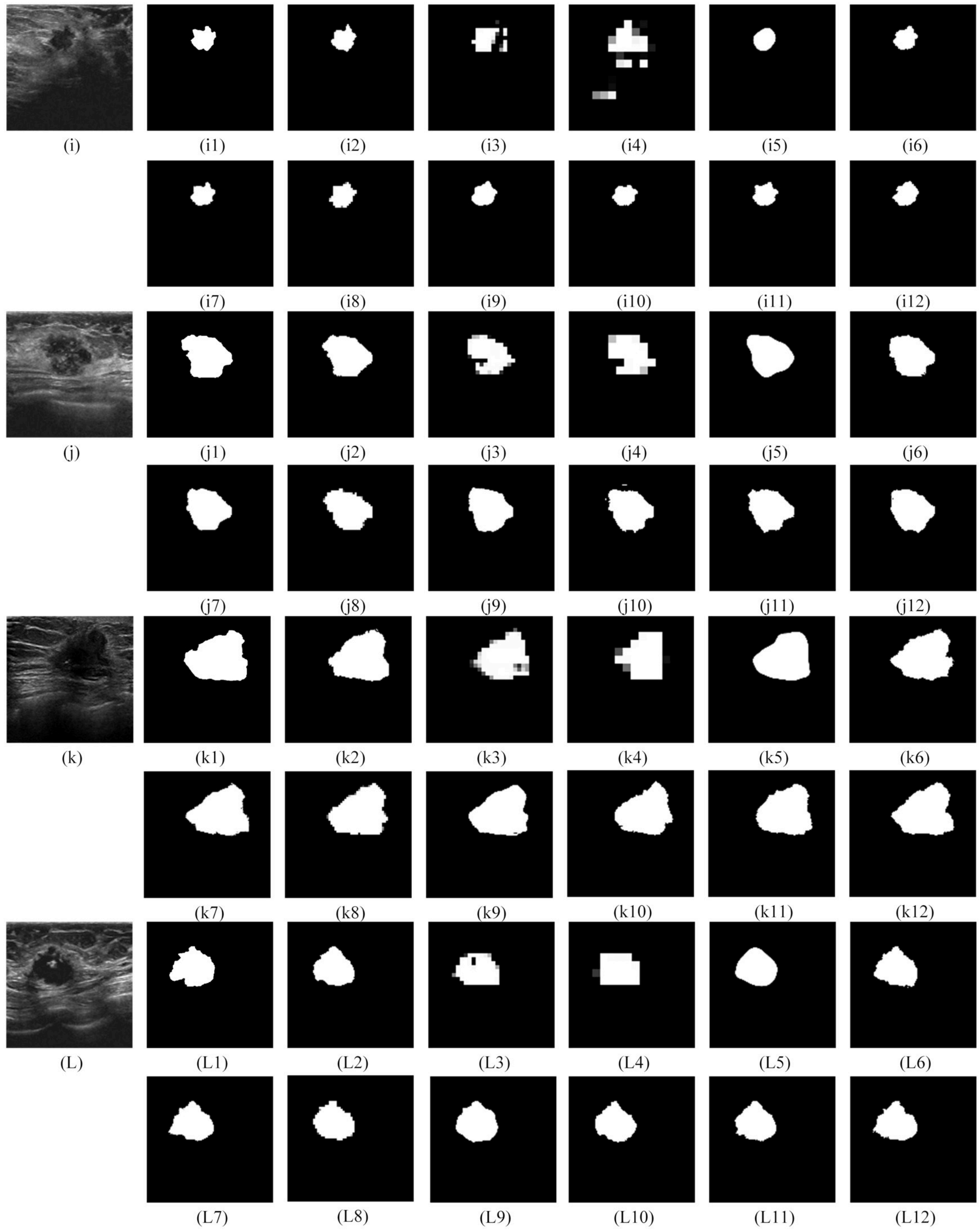


Fig 13. Segmentation outputs for the BUS images from the test dataset. The test dataset was obtained from the Imaging Department of the First Affiliated Hospital of Shantou University. Fig 13(i—L) represents the outputs for the test images acquired from Imaging Department of the First Affiliated Hospital of Shantou University. (i1), (j1), (k1) and (L1) are the gold standard. (i2)—(i12), (j2)—(j12), (k2)—(k12) and (L2)—(L12) are the segmentation results from RDAU-NET, FCN8s, FCN16s, SegNet, U-Net, Residual U-Net, Squeeze U-Net, Dilated U-Net, RAU-NET, DAU-NET, RDU-NET respectively.

https://doi.org/10.1371/journal.pone.0221535.g013

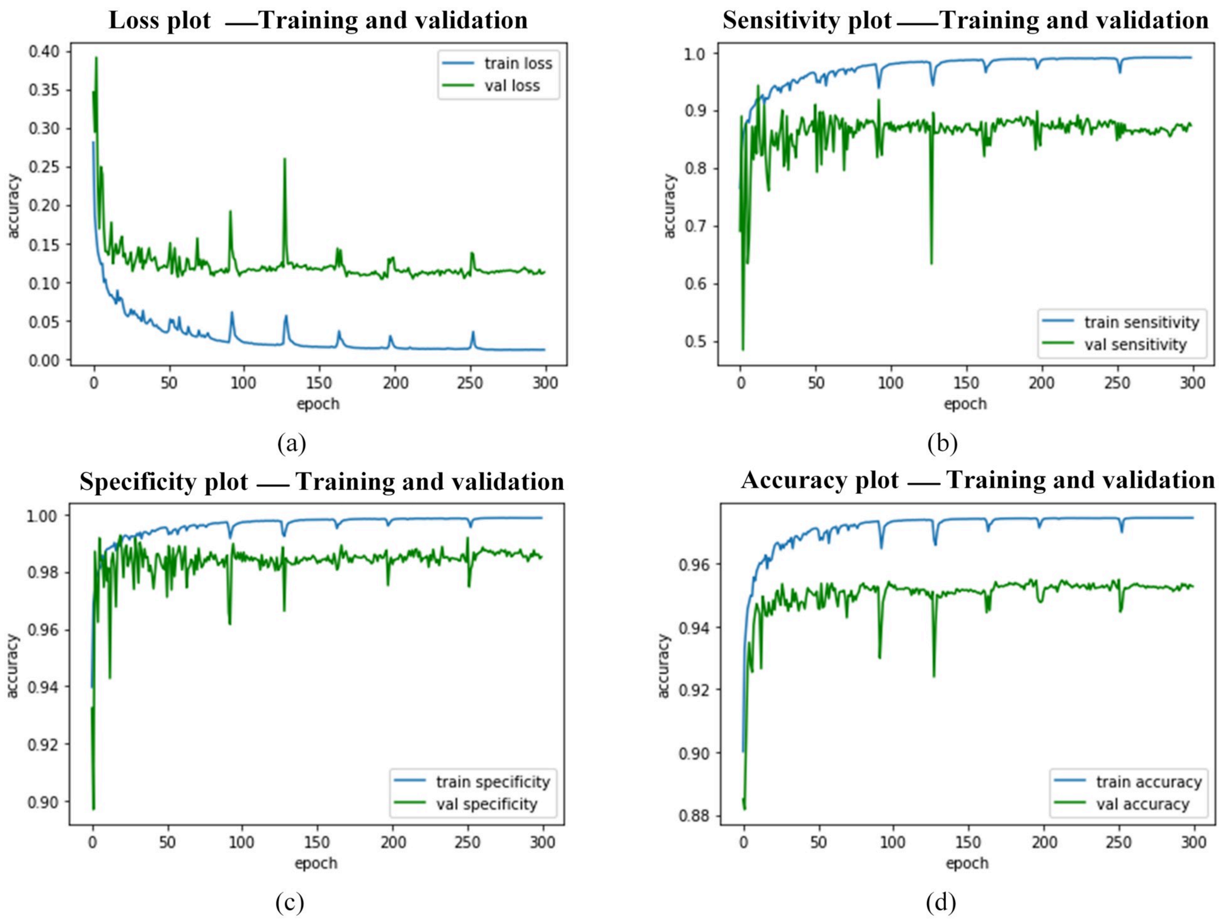


Fig 14. RDAU-NET performance indicators for training, validation. Here the plots (a—d) represent the performance metrics during training and validation.

https://doi.org/10.1371/journal.pone.0221535.g014

The qualitative comparison presents the following conclusions:

1. The segmentation results of FCN8s and FCN16s are rough, with details being neglected, especially at the edges which show jagged contours leading to poor segmentation outputs.
2. Squeeze U-Net, RAU-NET, DAU-NET, RDU-NET present segmentation outputs better than SegNet and U-Net models.
3. The RDAU-NET model presents visually better segmentation results than other models and the final segmentation outputs are close to the gold standards. Also, the segmentation outputs of RDAU-NET model are superior when compared to Residual U-Net and Dilated U-Net.

Further, Figs 14–16 presents the performance curves obtained during the simulation the RDAU-NET during training, validation and testing process.

Quantitative comparison with other models. For the quantitative evaluation, a comparison was performed based on Eqs (9) to (12) between the segmented results of the proposed model and those obtained for the FCN8s, FCN16s, SegNet, U-Net, Residual U-Net, Squeeze U-Net, Dilated U-Net, RAU-NET, DAU-NET, RDU-NET, and RDAU-NET. The evaluation results are tabulated in Table 4.

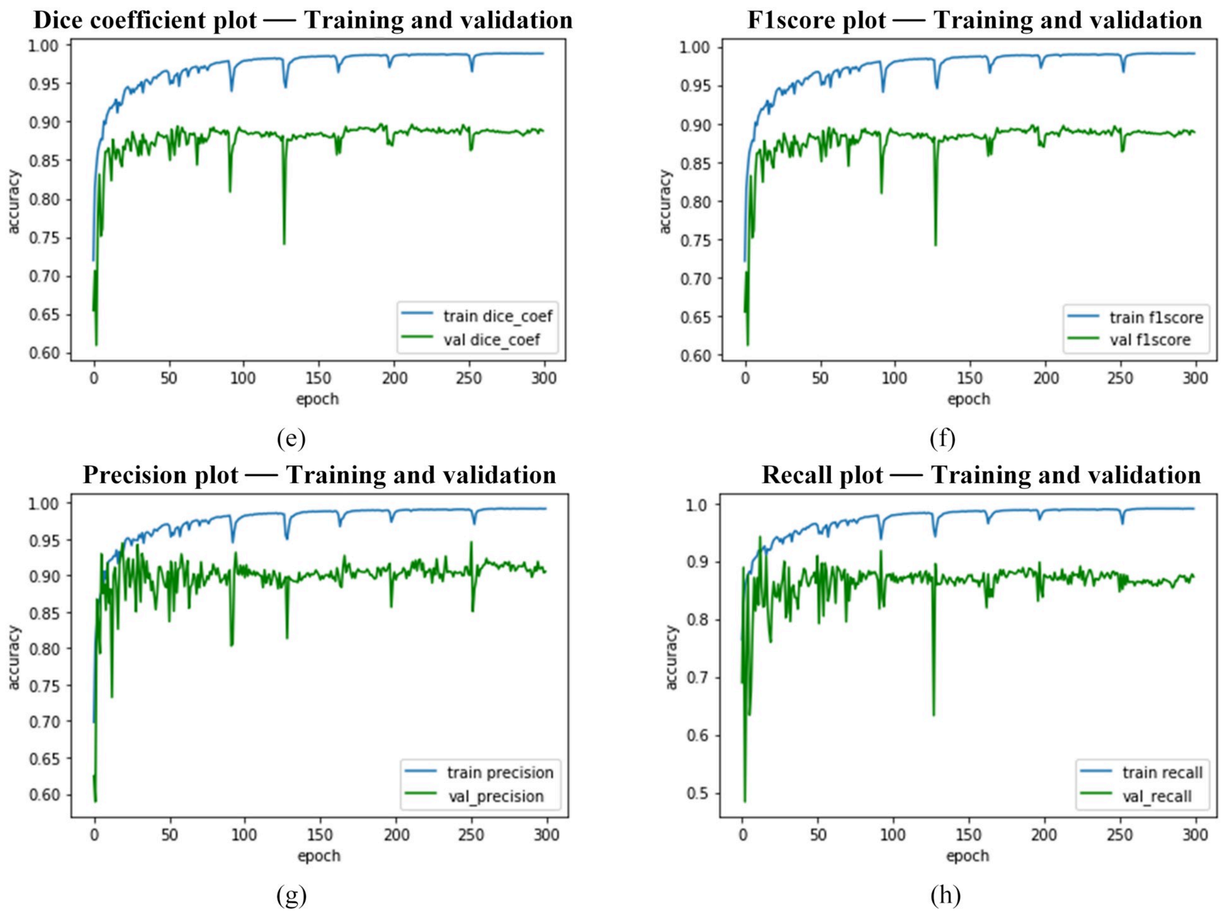


Fig 15. RDAU-NET performance indicators for training, validation. Here the plots (e—h) represent the performance metrics during training and validation.

https://doi.org/10.1371/journal.pone.0221535.g015

The quantitative comparison presents the following conclusions from Table 4:

1. The segmentation performance of traditional U-Net is better than FCN8s, FCN16s, SegNet.
2. The segmentation results are comparatively better for Residual U-Net, Squeeze U-Net, and Dilated U-Net when compared with traditional U-Net. The improvement can be attributed to additional modules that are integrated into U-Net architecture.
3. In most of the evaluation parameters, the proposed RDAU-NET outperforms other models, and thus combing the three modules (Residual unit, Dilation unit, and Attention Gate) has provided accurate segmentation of lesions in BUS images.

Conclusion

Though U-Net is a widely used model in medical image segmentation, it has not achieved the expected outcomes in BUS tumor segmentation. This is mainly due to the high noise, low contrast and weak boundary of ultrasound images. Therefore to achieve accurate segmentation, the model requires more powerful feature extraction and classification abilities. The new model, RDAU-NET proposed in the paper employed residual units dilated convolution and attention gate system top effectively segment the tumor region in BUS images.

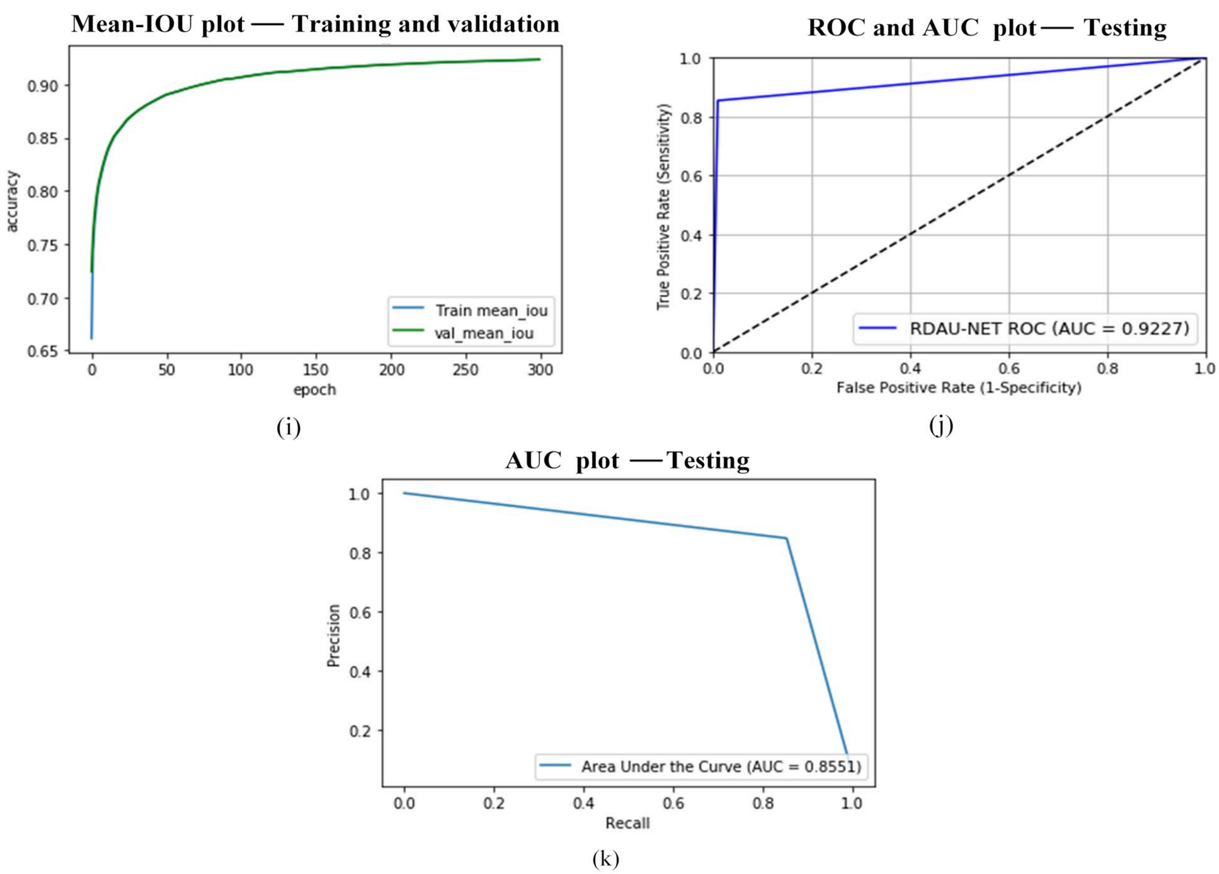


Fig 16. RDAU-NET performance indicators for training, validation, and testing. Here the plots (i) represent the performance metrics during training and validation and plots (j and k) specify the performance during testing: Fig 16(j) denotes ROC curve and AUC with respect to True Positive Rate and False Positive Rate and Fig 16(k) illustrate the AUC in relation to Precision and Recall.

https://doi.org/10.1371/journal.pone.0221535.g016

The experimental results show that the RDAU-NET model can accurately and efficiently segment the tumor region, and the final test results are superior to the traditional convolution neural network segmentation models, and hence has a great prospect for clinical application.

Table 4. Quantitative segmentation results for different models based on the testing dataset.

Experimental Model	Loss	Acc	DC	Sen	Sp	F1	Pc	M-IOU	AUC
FCN8s	0.3676	0.9530	0.6323	0.7040	0.9729	0.6333	0.6085	0.7013	0.9500
FCN16s [33]	0.4507	0.9348	0.5492	0.7018	0.9528	0.5498	0.4842	0.6642	0.9147
SegNet [34]	0.1829	0.9752	0.8170	0.8395	0.9883	0.8171	0.8141	0.7914	0.9276
U-Net [22]	0.1795	0.9757	0.8204	0.8466	0.9891	0.8211	0.8185	0.7983	0.9269
Residual U-Net [35]	0.1746	0.9778	0.8253	0.8165	0.9930	0.8255	0.8670	0.7933	0.9181
Squeeze U-Net [36]	0.2077	0.9745	0.7922	0.7801	0.9909	0.7924	0.8425	0.7863	0.9301
Dilated U-Net [37]	0.1905	0.9740	0.8094	0.8433	0.9877	0.8098	0.8084	0.7784	0.9487
RAU-NET	0.1925	0.9768	0.8074	0.7847	0.9929	0.8081	0.8680	0.8023	0.9070
DAU-NET	0.1576	0.9781	0.8423	0.8392	0.9925	0.8431	0.8659	0.8035	0.9210
RDU-NET	0.1650	0.9784	0.8349	0.8107	0.9936	0.8356	0.8896	0.8087	0.9148
RDAU-NET	0.1530	0.9791	0.8469	0.8319	0.9934	0.8478	0.8858	0.8067	0.9227

https://doi.org/10.1371/journal.pone.0221535.t004

Acknowledgments

We would like to acknowledge the following for providing the Ultrasound Datasets.

Department of Radiology, Gelderse Vallei Hospital, Ede, the Netherlands.

UDIAT-Centre Diagnostic, Corporacio Parc Tauli Sabadell (Spain)—Dr. Robert Marti and Dr. Moi Hoon Yap, Principal authors of the paper "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks" IEEE journal of biomedical and health informatics. DOI: 10.1109/JBHI.2017.2731873 for providing us the Breast Ultrasound Lesions Dataset (Dataset B).

Dr. Shunmin Qiu, Imaging Department, First Hospital of Medical College of Shantou University, Shantou, Guangdong, China.

Author Contributions

Conceptualization: Zhemin Zhuang, Nan Li.

Formal analysis: Zhemin Zhuang, Nan Li, Alex Noel Joseph Raj.

Funding acquisition: Zhemin Zhuang.

Methodology: Zhemin Zhuang, Nan Li, Alex Noel Joseph Raj.

Resources: Shunmin Qiu.

Software: Nan Li.

Validation: Vijayalakshmi G. V. Mahesh, Shunmin Qiu.

Writing – original draft: Nan Li.

Writing – review & editing: Zhemin Zhuang, Alex Noel Joseph Raj, Vijayalakshmi G. V. Mahesh.

References

1. DeSantis CE, Siegel RL, Sauer AG, Miller KD, Fedewa SA, Alcaraz KI, et al. Cancer statistics for African Americans, 2016: progress and opportunities in reducing racial disparities. CA: a cancer journal for clinicians. 2016; 66(4):290–308.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA: a cancer journal for clinicians. 2015; 65(2):87–108.
3. Horsch K, Giger ML, Venta LA, Vyborny CJ. Automatic segmentation of breast lesions on ultrasound. Medical Physics. 2001; 28(8):1652–1659. https://doi.org/10.1118/1.1386426 PMID: 11548934
4. Xu Y, Nishimura T. Segmentation of breast lesions in ultrasound images using spatial fuzzy clustering and structure tensors. World Academy of Science, Engineering and Technology. 2009; 53.
5. Gomez W, Leija L, Alvarenga A, Infantosi A, Pereira W. Computerized lesion segmentation of breast ultrasound based on marker-controlled watershed transformation. Medical physics. 2010; 37(1):82–95. https://doi.org/10.1118/1.3265959 PMID: 20175469
6. Daoud MI, Baba MM, Awwad F, Al-Najjar M, Tarawneh ES. Accurate segmentation of breast tumors in ultrasound images using a custom-made active contour model and signal-to-noise ratio variations. In: 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems. IEEE; 2012. p. 137–141.
7. Virmani J, Agarwal R, et al. Assessment of despeckle filtering algorithms for segmentation of breast tumours from ultrasound images. Biocybernetics and Biomedical Engineering. 2019; 39(1):100–121. https://doi.org/10.1016/j.bbe.2018.10.002
8. Chan TF, Vese LA. Active contours without edges. IEEE Transactions on image processing. 2001; 10 (2):266–277. https://doi.org/10.1109/83.902291 PMID: 18249617
9. Daoud MI, Atallah AA, Awwad F, Al-Najjar M, Alazrai R. Automatic superpixel-based segmentation method for breast ultrasound images. Expert Systems with Applications. 2019; 121:78–96. https://doi.org/10.1016/j.eswa.2018.11.024

10. Panigrahi L, Verma K, Singh BK. Ultrasound image segmentation using a novel multi-scale Gaussian kernel fuzzy clustering and multi-scale vector field convolution. Expert Systems with Applications. 2019; 115:486–498. https://doi.org/10.1016/j.eswa.2018.08.013

11. Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on pattern analysis and machine intelligence. 1990; 12(7):629–639. https://doi.org/10.1109/34.56205

12. Zhuang Z, Lei N, Raj ANJ, Qiu S. Application of fractal theory and fuzzy enhancement in ultrasound image segmentation. Medical & biological engineering & computing. 2019; 57(3):623–632. https://doi.org/10.1007/s11517-018-1907-z

13. Yap MH, Pons G, Martí J, Ganau S, Sentís M, Zwiggelaar R, et al. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE journal of biomedical and health informatics. 2018; 22(4):1218–1226. https://doi.org/10.1109/JBHI.2017.2731873 PMID: 28796627

14. http://ultrasoundcases.info/category.aspx?cat=67/.

15. Xu Y, Wang Y, Yuan J, Cheng Q, Wang X, Carson PL. Medical breast ultrasound image segmentation by machine learning. Ultrasonics. 2019; 91:1–9. https://doi.org/10.1016/j.ultras.2018.07.006 PMID: 30029074

16. Lian S, Luo Z, Zhong Z, Lin X, Su S, Li S. Attention guided U-Net for accurate iris segmentation. Journal of Visual Communication and Image Representation. 2018; 56:296–304. https://doi.org/10.1016/j.jvcir.2018.10.001

17. Xia X, Kulis B. W-net: A deep model for fully unsupervised image segmentation. arXiv preprint arXiv:171108506. 2017.

18. Tong G, Li Y, Chen H, Zhang Q, Jiang H. Improved U-NET network for pulmonary nodules segmentation. Optik. 2018; 174:460–469. https://doi.org/10.1016/j.ijleo.2018.08.086

19. Schildkraut J, Prosser N, Savakis A, Gomez J, Nazareth D, Singh A, et al. Level-set segmentation of pulmonary nodules in megavolt electronic portal images using a CT prior. Medical physics. 2010; 37 (11):5703–5710. https://doi.org/10.1118/1.3495538 PMID: 21158282

20. Ye X, Beddoe G, Slabaugh G. Graph cut-based automatic segmentation of lung nodules using shape, intensity, and spatial features. In: Workshop on Pulmonary Image Analysis, MICCAI. Citeseer; 2009.

21. Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net. IEEE Geoscience and Remote Sensing Letters. 2018; 15(5):749–753. https://doi.org/10.1109/LGRS.2018.2802944

22. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer; 2015. p. 234–241.

23. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–778.

24. Chen Y, Guo Q, Liang X, Wang J, Qian Y. Environmental sound classification with dilated convolutions. Applied Acoustics. 2019; 148:123–132. https://doi.org/10.1016/j.apacoust.2018.12.019

25. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:151107122. 2015.

26. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: learning where to look for the pancreas. arXiv preprint arXiv:180403999. 2018.

27. Khened M, Kollerathu VA, Krishnamurthi G. Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Medical image analysis. 2019; 51:21–45. https://doi.org/10.1016/j.media.2018.10.004 PMID: 30390512

28. Roth HR, Oda H, Hayashi Y, Oda M, Shimizu N, Fujiwara M, et al. Hierarchical 3D fully convolutional networks for multi-organ segmentation. arXiv preprint arXiv:170406382. 2017.

29. Roth HR, Lu L, Lay N, Harrison AP, Farag A, Sohn A, et al. Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. Medical image analysis. 2018; 45:94–107. https://doi.org/10.1016/j.media.2018.01.006 PMID: 29427897

30. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: Learning to leverage salient regions in medical images. Medical image analysis. 2019; 53:197–207. https://doi.org/10.1016/j.media.2019.01.012 PMID: 30802813

31. Abraham N, Khan NM. A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation. arXiv preprint arXiv:181007842. 2018.

32. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems; 2012. p. 1097–1105.

33. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 3431–3440.

34. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence. 2017; 39 (12):2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615 PMID: 28060704

35. Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:180206955. 2018.

36. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and¡ 0.5 MB model size. arXiv preprint arXiv:160207360. 2016.

37. Rad RM, Saeedi P, Au J, Havelock J. Blastomere cell counting and centroid localization in microscopic images of human embryo. In: 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP). IEEE; 2018. p. 1–6.

38. Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945; 26(3):297–302. https://doi.org/10.2307/1932409

39. Hariharan B, Arbeláez P, Girshick R, Malik J. Simultaneous detection and segmentation. In: European Conference on Computer Vision. Springer; 2014. p. 297–312.

sensors

MDPI

Article

A Multi-Sensor System for Silkworm Cocoon Gender Classification via Image Processing and Support Vector Machine

Alex Noel Joseph Raj [1], Rahul Sundaram [2], Vijayalakshmi G.V. Mahesh [3], Zhemin Zhuang [1] and Alessandro Simeone [4,*]

1 Department of Electronic Engineering, Shantou University, Shantou 515063, China; jalexnoel@stu.edu.cn (A.N.J.R.); zmzhuang@stu.edu.cn (Z.Z.)
2 Department of Agricultural and Food Engineering, Indian Institute of Technology Kharagpur, West Bengal 721302, India; rahulks@iitkgp.ac.in
3 Department of Electronics and Communication Engineering, Dr T Thimmaiah Institute of Technology, Kolar Gold Fields, Karnataka 563120, India; vijayalakshmi@drttit.edu.in
4 Intelligent Manufacturing Key Laboratory of Ministry of Education, Shantou University, Shantou 515063, China
* Correspondence: simeone@stu.edu.cn; Tel.: +86-134-1190-3575

Received: 16 April 2019; Accepted: 10 June 2019; Published: 12 June 2019

check for updates

Abstract: Sericulture is traditionally a labor-intensive rural-based industry. In modern contexts, the development of process automation faces new challenges related to quality and efficiency. During the silkworm farming life cycle, a common issue is represented by the gender classification of the cocoons. Improper cocoon separation negatively affects quantity and quality of the yield resulting in disruptive bottlenecks for the productivity. To tackle this issue, this paper proposes a multi sensor system for silkworm cocoons gender classification and separation. Utilizing a load sensor and a digital camera, the system acquires weight and digital images from individual silkworm cocoons. An image processing procedure is then applied to extract significant shape-related features from each image instance, which, combined with the weight data, are provided as inputs to train a Support Vector Machine-based pattern classifier for gender classification. Subsequently, an air blower mechanism and a conveyor system sort the cocoons into their respective bins. The developed system was trained and tested on two different types of silkworm cocoons breeds, respectively CSR2 and Pure Mysore. The system performances are finally discussed in terms of accuracy, robustness and computation time.

Keywords: multi-sensor; image processing; support vector machine; pattern recognition

1. Introduction

Silk is the most distinguished textile in the world. It has a natural luster and is popularly known as the "Queen of Textiles" [1]. Silk is a product of sericulture, which is produced from silkworm or Bombyx mori, which means "silkworm of the black mulberry tree" [2,3]. Like most insects, the silkworm life cycle has four stages of development, respectively egg, larva, pupa, and adult moth (Figure 1a). A pair of male and female fully-grown adult moth mate with each other, and the female subsequently lays eggs and dies. The egg hatches and emerges out as a larva (also called a caterpillar), which feeds on mulberry leaves and grows for a period of 4 weeks. Once the larva stops feeding it enters the pupal stage, where it uses its secretion to form a protective outer shell called a cocoon [4]. After 3 weeks, the chrysalis emerges from the cocoon as a moth, it mates, and the female lays eggs permitting the life cycle to restart. Among these four stages, cocoons are of commercial importance since a continuous filament of raw silk is directly produced from cocoons by terminating the growth of the caterpillar while inside the cocoon.

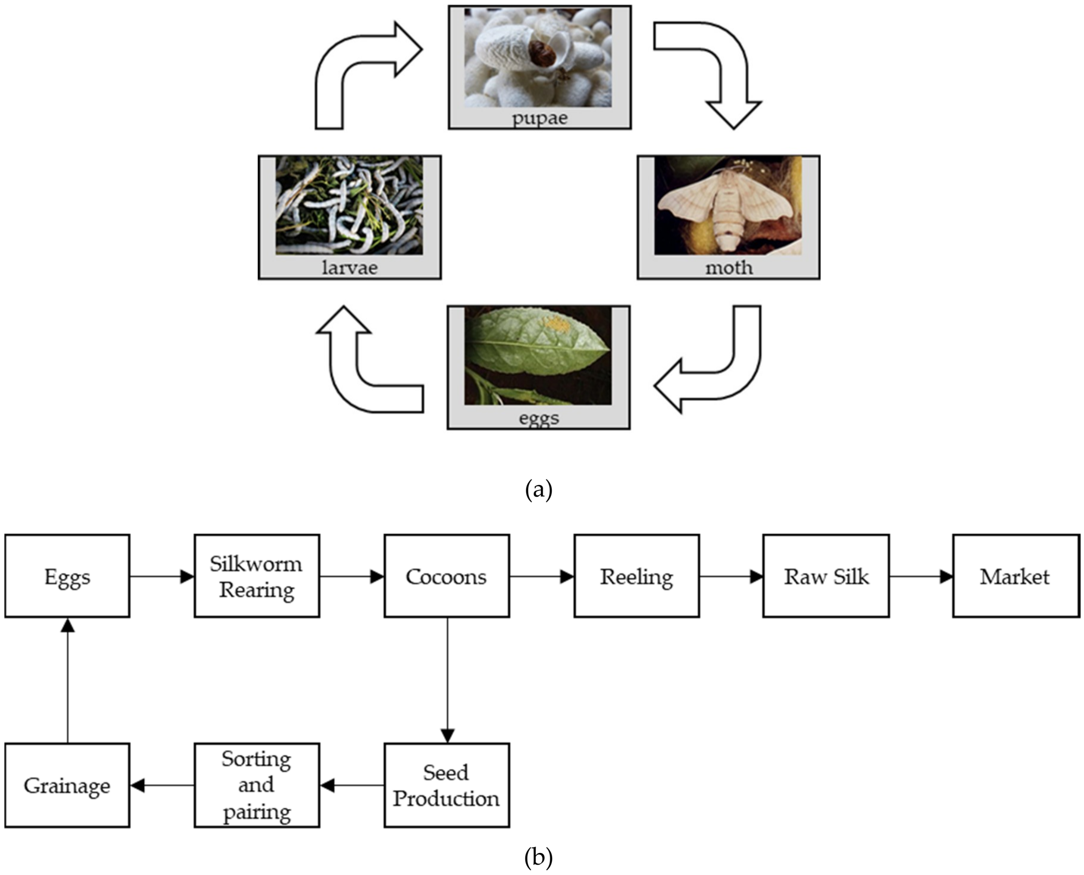


Figure 1. (**a**) Silkworm life cycle, (**b**) silkworm process flow for commercial usage.

The sericulture industry is labor intensive, mostly rural-based, and multidisciplinary in nature. It involves on-farm activities such as mulberry cultivation, egg production, silkworm rearing, cocoon production, and off-farm activities like raw silk reeling, spinning, throwing, and weaving. Grainage centers [5] separate the cocoons based on their gender and allow pairs to mate and produce silkworm eggs (also called seeds). These seeds are utilized by the farmers for cocoon production and later based on the quality and requirements, the cocoons are either sent for reeling, to obtain raw silk or made available to the grainage centers for increased seed production.

In terms of productivity, the success of the silk industry depends mainly on the pre-cocoon stages of the silkworm life-cycle [6], however, literature and industrial practices surveys highlight a gap in the automation development in the pre-cocoon stage compared to post-cocoon stage.

Table 1 summarizes the commercially-available machinery for both pre- and post-cocoon stages, the latter of which can rely on automated machines to reduce the manual labor involved and to improve the production yield.

In terms of quality and quantity, best practices in seed production require (a) sorting the silkworms based on gender and (b) and allowing the best pairs to mate in a constrained environment [5]. Although silkworms can be sorted at different stages during their life cycle (see Table 2), the cocoon stage is the most effective since it causes minor damages to the pupa [6].

Currently, at the grainage centers [7], the cocoons which contain the live chrysalis are weighed and separated as males or females. Then the best cocoon pairs are kept in circular cubicles (Figure 2a) to subsequently allow the moth to emerge, mate and lay the eggs. This manual cocoon-sorting is made possible since females are bulkier and heavier than the male chrysalises. However, this task is still mainly performed manually by employing highly skilled professionals, since a low accuracy in cocoon separation according to the gender may lead to selfing [8,9], decreasing the quantity of the laid eggs [10].

Table 1. Available machinery for pre and post-cocoon stage.

Pre-Cocoon Stage	Post-Cocoon Stage
Machine for crushing shoots	Cocoon de-flossing machine
Mulberry pruning machine	Denier detecting device in silk reeling
Litter separation machine	Long skein silk book making machine
	Pedal-operated reeling twisting machine for muga and tasar silk Reeling
	Reeling and twisting machine
	Solar-operated spinning machine
	Motorized pedal-operated spinning machine
	Wet reeling machine

(a) (b) (c)

Figure 2. (**a**) Image of cubicles used for egg production. (**b**) Eggs produced by incorrect separation of cocoons. (**c**) Eggs produced by pairing best male and female cocoons.

Table 2. Methods for silkworm gender classification for different stages [11–15].

Stages	Methods	Remarks
Chromosome	Presence or absence of the "W" chromosome Female—ZW chromosome Male—ZZ chromosome	Not practical—high cost
Egg	Color of the egg Males are usually light yellow Females are dark brown in color	Not practical—need for skilled workers
Larval	Markings are exhibited on the larval body Female—crescent marking Males—plain	Sex separation is possible only on the 1st day of 5th instar. Process is laborious and too slow operation, larvae may get injured
Cocoon	Color and weight Color—females are golden yellow/ Males white—CSR2 Weight—females are heavier than males	Color depends on various silkworm breeds Weight—each cocoon is weighed individually and sorted—presently followed in grainages—non-destructive
Pupa	Males are smaller in size whereas females are plumper	Reliable/low error—cocoons have to be cut open to remove the pupae, which may cause injury to pupae
Moth	Males are small, slender active moving in semi-circles with bent abdomen/females are bigger with bloated abdomen and rather lethargic	Males and females are easily separated. Selfing takes place affecting the quality of the eggs, health hazards from moth dust

Moreover, the quantity of eggs depends on the proper selection of cocoons quality, in terms of weight and shape, which should fall in specific ranges indicated by industrial standards [6,10].

According to the industrial partner best practices, the weight of Pure Mysore-breed silkworm cocoons ranges from 0.8 to 1.1 g for male and from 1.2 to 1.4 g for female cocoons. The CSR 2 breed cocoons weight ranges from 0.7 to 1.4 g for male and 1.5 to 2.0 g for female cocoons. Moreover, overall circumference of the female cocoon is larger when compared to the male cocoons.

Figure 2b shows the effect of selfing due to incorrect classification on the quantity of laid eggs compared to the ideal separation (Figure 2c).

The literature review provides detailed information about various methods and techniques used for gender separation of silkworm cocoons with a focus on feasibility of usage at grainage centers by untrained professionals. The available techniques can be broadly classified as destructive and non-destructive methods. Destructive methods effectively differentiate the silkworms, but cause permanent damage either to egg, larva, cocoon, or pupae and therefore cannot be further used for seed production or reeling. Non-destructive methods cause less or minimal damage to the silkworms and allow the insect (egg, larva, cocoon, or pupae) to be used in subsequent process. Figure 3 illustrates these methods and the following paragraph reports them in detail.

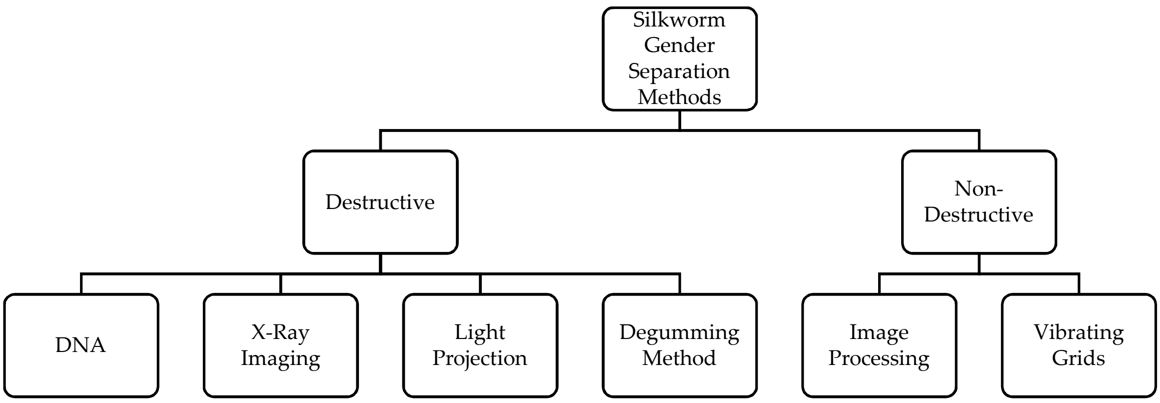


Figure 3. Methods of silkworm gender separation process.

A DNA based gender separation method for sorting silkworm at larvae stage was developed by Tang Pei [15]. It is a chemical based method which required highly trained professionals and suitable only for constrained working environments. X-ray imaging-based gender separation techniques were presented in [14,16]. Calvin J. Witdouck [14] passed X-rays through the silkworm larvae (caterpillar) to identify its gender. Cai et al. [16] demonstrated a prototype with an appropriate classification tool for successfully discriminating the gender of silkworm by X-ray imaging the cocoons. Shape features such as major axis, minor axis, ratio of major axis to minor axis, eccentricity, roundness, rectangularity, complexity, concave and convex characteristics of the chrysalis are extracted from X-ray images and inputted to pre-trained classifiers such as k-Nearest Neighbor (kNN) [17], Linear Discriminant Analysis (LDA) [18], Neural Networks (NN) [19] and Support Vector Machine (SVM) [20] to accurately classify the cocoons as male or female. The authors considered 1071 samples from three hybrid breeds and have reported an accuracy of 93.68% with kNN classifier. Although the X-ray-based methods provide considerably high classification accuracy, constant exposure of the silkworm cocoons to the X-rays may again result mutations leading to poor egg lays, if the cocoons are used in grainages for seed production [21].

A number of light-based sorting systems were developed by Shinji Hayashizaki et al. [22], Raie et al. [23], Yang Bin et al. [24]. In [22] the authors reported light inspection system where the cocoon was cut-open and each pupa was examined under visible light and near infrared light of wavelength 600 nm to 1100 nm. Further the frequency data reflected from the pupa are analyzed and compared with the predefined threshold, based on which the pupa was classified as male or female. Further, in [23], each pupa was again examined under a stereo-microscope. These images were transferred to a computer for processing and image analysis. Under analysis, the gender gland of the silkworm was observed, based on which the sorting was carried out. In both cases, the cocoons are opened to take out the pupa resulting in a high chance of the pupa getting damaged, and further cut cocoons cannot be used for reeling. Moreover, the method is slow and requires trained professionals to examine each pupa accurately.

Zhang et al. [25] developed an apparatus to classify male and female cocoons by radiating ultraviolet rays on the cocoons. The cocoon being tested presented different fluorescence characteristics based on gender. The male cocoon showed yellow (wavelength in the range of 577–597 nm) and female cocoon showed purple (wavelength 390–455nm) fluorescence characteristics respectively. The entire process has to be carried out in a dark room and the assessment is based on the human vision. Additionally, this method is labor intensive and requires additional overhead of 200 W power source and 3600 A wavelength UV light source. This method was further improved by Yang Bin et al. [24] who automated the process based on the color assessment of the fluorescence characteristics provided by male cocoons, integrating a photosensitive yellow filter with the ultraviolet filter for detecting and separating the male cocoon from the female ones.

Yu Xiaohua [26] developed a method for identifying male silkworm cocoon by degumming, i.e., soaking the cocoon in hot water and extracting the fibers. Later the fibers are subject to a chemical process to assess the methionine content and aspartic acid value of three amino acids, based on which the gender can be determined. This is again a non-automated chemical-based technique, which requires heating of the cocoon which can damage the pupa.

Very few non-destructive methods were developed for gender separation of silkworms. An MRI imaging technique operating at 20MHz was used by C. Liu, Z.H. Ren [27] to determine the gender of the silkworm. The MRI image of the cocoon along with live pupa is acquired and later transformed by fast Fourier transform and T2 weighted images (to accurately reflect the tissue contrast into picture contrast) were obtained that aids in distinguishing the gender of the silkworms. Although the method is non-destructive and causes minimal damage to cocoon and the pupa, the imaging process is expensive and practically unsuitable for grainage centers.

Udaya et al. [28] developed a prototype which included metal grids of various sizes which vibrate when connected to an electric motor. The cocoons were sorted based on the size of the cocoons. The female cocoons are bulkier whereas males are thin and slender. The cocoons are transferred into the vibrating grids of the sorting machine which are of varying size to separate the cocoons. Though the system was able to achieve an accuracy of 96% in sorting, the device is not meant for gender separation but mainly used for grading the quality of Tasar variety of cocoons. The graded cocoons are later sent to reeling where raw silk is extracted from the graded cocoons.

Further, Mahesh et al. [20] proposed a novel non-destructive vision-based system to classify the cocoons. The methodology integrated the weight, volume, and ZM-based shape features of the cocoons to form an integrated feature vector for training kNN, LDA, NN, and SVM classifiers. To validate the integration of these features, the performance was compared with the one obtained from integration of geometric shape features and integration of weight and volume with geometric shape features. The method used CSR2 and pure Mysore breeds of cocoons to conduct the experiment. The results indicated a better performance of NN and SVM classifiers. An accuracy of 91.3% was achieved from CSR2 cocoon with NN classifier and 100% from pure Mysore cocoons via SVM-based classifier.

This detailed literature review indicates the existing technologies used for silkworm gender separation at different stages of their life cycle. Separation of silkworm at stages such as eggs, larvae, or pupae is not feasible in grainage centers [27,29]. X-ray or MRI images of the cocoon are high-cost alternatives which provide accurate classification, but the radiation can damage the pupa inside the cocoon. Currently, at grainage centers, the sorting process is manual, where the cocoons' weight (which includes the live pupa) and shape are used as features to distinguish their gender.

Taking into account literature and industrial practice gaps, this paper presents the design and development of a novel non-destructive multi-sensor-based system to classify silkworm cocoons according to their gender. The system extracts the features of cocoons (weight and shape) individually and provides them as inputs to a pre-trained pattern classification model which in turn classifies the cocoons as male or female. Subsequently, a pair of air blowers and a conveyor system sort the cocoons into their respective bins. The developed system was trained and tested on two types of silkworm cocoons breeds, namely CSR2 and Pure Mysore, both provided by Central Silk Board Registered Grainage Center, Karnataka, India.

The prominent advantages of the developed system are (a) elimination of human intervention in separation process, (b) reduction in mis-classification error, (c) good repeatability when compared to manual separation process, and (d) overall increase in speed of separation process.

2. Design and Development of Silkworm Cocoon Gender Classification Multi-Sensor System

The multi-sensor system was designed and prototyped with the aim of performing automatic silkworm cocoons gender classification process. A schematic diagram of the proposed system is shown in Figure 4. The corresponding 3D model and developed prototype are shown in Figures 5 and 6.

The cocoons were initially stored in the hopper, then individually picked by a vertical conveyor module (VCM). The cocoons entered one by one into the feature extraction module (FEM), where each cocoon was analyzed and their features such as shape and weight were extracted. A dedicated software, which executes on a standalone workstation, acquired a digital image and weight of each cocoon, and subsequently extracted significant shape-related features from image instances. Image features and weight data were then combined in an input feature vector, which was inputted to a pre-trained pattern recognition classifier for decision-making on gender classification. Eventually, individual cocoons were transported through a horizontal conveyor module (HCM) which performed the physical sorting of the cocoons and disposed them into dedicated male or female bins. Each module is illustrated in detail in the remainder of this paper.

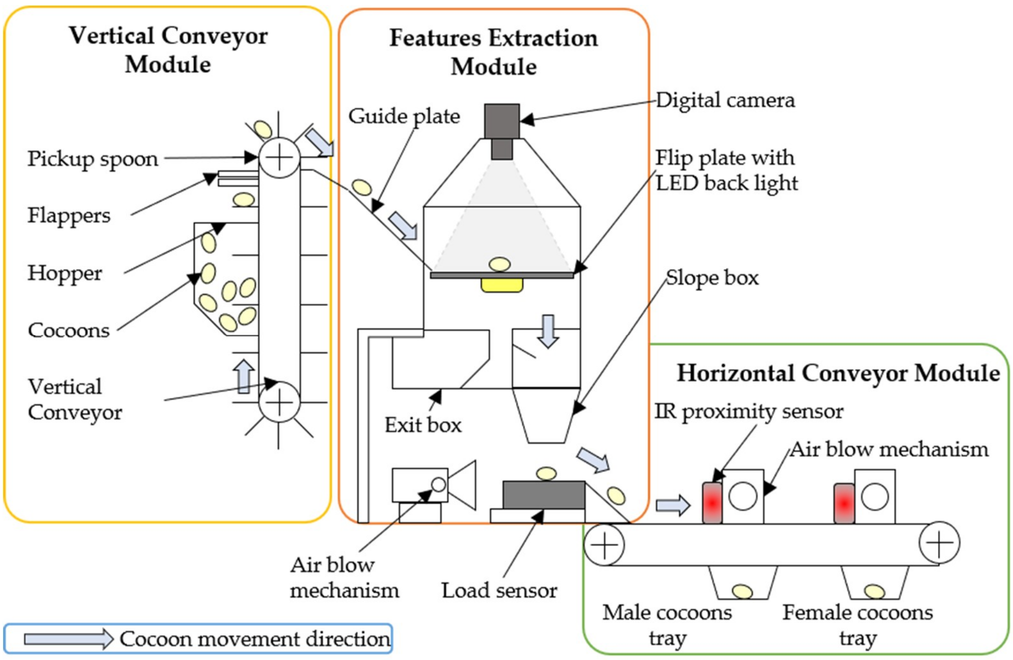


Figure 4. Schematic of silkworm cocoons separating machine.

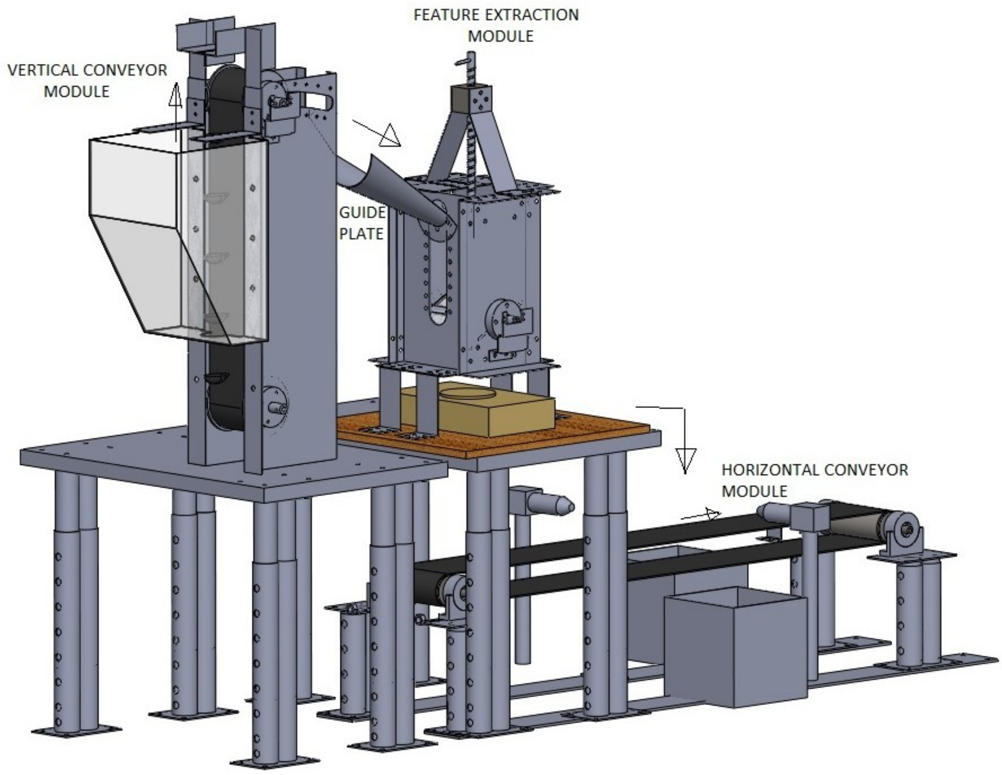


Figure 5. 3D model of silkworm cocoons separating machine.

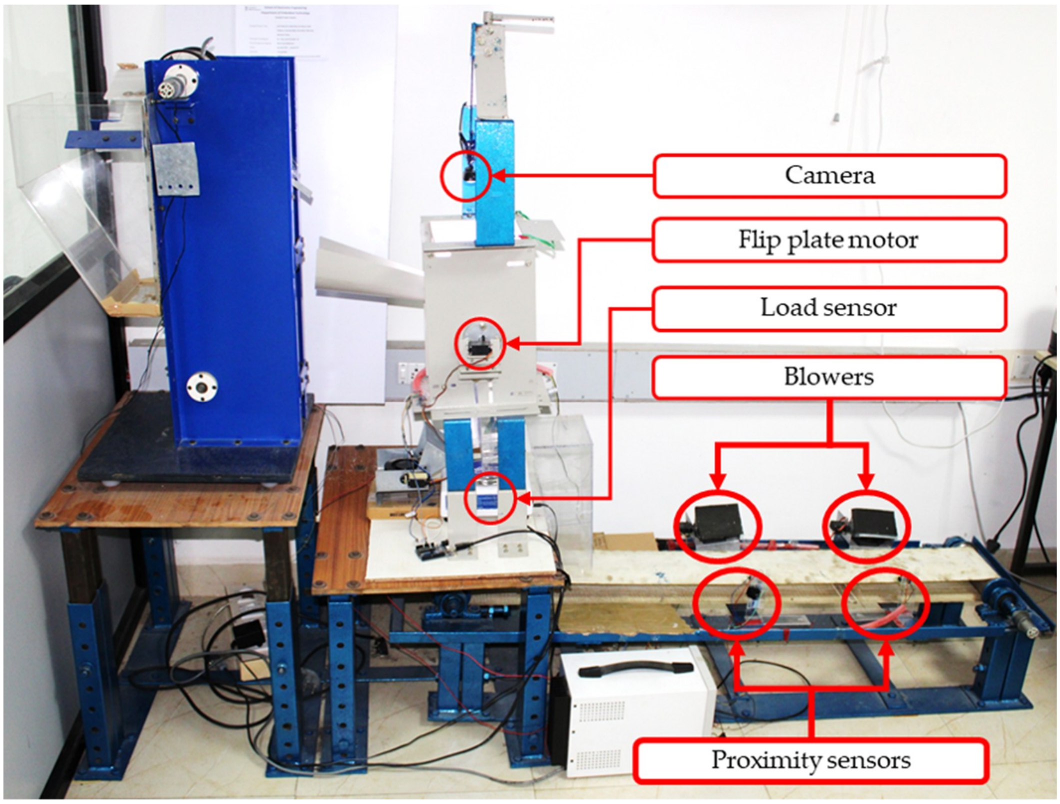


Figure 6. Developed prototype of silkworm cocoons multi-sensor classification system.

2.1. Vertical Conveyor Module (VCM)

The purpose of VCM is to pick individual cocoons from the hopper and feed them into the feature extraction module at a constant velocity without causing major physical damage to cocoons. It consists of a 60 cm-long conveyor belt, which passes through a 12 cm-diameter pulley mounted on the frame plates with the help of bearing support (Figure 7a,c). A 12 V, 10 rpm DC motor drives the pulley which allows each cocoon to travel at a speed of 6.3 cm/s. A loading hopper made from acrylic, which can accommodate up to 1 kg of cocoons (approximately 770 specimens), is rigidly mounted on frame plates. The VCM is endowed with 16 specially-designed concave-shaped spoons that can accommodate one cocoon at a time. Such spoons are riveted on the conveyor belt as shown in Figure 7a,c. The distance between two consecutive spoons is 10 cm. Spoon edges are smoothed to avoid sticking to the cocoon fibers. Two pairs of flappers are mounted on both sides of the metal frame (Figure 7c). The position of flappers was determined experimentally to align the cocoons with the concavity of the spoons and to avoid clinging to the fibrous outer shells of the other cocoons. Once the cocoons were picked up by these spoons, they were transferred to FEM for analysis. In this respect, a guide plate (Figure 8a,b) is positioned on rear side of VCM to enable smooth transfer of cocoons from VCM to FEM.

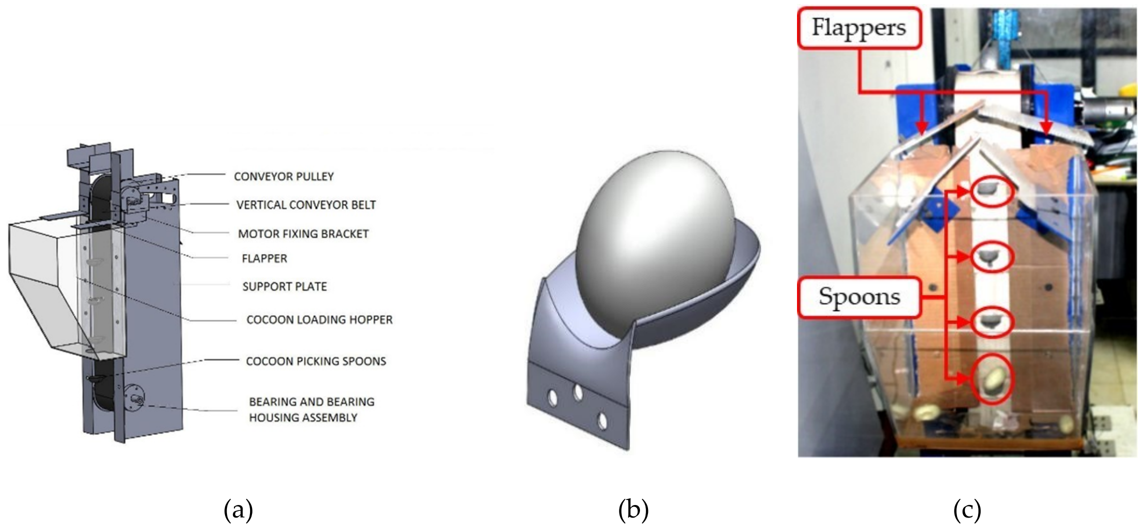


Figure 7. (**a**) 3D model of vertical conveyor module (VCM) module; (**b**) 3D model of spoon carrying a cocoon; (**c**) prototype of VCM module.

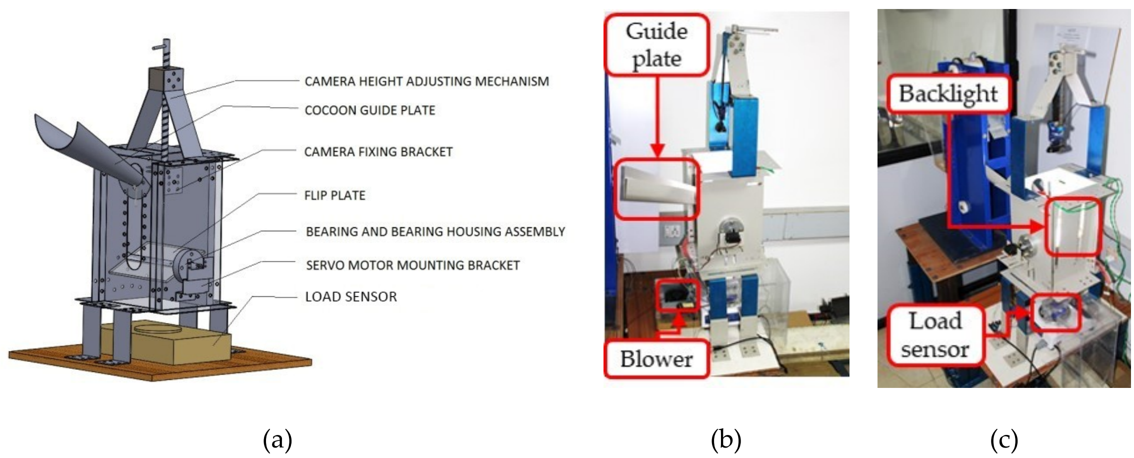
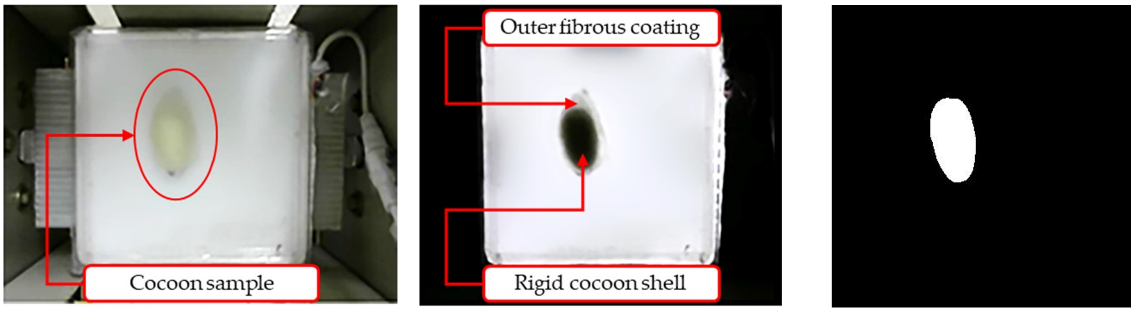


Figure 8. (**a**) 3D model of FEM; (**b**,**c**) prototype of FEM shown in two different angles.

2.2. Features Extraction Module (FEM)

The FEM shown in Figure 8 consists of a camera support bracket, flip plate, slope box, exit box, a detachable load sensor, and an air blower mechanism. The module extracts both the shape features of cocoon via image processing and the weight of the cocoon (in g) via the load sensor and feeds such information into a binary classifier to determine the cocoon gender. A 5 mega pixel digital camera is mounted on the support bracket and a flip plate made of acrylic is mounted below the camera assembly. The structure allows continuous acquisition of objects on the flip plate. The flip plate is attached to a rotating shaft which is powered by a servo motor to allow the plate to be positioned at three different orientations (a) 0° (horizontal), (b) 90° (clockwise), and (c) −90° (counter clockwise). When the system is switched on, the flip plate is positioned horizontally to receive cocoons from the VCM.

Silkworm cocoons have a hard shell that is covered by fibrous outer coating as shown in Figure 9a,b. The accuracy of the classifier depends on how well the shape features are extracted from the cocoon rigid shell images by eliminating the negative effect of fibrous coating. Practically, it is not possible to remove the fibers from each cocoon manually at grainage centers before loading cocoons into the hopper. To tackle this issue, the FEM is endowed with an 18 W square LED panel light (shown in Figure 9a,b) attached to the flip plate. Such an illumination system provides the necessary back-lighting to capture the silhouette of the hard shell of the cocoon enabling accurate area calculation by image thresholding techniques [30,31].



(a) (b) (c)

Figure 9. (**a**) RGB image of the cocoon on flip plate captured by FEM camera without backlight, (**b**) RGB image of the cocoon on flip plate captured by FEM camera with backlight illumination, and (**c**) binarized image of the cocoon shell with fibrous outer surface removed.

By comparing the two images, the cocoon sample in normal light conditions without backlight (Figure 9a) and the same cocoon sample placed on the flip plate with backlight (Figure 9b), the advantage of the adopted illumination system is evident in the results.

After image acquisition, the cocoon is transferred to the load sensor by letting it fall through a slope box (see Figure 4), where the velocity of falling cocoon is attenuated by travelling through number of inclined slopes. Once fallen on the load sensor unit, the cocoon weight data are acquired with a resolution of ±0.01 g and transferred to the workstation.

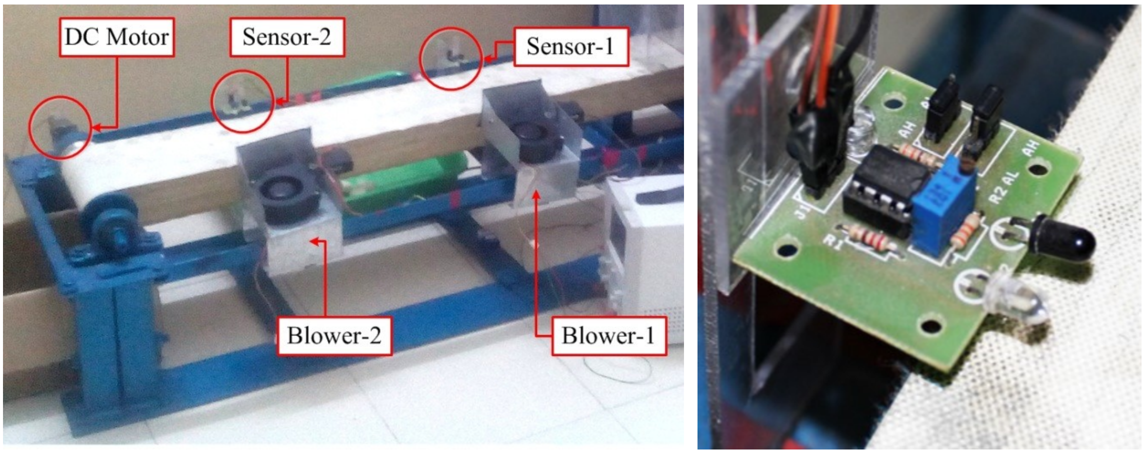
Following the weight data acquisition, an air blower mechanism (Figure 10) is employed to transfer the cocoon from the load sensor unit to the HCM. The blower mechanism consists of an air blower which continuously provides a compressed air supply and a freely rotating swivel arm which is used to stop the air flow instantly. One end of the swivel arm is coupled with a servo motor and the other end is fixed on the side wall by a freely rotating cylindrical pin joint. Figure 10a shows the mechanism in closed position (the air stream is blocked). Figure 10b shows the mechanism in open position (the air stream is directed to the cocoon). Normally, the air blower mechanism is in the closed position, hence no air flow is directed on the cocoon. Once the weight data are acquired, the system sends a command to the servo motor and the swivel arm opens for 2 s allowing for the cocoon to be transferred from the load sensor to the HCM. As regards the FEM power requirement, the major contributions are represented by the back light: 18 W, the flip plate servo motor: 5 V × 0.9 A = 4.5 W and the digital camera: 5 V × 0.5 A = 2.5 W, for a total power requirement equal to 25 W.

(a) (b)

Figure 10. (**a**) 3D model of air blower frames; (**b**) developed air blower mechanism.

2.3. Horizontal Conveyor Module

Once the shape and weight features are extracted by the FEM, the cocoon moves to next module, The Horizontal Conveyor Module (HCM) consists of a rotating conveyor belt, an infra-red (IR) proximity sensors and 2 blower mechanism units (Figure 11). The conveyor belt is 2-meter-long and actuated by a 12 V DC motor (10 rpm and 120 kg·cm torque) to rotate continuously at a speed of approximately 8 cm/s. A pair of IR sensors and a pair of air blowers are placed at a distance of 40 cm along the conveyor as shown in Figure 11.



(a) (b)

Figure 11. (**a**) Horizontal Conveyor Module; (**b**) IR proximity sensor close-up.

Their positions were determined empirically based on the computation time required by the workstation to provide the classification index. The HCM classifies the cocoons based on the index obtained from the workstation. As the cocoon crosses the first sensor–blower pair, its index is retrieved from the workstation. Based on the predicted index, the blowers in each pair activate/deactivate and transfer the cocoon to their respective trays.

2.4. Communication and Synchronization of Modules

The prototype consists of three modules which contain several individual components. Each component performs an individual function to carry out the cocoon sorting process. These components need to be synchronized in order perform automated operation. Figure 12 presents the flow diagram of silkworm cocoons gender sorting machine.

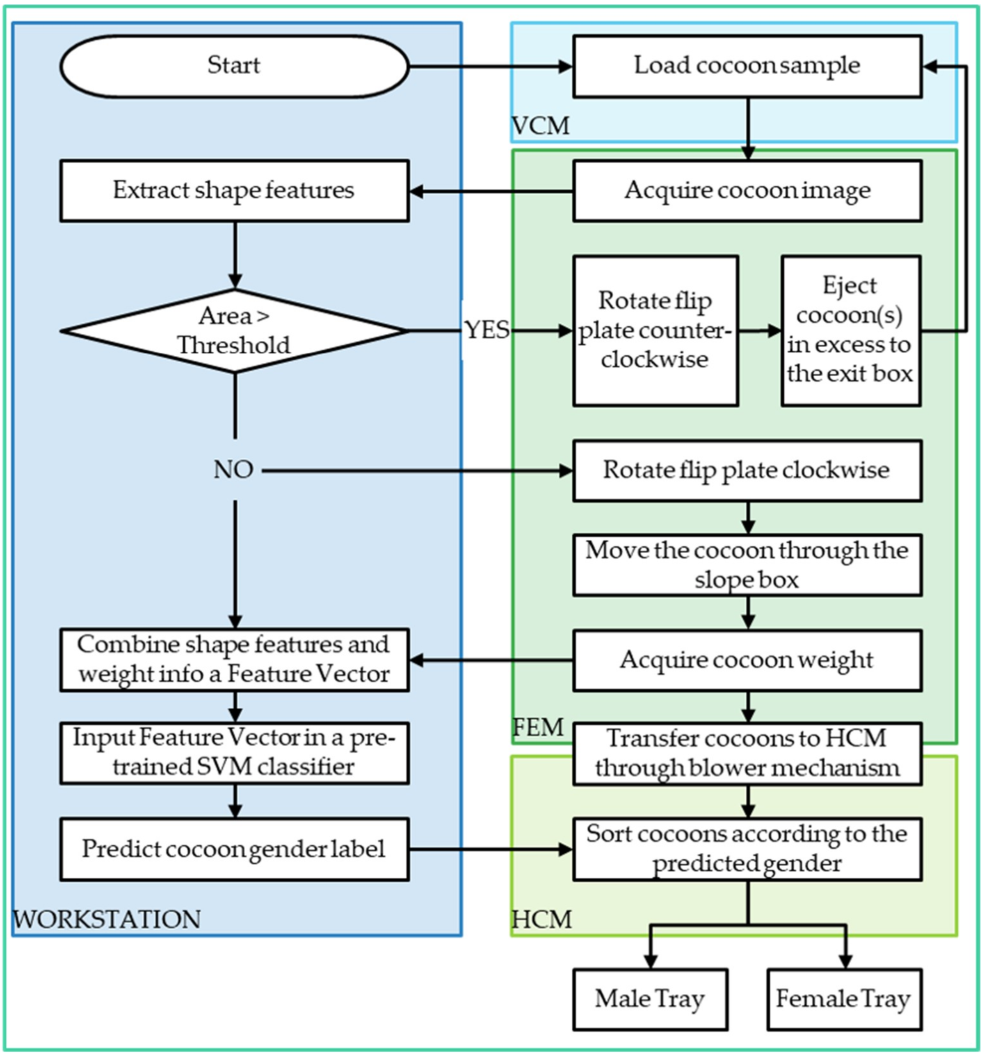


Figure 12. Flow diagram of silkworm cocoons gender classification machine.

The acquired cocoon image in FEM module is sent to the workstation, where the shape features (area, perimeter, major axis length, minor axis length, etc.) are computed. At times, there is a chance of entering more than one cocoon into the FEM module. This condition is detected by computing the cocoon area from the binarized image and comparing it to an empirical threshold value. The experimental setup was designed to achieve the most favorable experimental conditions in terms of image quality and illumination, for this reason, the thresholding operation for image binarization is carried out on the acquired image utilizing the Otsu's algorithm [32]. If the computed area exceeds the threshold, then exceeding cocoons are ejected by rotating the flip plate in counterclockwise direction. Ejected cocoons move out of the module through exit box (Figure 4) to be fed back to the hopper. If the binarized image area results within the threshold limit (i.e., only one cocoon present on the flip plate), a signal is provided to the microcontroller to rotate the flip plate in clockwise direction to transfer the cocoon to the load sensor smoothly through slope box. At this point, the sample cocoon weight is acquired and provided to the algorithm in the workstation. Shape and weight features are then combined and fed to a pre-trained SVM to determine the gender of cocoon under examination. Further, the cocoon index and its corresponding predicted label are stored in the workstation and the cocoon present on the load sensor is moved to HCM module by the air blower mechanism.

As the cocoon moves along the HCM, the IR proximity sensors (Figure 11) provide an input signal to microcontroller-2 which in turn retrieves the classification label of the current cocoon obtained from the workstation. The label is used to control the respective blowers. In this respect, if the predicted label is "male", the first blower is triggered, and the cocoon is pushed on to the "male cocoon tray". Conversely, if the cocoon label is "female", the second blower is triggered and pushes the cocoon into the "female cocoon tray".

3. Experimental Methodology

The experimental campaign was carried out on two silkworm cocoons breeds, namely CSR2 and Pure Mysore, both provided by the industrial partner. Prior to the experimental tests' commencement, the cocoons were manually labelled as male and female by highly trained and skilled professionals using a weight threshold as discriminating parameter. The cocoon weight is an important factor since it is highly correlated to the cocoons gender (i.e., cocoons above the weight threshold are considered as females and ones below are males) and is most commonly used gender separation method employed in the grainage centers. A weight threshold of 1.4 g for CSR2 breed and 1.1 g for Mysore breed was used to separate the cocoons as male and female to build the ground truth used for benchmarking. Besides cocoon weight, shape features are also significant in discriminating the cocoons based on gender [20].

A total number of 167 cocoons was used to build the dataset, which included 76 Pure Mysore and 91 CSR2 breeds. For Pure Mysore breed, there were 35 male and 41 female specimens; similarly, CSR2 breed contained 47 males and 44 females. The training set was used to pre-train the SVM classifier for decision-making on gender classification. The dataset subdivision was carried out using the hold-out method [33] with the following proportions: 60% for training and 40% for testing, as shown in Table 3.

Table 3. Dataset for silkworm cocoons.

	Training Set		**Testing Set**	
	M	**F**	**M**	**F**
CSR2	28	26	19	18
Pure Mysore	21	24	14	17

M: Number of male specimens; F: number of female specimens.

The training set was used to pre-train the Support Vector Machine (SVM) classifier. SVM is based on statistical learning theory aimed at determining the location of decision boundaries yielding the optimal separation of classes [34]. For a binary pattern recognition problem in which the classes are linearly separable the SVM selects from among the infinite number of linear decision boundaries the one that minimizes the generalization error. Thus, the selected decision boundary will be one that leaves the greatest margin between the two classes, i.e., the sum of the distances to the hyperplane from the closest points of the two classes [35]. The data points that are closest to the hyperplane are used to measure the margin; hence these data points are termed "support vectors" [29].

If the two classes are not linearly separable, the SVM tries to find the hyperplane that maximizes the margin while, at the same time, minimizing the misclassification errors. SVM can also be extended to handle nonlinear decision surfaces by projecting the input data onto a high-dimensional feature space using kernel functions [36] and formulating a linear classification problem in that feature space. In this research work a linear kernel has been utilized to train the SVM.

In order to train the classifier, the separated training cocoons were labelled and indexed manually prior to being loaded into the VCM hopper. Once the cocoon was transferred from the VCM to FEM, the cocoon's silhouette was acquired by camera and passed to the workstation. If the camera was rigidly fixed at distance of 18 cm from the flip plate, the area of an individual cocoons ranged between 500 and 550 pixels. If the area was greater than this interval, the system assumed that the VCM has transferred more than one cocoon to FEM, therefore the exceeding cocoons were ejected back to the hopper through the exit box. Once the FEM ensured the feeding of a single cocoon, a number of shape features were computed and extracted from the silhouette image as reported in Table 4.

Table 4. Shape-related features extracted from the silhouette binary image.

Parameter	Description
Area (A)	Describes the number of pixels in the region of the shape
Perimeter (P)	Provides the number of pixels in the boundary of the shape
Major axis length (λ_1)	Specifies the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region
Minor axis length (λ_2)	Specifies the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region
Eccentricity (E)	Measure of the aspect ratio. Computed using minimum bounding box (smallest rectangle containing every point in the shape) method. $$E = \frac{L_b}{W_b}$$ where L_b = Length of the bounding box and W_b = Width of the bounding box
Circularity/Roundness (C)	Circularity ratio represents how a shape is similar to a circle. It is given by the ratio of the area of a shape to the shape's perimeter square. $C = \frac{A}{P^2}$
Rectangularity (R)	Represents how rectangular a shape is, i.e. how much it fills its minimum bounding rectangle. It is given by: $$R = \frac{A}{A_r}$$ where A_r = Area of the minimum bounding rectangle
Solidity (S)	Describes the extent to which the shape is convex or concave. It is given by: $$S = \frac{A}{H}$$ where H = is the convex hull area of the shape. The solidity of a convex shape is always 1
Convex area (A_C)	Specifies the number of pixels in convex image. It is given by: $$A_C = \frac{\text{Perimeter of the convex hull}}{P}$$ where convex hull of a region is the smallest convex region including it.

Once the features were computed, the flip plate rotated clockwise and weight of the cocoon (W) was obtained from the weight sensor placed below and serially transferred to work station. Later, using the air-blower mechanism explained in section, the cocoon moved to the next stage of the pipeline. At the work station extracted shape features were integrated with the weight to form an integrated feature vector (*IFV*) as shown in Equation (1):

$$IFV = \left\{W,\ A,\ P, \left(\frac{\lambda_1}{\lambda_2}\right),\ E,\ C,\ R,\ S,\ A_C\right\} \quad (1)$$

The IFV was further normalized to standardize the range of obtained features using Z-score normalization [37], given by:

$$IFV_{norm} = \frac{IFV_i - \mu_i}{\sigma_i},\ i = 1, \ldots, L \tag{2}$$

where L is the length of the IFV, μ is the mean of the feature and σ the standard deviation of the feature. The normalized IFV is labeled for supervised machine learning. Label "1" indicates category "Male" and "0" indicated category "Female".

4. Results and Discussion

The proposed pattern classifier performance is assessed in terms of accuracy, robustness and computation speed.

4.1. Classifier Accuracy

The performance assessment of the SVM classifier is carried out using the hold-out method [33], where the labelled training set is used in training the classifier to create an optimal model. The model is further evaluated using the testing data set. The results of the training process are displayed using a confusion matrix (CM) to calculate the performance metrics of the training phase [38]. Such matrices show the True Male, True Female, False Male, and False Female. With reference to the SVM training process, the classification results CMs for the CSR2 and Pure Mysore breeds cocoons are reported in Figure 13.

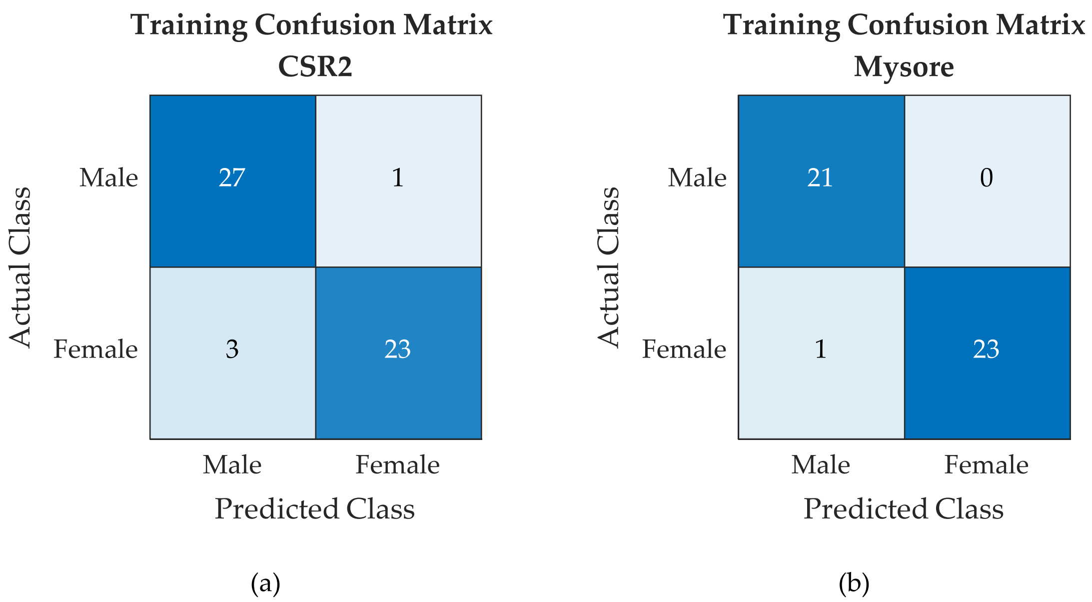


Figure 13. SVM training confusion matrices for CSR2 (**a**) and Pure Mysore cocoons (**b**).

To validate the accuracy of the prototype, unknown cocoon samples are indexed and loaded into the VCM's hopper. As the cocoon travel from VCM to FEM, its features are extracted and transferred to the workstation where the pre-trained SVM provides the predicted classification label. The label and the corresponding cocoon indices are stored as look up table within the workstation. When the cocoon moves along the HCM, the IR proximity sensors and with the microcontroller query the workstation to provide the classification label of the current cocoon. This label is used by the sensor to blow the cocoons to their respective trays as explained in Section 2. This process utilizes all the cocoons present in the testing dataset and the performance of the prototype with SVM model is evaluated similar to that of the training process. With reference to the SVM test process, the classification results CMs for the CSR2 and Pure Mysore breeds cocoons are reported in Figure 14.

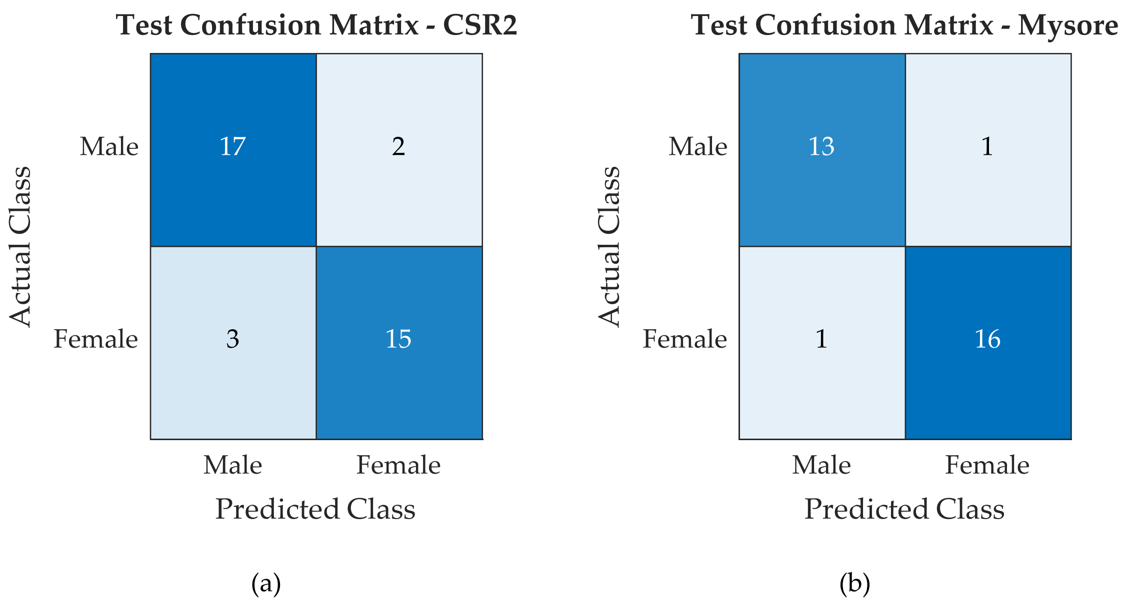


Figure 14. SVM test confusion matrix for CSR2 cocoons (**a**) and Pure Mysore cocoons (**b**).

From the confusion matrices reported in Figures 13 and 14, a number of performance metrics were computed, namely Accuracy, True Male Rate (TMR), True Female Rate (TFR), Male Predictive Rate (MPR), Female Predictive Rate (FPR), and F1 score [39,40]. Such performance metrics are reported in Table 5 with reference to both training and test phases for CSR2 and Pure Mysore cocoons respectively:

Table 5. Performance metrics (PM) obtained for CSR2 and pure Mysore cocoons from SVM training and testing.

	Training		Testing	
PM	**CSR2**	**Pure Mysore**	**CSR2**	**Pure Mysore**
Accuracy:	0.9259	0.9778	0.8649	0.9355
True Male Rate	0.9642	1.0000	0.8947	0.9286
True Female Rate	0.8846	0.9583	0.8333	0.9412
Male Predictive Value	0.9000	0.9545	0.85	0.9286
Female Predictive Value	0.9583	1.0000	0.8824	0.9412

4.2. Robustness and Computation Speed

To validate the prediction robustness, 50 cocoons were randomly selected from CSR2 breed and four trials were conducted. The cocoons were indexed, and their classification labels recorded prior to the trial conduction. The prediction results from protype each cocoon for all the four trials are illustrated in Figure 15. The chart shows that 44 cocoons were correctly predicted in all the four trials. Hence the repeatability of the machine is calculated as the ratio of the number of correctly classified cocoons in all the four trials over the total number of tested cocoons. The repeatability of cocoons separation through the fusion of weight and shape features using the prototype resulted to be 44/50 = 88%.

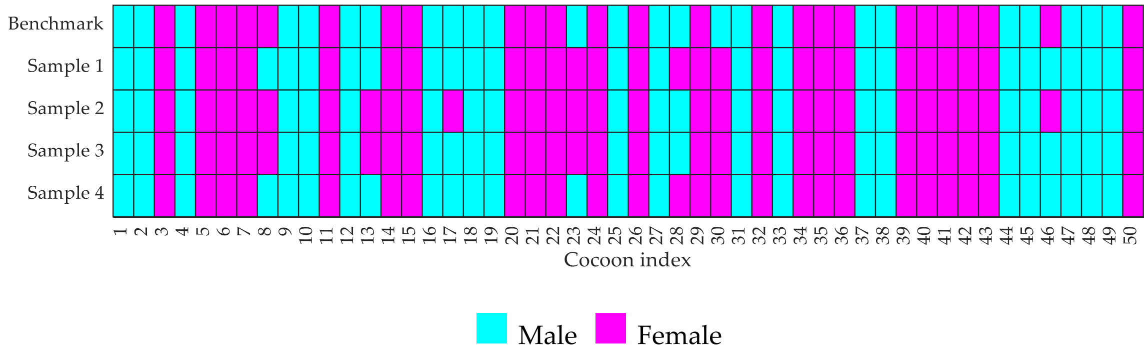


Figure 15. Graph of cocoon index vs. predicted gender for the sample batch.

The average time required for a cocoon to reach the FEM from VCM is 4.6 s. The cocoon then lays on the flip plate for about 1 s to allow for the image acquisition. The cocoon is subsequently transferred to the load sensor, where the weight feature extraction requires 1.2 s. In total, the cocoon stays in the FEM module for about 2.2 s. The cocoon reaches the collection tray through the HCM in about 3.6~4.1 s. Thus, the maximum time taken for a cocoon to reach the tray from hopper is 10.9 s.

From the observed results, the proposed system classifies approximately 5.5 cocoons in a minute and thus it can classify 330 cocoons in an hour. For eight hours shift, the prototype can classify about 2640 cocoons. Considering an average weight of a cocoon equal to 1.3 g, the presented system can classify approximately 3.4 kg of cocoons in eight hours shift yielding an accuracy ranging from 86.48% to 93.54% depending on the breed, whereas highly experienced staff working in a grainage center can probably sort a similar daily amount of cocoons bearing in mind that manual classification is prone to human error over prolonged working hours and may lead to serious health and safety issues [41,42].

5. Conclusions

The work presented in the paper represents a kick-off in modern sericulture automation to eliminate human intervention in classifying silkworms based on gender in the cocoon stage without damaging the shell. The developed system has the potential to boost the productivity of grainage centers who are currently carrying out the gender classification process manually. The results obtained during testing showed a maximum accuracy of 93.54% with a repeatability of 88%, demonstrating a potential suitability of the proposed method for industrial applications.

Future research efforts need to be focused on the following critical aspects for improving the industrial suitability of the system:

- design optimization to reduce the overall dimensionality and operation speed from a hardware perspective in terms of more powerful workstation and more efficient blower mechanism;
- endowing the VCM with a deflossing unit [5] in order to remove the fibers of the cocoon to avoid clinging phenomena which drastically reduce the system speed;
- extend the experimental campaign to a wider variety of cocoon breeds to improve the system generalization and to increase the system versatility.

Author Contributions: Conceptualization, A.N.J.R.; methodology, A.N.J.R. and R.S.; software, R.S. and V.M.; validation, A.N.J.R. and A.S.; formal analysis, A.N.J.R., A.S., and Z.Z.; investigation, A.N.J.R., R.S., and V.M.; resources, A.N.J.R.; data curation, R.S. and V.M.; writing—original draft preparation, A.N.J.R. and A.S.; writing—review and editing, A.N.J.R. and A.S.; visualization, A.N.J.R. and A.S.; supervision, A.N.J.R., A.S., and Z.Z.; project administration, A.N.J.R. and Z.Z.; funding acquisition, A.N.J.R.

Funding: This research was funded by the Science for Equity, Empowerment and Development (SEED) Division Department of Science and Technology, India (No. SP/YO/005/2015), the Foundation of China (No. 61471228), the Key Project of Guangdong Province Science & Technology Plan (No. 2015B020233018), the Research Start-up Fund Subsidized Project of Shantou University, China, (No. NTF17016 and NFT17004).

Acknowledgments: The authors gratefully acknowledge the industrial partner Central Silk board Registered Grainage center, Karnataka, India (registration number KA/01/RSP/0147) for providing the cocoons samples used in the experimental campaign.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Asakura, T.; Kametani, S.; Suzuki, Y. *Encyclopedia of Polymer Science and Technology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2018; pp. 1–19.
2. Rockwood, D.N.; Preda, R.C.; Yücel, T.; Wang, X.; Lovett, M.L.; Kaplan, D.L. Materials fabrication from bombyx mori silk fibroin. *Nat. Protoc.* **2011**, *6*, 1612. [CrossRef] [PubMed]
3. Sprague, K.U. Bombyx mori silk proteins: Characterization of large polypeptides. *Biochemistry* **1975**, *14*, 925–931. [CrossRef] [PubMed]
4. Matsuura, S.; Morimoto, T.; Nagata, S.; Tashiro, Y. Studies on the posterior silk gland of the silkworm, bombyx mori. II. Cytolytic processes in posterior silk gland cells during metamorphosis from larva to pupa. *J. Cell Biol.* **1968**, *38*, 589–603. [CrossRef] [PubMed]
5. Murugesh Babu, K. Introduction to silk and sericulture. In *Silk*; Elsevier: Amsterdam, The Netherlands, 2013; pp. 1–32.
6. Lee, Y.-W.; Food and Agriculture Organization of the United Nations. *Silk Reeling and Testing Manual*; Food and Agriculture Organization of the United Nations: Rome, Italy, 1999; ISBN 9251042934.
7. Govt. of India Central Silk Board—Ministry of Textiles Registered Seed Producers. Available online: http://csb.gov.in/wp-content/uploads/2019/04/FRESH-STATE-WISE-LIST-OF-RSPs-AS-ON-MARCH-2019.pdf (accessed on 4 April 2019).
8. Sreekumar, S.; Kadono-Okuda, K.; Nagayasu, K.; Hara, W. Identification of 2nd chromosome region translocated onto the W chromosome by RFLP with EST-cDNA clones in the gensei-kouken strains of the mulberry silkworm, bombyx mori L. *Genet. Mol. Biol.* **2009**, *33*, 27–35. [CrossRef] [PubMed]
9. Bindroo, B.B.; Manthira Moorthy, S. Genetic divergence, implication of diversity, and conservation of silkworm, bombyx mori. *Int. J. Biodivers.* **2014**. [CrossRef]
10. Narasimhanna, M.N. *Manual on Silkworm Egg Production*; Central Silk Board: Karnataka, India, 1988.
11. Sakai, H.; Aoki, F.; Suzuki, M.G. Identification of the key stages for sex determination in the silkworm, bombyx mori. *Dev. Genes Evol.* **2014**, *224*, 119–123. [CrossRef] [PubMed]
12. Suzuki, M.G. Sex determination: Insights from the silkworm. *J. Genet.* **2010**, *89*, 357–363. [CrossRef] [PubMed]
13. Ohbayashi, F.; Suzuki, M.G.; Shimada, T. Sex determination in bombyx mori. *Curr. Sci.* **2002**, *83*, 466–471.
14. Calvin, J.; Witdouck, C.J. System and Method for Sorting Larvae Cocoons. U.S. Patent 7,737,379, 15 June 2010.
15. Tang, P. Sex identification method for silkworm larva. CN Patent CN1990881A, 4 July 2007.
16. Cai, J.R.; Yuan, L.M.; Liu, B.; Sun, L. Nondestructive gender identification of silkworm cocoons using X-ray imaging with multivariate data analysis. *Anal. Methods* **2014**, *6*, 7224–7233. [CrossRef]
17. Wang, G.-Z.; Li, J.; Hu, Y.-T.; Li, Y.; Du, Z.-Y. Fault identification of chemical processes based on k-NN variable contribution and CNN data reconstruction methods. *Sensors* **2019**, *19*, 929. [CrossRef] [PubMed]
18. Jia, W.; Liang, G.; Tian, H.; Sun, J.; Wan, C. Electronic nose-based technique for rapid detection and recognition of moldy apples. *Sensors* **2019**, *19*, 1526. [CrossRef] [PubMed]
19. Segreto, T.; Simeone, A.; Teti, R. Principal component analysis for feature extraction and NN pattern recognition in sensor monitoring of chip form during turning. *CIRP J. Manuf. Sci. Technol.* **2014**, *7*, 202–209. [CrossRef]
20. Mahesh, V.G.V.; Raj, A.N.J.; Celik, T. Silkworm cocoon classification using fusion of zernike moments-based shape descriptors and physical parameters for quality egg production. *Int. J. Intell. Syst. Technol. Appl.* **2017**, *16*, 246–248. [CrossRef]
21. Subramanya, G. Effect of X-ray irradiation on the male moths of two tropical races of the silkworm bombyx mori and inheritance of induced Sterility in the progenies. *Int. J. Insect Sci.* **2010**, *2*, IJIS-S3655. [CrossRef]
22. Hayashizaki, S.; Machida, J.; Maeda, H. Method and device for determining sex of silkworm pupa. JP Patent JP2007071620A, 22 March 2007.

23. El-Raie, A.E.; El-Rahman, A.A.; Saker, H.E.; Hassan, H.E. Comparative light inspection to sort bombyx mori l. cocoons. *Misr J. Agric. Eng.* **2006**, *23*, 651–668.
24. Yang, B.; Lu, F.; Ren, J.; Zeng, W.; Tang, X.; Shi, X.; Liang, Y.; Li, Y.; Ding, Z.; Jia, Z. Automatic male and female sorting apparatus for cocoons judged by fluorescence color. CN Patent CN101403142B, 18 August 2010.
25. Zhang, Y.; Yu, X.; Shen, W.; Ma, Y.; Zhou, L.; Xu, N.; Yi, S. Mechanism of fluorescent cocoon sex identification for silkworms bombyx mori. *Sci. China Life Sci.* **2010**, *53*, 1330–1339. [CrossRef] [PubMed]
26. Yu, X. Method for identifying male silk of mulberry silkworm. CN Patent CN102590482A, 18 July 2012.
27. Liu, C.; Ren, Z.H.; Wang, H.Z.; Yang, P.Q.; Zhang, X.L. Analysis on gender of silkworms by MRI technology. In Proceedings of the 2008 International Conference on BioMedical Engineering and Informatics, Sanya, China, 28–30 May 2008.
28. Udaya, C.J.; Subrata, R.; Garageswari, N.R.; Kariyappa. An automatic Tasar cocoon separating machine. Indian Patent 2374/CHE/2008 A, 7 October 2011.
29. Sumriddetchkajorn, S.; Kamtongdee, C. Optical penetration-based silkworm pupa gender sensor structure. *Appl. Opt.* **2012**, *51*, 408–412. [CrossRef] [PubMed]
30. Abutaleb, A.S. Automatic thresholding of gray-level pictures using two-dimensional entropy. *Comput. Vis. Graph. Image Process.* **1989**, *47*, 22–32. [CrossRef]
31. Daniel Ratna Raju, P.; Neelima, G. Image segmentation by using histogram thresholding. *Int. J. Comput. Sci. Eng. Technol.* **2012**, *2*, 776–779.
32. Otsu, N. A Threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man. Cybern.* **1979**, *9*, 62–66. [CrossRef]
33. Ruppert, D. The elements of statistical learning: Data mining, inference, and prediction. *J. Am. Stat. Assoc.* **2004**. [CrossRef]
34. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
35. Pal, M.; Mather, P.M. Support vector machines for classification in remote sensing. *Int. J. Remote Sens.* **2005**, *26*, 1007–1011. [CrossRef]
36. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992.
37. Nawi, N.M.; Atomi, W.H.; Rehman, M.Z. The effect of data pre-processing on optimized training of artificial neural networks. *Procedia Technol.* **2013**, *11*, 32–39. [CrossRef]
38. Ting, K.M. Confusion matrix. In *Encyclopedia of Machine Learning and Data Mining*; Springer: Boston, MA, USA, 2017; p. 260. ISBN 978-1-4899-7687-1.
39. Huddleston, S.H.; Brown, G.G. Machine learning. In *Informs Analytics Body of Knowledge*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2018; pp. 231–274. ISBN 9781119505914.
40. Bishop, C.M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*; Springer: New York, NY, USA, 2006; ISBN 9780387310732.
41. Araujo, L.M.L.; Filho, N.A.R.; Riedi, C.A. Respiratory allergy to moth: The importance of sensitization to bombyx mori in children with asthma and rhinitis. *J. Pediatr.* **2014**, *90*, 176–181. [CrossRef]
42. Reshma, S.; Adhikrao, D.J. Occupational health and physiological profile of sericulture industry workers with respect to Workplace environment. *Int. J. Pharm. Bio Sci.* **2016**, *7*, 137–143.

Implementation of Automated Vehicle Identity Recognition System

Hamsavahini. R[1] Soujanya A[2] Sushma N[3] Prashanti P Rao[4] Thanushree S[5]
[1]Assistant Professor
[1,2,3,4,5]Department of Electronics & Communication Engineering
[1,2,3,4,5]BMS Institute of Technology and Management, Bengaluru, Karnataka, India

Abstract— In this paper we have proposed and designed an automated vehicle identify recognition system using image processing. The captured image of the vehicle number plate is processed and characters are recognized using Convolutional neural network (CNN) and authenticated by implementing it on Raspberry Pi.

Key words: Image processing, CNN, Raspberry Pi

I. Introduction

The increase growth in the number of automobile has facilitated human life but it has lead to various issues like traffic congestion, traffic problem etc. The rapid growth in the number of vehicles demands for the efficient automatic vehicle identification system. So automatic license plate detection and Recognition is necessary. Automatic vehicle identity recognition is surveillance method that uses Convolutional Neural Network. They can also used at CC TVs or ones specifically designed for the task. The basic process involves image capturing, detection of number plate, segmentation and recognition of characters. In the proposed system Convolutional Neural Network and Raspberry pi is used and we have used Open Computer Vision and Python for Programming which is better than MATLAB. OpenCV can be used for Real Time Applications.

II. Methodology

A proposed automated vehicle identity recognition system is as shown in the figure (1). The images of Indian vehicle number plate with different font size, styles and background are processed using image processing. The convolutional neural network used is wide to recognise the characters and numbers in the plate without errors.

The car license plate recognised using many other methods previously like MATLAB and DSP. Those methods had complexity in getting accurate result, time taken, efficient rate etc. To overcome such complexity other method is proposed using OpenCV. The language used here is Python. Raspberry Pi used here is mainly for recognizing the captured image, and it acts as real time displaying system.

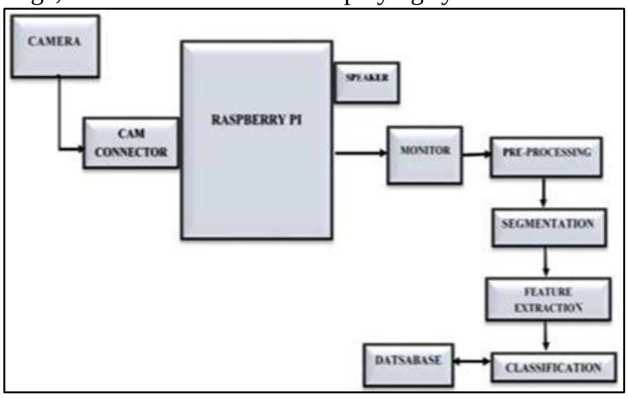


Fig. 1: Proposed Automated Vehicle Identity Recognition System

The hardware part consists of raspberry pi 3B, camera and LED. This is called as Raspberry pi camera module.

The recognition is done by raspberry pi from the image which camera captures in real time. This module is connected to a monitor, which is used for display. Raspberry pi is also connected to mouse and keyboard. Raspberry pi can also be connected to the PC monitor, laptop or TV monitor. In this case it is been connected to the PC monitor along with mouse and keyboard. Monitor is used to display the captured image and also the recognised character. Mouse and keyboard are the supporting features, which gives access to the monitor for controlling the raspberry pi recognition. After capturing the image the basic image processing steps are taken place. For each individual step the specific algorithm should be used based on the applications, because these algorithms can improve the recognition rate, efficiency rate and reduces the time taken.

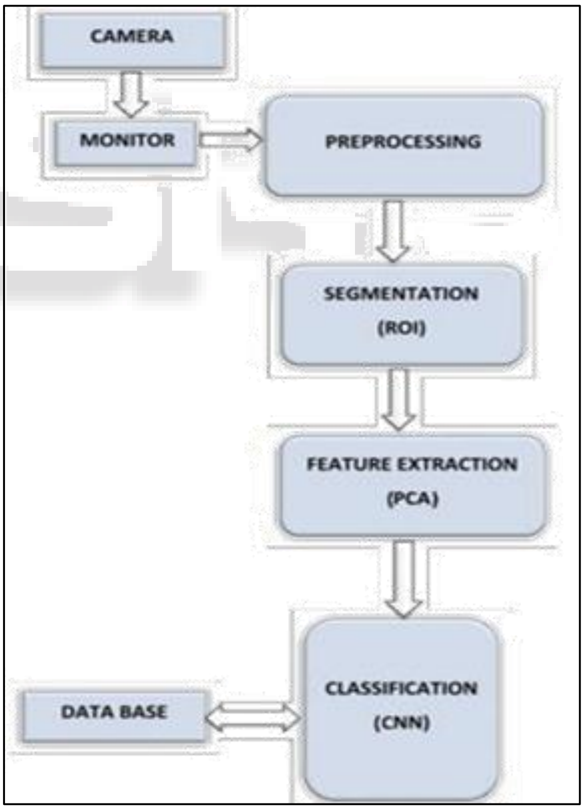


Fig. 2: Flow Chart of Processing of Image

The general image processing steps are used for this process, here we used many algorithms for each step. Classification is done by CNN algorithm. For segmentation separate region is parted and segmented after that classification step is taken place. CNN classifier is used to recognise the characters present on the vehicle number plate. These algorithms called out in the package. Inpython, programming code is grouped as package.

A. Real Time Input:

Camera, which is connected to Raspberry pi capture the image of the car in front and to process vehicle number plate detection and recognition. We give an image of car as input that goes through the image pre-processing stages that improves the image quality leads to better result in later stages.

B. Pre-processing:

Pre-processing is an important step in any image analysis. The most motive of pre-processing is to reinforce the standard of the image which will be processed for recognition .Various processes that we tend to area unit progressing to apply area unit changing, RGB image to Gray scale, noise reduction and binarization of an image. In python many packages are available to perform pre-processing.

C. Segmentation:

Character segmentation is carried on the binary image of the extracted number plate. The algorithm used here is horizontal scanning that makes use of a scanning line that finds the conditions satisfying the beginning and end position of the character.
Features will be extracted from segmented number plate.

D. Classification:

To recognize the segmented characters efficiently, we used convolutional neural network training to train our system over a dataset downloaded. After the training, we used the same neural network model for recognizing the characters.

III. CONVOLUTIONAL NEURAL NETWORK

A Convolutional Neural Network (CNN) is a Deep Learning algorithm which can take in a input image, provide significance to different perspectives/questions in the picture and have the capacity to separate one from the other. The pre-preparing required in a CNN is much lower when contrasted with other order algorithms. While in crude techniques channels are hand-built, with enough preparing, CNN can gain proficiency with these qualities.

The architecture of a CNN is comparable to that of the availability example of Neurons in the Human Brain and was propelled by the association of the Visual Cortex. Singular neurons react to improvements just in a confined locale of the visual field known as the Receptive Field. An accumulation of such fields Overlap to cover the whole visual territory.

A. Layers of Convolutional layer Network

The fundamental layers present in a Convolution Neural Network are as recorded underneath. The CNN's layers each play out a predefined activity that prompts satisfying the general undertaking of the CNN. The CNNs are commonly utilized for classification and object recognition purposes and accordingly, the layers present in the system help in deciding the set features of the images gained.
The vital layers present are as per the following:

1) Convolutional layer
2) Relu layer
3) Cross channel Normalization layer
4) Pooling layer
5) Dropout layer
6) Fully Connected layer
7) Loss layer

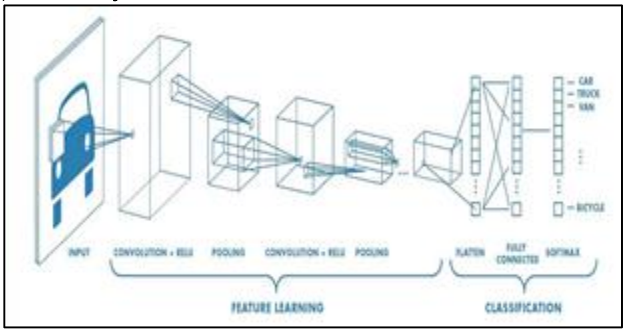


Fig. 3: Convolutional Neural Network

IV. HARDWARE & SOFTWARE REQUIREMENTS

A. Hardware

Raspberry Pi 3model B is used for processing, which has 1GB RAM, 4USB ports, SD card is used for memory, 40 pins. It is more powerful and efficient processor when compared other models of Raspberry Pi. This also has inbuilt Wi-Fi and Bluetooth models. The recommended programming language for Raspberry Pi is python.

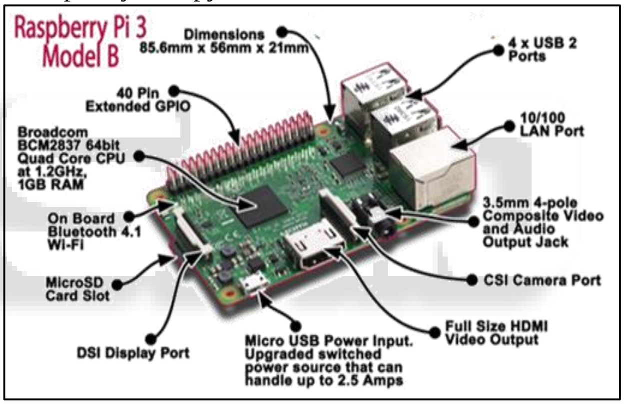


Fig. 4: Raspberry Pi 3 Model B

B. Software

– Keras is the open source library written in python for Neural Network and supports CNN.
– Tensorflow is the open source platforms used for building models in Machine Learning.
– OPenCV consist of library of programming functions for real time applications. It supports deep learning frame works like Tensorflow.
– Tkinter is the standard python interface for Graphical User Interface (GUI).

V. RESULTS

The performance of this proposed algorithm has been tested on several vehicle plates and provides very satisfactory results.

In our project we have classified the vehicles into two category, namely Registered vehicles for which Parking is allowed and Not Registered vehicles are allocated for visitors parking.

VI. CONCLUSION

The proposed work shows that free and open source technologies are matured enough for scientific computing domains. The system works satisfactorily for wide variations in illumination conditions and different types of number plates commonly found in India. When compared to simulation process it is for sure, a better alternative to the existing proprietary systems. Currently, we have proposed the algorithms for automated vehicle identity recognition system. We have implemented this system on Open CV library. The algorithms which is suitable for this application is been chosen and the process were performed and executed successfully.

REFERENCES

[1] N.Abirami, Dr.J.S.Leena Jasmine. Accurate Vehicle Number Plate Recognition And Real Time Identification Using raspberrypi. International Research of Engineering and Technology(IRJET). E-ISSN:2395-00056, p-ISSN:2395-0072, Volume:05,Apr-2018.
[2] Aishwarya Agrawal, Nikita Pardakhe. Automatic licence plate recognition using raspberry pi. International Interdisciplinary Conference on Science Technology Engineering Management Pharmacy and Humanities, ISBN: 9780998900001, 22nd – 23rd April 2017.
[3] Prof. Kumthekar a.v., ms. Sayali owhal, ms. Snehal supekar, ms. Bhagyashri tupe. Recognition of vehicle number plate using raspberry pi. International research journal of engineering and technology (IRJET). E-ISSN: 2395-0056,p-ISSN: 2395-0072.Volume: 05, apr-2018.
[4] D.Lavanya,C.V.Keerthi latha,Nirmala. License Plate Extraction Of Images Using Raspberry Pi. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET). Volume 4 Issue 1, January 2015.
[5] Radhika S. Dangare, Prof. G. D. Dalvi. "Automatic Number Plate Recognition for use in Different Countries using Raspberry pi" (Real Time Approach). International Journal of Science and Research (IJSR). ISSN (Online): 2319-7064, Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391.
[6] Ayush Agarwal, Aman Saluja, License plate recognition for remote monitoring of traffic. International journal of industrial electronics and electrical engineering. ISSN: 2347-6982, volume-5, issue-5, may-2017.

ISSN 2321 3361 © 2019 IJESC

Research Article Volume 9 Issue No. 7

Multi Lingual Speech to Text Conversion using Matlab

Hamsavahini. R[1], Mamatha. K. R[2], Saneesh Cleatus. T[3]
Department of Electronics and Communication Engineering
BMS Institute of Technology and Management, Bangalore, India

Abstract:
This paper presents a multilingual speech to text conversion system. Conversion is based on information present in speech signal. Speech is the natural and most important form of communication for human being. Speech-To-Text (STT) system takes a human speech utterance as an input and requires a string of words as output. The objective of the system is to extract, characterize and recognize the information about speech. Mel Frequency Cepstrum Coefficients (MFCCs) technique is used to extract the features from a speech signal. The proposed recognition system was designed and implemented using MFCC algorithm and k-means clustering in MATLAB. This system helps to understand the unknown language easily by displaying the text.

Keywords: Speech to text (STT), Mel frequency cepstral coefficients (MFCC), k-means clustering.

I. INTRODUCTION

Speech is an easiest way to communicate with each other. Speech processing is widely used in many applications like security devices, cellular phones, ATM machines and computers. Speech to text conversion (STT) systems have a lot of benefits for the deaf people. Also we see manyPeople finding difficult to understand other languagesWhich they don't know.Humans interact with each other in several ways such as facial expression, eye contact, gesture and mainly speech. Speech is the primary mode of communication among humans and also most natural and efficient form of exchanging information among us. Speech to text conversion is used in many applications. It is widely usedin educational fields to help deaf people. Speech is the most natural form of communication for humans and the speech of everyindividual is unique.This is mainly due to biological factors such as size of vocal tract or due to behavioral characteristics such as accent, speaking speed etc. A voice recognition system can also be used as a biometric system that can identify an individual via the unique acoustic characteristics of the individual's voice. This system has a wide variety of applications in the field of security. In speech to text conversion feature extraction is the main part. There are various feature extraction techniques. Some of the most commonly used frequency extraction techniques are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Cepstral Analysis, MFCC Calculations etc. In this paper we are using Mel frequency cepstral coefficient (MFCC)method .It is based on the characteristics of the human ear's hearing, which uses a nonlinear frequency unit to simulate the human auditory system. Mel frequency scales are used to extract features of the speech. After MFCC we do K-means clustering.

K- means clustering is a partitioning method. The function K-means partitions data into k mutually exclusive clusters and returns the index of clusters to which it has assigned each observation. K-means treat each observation in your data as an object having a location in space. After this we do distance measurement and identify the input speech signal and convert it to speech.

II. PROPOSED SYSTEM

The proposed system does the following operations:
Pre-processing, Speech recognition, Feature extraction, MFCC calculations.K-means clustering,Distance measurement and Displaying the text.

The flow diagram of the proposed system is as given Below:

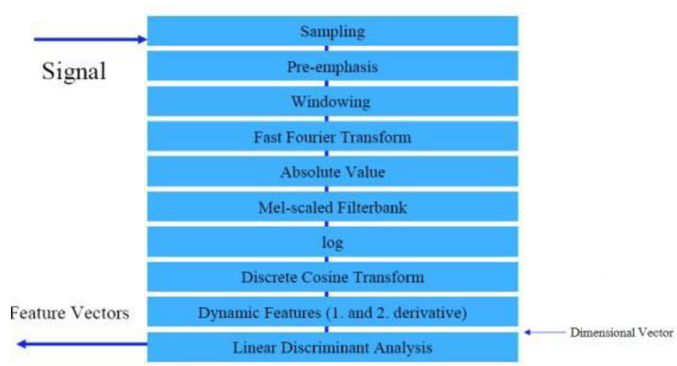


Figure.1. Flow Diagram of Proposed system

i. Pre-Processing: The input speech signalis sampled andcut into frames. Speech is a highly variable signal and is difficult to analyze as it is. However, in very short intervals of time (in milliseconds) the signal appears to be stationary thereby facilitating analysis. After cutting into frames, the signal is then windowed to remove any discontinuities between the frames and reduce spectral leakages.

ii. Speech recognition: Speech analysis is done in this stage. During this stage it identifies the speech data which contains different types of information and can also be used for storing the speaker identity information.

iii. Feature extraction: The unique characteristics of an individual's voice is extracted and stored in a database for future referencing. The main goal here is to simplify the recognition by summarizing the vast amount of speech data without losing the acoustic properties that defines the speech. Feature Extraction is used in both training and testing phases.

iv. MFCC calculations: The short-term power spectrum of a sound based on a linear cosine transform of a log power spectrum on a non-linear Mel scale. The coefficients that make up thisCepstrum are called Mel Frequency Cepstral Coefficients (MFCCs). Extracting MFCCs from the speech signal consists of a few processes as Fast Fourier Transform, Mel Freq uency Warping and Cepstrum which does the inverse FFT of the logarithm of an estimated spectrum of a signal.

v. K-means clustering: K-means clustering is an elementary but very popular approximate method that can be used to simplify and accelerate convergence. The k-means algorithm divides a set of N samples X into K disjoint clusters C, each described by the mean µj of the samples in the cluster. The means are commonly called the cluster "centroids"; note that they are not, in general, points from X,although they live in the same space. The K-means algorithm aims to choose centroids that minimise the *inertia*, or within-cluster sum of squared criterion:

$$\sum_{i=0}^{n} \min_{\mu_j \in C}(||x_j - \mu_i||^2) \text{ ------------------------(1)}$$

Where µj=mean of the sample.

K-means can be used for vector quantization. This is achieved using the transform method of a trained model of K-means.

vi. Distance measurement: In this stage the distance is measured between the already existing database and the test sample that is recorded and saved in .wav format and matches with the nearest database.

Vii.Displaying the text: The most appropriate database that is matched in distance measurement stage is displayed in the form of required text on the display. [7]

III. WORKING

First the voice samples are recorded and saved in .wav format. These voice samples are pre-processed. Pre-processing involves sampling. After sampling, we do framing and windowing Here we are using hamming window. Hamming window is also called as raised cosine window. The equation and plot for the hamming window is shown below. In a window function there is a zero valued outside of some chosen interval. For example, a function that is stable inside the interval and zero elsewhere is called a rectangular window that illustrate the shape of its graphical representation. When signal or any other function is multiplied by a window function, the product is also zero-valued outside the interval. The windowing is done to avoid problems due to truncation of the signal. Window function has some other applications such as spectral analysis, filter design and audio data compression such as vorbis.

$$w(i)=0.54+0.46*\cos\left(\frac{2\Pi i}{N}\right)\text{..............................(2)}$$

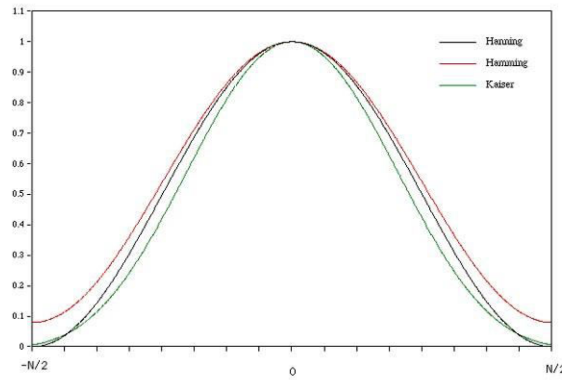


Figure.2. Graphs of various windows

After sampling feature extraction is done. Feature extrac tion is done using MFCC. Mel frequency Cepstral Coefficients are coefficients that represent audio based on perception. This coefficient has a great success in speaker recognition application. In the Mel Frequency Cepstral Coefficients the calculation of the Mel Cepstrum is same as the real Cepstrum except the Mel Cepstrum's frequency scale is warped to keep up a correspondence to the Mel scale. The Mel scale was projected by Stevens, Volkmann and Newman in 1937.The Mel scale is mainly based on the study of observing the pitch or frequency perceived by the human. The scale is divided into the units Mel.

$$m=1127.01048\log_e\left(1+\frac{f}{700}\right)\text{...............(3)}$$

$$f=700\left(e^{m/1127.01048}-1\right)\text{..................(4)}$$

Where f= normal frequency
m=Mel-frequency

The equations (3) above shows the mapping the normal frequency into the Mel frequency and equation (4) is the inverse, to get back the normal frequency.

The block diagram of MFCC is as shown below:

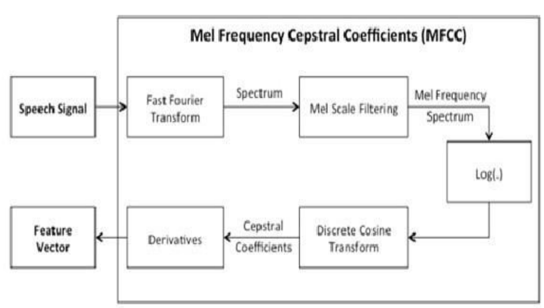


Figure.3. Block diagram of MFCC

The input for MFCC is the speech signal in time domain. Then using Fourier transform we convert the time domain signal to frequency domain signal .After converting it to frequency domain we use Mel scale filtering to get Mel frequency spectrum. To the obtained Mel frequency spectrum we apply logarithm. Logarithm is used in order to mimic the human perception of loudness because it is experimentally proved that humans perceive loudness on logarithmic scale. After logarithm we perform discrete cosine transform to get cepstral coefficients. Cepstral coefficients are used to eliminate speaker dependent characteristics. The cepstral coefficients can be calculated using the following formula:

$$c_{T,j}^{(4)} = \sum_{j=1}^{Nd} c_{T,j}^{(3)} \cos\left[\frac{k(2j-1)\Pi}{2Nd}\right] \ldots\ldots\ldots\ldots (5)$$

where k=0,1,..........,Nmnc<Nd

Where Nmc is the no of chosen cepstral coefficients. Next step is calculating derivatives. To represent the dynamic nature of speech first and second order derivatives of the cepstral coefficients extend the feature vector. The final feature vector is

$$C_T = [C_{T,j}^{(4)}, \Delta C_{T,j}^{(4)}, \Delta\Delta C_{T,j}^{(4)}] \ldots\ldots\ldots\ldots (6)$$

Where $C_{T,j}^{(4)}$, $\Delta C_{T,j}^{(4)}$, ,$\Delta\Delta C_{T,j}^{(4)}$ are derivates of cepstral coefficients.

After calculating MFCC coefficients we do k-means clustering. K-means clustering partitions data into k number of mutually exclusive clusters. This techniqueassigns each observation to a cluster by minimizing the distance from the data point to mean or median location of its assigned cluster respectively. The process of k means algorithm used least-squares partitioning method to dividethe input vectors into k initial sets. It then calculates the mean point or centroid of each set. It constructs a new partition by associating each point with the closest centroid. Then thecentroids are re calculated for the new clusters, and algorithm repeated until when the vectors no longer switch clusters or alternatively centroids are no longer changed. After this the given voice sample is compared with the data base samples and which of the data sample in the data base has less distance to the given voice sample is displayed on the screen.

IV.SIMULATION RESULTS

Below shown figures are the simulation results of our system. They show the graphs of the input signal and the database signals and the converted text.

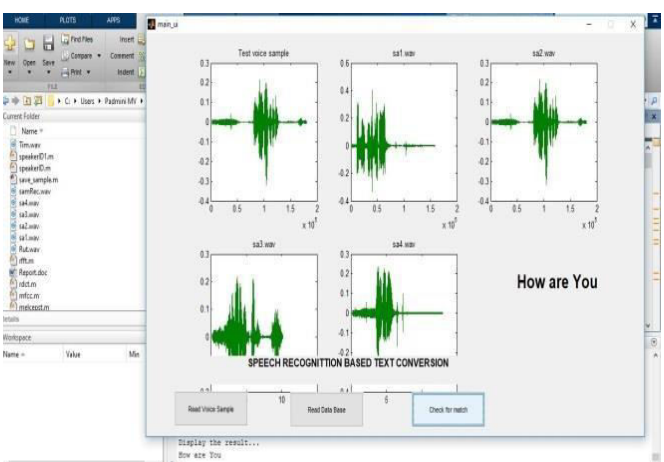


Figure.4. Graphof first sample matching with the existing database

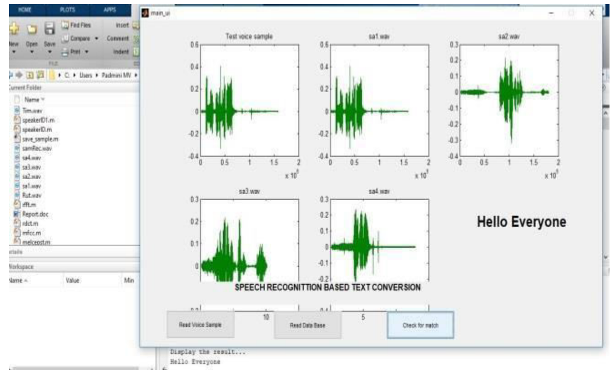


Figure.5. Graph of second sample matching with the existing Database

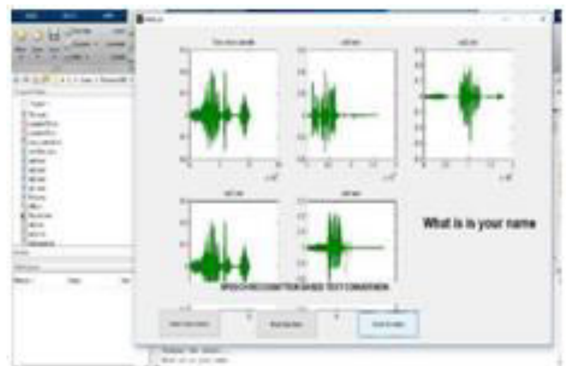


Figure.6. Graph of third sample matching with the existing database.

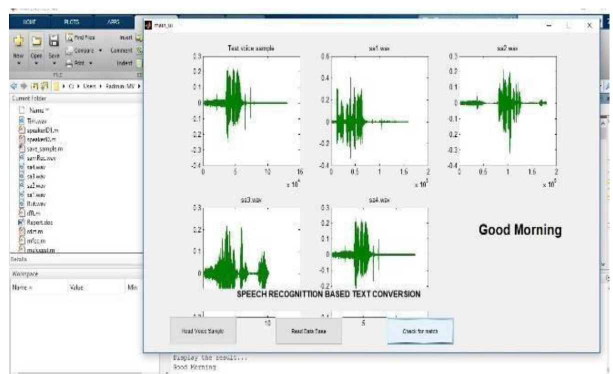


Figure.7. Graphof fourth sample matching with the existing database

IV. CONCLUSION

The voice recognition system was successfully designed and implemented using matlab. The system was successfully able to identify speakers with high accuracy in low noise environments and with an accuracy compared to moderate noise environments. Higher accuracy rates could be achieved by either using two factor authentication or by using different feature matching techniques such as Hidden Markov Models that would yield high accuracy but potentially at the cost of computation time and memory.

V. FUTURE SCOPE

This system can take only four inputs. In future the no.of inputs can be increased and can also be made real time. This system converts Telugu, Kannada, and Hindi language to

English. This can beextended for other regional languages. This connected word speech recognition system is developed only for speaker independent system.

VI. REFERENCES

[1]. Karpagavalli S and Chandra E A Review on Automatic Speech Recognition Architechtures Approaches International Journel of Signal Processing, Image Processing and Pattern Recognition Vol.9,No.4,(2016)

[2].Amit Kumar Singh,Rohan Singh.Ashutosh d," Mel Frequency Cepstral Coefficients Based Text Independent Automatic Speaker Recognition Using Matlab",2014 International Conference on Reliability ,optimization and Information Technology-ICROIT 2014.

[3].Priyanka P. Patil, Sanjay A. Pardeshi,"Marathi Connected Word SpeechRecognition System IEEE First International Conference on NetworksSoft Computing pp 314-318,Aug 2014.

[4]. Yogita H. Ghadage, Sushama D. Shelke," Speech to Text Conversion for Multilingual Languages", International Conference on Communication and Signal Processing, April 6-8, 2016, India.

[5]. V.Udayashankara, Swapna Havalgi," Speech Therapy System to Kannada Language", 2016 Second International Conference on Cognitive Computing and Information Pro cessing (CCIP).

[6]. Ashwin Nair Anil Kumar, Senthil Arumugam Muthu kumaraswamy," Text Dependent Voice Recognition System using MFCC and VQ for Security Applications", International Conference on Electronics, Communication and Aerospace Technology ICECA 2017.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1742-7371.htm

Self-improved grey wolf optimization for estimating carrier frequency offset in SCM-OFDM systems

Received 7 March 2019
Revised 2 July 2019
18 November 2019
Accepted 18 November 2019

Rashmi N.
Department of Electronics and Communication Engineering, BMS Institute of Technology and Management, Bangalore, India, and
Mrinal Sarvagya
School of Electronics and Communication Engineering, Reva University, Bangalore, India

Abstract

Purpose – The purpose of this paper is to demonstrate a proficiency for accomplishing optimal CFO and keep down the error among the received and transmitted signal. Orthogonal frequency-division multiplexing (OFDM) is considered as an attractive modulation scheme that could be adopted in wireless communication systems owing to its reliability in opposition to multipath interruptions under different subchannels. Carrier frequency offset (CFO) establishes inter-carrier interference that devastates the orthogonality between the subcarriers and fluctuates the preferred signal and minimizes the effectual signal-to-noise ratio (SNR). This results in corrupted system performance. For sustaining the subcarriers' orthogonality, timing errors and CFOs have to be approximated and sufficiently compensated for. Single carrier modulation (SCM) is a major feature for efficient OFDM system.

Design/methodology/approach – This paper introduces a novel superposition coded modulation-orthogonal frequency-division multiplexing (SCM-OFDM) system with optimal CFO estimation using advanced optimization algorithm. The effectiveness of SCM-OFDM is validated by correlating the transmitted and received signal. Hence, the primary objective of the current research work is to reduce the error among the transmitted and received signal. The received signal involves CFO, which has to be tuned properly to get the signal as closest as possible with transmitted signal. The optimization or tuning of CFO is done by improved grey wolf optimization (GWO) called GWO with self-adaptiveness (GWO-SA). Further, it carries the performance comparison of proposed model with state-of-the-art models with the analysis on bit error rate (BER) and mean square error (MSE), thus validating the system's performance.

Findings – From the analysis, BER of the proposed and conventional schemes for CFO at 0.25 was determined, where the adopted scheme at 10th SNR was 99.6 per cent better than maximum likelihood, 99.6 per cent better than least mean square (LMS), 99.3 per cent better than particle swarm optimization (PSO), 75 per cent better than genetic algorithm (GA) and 25 per cent better than GWO algorithms. Moreover, MSE at 1st SNR, the proposed GWO-SA scheme, is 4.62 per cent better than LMS, 60.1 per cent better than PSO, 37.82 better than GA and 67.85 per cent better than GWO algorithms. Hence, it is confirmed that the performance of SCM-OFDM system with GWO-SA-based CFO estimation outperformed the state-of-the-art techniques.

Originality/value – This paper presents a technique for attaining optimal CFO and to minimize the error among the received and transmitted signal. This is the first work that uses GWO-SA for attaining optimal CFO.

Keywords SCM, GWO, Bet error rate, Carrier frequency, MIMO-OFDM, MSE analysis

Paper type Research paper

International Journal of Pervasive Computing and Communications

1742-7371
DOI 10.1108/IJPCC-03-2019-0020

Nomenclature

OFDM	= Orthogonal Frequency-Division Multiplexing;
MIMO	= Multiple-Input Multiple-Output;
SCM	= Superposition Coded Modulation;
SC-SCM	= Single-Code SCM;
MCSCM	= Multi-Code SCM;
PAPR	= Peak-to-Average Power Ratio;
ML	= Maximum Likelihood;
OFDMA	= OFDM Access;
BS	= Base Station;
JMAP	= Joint Maximum a posteriori;
MJMAP	= Modified JMAP;
DSC	= Doubly Selective Channel;
OQAM	= Offset Quadrature Amplitude modulation;
NDA	= Non-Data Aided;
STO	= Symbol Timing Offset;
I/Q	= In-phase/quadrature;
AF	= Amplify-and-Forward;
BER	= Bit Error Rate;
SKF	= Schmidt–Kalman Filtering;
AWGN	= Additive White Gaussian Noise;
GPF	= Gaussian particle filtering;
SR-CFO	= Sparse Recovery assisted CFO;
SM-OFDM	= Spatial Modulation OFDM;
DFT	= Discrete Fourier Transform;
ECC	= Error Control Coding;
CM	= Coded Modulation;
MLSE	= Maximum Likelihood Sequence Estimation;
BCM	= Block Coded Modulation;
UEP	= Unequal Error Protection;
FFT	= Fast Fourier Transform;
PSO	= Particle Swarm Optimization;
MUI	= Multi-User Interference;
HPSO	= Hybrid Particle Swarm Optimization;
SVM	= Support Vector Machine; and
PU	= Primary User.

1. Introduction

Multiple-input, multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) has been adopted by several wireless communication systems owing to its increased spectral robustness and effectiveness to multipath fading. Orthogonal frequency-division multiplexing (OFDM)-dependent systems ((Jose, 2017; Liu *et al.*, 2018; Xu *et al.*, 2015b)) are vulnerable to performance deprivation owing to impairments such as carrier frequency offset (CFO; Daoqi *et al.*, 2016; Abdzadeh-Ziabari *et al.*, 2018). Currently, the estimation of CFO is exploited to minimize the effect of frequency offset. The technology for CFO (Huang *et al.*, 2017; Zhang *et al.*, 2016; Zhao *et al.*, 2017) estimation could minimize the necessity of central wavelength positioning precision for a local oscillator laser and transmitting laser in the consistent receiver. It takes place owing to the frequency differentiations found among RF oscillators deployed in the OFDM receiver and transmitter, and channel-based Doppler shifts.

The OFDM systems performance will be depreciated if the channel and CFO (Kang *et al.*, 2018; Saeedi-Sourck, 2017; Gul *et al.*, 2014) were not approximated and remunerated correctly.

In addition to CFO estimation, single carrier modulation (SCM) is a well-known transmission model. Accordingly, two kinds of SCM models exist, i.e. MCSCM and SC-SCM (Tong and Ping, 2010; Damrath *et al.*, 2017). The major intention of SCM systems is to decrease the peak-to-average power ratioof the signal that was transmitted. Moreover, a characteristic property of SCM (Chau *et al.*, 2018) is its randomness, simplicity, performance and flexibility. SCM could be generated by logically superimposing the waveforms from various relay/antennas/users nodes. This feature is much constructive for multiple user and *ad hoc* environments. Moreover, SCM (Krikidis and Thompson, 2012; Cai and Shen, 2017) was often deployed as a hypothetical concept in such an environment.

So far, numerous techniques have been established for the reimbursement of the degrading influences of CFO (Lin *et al.*, 2012; Su *et al.*, 2009; Saeedi-Sourck *et al.*, 2011; Fusco *et al.*, 2008) and channel. Precise approximations of CFO and symbol timing using a training sequence would be attained together with maximum likelihood (ML) joint estimator. For several years, intensive studies were performed on CFO (Fusco *et al.*, 2008) estimation for OFDM access systems. It is known that OFDMA systems are vulnerable to CFO (Li *et al.*, 2008; GUAN *et al.*, 2010) and for having acceptable performance; thus, OFDMA receivers have to offer a better estimation of CFO (Li *et al.*, 2007). However, while performing uplink transmission in OFDMA, various clients convey signals to base station. Hence, CFO estimation (Chopra *et al.*, 2017) turns out to be more challenging because various users have dissimilar CFOs. Several optimization algorithms are considered to optimize the parameters and obtain better results (Shankar and Jaisankar, 2016; Singh *et al.*, 2018; Shareef and Rao, 2018).

This paper contributes a novel superposition coded modulation-orthogonal frequency-division multiplexing (SCM-OFDM) system with optimal CFO by deploying an enhanced optimization model. The efficiency of SCM-OFDM is confirmed by associating the transmitted and received signal. The major objective of this work is to reduce the error among the transmitted and received signal. In the received signal, CFO is integrated and has to be appropriately attuned to obtain the signal as nearer to the transmitted signal. The tuning or optimization of CFO is made by improved meta-heurisc algorithm called grey wolf optimization with self-adaptiveness (GWO-SA). A metaheuristic is an advanced technique or heuristic designed to locate, create or select a heuristic that may provide adequately superior result to an optimization problem, particularl with partial or imperfect information. Metaheuristic sample a set of solutions that is too large to be completely sampled. Metaheuristic may make few assumptions about the optimization problem being solved; therefore, they may be usable for various problems. Moreover, the adopted scheme is distinguished with conventional schemes with respect to BER and MSE and the system performance is successfully validated. The paper is organized as follows. Section II describes the related works done under this topic. Section III portrays the superposition coded modulation and Section IV describes the mathematical model on CFO estimation in superposition coded modulation. In addition, Section V explains the results and Section VI provides the conclusion for this study.

2. Literature review

2.1 Related works

In 2017, Jose (2017) have suggested a novel technique, which deploys the joint maximum a posteriori(JMAP) approach for the evaluation of CFO and channel in OFDM by exploiting the previous arithmetic information of the channel. To decrease the complication of the

JMAP estimator, a modified joint maximum a posteriori scheme that has no grid explorations was implemented. The estimated techniques were examined by arithmetical simulations and resulting conclusions, which authenticate the improved performance of the proposed technique while distinguishing with the conventional schemes.

In 2018, Liu *et al.* (2018) have introduced a scheme that handles with the issues of blind synchronization for OFDM depending on OFDM/OQAM systems with no previous information regarding the response of the channel. A novel NDA joint estimation technique of CFO and symbol timing offset (STO) was considered in multipath fading channels condition. Finally, numerous extensive experiments were performed to assess the high efficiency and effectiveness of the introduced scheme.

In 2015, Xu *et al.* (2015b) had established a new approach based on a blind estimator of CFO and I/Q imbalance in OFDM system that does not necessitate any previous acquaintance of channel side information or training symbols. The most important reason behind this approach was the null sub-carriers that were generally found in numerous OFDM systems. At last, execution outcomes were offered to substantiate the efficacy of the adopted model and performance evaluation was carried out with conventional schemes.

In 2016, Daoqi *et al.* (2016) have introduced a new approach that handles with the performance examination of OFDM amplify-and-forward (AF) system in the existence of CFO. Even though SCM was examined comprehensively in OFDM systems; however, system BER with CFO was not obtainable in the conventional works. At last, its robustness was evaluated usig widespread experimentations.

In 2018, Abdzadeh-Ziabari *et al.* (2018) have adopted a scheme, which offers two new techniques for DSC estimation and joint CFO systems. Here, two new-fangled schemes were subsequently implemented depending on SKF. The initial technique exploits Schmidt–Kalman filtering for every user to approximate block element modifier (BEM) and CFO coefficients. The subsequent approach exploits Gaussian particle filtering together with SKF to approximate BEM and CFO coefficients of every user. Thus, the proposed model was found to establish better outcomes from the experimental analysis.

In 2017, Huang *et al.* (2017) have established a novel blind CFO estimator for the uplink OFDMA systems. On utilizing the sparsity entrenched in the OFDMA data, the CFO was manipulated as a sparse recovery optimization issue with increased resolution. Moreover, the behavior of established technique together with other conventional estimators was analyzed, and the arithmetical outcomes demonstrate that the established methodology was better than the conventional estimators with respect to consistency.

In 2016, Zhang *et al.* (2016) have suggested a novel computationally proficient blind CFO estimator for MIMO-OFDM systems. In addition, a cost function was cautiously modeled, and it could be precisely articulated as the superposition of certain cosine waves with the impact of noise. Finally, both arithmetical outcomes and hypothetical performance were offered to substantiate the established scheme. Moreover, the performance of the suggested technique was examined and compared with three conventional techniques to analyze the performance of the technique.

In 2017, Zhao *et al.* (2017) have adopted a novel proficient detection system to alleviate the consequence of the CFO in SM-OFDM systems. Particularly, the implemented technique deploys a novel detector to SM-OFDM models with several CFOs in which zero symbols were measured as the constellation points. Moreover, depending on the outcomes attained by a variety of conventional algorithms, the adopted scheme was found to offer better results.

In 2015, Xu *et al.* (2015a) have introduced hybrid particle swarm optimization (HPSO) to enhance the performance of OFDM network. HPSO is the combination of both PSO and SVM. This was to solve the problem of resource allocation. A resource allocation problem over the power and subcarrier allocation based on chance-constrained programming was formulated to maximize the average weighted sum-rate throughput and guarantee the probabilistic interference constraint condition for PU. To solve the above resource allocation problem, the probabilistic interference constraint condition was computed using SVM and HPSO.

In 2019, Kaur and Sharma (2019) have proposed a discrete wavelet transform (DWT) approach with conventional MIMO–OFDM-RoF system to attain the threshold bit error rate (BER). The proposed model was also incorporated with Alamouti-Space–time Block Codes (A-STBC) MIMO–OFDM based Radio over Fibre (RoF) system and the resultant signal was transported over a RoF link of 10 km in the presence of non-linearity. At the radio access point (RAP), the RF received signals are transmitted over Rayleigh faded wireless channel in the presence of additive white Gaussian noise (AWGN) and phase offset arising because of imperfect carrier phase estimation.

In 2018, Ali *et al.* (2018) have experimented a quasi-gapless integrated with visible light communication and positioning (VLCP) system. While comparing with orthogonal frequency division multiplexing-based SCM (OFDM-SCM), FBMC-SCM has lower out-of-band interference (OOBI) and therefore requires much smaller guard band (GB) spacing. A low-complexity transmit diversity transmission scheme was adopted here, where all the light emitting diode (LED) lamps transmit the same communication signal. Moreover, phase difference of arrival (PDOA) was applied for positioning. The experimental results show that, in a coverage area of $1.2 \times 1.2\,m^2$ with a height of 2.1 m, the mean position errors using OFDM–SCM and FBMC-SCM for a GB of 0.7 MHz are 10.91 and 6.08 cm, respectively. Moreover, a comparable BER performance can be achieved for both OFDM–SCM and FBMC-SCM. Due to the negligible GBs when using FBMC-SCM, the effective bandwidth utilization ratio of the integrated VLCP system was improved from 72 to 98 per cent when OFDM-SCM was replaced by FBMC-SCM.

2.2 Review

Table I shows the methods, features and challenges of conventional techniques based on the CFO in MIMO-OFDM systems. At first, MJMAP algorithm was introduced in Jose (2017), which offers the reduced complexity and offers minimized error. However, it needs a contemplation on grid searches, having possibilities for omission error. The non-data aided (NDA) method, which was a feasible and effective model, was exploited in Liu *et al.* (2018), and it offered reduced noise effects, but the increase in normalized doppler shift affected the performance. In addition, DFT was deployed in Xu *et al.* (2015b) and provided reduced complexity and offered a precise estimation of CFO. However, it required more contemplation on error constraints. Similarly, the AF model was exploited in Daoqi *et al.* (2016), which offers a high signal-to-noise ratio (SNR) and reduced BER. However, MSE has to be more concerned. Also, SKF was used in Abdzadeh-Ziabari *et al.* (2018), which offers enhanced MSE. It also minimizes the unknown constraints; however, reducing the count of particles may result in poor PDF estimation. Note that sparse recovery-assisted (SR-CFO) was exploited in Huang *et al.* (2017) and offered high reliability and provided minimum MSE; however, it seems to be more complex. Blind estimator was implemented in Zhang *et al.* (2016), which offered improved MSE and offered minimum cost function; however, the performance gap decreased with a reduction in block count. In addition, GA at last, SM-OFDM was suggested in Zhao *et al.* (2017) that offers reduced BER and provides minimum complexity. However, it requires more contemplation on generalized SM systems. Therefore, these limitations have to be considered for improving the CFO in SCM-OFDM systems effectively in the current research work.

Table I. Review of state-of-the-art CFO estimation models in MIMO-OFDM systems

Author [citation]	Adopted methodology	Features	Challenges
Jose (2017)	MJMAP algorithm	Reduced complexity Minimized error	Needs contemplation on grid searches
Liu *et al.* (2018)	NDA method	Feasible and effective model Reduced noise effects	Increase in normalized doppler shift affects the performance
Xu *et al.* (2015b)	DFT	Reduced complexity Offers precise estimation of CFO	Needs more contemplation on error constraints
Daoqi *et al.* (2016)	AF model	High SNR ratio Reduced BER	MSE has to be concerned more
Abdzadeh-Ziabari *et al.* (2018)	SKF	Enhanced MSE Minimizes the unknown constraints	Reducing the count of particles may result in poor PDF estimation
Huang *et al.* (2017)	SR-CFO	Better reliability Minimum MSE	More complex
Zhang *et al.* (2016)	Blind Estimator	Improved MSE Minimum cost function	Performance gap decreases with the reduction in block count
Zhao *et al.* (2017)	SM-OFDM	Reduced BER. Minimum complexity	Requires contemplation on generalized SM systems
Xu *et al.* (2015a)	HPSO	It provides better average weighted sum-rate throughput and satisfy the probabilistic interference constraint condition	Large computational complexity Requires large feedback overhead
Kaur and Sharma (2019)	A-STBC MIMO–OFDM	Power efficient L:ow power modulation	High computation
Ali *et al.* (2018)		Achieved effective band width Minimum complexity	High BER achieved

3. Superposition coded modulation

Modulation processes and error control coding have to be included in a band-limited sector. It was revealed that the integrated methods, known as CM, offer both bandwidth and power efficient communication by plotting information sequences onto an extended set of channel signals with the assistance of ECCs. The method is known as BCM when the component codes of the CM are selected from block codes. For BCM arrangements of reduced block lengths, the optimum decoding process, termed as MLSE, is feasible for execution owing to their lattice arrangement (Seshadri and Sundberg, 1991). Usually, a 2’-ary CM system shows “/” varied protection levels. Furthermore, unequal error protection (UEP) capacity next to intrinsic levels of CM scheme could be attained with various schemes. A flexible technique to sum up UEP to CM is known as SCM (Bergmans and Cover, 1974). Equation (2) is exploited in a standardized model to describe SCM.

Assume that p show the count of UEP levels and $C_1, C_2 \ldots C_p$ indicates bit streams in reducing the order of significance. Assume the energy transmitted for each bit of the bit stream C_i, indicated by e_i, and the relevant modulation constellation is referred to as U_i. Consider $\psi_1, \psi_2, \ldots \psi_p$ indicates the relevant channel codes and $w_1, w_2, \ldots w_p$ specifies their corresponding code rates. Then such rates are specified concerning the bit rate information by the entire count of modulation intervals. Assume that y_i indicates the codeword of the code ψ_i from the constellation U_i. Then, equation (1) denotes the transmitted sequence and the entire code rate could be evaluated by equation (2), and the model of SCM transmitter is given by Figure 1:

$$t = \sum_{i=1}^{p} y_i \tag{1}$$

$$w = \sum_{i=1}^{p} w_i \tag{2}$$

The notable features of SCM is given as follows:

- The detection of SCM can be done at low cost.
- SCM is highly suitable for adaptive modulation.

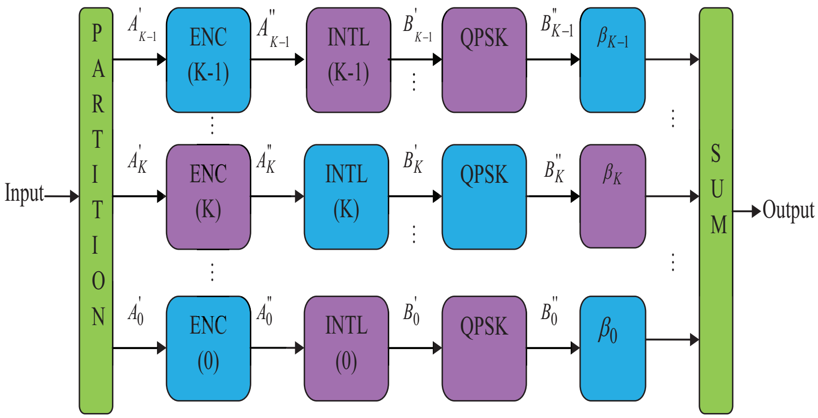


Figure 1. Model of an SCM transmitter

- SCM provides a remedy to the PAPR problem in OFDM.
- The linear superposition of several SCM signals naturally results in an SCM signal. These signals can be encoded distributively at different users/antennas/relay nodes. The receiver structure remains roughly the same for different variations. This property is very useful in multiple access/relay/ad hoc networks, where a challenging issue is how to separate signals from different sources after they are superimposed.
- There is a posibility to build space-time codes by SCM, which can attain high multiplexing gains and diversity in MIMO systems.

4. Mathematical model on carrier frequency offset estimation in superposition coded modulation

4.1 Carrier frequency offset estimation model

The main cause of CFO among the BS and users are attributed to the local oscillator mismatch and/or the mobility of users. In fact, it demolishes the orthogonality among the sub-carriers. In addition, ICI and MUI caused by these CFOs severely degrade the system performance. Under such circumstances, carrier synchronization is a fundamental portion of the BS receiver design. Carrier synchronization is performed both on the downlink and the uplink of a cellular communication system. On the downlink, a mobile user tries to synchronize to the corresponding BS.

Consider the data after constellation mapping and serial-to-parallel conversion are not dependent and are uniformly distributed (Shah *et al.*, 2013); hence, $E\left\{b^P_{o,m}.b^P_{o',m'}\right\} = E\left\{b^J_{o,m}.b^J_{o',m'}\right\} = \sigma^2_a.\delta(o-o').\delta(m-m')$, $E\left\{b^P_{o,m}.b^J_{o',m'}\right\} = 0, E\left\{b^J_{o,m}.b^P_{o',m'}\right\} = 0$ and $E\left\{|\mu_k|^2\right\} = \sum_{n=0}^{N-1} |g_n|^2 = \sigma^2_\mu$, in which μ_k indicates the attenuation factor, g_n denotes channel coefficient of n^{th} path, $b^J_{o,m}$. and $b^P_{o,m}$ indicate the imaginary and real parts of complex symbols of data that are transmitted on o^{th} subcarrier throughout m^{th} OFDM/OQAM symbols, respectively, and N specifies the order of multi-path channel.

The autocorrelation function of received OQAM/OFDM signals is indicated by equation (3), in which $\otimes$ denotes the symbol of convolution function. Here, σ denoting the spectral density f_e indicates the normalized CFO, m_e denotes STO and $h[l]$ indicates the prototype filter and M refers to the number of sub-channels (Mohan *et al.*, 2016):

$$
\begin{aligned}
B[l,\tau] &= E\left\{r[l].r^*[l-\tau]\right\}. \\
&= \sigma^2_a.\,\sigma^2_\mu.e^{j\frac{2\pi}{M}f_e\tau}\,.\sum_{o=0}^{M-1}|w_o|^2\,e^{j\frac{2\pi}{M}.o\tau}\,.\,.h[l]h^*[l-\tau]\otimes\left(1+\delta\left[l-\frac{M}{2}\right]\right) \\
&\quad \otimes\delta[l-m_e]\otimes\sum_{m=-\infty}^{\infty}\delta[l-mM]+\sigma^2
\end{aligned}
\tag{3}
$$

As per (3), the CFO data could be obtained from phase without difficulty. For a predetermined lagτ, the autocorrelation function is extended into a Fourier series with

Fourier coefficients known as "cyclic-cumulant." Based on the below properties of DFT, Equations (4), (5) and (6) are formulated, in which $FFT\{.\}$ signifies for FFT:

$$FFT\left\{\sum_{m=-\infty}^{\infty}\delta[l-mM]\right\}=\frac{1}{M}\sum_{m=-\infty}^{\infty}\delta\left[\beta-\frac{m}{M}\right] \tag{4}$$

$$FFT\left\{\delta\left[l-\frac{M}{2}\right]\right\}=e^{j2\pi.\beta.\frac{M}{2}=(-1)\beta M} \tag{5}$$

$$FFT\{\delta[l-m_e]\}=e^{-j2\pi.\beta.m_e} \tag{6}$$

The "cyclic-cumulant" of received OQAM/OFDM signals is specified by Vucic *et al.* (2017) in which $B(\tau)=\sum_{o=0}^{M-1}e^{j\frac{2\pi}{M}o\tau}|w_k|^2.$, $H(\beta,\tau)=\sum_{l=-\infty}^{\infty}h[l]h^*[l-\tau]e^{-j2\pi\beta l}$ denotes the cyclic-cumulant of prototype filter function, as shown by equation (7):

$$B(\beta,\tau)=e^{j\frac{2\pi}{N}.f_e\tau}e^{-j2\pi.\beta.m_e}.2\pi_b^2.\pi_\mu^2.D(\tau).H(\beta,\tau)+\pi^2.\delta(\beta) \tag{7}$$

where β indicates cycle frequency, $\beta=\frac{even}{M}$, and the term "even" indicates an even integer that is minimum than M.

The effect of inactive noise present in estimator is reduced by assigning $\beta\neq 0$. Thus, CFO and STO could be jointly evaluated by cyclic-cumulant with varied τ and β. Moreover, subcarrier weighting could increase the range of carrier frequency attainment (Bolcskei, 2001). If every subcarrier weighting parameter has a similar value, Equations (8) and (9) can be simply confirmed:

$$B(\tau)=\sum_{o=0}^{M-1}e^{j\frac{2\pi}{M}o\tau}|w_k|^2.=|w_k|^2\sum_{o=0}^{M-1}e^{j\frac{2\pi}{M}o\tau}=0 \tag{8}$$

$$B(\beta,\tau)=\pi^2.\delta(\beta) \tag{9}$$

Under such circumstances, there is no data of f_e and m_e in $B(\beta,\tau)$. Thus, the subcarrier weighting constraints could not be fixed to a similar value.

As $h[l]$ and $H(\beta,\tau)$ are recognized in the receiver, the impact of f_e, $\pi_b^2.\pi_\mu^2.$ and $B(\tau)$ are eradicated by evaluating the proportion of "cyclic-cumulant" with various cycle frequencies as shown by Equation (10), in which q denotes an integer in $\left[0,\frac{M}{2}-1\right]$, $\beta\neq 0$:

$$\frac{B\left(\beta+\frac{2q}{M},\tau\right)}{B(\beta,\tau)}=\frac{e^{j\frac{2\pi}{M}.f_e\tau}e^{-j2\pi\left(\beta.+\frac{2q}{M}\right).m_e.}.2\pi_b^2.\pi_\mu^2.B(\tau)H\left(\beta+\frac{2q}{N},\tau\right)}{e^{j\frac{2\pi}{M}.f_e\tau}e^{-j2\pi\beta.m_e.}2\pi_b^2.\pi_\mu^2.B(\tau)H(\beta,\tau)},$$
$$e^{-j\frac{4\pi}{M}.m_eq}\frac{H\left(\beta+\frac{2q}{N},\tau\right)}{H(\beta,\tau)} \tag{10}$$

As per (10), there is no data regarding CFO in $\frac{B\left(\beta+\frac{2q}{M},\tau\right)}{B(\beta,\tau)}$. The approximated value m_e remains only at the point of $e^{-j\frac{4\pi}{M}.m_e q}$. Considering $P_{me}(\beta,\tau,q)=e^{-j\frac{4\pi}{M}.m_e q}$ and $H_{me}(\beta,\tau,q)=\frac{H\left(\beta+\frac{2q}{M},\tau\right)}{H(\beta,\tau)}$. STO can be estimated by equation (11), in which $m_e \in Z$:

$$m_e = -\frac{M}{4\pi q}\arg\left\{\frac{P_{me}(\beta,\tau,q)}{H_{me}(\beta,\tau,q)}\right\} \tag{11}$$

To traverse the entire even cyclic frequencies and to enhance the accurateness of STO estimation, the value of q is set to one, and accordingly, equation (12) is attained:

$$m_e = -\frac{M}{4\pi}\frac{2}{M-2}\sum_{\beta=\frac{2}{M}}^{1-\frac{2}{M}}\arg\left\{\frac{P_{me}(\beta,\tau)}{H_{me}(\beta,\tau)}\right\} \tag{12}$$

As $H(\beta,\tau)$ and w_O are identified in the receiver, the impact of m_e, $\pi_b^2.\pi_\mu^2$ could be discarded by evaluating the proportion of cyclic-cumulant with varying delays in time as given by equation (13), in which ρ refers to time lag:

$$\begin{aligned}\frac{B(\beta,\tau+\rho)}{B(\beta,\tau)} &= \frac{e^{j\frac{2\pi}{M}.f_e(\tau+\rho)}\,e^{-j2\pi\beta.m_e}.2\pi_b^2.\pi_\mu^2.B(\tau+\rho)H(\beta,\tau+\rho)}{e^{j\frac{2\pi}{M}.f_e\tau}e^{-j2\pi\beta.m_e}.2\pi_b^2.\pi_\mu^2.B(\tau)H(\beta,\tau)} \\ &= e^{j\frac{2\pi}{M}.f_e(\tau+\rho)}\,\frac{.B(\tau+\rho)H(\beta,\tau+\rho)}{B(\tau)H(\beta,\tau)}\end{aligned} \tag{13}$$

According to (13), there is no data regarding STO in $\frac{B\left(\beta+\frac{2q}{M},\tau\right)}{B(\beta,\tau)}$. The approximated value f_e remains only in the item of $e^{j\frac{2\pi}{M}.f_e p}$. On considering $P_{fe}(\beta,\tau,\rho)=\frac{B(\beta,\tau+\rho)}{B(\beta,\tau)}$ and $H_{fe}(\beta,\tau,\rho)=\frac{B(\tau+\rho)H(\beta,\tau+\rho)}{B(\tau)H(\beta,\tau)}$, CFO can be estimated by Equation (14), in which, $f_e \in [-1/2, 1/2]$:

$$f_e = -\frac{M}{2\pi\rho}\arg\left\{\frac{R_{fe}(\beta,\tau,\rho)}{G_{fe}(\beta,\tau,\rho)}\right\} \tag{14}$$

4.2 Objective function

The minimum interference could be attained by maximizing the SNR value as per equation (15) and $SNR=\sigma_s^2/\sigma_n^2$, $\sigma_s^2\underline{\Delta}Z\left\{|s(k)|^2\right\}$ and $\sigma_n^2\underline{\Delta}Z\left\{|n(k)|^2\right\}$, where the difference between SNR and SNR(f_e) is a measure of the sensitivity to a frequency offset f_e. Here, $s(k)$ indicates a complex Gaussian process and $n(k)$ denotes complex AWGN and ε indicates the variation in transmitter and receiver oscillators (van de Beek *et al.*, 1997). However, the process is complicated; hence, to simplify the process, the received signal is split into real and imaginary parts, as shown in Equations (16) and (17). In Equation (16), Rx_{re} indicates the real part of the received signal, (Z_{Im}, Z_{re}) denotes the channel coefficients of the AWGN channel and z indicates the frame length and *ulayer* indicates the total number of layers. In Equation (17), Rx_{Im} denotes the imaginary part of the received signal:

$$SNR(f_e) \geq \frac{SNR}{1+0.5947SNR\sin^2\pi f_e}\left(\frac{\sin\pi f_e}{\pi f_e}\right)^2 \tag{15}$$

$$Rx_{re}(z) = \begin{pmatrix} Rx_{re1}(z) + (Z_{re} \times T \times \mathrm{Re}(ulayer, z)) - \\ (Z_{\mathrm{Im}} \times T \times \mathrm{Im}(ulayer, z)) \end{pmatrix} \times f_e \tag{16}$$

$$Rx_{\mathrm{Im}}(z) = \begin{pmatrix} Rx_{\mathrm{Im}1}(z) + (Z_{re} \times T \times \mathrm{Im}(ulayer, z)) - \\ (Z_{\mathrm{Im}} \times T \times \mathrm{Re}(ulayer, z)) \end{pmatrix} \times f_e \tag{17}$$

Thus, the proposed objective function intends to minimize the error by evaluating the difference between the received signal and transmitted signal for both the real and imaginary part, as shown in Equation (18):

$$Min\ \ e = \frac{(Rx_{re} - Tx_{re}) + (Rx_{\mathrm{Im}} - Tx_{\mathrm{Im}})}{2} \tag{18}$$

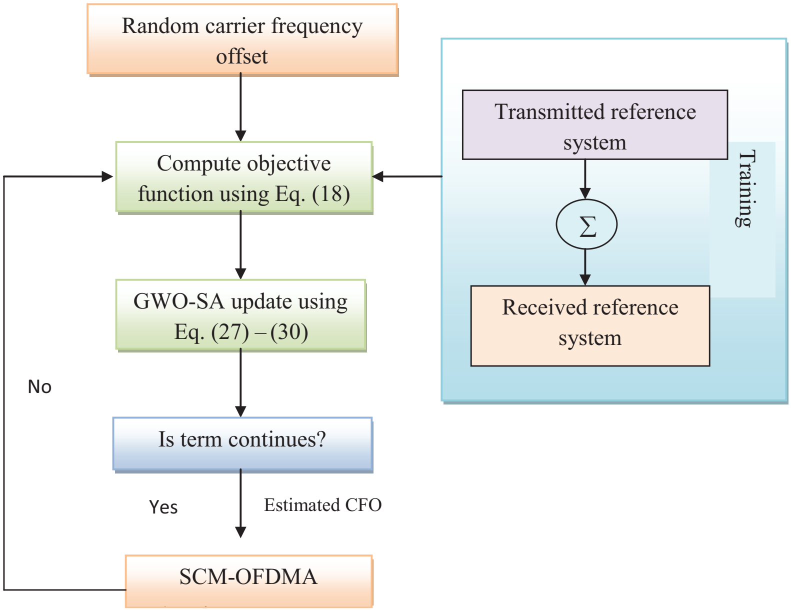
The proposed objective function is the error measurement between the real and imaginary parts of the received and the transmitted signals. According to Equations (16) and (17), the real and imaginary parts of the received signal is a function of the real and imaginary parts of the transmitted signal and the frequency offset function, respectively. Because the proposed methodology intends to estimate the frequency offset, the error between the transmitted signal and the received signal is defined as the objective function. However, the transmitted signal is unknown; therefore, the semi-blind approach is exploited here. As per the semi-blind concept, the training symbols are declared under which the transmitted signals of the training sequence are known. Based on the training signals, the error metric is minimized and the resultating the frequency offset parameter is declared as the optimal value in the system for further data transmission process.

4.3 Solution encoding

In this proposed model, CFO is estimated with the aid of a modified optimization algorithm called GWO-SA. Hence, parameter CFO, which is given as input for solution encoding, has to be tuned optimally. The tuning or optimization of is done in such a way that the difference between the transmitted signal and the received signal is minimum, where, i = 1, 2. . .nq. The diagrammatic representation of the solution encoding process is given by Figure 2, where nq indicates the number of CFO.

4.4 Grey wolf optimization

The mechanism of GWO (Sharma *et al.*, 2016) algorithm portrays the grey wolves' hunting characteristics and its headship hierarchy. There exist four categorizations of grey wolves such as α, η, ζ, and ω, which are exploited for carrying out the leadership hierarchy. Penetrating, circling and attacking the food are the three foremost courses of actions in hunting that are used to develop optimization.



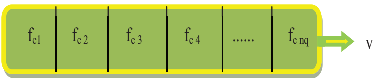


Figure 2.
Solution encoding

The wolves such as, α, η and ζ are considered as the foremost wolves, which are concerned in the process of hunting. Among these wolves, α is allocated as the leader, which makes a declaration relating to the hunting process, sleeping place and time to awake. However, η and ζ hold a second and third level that assists α in captivating choices. Furthermore, the final level of the wolves is regarded as ω that is allocated only for eating. The encircling characteristics of the wolves are designed using Equations (20) and (21), in which X and Y represents the coefficient vectors, v_p refers to the position vectors of prey, t denotes the current iteration and v indicates the grey wolves' position vectors. The model for X and Y is specified by equations (22) and (23), respectively, in which a is a parameter that is steadily minimized from 2 to 0 for the whole iterations. The standard formulation of a is given by Equation (19) and r_1 r_2 indicates the arbitrary vectors that are distributed constantly among [0, 1]:

$$a_1 = 2 - 1 * \left(\frac{2}{t_{\max}} \right) \tag{19}$$

$$Z = |Y.v_p(t) - v(t)| \tag{20}$$

$$v_p(t+1) = v_p(t) - X.Z \tag{21}$$

$$X = 2a.r_1 - a \tag{22}$$

$$Y = 2r_2 \tag{23}$$

The arithmetical formula for portraying the hunting nature of the wolves is specified in Equations (24)-(29), in which the final modified position of wolves is offered in Equation (30), which offers the updated a and v:

$$Z_\alpha = |Y_1 - v_\alpha - v| \tag{24}$$

$$Z_\eta = |Y_2 - v_\eta - v| \tag{25}$$

$$Z_\zeta = |Y_3 - v_\zeta - v| \tag{26}$$

$$v_1 = v_\alpha - X_1.(Z_\alpha) \tag{27}$$

$$v_2 = v_\eta - X_2.(Z_\eta) \tag{28}$$

$$v_3 = v_\zeta - X_3.(Z_\zeta) \tag{29}$$

$$v(t+1) = \frac{v_1 + v_2 + v_3}{3} \tag{30}$$

4.5 Proposed grey wolf optimization with self-adaptiveness algorithm

Despite the amazing facts of the conventional GWO algorithm, it includes certain shortcomings such as low precision and low rate of convergence. In conventional GWO, the value of a is determined based on Equation (19). As a modification, the value of a in GWO-SA is portrayed as per equation (27), where τ refers to the change of fitness as given by equation (28), under the condition, if $t \neq 1$, else, the value of τ. In Equation (28), $f(t-1)$ denotes the previous iteration and f(t) indicates the present iteration. Because the position of the wolf is adaptively changed based on a fitness function, it is known for having a self-adaptive principle:

$$a = \tau * \left(2 - 1 * \left(\frac{2}{t_{\max}}\right)\right) \tag{27}$$

$$\tau = \frac{\left(f(t-1) - f(t)\right)}{f(t-1)} \tag{28}$$

The pseudo code of the proposed GWO-SA model is demonstrated in Algorithm 1:

```
Algorithm 1: Proposed GWO-SA Algorithm
  Initialize the population of grey wolves
  Initialize a, X and Y
```

```
Compute the fitness of all search agents
Allocate v_α as the best search agent
Allocate v_η as the second best search agent
Allocate v_ζ as the third best search agent
While (t < t_max), t_max indicates the maximum iteration
  For each wolf
  If t ≠ 1
    Determine τ as per equation (28)
    else
    τ = 2
    End if
    Determine a as per equation (27)
    Update the position as per equation (30)
    End for
    Update a, X and Y
    Evaluate the fitness for whole search agents
    Update v_α, v_η and v_ζ
    t = t + 1
End while
Return v_α
```

5. Results and discussions

5.1 Simulation procedure

The proposed GWO-SA-based CFO estimation in SCM-OFDM model was simulated in MATLAB, and the corresponding outcomes were attained. Because the literature lags with the CFO estimate using meta-heuristic search algorithms, which is the primary contribution here, we have compared our proposed algorithm with state-of-the-art meta-heuristic algorithms and standard filtering algorithms such as ML (Bai and Zhang, 2018), LMS (Lorito, 1995), PSO (Zhang and Xia, 2017), GA (McCall, 2005) and GWO (Mirjalili *et al.*, 2014) for CFO estimation; it is held with respect to BER and MSE for ten values of SNR. In addition, results were taken for varying values of weight, W adopted in GWO-SA algorithm. In addition, the analysis was held with respect to BER and MSE for varying values of CFO at 0.25 and 0.5, respectively, and accordingly the performance was analyzed.

5.2 Bit error rate analysis

The BER analysis of the adopted GWO-SA scheme for enhancing CFO estimation in SCM-OFDM systems is shown in Figure 3. From Figure 3(a), the BER analysis of presented and conventional schemes for CFO at 0.25 can be determined, where the adopted scheme at 10th SNR is 99.6 per cent $((10^{-15}\text{-}10^{-38})/10^{-38})$ better than ML, 99.6 per cent better than LMS, 99.3 per cent better than PSO, 75 per cent better than GA and 25 per cent better than GWO algorithms. Similarly, the BER analysis of the implemented and conventional techniques for CFO at 0.5 can be determined from Figure 3(b), in which the suggested GWO-SA method at 10th SNR is 60 per cent superior to ML, 20 per cent superior to LMS, 90 per cent superior to PSO, 97.5 per cent superior to GA and 99.4 per cent superior to GWO algorithms. In addition, the random variables r_1 and r_2 indicated in Equations (18) and (19) are multiplied with varying weight parameter W and the results are obtained. Figure 3(c) demonstrates the BER analysis of the proposed model with respect to the varying weights for

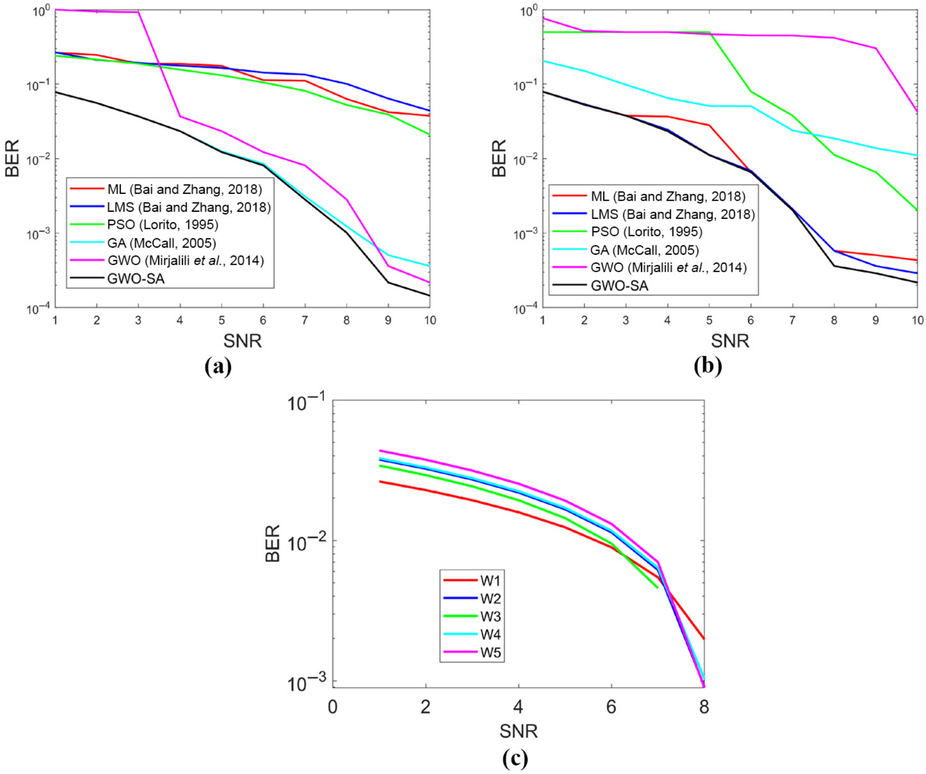


Figure 3. Error analysis of the proposed and conventional models in terms of varying (a) CFO = 0.25 (b) CFO = 0.5 and (c) proposed method in terms of varying weights

$W1 = 0.2$, $W2 = 0.4$, $W3 = 0.6$, $W4 = 0.8$ and $W5 = 1$. From the analysis, $W3$ is found to offer reduced error compared with other weighting factors. Thus, the enhancement of the proposed GWO-SA model in minimizing the BER has been substantiated in a better way.

5.3 Mean square error analysis

This analysis shows the difference among the transmitted and the received signal. The MSE analysis of the presented GWO-SA-based CFO estimation in SCM-OFDM systems is described in this section. Table II describes the MSE analysis of the proposed and conventional schemes at CFO = 0.25. From the analysis, at 1st SNR, the proposed GWO-SA scheme is 4.62 per cent better than LMS, 60.1 per cent better than PSO, 37.82 better than GA and 67.85 per cent better than GWO algorithms. On considering the 2nd SNR, the presented GWO-SA method is 0.6 per cent superior to LMS, 67.36 per cent superior to PSO, 40.59 better than GA, and 67.93 per cent superior to GWO algorithms. Likewise, the suggested scheme at 3rd SNR is 72.47 per cent better than PSO, 37.95 per cent better than GA and 72.47 per cent better than GWO algorithms. Accordingly, Table III describes the MSE analysis of the proposed and conventional schemes when CFO = 0.5. From the analysis, the presented GWO-SA method at 4th SNR is 20.26 per cent superior to ML, 2.24 per cent superior to LMS, 78.34 per cent superior to PSO, 40.02 per cent superior to GA and 78.34 per cent superior to GWO algorithms. In the same way, at 5th SNR, the implemented scheme is 36.99 per cent better than ML, 0.32 per cent better than LMS, 85.04 per cent better than PSO, 53.23 per cent superior to GA and 84.57 per cent better than GWO algorithms. In addition, at 7th SNR, the adopted GWO-SA scheme is 1.74 per cent superior to ML, 1.74 per cent superior to LMS, 76.84 per cent superior to PSO, 70.87 per cent superior to GA and 93.29 per cent superior to GWO algorithms.

Table IV demonstrates the MSE analysis of the proposed GWO-SA model for varying weights. From the analysis, for the third iteration, the W3 is 6.65 per cent better than W1, 22.13 per cent superior to W2, 30.06 per cent superior to W4 and 52.22 per cent superior to W5. Moreover, for the fifth iteration, the W3 is 28.74 per cent superior to W1, 23.49 per cent superior to W2, 20.42 per cent superior to W4 and 18.25 per cent superior to W5. Therefore, the improved MSE outcomes of the proposed GWO-SA scheme have been validated in the proposed CFO estimation in SCM-OFDM system from the simulated outcomes.

5.4 Cost function analysis

The analysis on fitness function or cost function from Eq. (18) for both CFO = 0.25, and CFO = 0.5 is shown in Tables V and VI, respectively. Here, the cost function of the proposed GWO-SA model is 45.57 per cent better than PSO, 78.45 per cent better than GA, and 44.85 per cent better

Table II. MSE Analysis of the proposed and coventional models at CFO = 0.25

SNR	ML (Bai and Zhang, 2018)	LMS (Lorito, 1995)	PSO (Zhang and Xia, 2017)	GA (McCall, 2005)	GWO (Mirjalili *et al.*, 2014)	GWO-SA
1	0.51504	0.51686	0.48999	0.27951	0.99949	0.27951
2	0.49615	0.45928	0.46234	0.23616	0.97171	0.23616
3	0.4341	0.43874	0.43418	0.19207	0.96014	0.19207
4	0.43326	0.42047	0.39519	0.15286	0.19207	0.15286
5	0.41926	0.40621	0.36264	0.11187	0.15286	0.11057
6	0.33636	0.37798	0.32454	0.09239	0.11057	0.09001
7	0.33344	0.36632	0.28489	0.055772	0.09001	0.053115
8	0.2513	0.31744	0.22917	0.035068	0.053115	0.031823
9	0.20518	0.25288	0.19764	0.022503	0.019018	0.014731
10	0.19376	0.20972	0.14484	0.019018	0.014731	0.012028

Table III. MSE Analysis of the proposed and coventional models at CFO = 0.5

SNR	ML (Bai and Zhang, 2018)	LMS (Lorito, 1995)	PSO (Zhang and Xia, 2017)	GA (McCall, 2005)	GWO (Mirjalili *et al.*, 2014)	GWO-SA
1	0.28119	0.28157	0.70588	0.45302	0.8762	0.2817
2	0.22995	0.23183	0.70588	0.3879	0.71842	0.23043
3	0.19395	0.19413	0.70588	0.31319	0.70588	0.19432
4	0.1917	0.15637	0.70588	0.25487	0.70588	0.15286
5	0.16753	0.10589	0.70588	0.22567	0.68386	0.10555
6	0.081134	0.082898	0.2817	0.22503	0.67105	0.081134
7	0.045802	0.045802	0.19432	0.1545	0.67024	0.045005
8	0.024056	0.024056	0.10589	0.13661	0.64651	0.019018
9	0.022503	0.019018	0.081134	0.11754	0.54995	0.01701
10	0.020833	0.01701	0.045005	0.1052	0.20571	0.014731

than GWO at SNR = 10 dB. Moreover, at SNR = 8 dB, and CFO = 0.5, the cost function of the proposed models is 46.18 per cent, 22.56 per cent and 82.01 per cent better than PSO, GA and GWO, respectively, which proves the efficiency of the developed method.

5.5 Performance analysis using benchmark functions

To test the performance of the proposed model, bench mark functions such as Ackley, Beale, leon, and rastrigin are evaluated, with respect to the cost function. Figure 4 shows that the proposed model shows minimized in cost function by varying the benchmark mark functions and iterations.

6. Conclusion

This paper has presented a novel SCM-OFDM system with optimal CFO by exploiting an improved optimization scheme. The effectiveness of SCM-OFDM was substantiated by relating the transmitted and received signal. Consequently, the most important objective here was to minimize the error among the received and transmitted signal. At the receiver side, CFO was integrated and has to be attuned suitably to acquire the received signal closer to the

SNR	W1	W2	W3	W4	W5
1	0.70588	0.70588	0.70588	0.70588	0.70588
2	0.36124	0.70588	0.66291	0.70588	0.70588
3	0.32598	0.39078	0.30429	0.4351	0.63692
4	0.29926	0.27573	0.26021	0.2808	0.27951
5	0.27163	0.25302	0.19357	0.24325	0.23677
6	0.25769	0.20885	0.18002	0.19264	0.19395
7	0.20764	0.1991	0.12237	0.17554	0.14902
8	0.18634	0.12871	0.090411	0.16448	0.12178
9	0.12955	0.097717	0.059536	0.13845	0.069096
10	0.078874	0.073657	0.020833	0.024056	0.014731

Table IV. MSE Analysis of the proposed model with varying weights

Methods	SNR 1	2	3	4	5	6	7	8	9	10
PSO (Zhang and Xia, 2017)	1.407	0.109	1.33	0.148	0.38778	1.38	0.304	0.278	0.404	0.542
GA (McCall, 2005)	0.131	0.147	0.327	1.35	1.314	0.318	1.279	1.406	0.317	1.370
GWO (Mirjalili *et al.*, 2014)	1.387	1.088	1.016	0.980	0.612	1.366	0.253	1.324	1.398	0.535
GWO-SA	0.110	0.064	0.194	0.057	0.105	0.294	0.123	0.088	0.132	0.295

Table V. Const function analysis of proposed and conventional models at CFO = 0.25

Methods	SNR 1	2	3	4	5	6	7	8	9	10
PSO (Zhang and Xia, 2017)	0.321	0.458	0.265	0.133	0.275	0.755	0.354	0.472	0.269	0.461
GA (McCall, 2005)	0.967	0.143	0.280	0.195	0.183	0.446	1.273	0.328	0.642	0.407
GWO (Mirjalili *et al.*, 2014)	0.591	0.250	0.459	0.986	0.919	0.483	0.268	1.412	0.316	1.285
GWO-SA	0.251	0.053	0.103	0.125	0.052	0.415	0.156	0.254	0.072	0.123

Table VI. Const function analysis of proposed and conventional models at CFO = 0.5

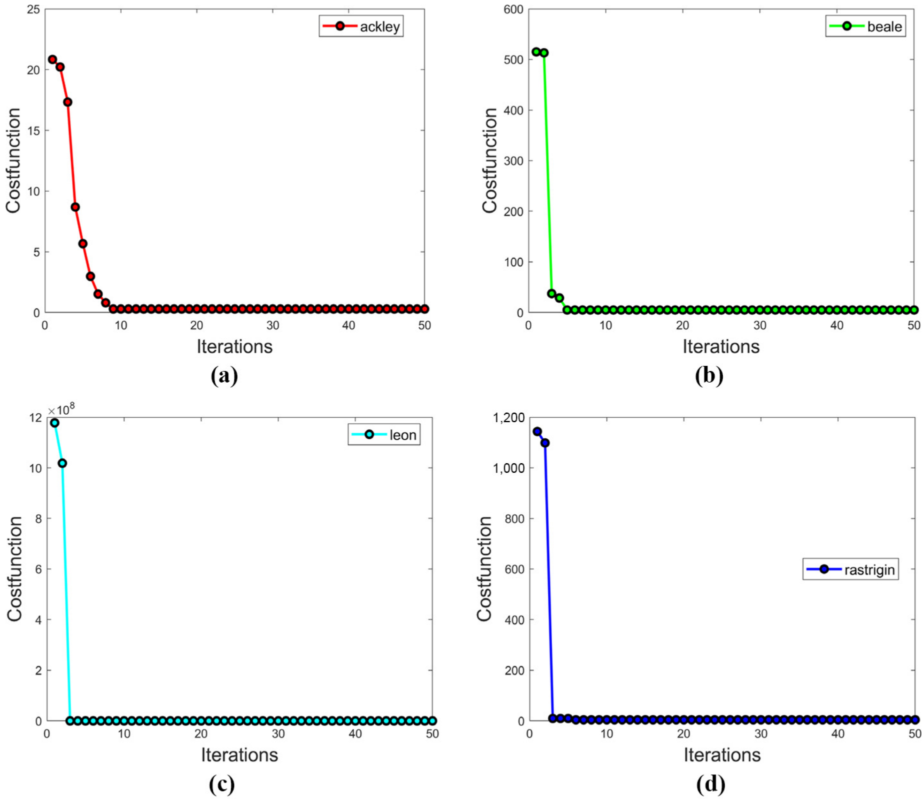


Figure 4.
Performance analysis of the proposed models in terms of varying bench mark functions (a) Ackley (b) Beale and (c) leon and (d) rastrigin functions in terms of varying the iterations

transmitted signal. The optimization of CFO was performed by deploying GWO-SA. Furthermore, the presented model was distinguished with traditional models in terms of BER and MSE, and the system performance has been accordingly confirmed. From the analysis, BER of the proposed and conventional schemes for CFO at 0.25 was determined, where the adopted scheme at 10th SNR was 99.6 per cent better than ML, 99.6 per cent better than LMS, 99.3 per cent better than PSO, 75 per cent better than GA and 25 per cent better than GWO algorithms. Moreover, MSE at 1st SNR, the proposed GWO-SA scheme, is 4.62 per cent better than LMS, 60.1 per cent better than PSO, 37.82 better than GA, and 67.85 per cent better than GWO algorithms. Hence, it is confirmed that the performance of SCM-OFDM system with GWO-SA-based CFO estimation outperformed the state-of-the-art techniques.

References

Abdzadeh-Ziabari, H., Zhu, W.P. and Swamy, M.N.S. (2018), "Joint carrier frequency offset and doubly selective channel estimation for MIMO-OFDMA uplink with kalman and particle filtering", *IEEE Transactions on Signal Processing*, Vol. 66 No. 15, pp. 4001-4012.

Ali, F.M., Ali, E.A.A.A. and Tarbul, M.G. (2018), "Performance analysis of radio over optical fiber system with OFDM using multiplexing techniques", *International Journal of Applied Engineering Research*, Vol. 13 No. 12, pp. 10831-10844.

Bai, C. and Zhang, Z. (2018), "A least mean square based active disturbance rejection control for an inertially stabilized platform", *Optik*, Vol. 174, pp. 609-622.

Bergmans, P.P. and Cover, T.M. (1974), "Cooperative broadcasting", *IEEE Transactions on Information Theory*, Vol. 20 No. 3, pp. 317-324.

Bolcskei, H. (2001), "Blind estimation of symbol timing and carrier frequency offset in wireless OFDM systems", *IEEE Transactions on Communications*, Vol. 49 No. 6, pp. 988-999.

Cai, J. and Shen, X. (2017), "Modified optical asymmetric image cryptosystem based on coherent superposition and equal modulus decomposition", *Optics and Laser Technology*, Vol. 95, pp. 105-112.

Chau, P., Shin, J. and Jeong, J.(P.). (2018), "Scalable video multicast using inter-layered superposition and network-coded cooperation over MIMO relay systems", *Computer Communications*, Vol. 120, pp. 22-31.

Chopra, R., Ghosh, D. and Mehra, D.K. (2017), "Spectrum sensing for OFDM signals using pilot induced cyclostationarity in the presence of cyclic frequency offset", *Physical Communication*, Vol. 24, pp. 182-194.

Damrath, M., Hoeher, P.A. and Forkel, G.J.M. (2017), "Symbol detection based on voronoi surfaces with emphasis on superposition modulation", *Digital Communications and Networks*, Vol. 3 No. 3, pp. 141-149.

Daoqi, H., Jianhua, Z., Hao, Y. and Mengmeng, L. (2016), "Performance analysis of frequency synchronization in MIMO-OFDM systems with ST/MRC scheme", *The Journal of China Universities of Posts and Telecommunications*, Vol. 23 No. 5, pp. 10-39.

Fusco, T., Marrone, F. and Tanda, M. (2008), "Semiblind techniques for carrier frequency offset estimation in OFDM systems", *Signal Processing*, Vol. 88 No. 3, pp. 704-718.

Fusco, T., Petrella, A. and Tanda, M. (2008), "Non-data-aided carrier-frequency offset estimation for pulse-shaping OFDM/OQAM systems", *Signal Processing*, Vol. 88 No. 8, pp. 1958-1970.

Guan, Q-y., Zhao, H-L. and Guo, Q. (2010), "Parallel interference cancellation for carrier frequency offset in OFDM assisted with recursive least squares algorithm", *The Journal of China Universities of Posts and Telecommunications*, Vol. 17 No. 3, pp. 13-19.

Gul, M.M.U., Lee, S. and Ma, X. (2014), "Carrier frequency offset estimation for OFDMA uplink using null Sub-carriers", *Digital Signal Processing*, Vol. 29, pp. 127-137.

Huang, M., Huang, L., Guo, C., Zhang, P., Zhang, J. and Yang, L.L. (2017), "Carrier frequency offset estimation in uplink OFDMA systems: an approach relying on sparse recovery", *IEEE Transactions on Vehicular Technology*, Vol. 66 No. 10, pp. 9592-9597.

Jose, R. (2017), "Estimation of channel and carrier frequency offset in OFDM systems using joint statistical framework", *Physical Communication*, Vol. 25, pp. 139-147.

Kang, C., Xingguang, L., Shuang, Z. and Lei, C. (2018), "Carrier frequency offset estimation for BPSK signals in coherent optical communications", *Procedia Computer Science*, Vol. 131, pp. 1128-1133.

Kaur, J. and Sharma, V. (2019), "A-STBC incorporated power-efficient radio over fibre system", *Optics Communications*, Vol. 441, pp. 84-89.

Krikidis, I. and Thompson, J.S. (2012), "MIMO two-way relay channel with superposition coding and imperfect channel estimation", *Journal of Network and Computer Applications*, Vol. 35 No. 1, pp. 510-516.

Li, J., Liu, P-Z. and Sun, H. (2008), "Joint channel and carrier frequency offset estimation for multi-user MIMO-OFDM uplink transmissions", *The Journal of China Universities of Posts and Telecommunications*, Vol. 15 No. 1, pp. 75-79.

Li, P., Zhang, X., Su, Y. and Wu, L. (2007), "A novel minimum output variance estimator for carrier frequency offset in OFDM systems", *Tsinghua Science and Technology*, Vol. 12 No. 6, pp. 674-677.

Lin, K.-H., Chiu, W.-H. and Tseng, J.-D. (2012), "Low-complexity architecture of carrier frequency offset estimation and compensation for body area network systems", *Computers and Mathematics with Applications*, Vol. 64 No. 5, pp. 1400-1408.

Liu, M., Wang, J. and Li, B. (2018), "Non-data aided joint estimation of symbol timing offset and carrier frequency offset for OFDM/OQAM systems", *AEU – International Journal of Electronics and Communications*, Vol. 87, pp. 164-172.

Lorito, F. (1995), "Dynamical properties of the recursive maximum likelyhood algorithm for frequency estimation", *IFAC Proceedings Volumes*, Vol. 28 No. 13, pp. 143-148.

McCall, J. (2005), "Genetic algorithms for modelling and optimisation", *Journal of Computational and Applied Mathematics*, Vol. 184 No. 1, pp. 205-222.

Mirjalili, S., Mirjalili, S.M. and Lewis, A. (2014), "Grey wolf optimizer", *Advances in Engineering Software*, Vol. 69, pp. 46-61.

Mohan, Y., Chee, S.S., Xin, D.K.P. and Foong, L.P. (2016), "Artificial neural network for classification of depressive and normal in EEG", *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), IEEE, Kuala Lumpur*.

Saeedi-Sourck, H. (2017), "Multiple carrier frequency offsets and multipath channels estimation for coordinated multi-point FBMC systems", *Physical Communication*, Vol. 25, pp. 511-518.

Saeedi-Sourck, H., Wu, Y., Bergmans, J.W.M., Sadri, S. and Farhang-Boroujeny, B. (2011), "Sensitivity analysis of offset QAM multicarrier systems to residual carrier frequency and timing offsets", *Signal Processing*, Vol. 91 No. 7, pp. 1604-1612.

Seshadri, N. and Sundberg, C.-E.W. (1991), "Multi-level block coded modulations for the rayleigh fading channel", *IEEE Global Telecommunications Conference GLOBECOM' 91: Countdown to the New Millennium. Conference Record, IEEE, Phoenix, AZ*, pp. 47-51.

Shah, H.K., Dasgupta, K.S. and Soni, H. (2013), "Low complexity scheme for carrier frequency offset estimation in orthogonal frequency division multiple access uplink", *IET Communications*, Vol. 7 No. 13, pp. 1405-1411.

Shankar, A. and Jaisankar, N. (2016), "A novel energy efficient clustering mechanism in wireless sensor network", *Procedia Computer Science*, Vol. 89, pp. 134-141.

Shareef, S.M. and Rao, R.S. (2018), "Optimal reactive power dispatch under unbalanced conditions using hybrid Swarm intelligence", *Computers and Electrical Engineering*, Vol. 69, pp. 183-193.

Sharma, S., Bhattacharjee, S. and Bhattacharya, A. (2016), "Grey wolf optimisation for optimal sizing of battery energy storage device to minimise operation cost of microgrid", *IET Generation, Transmission and Distribution*, Vol. 10 No. 3, pp. 625-637.

Singh, G., Jain, V.K. and Singh, A. (2018), "Adaptive network architecture and firefly algorithm for biogas heating model aided by photovoltaic thermal greenhouse system", *Journal of Energy Environment*, Vol. 29 No. 7, pp. 1-25.

Su, H., Zhang, C. and Lu, J. (2009), "Interference-cancellation-based carrier frequency offset estimation for OFDMA uplink transmissions", *Tsinghua Science and Technology*, Vol. 14 No. 4, pp. 487-491.

Tong, J. and Ping, L. (2010), "Performance analysis of superposition coded modulation", *Physical Communication*, Vol. 3 No. 3, pp. 147-155.

van de Beek, J.J., Sandell, M. and Borjesson, P.O. (1997), "ML estimation of time and frequency offset in OFDM systems", *IEEE Transactions on Signal Processing*, Vol. 45 No. 7, pp. 1800-1805.

Vucic, D., Vukotic, S. and Erić, M. (2017), "Cyclic spectral analysis of OFDM/OQAM signals", *AEU-International Journal of Electronics and Communications*, Vol. 73 No. 3, pp. 139-143.

Xu, W., Wang, Y. and Hu, X. (2015b), "Blind joint estimation of carrier frequency offset and I/Q imbalance in OFDM systems", *Signal Processing*, Vol. 108, pp. 46-55.

Xu, L., Wang, J., Li, Y-P., Li, Q. and Zhang, X. (2015a), "Resource allocation algorithm based on hybrid particle swarm optimization for multiuser cognitive OFDM network", *Expert Systems with Applications*, Vol. 42 No. 20, pp. 7186-7719, doi: 10.1016/j.eswa.2015.05.012.

Zhang, J. and Xia, P. (2017), "An improved PSO algorithm for parameter identification of nonlinear dynamic hysteretic models", *Journal of Sound and Vibration*, Vol. 389, pp. 153-167.

Zhang, W., Yin, Q. and Gao, F. (2016), "Computationally efficient blind estimation of carrier frequency offset for MIMO-OFDM systems", *IEEE Transactions on Wireless Communications*, Vol. 15 No. 11, pp. 7644-7656.

Zhao, Y., Xiao, Y., Dong, B., Yang, P., Xiao, L. and Xiang, W. (2017), "Efficient detection of spatial modulation OFDM systems with multiple carrier frequency offsets", *IEEE Communications Letters*, Vol. 21 No. 2, pp. 426-429.

Corresponding author
Rashmi N. can be contacted at nrashmi814@gmail.com

International Journal on Emerging Technologies 10(3): 260-267(2019)

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Power Efficiency Enhancement using Hybrid Techniques for OFDM

Rashmi N.[1] and Mrinal Sarvagya[2]
[1]Assistant Professor, Department of Electronics and Communication Engineering,
BMS Institute of Technology and Management, Karnataka (State), India.
[2]Professor, Department of Electronics and Communication Engineering,
Reva University, Karnataka, (State), India.

(Corresponding author: Rashmi N.)
(Received 27 June 2019, Revised 30 August 2019 Accepted 28 September 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: With the increased number of internet traffic users there is a huge demand for high data rate and power efficient wireless communication system. The physical layer of any wireless communication systems plays a vital role in building the foundation for data rate enhancement. Using multicarrier modulation techniques multiple users can be accommodate having varies data rate requirements. The Orthogonal frequency division multiplexing system has been proved as best multicarrier modulation technique to fulfill the user requirement. Although the OFDM system is set to be the best example of multi carrier system it under goes sever problem of PAPR. As a consequence it induces nonlinear distortions in the system thus limiting the power amplifiers to work in linear regions only. In this article we propose a unique solution to the above mentioned problem that combats the nonlinearity in the communication systems. The proposed solution is a combination of OFDM and Superposition coded modulation. The comparative analysis is presented with the existing techniques used for PAPR reduction.

Keywords: Clipping technique, Selective mapping, Partial Transmit sequence, Discrete Hartley Transform (DHT), Superposition Coded Modulation (SCM)-OFDM, Modified µ law Companding (MMC), Peak to Average Power Ratio (PAPR), Bit Error Rate (BER) , Complementary Cumulative Distribution Function(CCDF).
Abbreviations: OFDM, Orthogonal Frequency Division Multiplexing; SCM, Superposition Coded Modulation; PTS, Partial Transmit Sequence; TR, Tone Reservation; **DHT, D**iscrete Hartley Transform.

I. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) is an assured technique for systems that require a high data transmission rate, such as 4G LTE-Advanced and Wimax. Several researchers and academicians have studied the OFDM technique, which can cater to high data rate applications. OFDM offers several advantages over channels that experience frequency selectivity and time variations. Furthermore, the OFDM technique allows each subcarrier to independently select the constellation and coding scheme. OFDM offers robustness to the multipath fading channel and has a low- implementation complexity. Although OFDM exhibits several advantages, OFDM suffers from various technical challenges, such as high Peak to Average Power Ration (PAPR) and Carrier Frequency Offset (CFO).The complexity of Digital-to-Analog Converter (DAC) increases with PAPR, that leads to the degradation of power amplifier efficiency in OFDM system. The transmission of high-PAPR OFDM signals through nonlinear power amplifier causes spectral broadening, that will expand the dynamic range of the DAC. Superposition Coded Modulation (SCM) is emerging as non-bijective, bandwidth and power efficient coded modulation with Gaussian Quadrature components. The shaping gain in SCM is inherent without any additional complexity at the transmitter side. At the receiver side the complexity is low as O(K) for K-layer SCM when equal power is allocated to the symbols. Hybrid combination of multi layer SCM and OFDM can achieve greater data rates compared to single carrier OFDM system.

The main aim of this study is to decrease the PAPR in OFDM system to reduce the range of DAC and thereby decrease the cost of the system. We propose the use of a novel technique called hybrid precoding using DHT and µ companding for SCM- OFDM system. SCM-OFDM system increases the data rate of OFDM system complementing the advantages of OFDM. Implementing SCM coding scheme for MIMO system reaches Shannon channel capacity. But, using SCM in OFDM system further increases the PAPR problem in OFDM but increases the data rate and capacity of the OFDM system. Several techniques for PAPR reduction have been proposed in literature, and are categorized into three categories: multiplicative schemes, coding techniques and additive schemes. In multiplicative schemes, phase sequences are multiplied with OFDM sequences, the popular techniques using multiplicative schemes such as Selective Mapping (SM) and Partial Transmit Sequence (PTS) [1–5]. In [6], a low complexity PAPR reduction with modified linear SM scheme was suggested for the OFDM system. The performance for modified linear SM scheme for PAPR reduction in [6] was insignificant compared with conventional schemes. Additive schemes include Tone Reservation (TR) and clipping and filtering techniques. TR technique with low-complexity tones with null subcarriers was proposed in

[8] to decrease the PAPR in the filter bank multi carrier In [9], Decision directed method proposed for PAPR reduction for optical OFDM. In [10] a tone rejection technique for PAPR reduction was proposed. In [11], clipping and quantization noise cancellation is proposed for low-complexity PAPR reduction. Clipping PAPR reduction technique causes in- and out-band interferences that take out symmetry among subcarriers of OFDM that result in Inter Carrier Interference (ICI). A PAPR reduction system including coding technique anticipated in [12] implemented Reed Muller codes are implemented with PTS scheme. In [13] code scrambling method are used to reduce PAPR. All these schemes reduce PAPR at cost of computational complexity and with additional information need to be transmitted to receiver that increases bandwidth requirement. Motivated by the limitations of clipping and other techniques, we propose Modified µ law companding techniques with **D**iscrete Hartley Transform (DHT) as precoding technique for PAPR reduction in SCM-OFDM system. The proposed technique does not require any side information as compared to multiplicative methods, precoding techniques scrambles the signal phase to reduce the PAPR of the OFDM system. Precoding with DHT demonstrates higher PAPR reduction compare to Discrete Fourier Transform (DFT) technique.DHT distorts phase of the signal there by reducing PAPR. DHT is less complex compared to DFT as indicated in Table 2. DHT brings down the multiplication complexity by a factor of two contrasts with DFT, since DHT matrix has only real elements. DHT provides higher spectral efficiency, since it does not require side information compared to various different PAPR reduction systems like PTS and SLM techniques.

This paper is organized as follows. In Section II, the SCM-OFDM system is discussed, and the PAPR and cumulative complementary distribution function (CCDF) are formulated. In Section III, the clipping technique for PAPR reduction for SCM-OFDM system, PTS and SM techniques were discussed. Section IV Simulation results were discussed, section V conclusion

II. SCM-OFDM SYSTEM IS DISCUSSED ???????

A. Characterization of the SCM-OFDM System

OFDM system has high bandwidth efficiency compared to other multiple access system. However, there is continuous demand for high data rate and high capacity to meet the requirements of various evolving systems combined with OFDM system.

technologies and applications. Therefore there is strong motivation to come up with systems and methods that achieve even higher bandwidth efficiency compared to traditional OFDM system which inherits various advantages of traditional OFDM system. Superposition Coded Modulation is the multicarrier system, which has the advantage of adaptive modulation and coded techniques. Combining SCM with OFDM system further increases channel capacity and the data rate of the OFDM system, preserving the advantages of traditional OFDM system. However use of SCM system for OFDM further increases the PAPR problem in OFDM system. To reduce PAPR in SCM-OFDM system we propose DHT which provides spectral efficiency to SCM-OFDM system. In SCM –OFDM system, K-layer SCM is defined over a 2^K constellation, the binary data sequence **P** is partitioned into K sub-sequences {$\mathbf{P_k}$}. The k-th layer, resulting in a coded bit sequence c_k = {$c_k(j)$} of length 2J. Where $c_k(j) \in \{0,1\}$ and J is the frame length. Coded sequence c_k is then mapped to QPSK sequence $x_k(j)$.

$$x^K_k = x^K_{re,j} + i\, x^K_{im,j} \qquad (1)$$

The superscripts 'Re' and 'Im' denote the real and imaginary components, of complex numbers, respectively.

QPSK sequences are linearly superimposed to form X_j, which is given as follows:

$$x(j)=\sum_{k=0}^{K-1}\rho_k\, x_k(j), \qquad j= 0,1,\ldots J-1 \qquad (2)$$

Where $\{\rho_k\}$ are constant weighting actor k. In this study, $\{\rho_k\}$ was obtained using the simulation-based power allocation method projected in [12].

The superimposed symbols are fed to a customary OFDM modulator unit that consists of an N-point inverse discrete Fourier transform (IDFT) unit pursued by an N_g-point cyclic prefix (CP) converts linear convolution to circular convolution, applied as a guard band to avoid Inter Symbol Interference. The IDFT output is obtained as vector

$D_i = [D_i(0), D_i(1),\ldots D_i(N-1)]^T$. The CP is appended to D_i such that

$$S_i(k) = \left\{\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} X_i(k)\, e^{j2\pi nk/N}\right\}, \qquad (3)$$

Where $X_i(k)=x(j)$ from Eqn. (2) is the superimposed modulated data symbol assigned to the sub carriers and N_g represents the length of CP ($-Ng < n < N-1$).

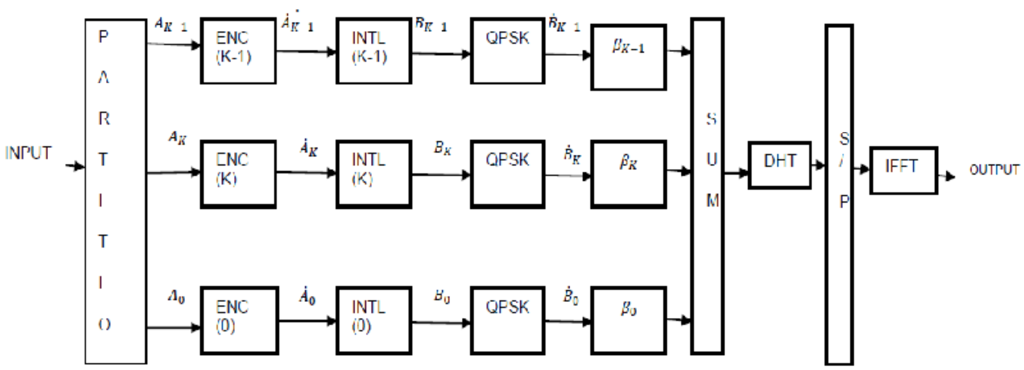


Fig. 1. The Proposed SCM-OFDM system for PAPR reduction.

B. Peak-to-Average Power Ratio (PAPR)

The PAPR is important for enabling the high-power amplifier to operate in the linear region. The SCM-OFDM signal is oversampled '*L*' times to better approximate the PAPR. The oversampled SCM-OFDM signal in time domain is given as follows:

$$S_i(k) = \left\{\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1} X_j(k)e^{j2\pi nk/LN}\right\}, 0 < n < LN \quad (4)$$

The PAPR is termed as the ratio of the maximum instantaneous power to the average power and is formulated mathematically as follows:

$$\text{PAPR}[S[n]] = \frac{max\,[|s[K]|^2]\ 0\leq n\leq LN-1}{E[|s[K]|^2]}, \quad (5)$$

where E [.] specifies the expectation operator.

C. Complementary Cumulative Distribution Function

The CCDF refers to the probability that the distribution of the output power of the SCM-OFDM signal exceeds the predefined threshold value. It is done by determining the CCDF for the PAPR values. The CCDF is formulated and expressed as follows:

$$\text{CCDF} = \Pr(\text{PAPR}_{SCM-OFDM} > \text{PAPR}_o), \quad (6)$$

Where $\text{PAPR}_{SCM-OFDM}$ and PAPR_o are the output of the SCM-OFDM system and the threshold value respectively.

III. RESULTS AND DISCUSSION

A. Clipping technique for reducing the PAPR in the SCM-OFDM system

In the SCM-OFDM system, the PAPR is maximum for a given $\{\rho_k\}$ when all θ_ks are equal. To reduce the PAPR, clipping of X_j to form $\overline{X_j}$ is followed based on the rule presented below:

$$\overline{X_j} = \begin{cases} s_j, |s_j| \leq A \\ \frac{As_j}{|s_j|}, |s_j| > A \end{cases} \quad (7)$$

where A > 0 is the clipping threshold.

From [13], we outline the clipping ratio as follows:

$$\gamma = A^2/E(|X_j^2|) \quad (8)$$

The PAPR of the transmitted signal with clipping is given as follows

$$\text{PAPR} = A^2/E(|\overline{X_j^2}|).$$

The received signal can be written as follows:

$$Y_j = \overline{X_j} + w_j, \quad (9)$$

where w_j is the complex Gaussian noise with a zero mean and σ^2 variance.

B. Selective Mapping

Selective Mapping (SM) is a multiplicative PAPR reduction technique. The SCM-OFDM signal is copied and each copy is multiplied with different phase sequences, as a result different PAPR values are captured. The SM signal is a product of SCM-OFDM and phase sequences and represented by X(1), X(2), X(3),… X(k), where k is the number of SCM-OFDM signals. Inverse Discrete Fourier transform is performed on X(k). X(k) with lowest PAPR is selected and transmitted. With increase in number of k the performance of SM raises but as the k increases number of IDFTs also increases, increasing the complexity of the system. The block diagram of the SM is presented in Fig. 2. This technique also requires side information. The side information are extra bits that carries the phase sequences to receiver increasing the bandwidth. SM technique is not only spectral inefficient but also complex system.

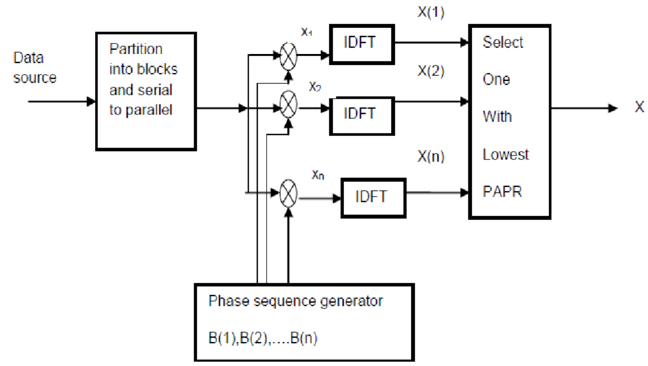


Fig. 2. Block diagram of Selective Mapping.

C. Partial Transmit Sequence

Partial transmit sequence is selection transmission technique with point-to-multipoint mapping, multiply single input data signals by a phase factor to map multiple candidate signals. Similar to SM technique, candidate signal with lowest PAPR is selected as the OFDM transmission signal. PTS also requires side information to be transmitted to receiver to retrieve which candidate signal was selected by the transmitter. Let X denote the random signal in frequency domain. Now X is partitioned into U disjoint subblocks represented by $\{X^{(u)}, u = 1, 2,\ldots U\}$ where $X^{(u)}$ is given by

$$X^{(u)} = [X_1^{(u)}, X_2^{(u)}, \ldots. X_{M-1}^{(u)}] \quad (10)$$

$$X = \sum_{u=1}^{U} X^u \quad (11)$$

The phase rotation factors are given by

$$C_u = e^{j\theta u}, u=1,2, \ U \quad (12)$$

IV. PROPOSED MMC-DHT PAPR REDUCTION TECHNIQUE FOR SCM-OFDM SYSTEM

The objective of this investigation is to realize PAPR reduction before OFDM modulation using Modified Companding and DHT precoding. Since the SCM symbols takes Gaussian distribution, it implies that there is no requirement of active shaping filters which is used to adapt the advanced modulation schemes such as QAM to Gaussian channel, thereby reducing the shaping filter burden on transmitter and receiver of the system. The Cumulative Distributed Function (CDF) of the un-companded signal is expressed as

$$F_x(.) = 1-\exp\left(\frac{x_0^2}{\sigma_x^2}\right), x_o>0 \quad (13)$$

Where x_o is that the discrete envelope of $x(n)$ and σ_x^2 is the variance.

The μ companding was introduced based on [15]

$$F_{EC}(x(n)) = sgn(x).\sqrt{3\sigma_x^2}.\,\text{erf}\left(\frac{|x|}{\sqrt{2\sigma_x^2}}\right), 0\leq x\leq 1 \quad (14)$$

The above equation is the fundamental error function due to companding transform.

μ law companding with constant value for μ applied to SCM-OFDM signal is given by

$$y_c(n) = F(x_j(n)) \quad (15)$$

$$= sgn(x_j(n))\ \frac{A_s \times \log\left(1+\mu\left|\frac{x_j(n)}{A_s}\right|\right)}{\log(1+\mu)}$$

where $sgn(x(n))=\frac{x(n)}{|x(n)|}$ is the phase and A_s=max($|x_j(n)|$.

The multi μ is simulated for SCM-OFDM system for better PAPR reduction .The fundamental problem in μ companding is, it enlarges the amplitudes of the lower amplitude signal keeping higher amplitude signals unchanged at the output of the compander. Thus increasing the average power and reduces the PAPR. However, this technique leads to unfair improvement of BER when compared to uncompanded signal. The companding profile known as peak ratio is presented in [15], it is articulated as the ratio of amplitude A of the signal specified in μ compander to the peak amplitude of the actual signal to be companded i.e.

$$K=\frac{\text{Peak amplitude of compressor}}{\text{Peak of actual signal}}=\frac{A}{x_{j\,\text{peak}}} \quad (16)$$

The transfer characteristics of the modified μ compander including new parameter peak ratio 'K' is expressed by substituting the above equation in (15).

$$y=K\times x_{jpeak}\times\frac{\log\left(1+\mu\frac{|x_{j\text{peak}}|}{K\times x_{j\text{peak}}}\right)}{\log(1+\mu)}sgn(x) \quad (17)$$

The above comapanding profile permits the $\mu-\mathrm{Law}$ companding profiles to be modified such that all amplitudes as well as the peaks of the input signal can be amplified by changing the value of K. Higher the values of K, greater is the gain for the peaks and much higher the gain for the lower amplitude signals.

OFDM signal is precoded using DHT precoder before IFFT block. The Hartley Transform of $S_i(n)$ is expressed as

$$H_s(\omega)=\int_{-\infty}^{\infty}f(t)Cas(\omega t)dt. \quad (18)$$

$$Cas(\omega t)=\cos(\omega t)+\sin(\omega t) \quad (19)$$

$$f(t)=\int_{-\infty}^{\infty}H_s(\omega)Cas(\omega t)d\omega \quad (20)$$

To reduce the number of functional computation $Cas(\omega t)$ can be expressed as:

$$\sqrt{2}\cos\left(\omega t-\frac{\pi}{4}\right)=\sqrt{2}\sin\left(\omega t+\frac{\pi}{4}\right) \quad (21)$$

For discrete signal Hartley Transform is expressed as:

$$H_s(k)=\frac{1}{\sqrt{2}}\sum_{n=0}^{NL-1}y_i(n)Cas(\frac{2\pi nk}{N}) \quad (22)$$

$$H_s(k)=\frac{\sqrt{2}}{\sqrt{N}}\sum_{n=0}^{NL-1}y_i(n)\cos(\frac{2\pi nk}{N}\frac{\pi}{4}) \quad (23)$$

k=0, 1,...... N-1.

The discrete Hartley Transform rotates the phase of input SCM symbols. While precoding, the signal is re-established to the single carrier. SCM-OFDM is analysed in AWGN channel. The received signal is as follows:

$$Y_j(n)=\left\{\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}H_s(k)e^{\frac{j2\pi nk}{LN}},\right\}+\mathrm{w(n)} \quad (24)$$

$0 < n < LN-1$

V. SIMULATION RESULTS

In this study, computer simulations were performed to evaluate the performance of the SCM-OFDM system. Cyclic prefix is added to the time domain signal to avoid the Inter Symbol Interference (ISI). Besides PAPR drawback in OFDM system, OFDM is additionally sensitive to spectral null problem over a frequency selective fading channel. By using Discrete Hartley Transform over the entire bandwidth the nulls are spread to increase the probability of correctly receiving the transmitted symbols.

Use of Precoding, further improves the PAPR performance of an OFDM multicarrier system. SCM-OFDM system is simulated in MATLAB 2017 environment. Each layer in multicarrier is coded using convolution encoder with rate ½. The coded code word is interleaved to remove any burst error present in the data stream. Each layer is modulated using different modulation schemes. We adopted QPSK and 16 QAM for the layer 1 and layer 2 respectively.

The simulation result of SCM scheme with two layers, where layer 1 adopted QPSK and for layer 2 16QAM and resultant superimposed signal is shown in Fig. 3. In SCM symbols are arranged such that the PAPR is reduced with different power factor $'\beta'$ .For the higher modulation scheme when SCM is simulated with 64QAM PAPR obtained is 6.4089dB, OFDM with 64QAM is 9.31dB. Compared to OFDM PAPR in SCM improved by 31.2%. However, PAPR in the hybrid system SCM-OFDM system is 13.53dB. PAPR in hybrid system increases due to nonlinearity in SCM; Since Gaussian like transmitted signal has a relatively high PAPR. SCM-OFDM with clipping resulted in reduction of PAPR by 2.87dB. Although PAPR reduction using clipping and filtering technique is reduced to 2.87dB, but BER performance of the same reduction scheme is poor as shown in Fig. 4 compared to BER performance of the proposed method. The BER performance is enhanced using proposed hybrid technique for SCM-OFDM system using DHT. PAPR for SCM-OFDM signal is 5.8971dB, thereby reducing PAPR by 4.7565dB compared to clipping and without reduction technique by 7.6329dB respectively, thereby improving performance of PAPR by 56.41%.

Table 1: Simulation Parameters

Bandwidth	20MHz
Carrier frequency	2.5GHz
Number of subcarriers	256
Subcarrier spacing	15KHz
Number of Cyclic prefix	64
Convolution encoder rate	½
Interleaver	Random Interleaver
Modulation for layer 1	256 QAM
Modulation for layer 2	256 QAM

With proposed hybrid technique the PAPR is further reduced to 1.7949 dB and PAPR performance improvement by 86.7%. The proposed system is also simulated with 256 QAM, PAPR in OFDM with 256 QAM is as high as 30dB due to nonlinearity in the 256 QAM. PAPR will increase up to a greater extent in OFDM system. Computer simulation is done and results are shown in figure 9. With the proposed hybrid scheme the PAPR is reduced significantly to 12.5dB. Fig. 3 shows the simulation of SCM symbols, considering K=2 layers the number of symbols in SCM is 2. Fig. 4 show that the PAPR reduction in SCM-OFDM system with DHT precoder and clipping as a PAPR reduction technique. DHT precoder outperforms clipping technique by 7.01dB. BER performance is better compared to clipping technique for SCM-OFDM system.

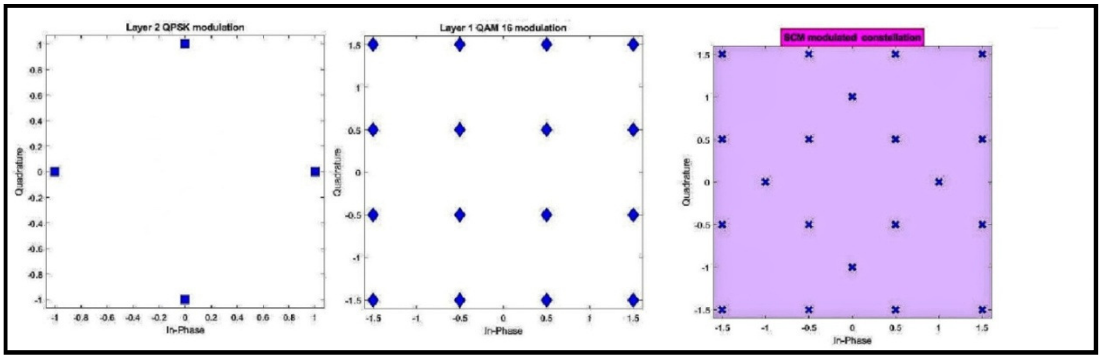


Fig. 3. Simulation result of Superposition Coded Modulation of two signals QPSK and 16 QAM modulated signal.

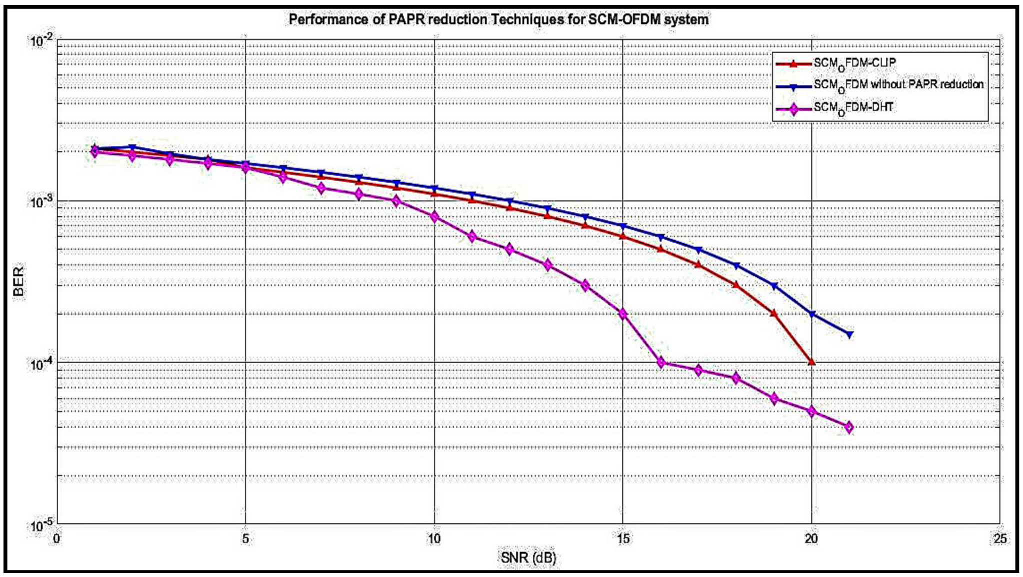


Fig. 4. BER performance of Clipping and DHT PAPR reduction techniques in SCM-OFDM system.

Fig. 5 present that when Modified µ law Compander with different value of µ, with µ=30 and above, the PAPR of SCM-OFDM system reduces to 3.8 dB. Increment in PR ratio to 2, increases the BER performance, compared to other µ-law companding technique as shown in Fig. 6. The BER performance in DHT precoder is better than the BER of µ-law companding techniques. The simulation uses different µ-law companding levels and is denoted as U_1 and U_2 in the result. However, the PAPR is more in DHT precoding technique compared to µ-law companding. The proposed system considers the advantages of both the techniques. Fig. 7 show different PAPR reduction technique for SCM-OFDM system, the proposed system outperforms the other reduction technique with PAPR of 1.79 dB. The BER performance of proposed system and its outperformance with the other PAPR reduction techniques is shown in Fig. 7. The Table 2 shows the comparison of computational complexity in DHT and DFT scheme for PAPR reduction. Since DHT requires less multiplications compared to DFT, proposed system is low computationally complex system. Table 3 compares the spectral efficiency of the DHT with other alternative techniques. PTS and SLM techniques use side information which increases bandwidth requirement. The proposed method is bandwidth efficient, since no side information is required. Fig. 8 provides BER comparative study of SCM –OFDM system with different PAPR reduction techniques. PAPR reduction for SCM-OFDM system with proposed reduction technique is 12.5dB for 256 QAM modulations. Table 4 provides the PAPR results and comparisons with SCM only, OFDM only and Hybrid combination of SCM-OFDM system with Clipping, SLM, PTS and MMC- DHT precoder for 64 QAM and 256 QAM. It is evident that from Table 4 with MMC-DHT precoder improves PAPR Performance of SCM-OFDM system.

Table 2: Comparison of DHT and DFT for computational complexity.

Schemes	Real multiplications	Complex additions
DHT	$2L^2$	L(L-1)
DFT	$4L^2$	L(L-1)

where L is order of DHT matrix and DFT matrix.

From the Table 3 it is observed that the computational complexity is reduced to 50% in case of DHT compared to DFT precoding. With use of DHT precoding for PAPR reduction it not only reduces PAPR but also reduces computational complexity in the system.
Where S is the number of subblocks and ϕ : is the number of phase factors.
The PAPR reduction techniques such as PTS and SLM requires side information which reduces the spectral efficiency of the system, on the contrary, DHT precoding requires no information for PAPR reduction in SCM-OFDM system improving spectral efficiency in the system as compared to existing reduction techniques.

The number od side information required for the above mention techniques are mentioned in Table 3.

Table 3: Comparison of DHT for spectral efficiency.

Scheme	Sideband information (bits/OFDM symbol)
DHT	None
PTS	$S \log_2 \phi$
SLM	$\log_2 \phi$

From the Table 4 it is evident that the proposed technique is efficient, 61% PAPR reduction compared to DHT alone for the SCM-OFDM system.

Table 4: Comparison of PAPR Reduction techniques for SCM-OFDM system.

PAPR reduction technique	SCM	OFDM	SCM-OFDM without PAPR reduction	SCM-OFDM with clipping	SCM-OFDM with DHT	SCM-OFDM with μ companding	SCM-OFDM with Proposed PAPR reduction for 64 QAM	SCM-OFDM with Proposed PAPR reduction for 256 QAM
PAPR (in dB)	6.4089	9.31	13.53	10.6536	5.8971	3.64	1.7949	12.5

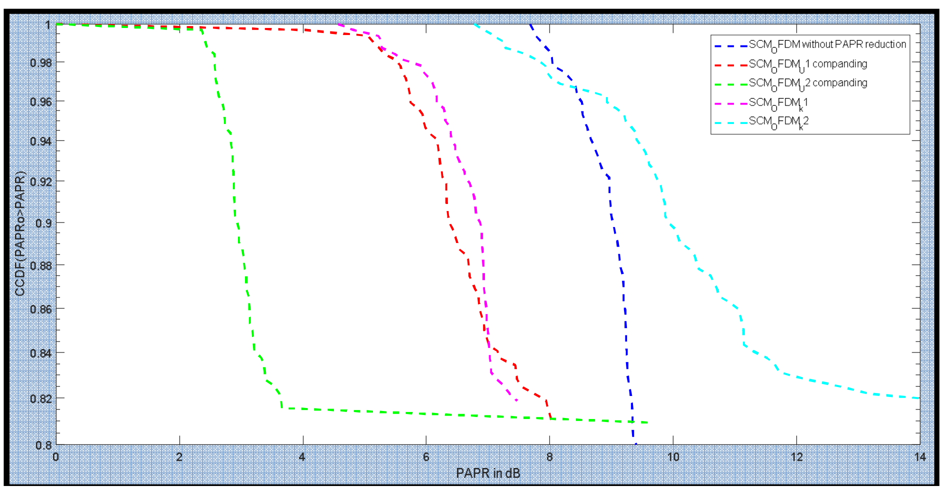


Fig. 5. CCDF Vs PAPR for different μ companding profiles (u_1,u_2) and modified μ companding techniques.

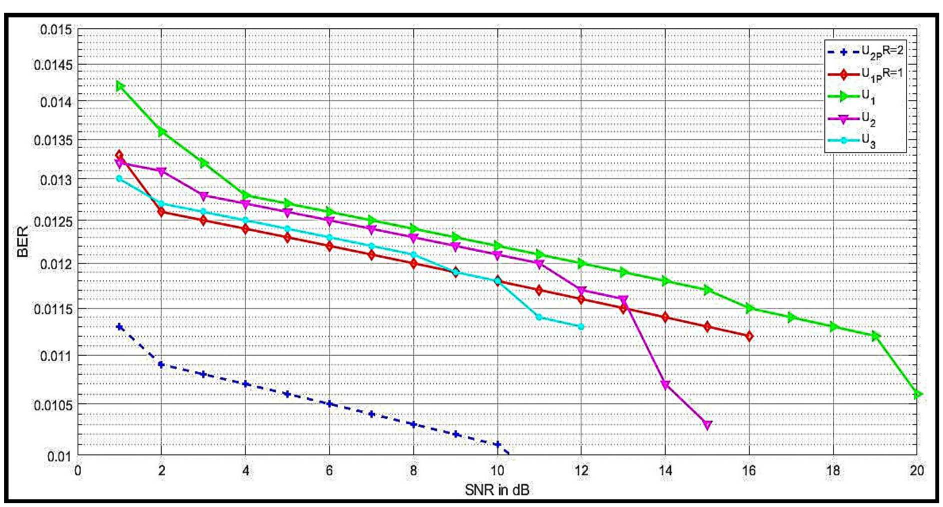


Fig. 6. BER performance of Modified µ law companding technique on SCM-OFDM system with different companding levels U_1,U_2 and U_3.Table IV: Comparison of PAPR Reduction techniques for SCM-OFDM system.

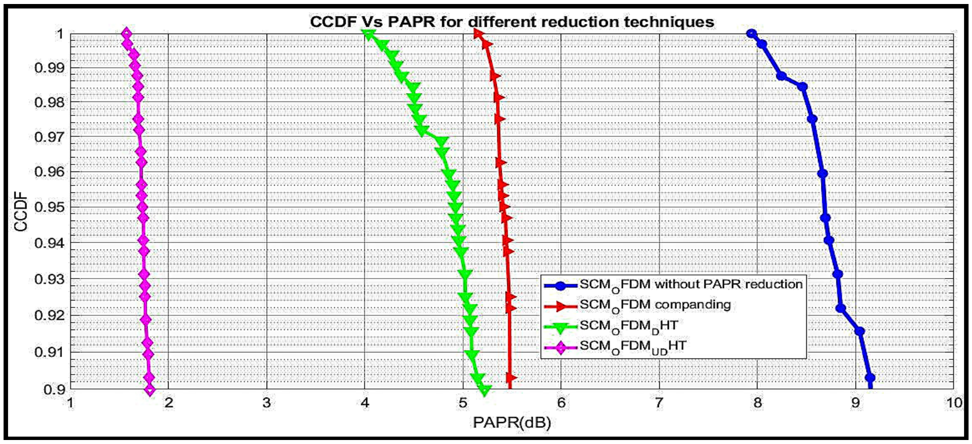


Fig. 7. CCDF Vs PAPR for different PAPR reduction techniques on SCM-OFDM system.

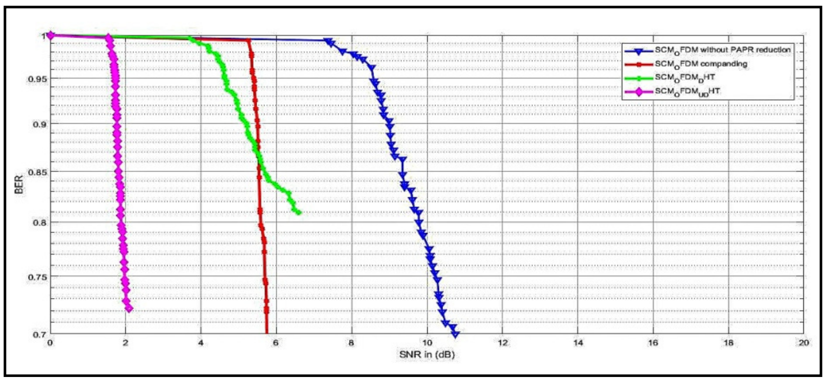


Fig. 8. BER performance of different PAPR reduction technique on SCM-OFDM system.

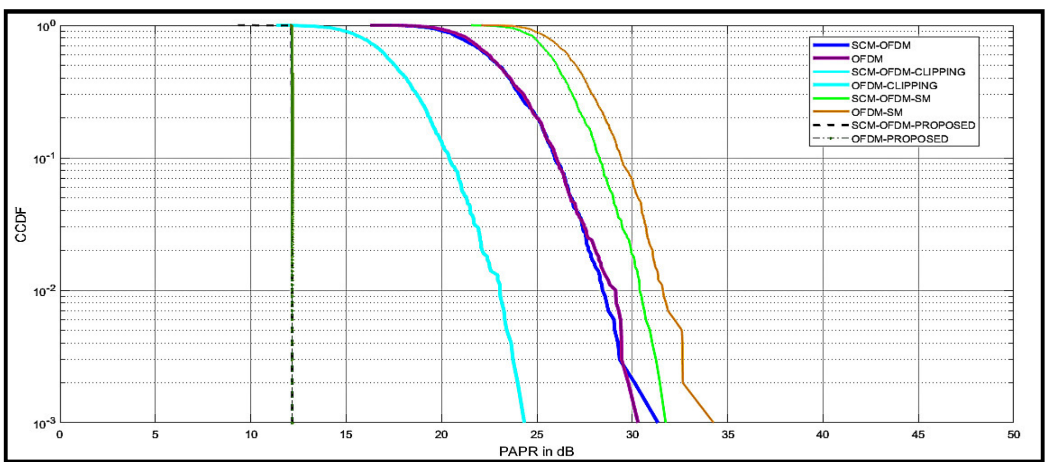


Fig. 9. Comparative study of PAPR on SCM-OFDM systems for 256 QAM modulation with different reduction techniques.

The SCM OFDM system is simulated for 64 QAM and 256 QAM respectively, PAPR for SCM signal and OFDM signal PAPR are 6.4089dB and 9.31dB respectively. With the combination of SCM-OFDM system PAPR increased to 13.53dB due to Gaussian like signal of SCM-OFDM signal and with equal power allotted for two layers of SCM. Clipping technique reduces PAPR by 2.87dB, where as DHT achieves PAPR reduction by 7.63dB. PAPR is further reduced by 9.89dB by μ companding technique. As observed from the simulation results DHT technique provides better BER performanc, µ companding technique provides better PAPR performance. Combining the advantages of both the technique we achieved reduction in PAPR by 11.73dB and improvement in PAPR by 86.7%. SCM-OFDM system is simulated with 256QAM to demonstrate the proposed method works better compared to well known PAPR technique such as PTS and Selective Mapping (SM). The Fig. 9 shows that PAPR in SCM-OFDM without PAPR reduction is 33dB, with SM technique PAPR is reduced to 32 dB. Some improvement is observed when PTS is applied to SCM-OFDM system. Significant result is obtained with the proposed technique, PAPR is reduced to 12.5dB, PAPR performance is improved by 62.1%.

VI. CONCLUSIONS

A novel PAPR reduction technique for the OFDM system with SCM is presented and elaborated in this paper. An investigation on performance of different reduction techniques on SCM-OFDM system is achieved through computer simulation. The results indicated that the proposed technique outperformed the clipping technique by 83.1%, PTS by 58.3% , SM by 62% and achieved PAPR of 1.7949 dB with 64QAM SCM-OFDM system, 12.5dB with 256QAM SCM-OFDM system. The proposed method is beneficial for high data rate systems, where many constellations are required to achieve a high data rate.

Future scope: For future research, time-varying channels can be studied; furthermore, different modulations and layers can be experimented with SCM-OFDM system.

Conflict of Interest: No conflict of interest.

REFERENCES

[1]. Khan, S., & Jai, S. (2018). Multiband-OFDM MIMO Based Ultra-Wideband Communication System. *IJET 9*(2): 27-30.

[2]. Singh, J., Dodiya., & Sharma, R. (2014) (PAPR Reduction of OFDM Signal with Improved Selected Mapping and Clipping Techniques. *IJET*, *3*(1), 27-30.

[3]. Rashmi, N., & Sarvagya, M. (2016). State Space Model Based Channel Estimation using Extended Kalman Filter for Superposition Coded Modulation OFDM System. *BVICA M's International Journal of Information Technology*, *8*(2), 1017.

[4]. Vittal, M. V. R., & Naidu, K. R. (2017). A novel reduced complexity optimized PTS technique for PAPR reduction in wireless OFDM systems. *Egyptian informatics journal*, *18*(2), 123-131.

[5]. Liang, H. Y., Chu, H. C., Lin, C. B., & Lin, K. H. (2014). A partial transmit sequence technique with error correction capability and low computation. *International Journal of Communication Systems*, *27*(12), 4014-4027.

[6]. El-Hamed, M. A., Zekry, A., Elagooz, S. S., & El-Samie, F. E. A. (2018). Blind selective mapping for single-carrier frequency division multiple access system. *Digital Signal Processing*, *75*, 25-37.

[7]. Hassan, E. (2016). PAPR Reduction Using Selective Mapping Scheme. In *Multi-Carrier Communication Systems with Examples in MATLAB* (pp. 84-113). CRC Press.

[8]. Satyavathi, K., & Rao, B. R. (2019). Modified Phase Sequence in Hybrid Pts Scheme for PAPR Reduction in OFDM Systems. In *Innovations in Electronics and Communication Engineering* (pp. 327-333). Springer, Singapore.

[9]. Taşpınar, N., & Yıldırım, M. (2015). A novel parallel artificial bee colony algorithm and its PAPR reduction performance using SLM scheme in OFDM and MIMO-OFDM systems. *IEEE Communications Letters*, *19*(10), 1830-1833.

[10]. Chandrakantbhai, P.V. (2014). PAPR Reduction in OFDM System using Tone Reservation Technique. *IJEECE*

[11]. Wang, H., Wang, X., Xu, L., & Du, W. (2016). Hybrid PAPR reduction scheme for FBMC/OQAM systems based on multi data block PTS and TR methods. *IEEE Access*, *4*, 4761-4768.

[12]. Azim, A. W., Le Guennec, Y., & Maury, G. (2017). Decision-directed iterative methods for PAPR reduction in optical wireless OFDM systems. *Optics Communications*, *389*, 318-330.

[13]. Deepika Khandre,& Sneha Jain (2018). Performance of the LTE Low Complexity PAPR Reduction for Coded Scrambling Technique. IJET 9(2): 36-41(2018).

[14]. Lee, J. H., & Chung, H. (2019). Exact and approximate log-likelihood ratio of M-ary QAM with two-time dimensions. *ICT Express*, *5*(3), 173-177.

[15]. Nadal, L., Moreolo, M. S., Fàbrega, J. M., & Junyent, G. (2014). Low complexity PAPR reduction techniques for clipping and quantization noise mitigation in direct-detection O-OFDM systems. *Optical Fiber Technology*, *20*(3), 208-216.

[16]. Chen, H., & Liang, H. (2007). PAPR reduction of OFDM signals using partial transmit sequences and Reed-Muller codes. *IEEE Communications Letters*, *11*(6), 528-530.

[17]. Liang, H. Y. (2015). Integrating CE and modified SLM to reduce the PAPR of OFDM systems. *Wireless Personal Communications*, *80*(2), 709-722.

[18]. Nagaraju, C. H., Sharma, A. K., & Subramanyam, M. V. (2018). REDUCTION OF PAPR IN MIMO-OFDM USING ADAPTIVE SLM AND PTS TECHNIQUE. *International Journal of Pure and Applied Mathematics*, *118*(17), 355-373.

[19]. Ghahremani, R., & Shayesteh, M. G. (2014, May). BER performance improvement and PAPR reduction in OFDM systems based on combined DHT and μ-law companding. In *2014 22nd Iranian Conference on Electrical Engineering (ICEE)* (pp. 1483-1487). IEEE.

[20]. Saied, O., Ghassemlooy, Z., Rajbhandari, S., & Burton, A. (2019). Optical single carrier-interleaved frequency division multiplexing for visible light communication systems. Optik, 194, 162910.

[21]. Khatri, A. K., & Tharani, L. (2019). A Novel Detected Peak Samples Combining Technique for PAPR Minimization of MC-CDMA Signal. *Wireless Personal Communications*, 1-12.

How to cite this article: Rashmi, N. and Sarvagya, M. (2019). Power Efficiency Enhancement using Hybrid Techniques for OFDM. *International Journal on Emerging Technologies*, **10**(3): 260–267.

Self-improved grey wolf optimization for estimating carrier frequency offset in SCM-OFDM systems

Rashmi N., Mrinal Sarvagya

International Journal of Pervasive Computing and Communications
ISSN: 1742-7371 (International Standard Serial Number)
Publication date: 2 January 2020

Abstract

Purpose

The purpose of this paper is to demonstrate a proficiency for accomplishing optimal CFO and keep down the error among the received and transmitted signal. Orthogonal frequency-division multiplexing (OFDM) is considered as an attractive modulation scheme that could be adopted in wireless communication systems owing to its reliability in opposition to multipath interruptions under different subchannels. Carrier frequency offset (CFO) establishes inter-carrier interference that devastates the orthogonality between the subcarriers and fluctuates the preferred signal and minimizes the effectual signal-to-noise ratio (SNR). This results in corrupted system performance. For sustaining the subcarriers' orthogonality, timing errors and CFOs have to be approximated and sufficiently compensated for. Single carrier modulation (SCM) is a major feature for efficient OFDM system.

Design/methodology/approach

This paper introduces a novel superposition coded modulation-orthogonal frequency-division multiplexing (SCM-OFDM) system with optimal CFO estimation using advanced optimization algorithm. The effectiveness of SCM-OFDM is validated by correlating the transmitted and received signal. Hence, the primary objective of the current research work is to reduce the error among the transmitted and received signal. The received signal involves CFO, which has to be tuned properly to get the signal as closest as possible with transmitted signal. The optimization or tuning of CFO is done by improved grey wolf optimization (GWO) called GWO with self-adaptiveness (GWO-SA). Further, it carries the performance comparison of proposed model with state-of-the-art models with the analysis on bit error rate (BER) and mean square error (MSE), per cent better than least mean square (LMS), 99.3 per cent better than particle swarm optimization (PSO), 75 per cent better than genetic algorithm (GA) and 25 per cent better than GWO algorithms. Moreover, MSE at 1st SNR, the proposed GWO-SA scheme, is 4.62 per cent better than LMS, 60.1 per cent better than PSO, 37.82 better than GA and 67.85 per cent better than GWO algorithms. Hence, it is confirmed that the performance of SCM-OFDM system with GWO-SA-based CFO estimation outperformed the state-of-the-art techniques.

Originality/value

This paper presents a technique for attaining optimal CFO and to minimize the error among the received and transmitted signal. This is the first work that uses GWO-SA for attaining optimal CFO.

Keywords

SCM | GWO | Bet error rate | Carrier frequency | MIMO-OFDM | MSE analysis

Citation

N., R. and Sarvagya, M. (2020), "Self-improved grey wolf optimization for estimating carrier frequency offset in SCM-OFDM systems", *International Journal of Pervasive Computing and Communications*, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/IJPCC-03-2019-0020

Download as .RIS

Publisher: Emerald Publishing Limited

International Journal of Advanced Science and Technology
Vol.133 (2019), pp.11-18
http://dx.doi.org/10.33832/ijast.2019.133.02

PERFORMANCE ENHANCEMENT OF SCM-OFDM SYSTEM USING CARRIER FREQUENCY OFFSET ESTIMATION AND CORRECTION

Rashmi N[1] and Mrinal Sarvagya[2]

[1] *Department of Electronics and Communication Engineering, BMS Institute of Technology and Management[1], India*
[2]*Department of Electronics and Communication Engineering, Reva university India*
[1] rashmiswamy@bmsit.in, [2]mrinalsarvagya@reva.edu.in

***Abstract*— OFDM system loses the orthogonality between subcarriers due to carrier frequency offset resulting in inter-carrier interference (ICI) in the OFDM system. CFO impedes the performance of the demodulation process in baseband receiver increasing BER in the OFDM system. In the proposed scheme, the estimation of CFO is performed using the Extended Kalman filter (EKF) and sigma pointed Kalman filter (SPKF) on the SCM-OFDM system. The simulation results show that the estimation using SPKF can achieve better results than other variants of the Kalman filter. The proposed system improves the BER performance of the OFDM system. Further, it is shown that SPKF outperforms than of EKF.**

***Keywords*— Carrier Frequency Offset, OFDM, Superposition Coded Modulation, Sigma pointed Kalman filter**

1. INTRODUCTION

The Orthogonal Frequency Division Multiplexing (OFDM) is a multicarrier bandwidth-efficient system, a promising technique for the systems which aspire for high rate transmission such as 4G-LTE advanced, Wimax, etc, Advantages of OFDM attracted several researchers and academicians to choose OFDM technique which can cater for high data rate applications. OFDM is also robust to multipath fading channels with low complexity implementation. OFDM uses higher modulation schemes such as PSK or QAM to achieve bandwidth efficiency. However, the Quadrature part of the PSK and QAM are not Gaussian distributed. To achieve channel capacity, the input to the channel must be Gaussian. Traditionally active signal shaping is performed on modulation schemes to achieve Gaussian distribution. Several methods were proposed by researchers and academicians to increase channel capacity [1]. With active signal shaping techniques the complexity at the transmitter and receiver increases. With APP decoder for bijective modulation (one to one mapping) schemes, the complexity increases with an increasing number of information bits as $O(2^N)$. In 1997, Daunt et al. designed a multilevel coded modulation scheme in [2-3] and pointed out that huge signal constellation and active signal shaping are indeed not mandatory to approach the channel capacity. Li Ping followed this work and did extensive analysis in [4]. Superposition coded modulation with one layer can achieve similar performance as presented by the authors of [4]. But the results are limited to 2 bps per signal dimension. Superposition Coded Modulation (SCM) is emerging as non-bijective, bandwidth and power efficient coded modulation with

Received:September 18, 2019
Reviewed: December 2, 2019
Accepted: December 10, 2019

ISSN: 2207-6360 IJAST

Gaussian quadrature components. The shaping gain in SCM is inherent without any additional complexity at the transmitter side. At the receiver side, the complexity is low as O(N) when equal power is allocated to the symbols. The hybrid combination of multi-layer SCM and OFDM can achieve greater data rates compared to a single carrier OFDM system. However, the challenge in OFDM is to maintain the orthogonality of the subcarrier. OFDM is highly prone to the timing errors and Carrier Frequency Offsets (CFO) which is caused mainly due to 1) mismatch between the transmitter and local oscillator frequencies at the receiver 2) Phase rotation caused due to Doppler Effect. The inaccurate CFO estimation destroys the orthogonality among subcarriers leads to Inter-Carrier Interference (ICI). Doppler effect is significant in an aeronautical communications application, for example, CFO is high as 0.54, at a 2.5GHz as carrier frequency, vehicle speed 120KM/Hr with 1024 subcarriers OFDM system. Hence apposite methods are imperative for the successful utilization of OFDM in such a high-speed environment. In this connection, several methods were proposed in literature such as the Successive Interference Cancellation (SIC) method proposed by [7]. In this method, at the transmitter for adjacent subcarriers are assigned with opposite symbols which are subtracted at the receiver, advanced SIC method, in which phase rotation is done before sending is proposed in [8]. To mitigate the ICI, Carrier Frequency Offset(CFO) is necessary, CFO estimation using the Maximum Likelihood estimator is proposed in[9], which increases the complexity of the receiver with an increase in the order of modulation. Several Kalman filter approaches are proposed in [10] to estimate CFO the results obtained were significant. Since CFO is nonlinear, linear approaches result in low performance of the OFDM system. EM algorithm proposed in [11] for estimation of CFO, exhibit computational complexity at the receiver, due to computationally intense detection at the receiver.

In [12] Particle filter was proposed, however, the particle approaches have occasionally particle insolvency (PI) problem. Due to the re-sampling process particle insolvency problem arises. In a dynamic system, if the state process noise is small, the PI problem increases. However, we can avoid the PI problem by modeling state process noise as zero in CFO estimation.

To estimate the CFO, the Extended Kalman filter is proposed in [13], but the computational complexity is more in the Extended Kalman filter, which results in the OFDM power inefficient receiver.

In this paper, the Unscented Kalman filter (UKF) also referred to as sigma point Kalman filter (SPKF) is proposed for investigating the performance of the system under consideration for the CFO problem [14-15]. UKF varies from CDKF in the calculation of sigma points and covariances.

The paper is structured as follows: the system model is presented in section II, we present a state-space model of CFO in SCM-OFDM in section III. In section IV we present a brief explanation of the Extended Kalman filter and outlines UKF using Sigma Points and section V the simulation results of the performance of the SCM-OFDM system with CFO estimation using UKF are presented and section VI is the conclusion.

2. SYSTEM MODEL

The serial data stream is divided into L- parallel data stream each layer is separately encoded, randomly interleaved and modulated with QPSK modulation. Quadrature phase-shift keying (QPSK) symbols.

$$x_l^L = x_{re,j}^L + i x_{im,j}^L \tag{1}$$

x_l^L : QPSK symbol

The QPSK sequences are linearly superimposed to form x_j.

$$X_j = \sum_{l=0}^{L-1} \rho_l \, e^{j\theta_l} x_l^L \quad (2)$$

Where $\rho_l > 0$ *is an amplitude factor*, $\theta_l \left(0 \leq \theta_l \leq \frac{\pi}{2}\right)$ for layer –l.

The superimposed symbols are fed to the OFDM modulator point consists of N-point Discrete Fourier Transforms (IDFT) unit followed N_g-point Cyclic Prefix(CP) to evade Inter Symbol Interference(ISI)[16][17]. The IDFT

output is collected as vector $D_i = [\, D_i(0),\ D_i(1), \ldots\ D_i\ (N-1)]^T$. A cyclic prefix is appended to D_i such that

$D_i(k) = D_i(k + N)$ for $-N_g \leq k \leq -1$.

$$D_i(k) = \left\{ \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_i(k) e^{j2\pi nk/N}, \right\};\ \ k=0,1,\ldots N-1 \quad (3)$$

Where $X_i(k)$ are superimposed modulated data symbols assigned to N subcarriers and with N_g as the length of CP, $-Ng < n < N-1$.

The SCM-OFDM signal propagates through the multipath channel, the L - tap channel impulse response is given by $h_l(l=0,1,\ldots.L-1)^T$.

The received signal can be articulated as:

$$y_k = e^{\frac{j2\pi\epsilon k}{N}} \left(\sum_{l=0}^{L-1} D_{k-l}\ h_l\right) + w_k \quad (4)$$

w_k is the AWGN

$'\epsilon'$ is a normalized carrier frequency offset. Where $'\varepsilon'$ needs to be estimated, to restore the orthogonality in the OFDM symbol. The CFO calculated with f_d as the Doppler frequency expressed as

$$f_d = \frac{CV_r F_c}{C} \quad (5)$$

Where v_r -the speed of the mobile station, C- the speed of light,

f_c- carrier frequency

$$\epsilon = f_d / \Delta f = N f_d / f_s \quad (6)$$

f_s is the subcarrier spacing of $D_i(k)$.

3. DYNAMIC STATE SPACE MODEL OF CFO

In slow time-varying channel carrier frequency offset is considered as constant for OFDM symbol duration.

The state and measurement equations for dynamic state system are as follows:

$$a_k = s_k[a_{k-1}] + u_k, \quad (7)$$

$$y_k = f_k\,[a_k] + w_k \quad (8)$$

Where $f_k(.)$ is the observation function a, $s_k(.)$ is the state transition function, u_k and w_k is the process noise and measurement noise respectively. Corresponding state and measurement equations are:

$$\epsilon_k = \epsilon_{k-1} \tag{9}$$

$$y_k = f_k(\varepsilon) + w_k \tag{10}$$

$$f_k(\varepsilon) = e^{\frac{j2\pi k\epsilon}{N}} \sum_{l=0}^{L-1} a_{k-1} h_l \tag{11}$$

Where ε the normalized CFO and w_k is additive noise. UKF estimates a_k in succession from the observed measurement y_k.

Another popular Kalman filter EKF propagates the Gaussian Random Variable (GRV) with the first-order linearization of non-linear equation (11) through the system dynamics [18-19]. The evaluation of mean and covariance matrices of GRV using EKF leads to errors when the first-order approximation used. In UKF Gaussian distribution is approximated with state distribution by a set of sigma points. Hence we propose the UKF for better convergence and higher accuracy in CFO estimation.

4. A) EXTENDED KALMAN FILTER

The Kalman filter is a recursive estimator, computes the estimate of the current state based on the knowledge of previously estimated state and current measurement.

The dynamic state model for the CFO in vector form is represented as follow:

$$\varepsilon(n) = \varepsilon(n-1) + u(n) \tag{12}$$

'ε' is the normalized CFO and u(n) is a Gaussian process with mean equals to zero and variance equals to σ^2. The extended Kalman filter minimizes and measurement noises.

The EKF equations are as follows:

Time update:

$$\epsilon(n) = \epsilon(n-1) + u(n)$$

$$\hat{\epsilon}(n|n-1) = \hat{\epsilon}(n-1|n-1) \tag{13}$$

$$P(n|n-1) = P(n-1|n-1) + \sigma^2 \tag{14}$$

Measurement update:

$$\hat{\epsilon}(n|n) = \hat{\epsilon}(n|n-1) + K(n)(R(n) - (\hat{\epsilon}(n|n-1)) \tag{15}$$

$$P(n|n) = [1-K(n)]P(n|n-1) \tag{16}$$

Kalman gain is given by

$$K(n) = P(n)[P(n)+1]^{-1} \tag{17}$$

Where 'K' is Kalman gain, P(n|n-1) and P(n|n) prior and posterior MSE $'\epsilon'$ normalized CFO to be estimated.

4. B) SIGMA POINTED KALMAN FILTER:

The gain and error variance from (9) and (10) are related as follows:

Let $\rho = 1$ be the length of the state vector.

$$P_i(n+1|n) = P_i(n|n) + \sigma_w^2 \tag{18}$$

$$P_i(n|n) = P_i(n|n-1) - K_i(n)\, P_i^r(n)\, K_i^*(n) \tag{19}$$

Where

$$K_i(n) = \{P_i^{er}(n)\}\{P_i^r(n)\}^{-1} \tag{20}$$

K_i is the gain of the UKF filter, $P_i^{er}(n)$ is the covariance between the state prediction error and innovation.

$$\breve{r}_i(n) = r_i(n) - \hat{r}_i(n), \tag{21}$$

where $\hat{r}_i(n) = \hat{R}_i(n)$

$$X_i(n|n-1) = \left[X_{0,i}(n|n-1)\;\; X_{1,i}(n|n-1)\;\; X_{2,i}(n|n-1)\right]$$

$$= \left[\hat{\epsilon}_i(n-1)\;\; \hat{\epsilon}_i(n-1) + \mu\sqrt{P_i}(n-1)\;\; \hat{\epsilon}_i(n-1) - \mu\sqrt{P_i}(n-1)\right] \tag{22}$$

Where $\mu = \sqrt{1+k}$, $k = \alpha^2 - 1$ and $10^{-3} \leq \alpha \leq 1$

$$\hat{\epsilon}_i(n|n-1) = \sum_{n=0}^{2} w_n X_{n,i}(n|n-1) \tag{23}$$

Where $w_0 = k/(1-k)$ and $w_n = 1/[2(1+k)]$

The sigma points are propagated to nonlinear function eq. (11) and rewritten as follows

$$Y_i(n|n-1) = f\{X_i(n|n-1)\} \tag{24}$$

The measurement update in time:

$$\hat{r}_i(n) = \sum_{n=0}^{2} w_n\, Y_{n,i}(n|n-1) \tag{25}$$

The sigma points are promulgated to nonlinear function(eq(11))is rewritten as follows

$$Y_i(n|n-1) = f\{X_i(n|n-1)\} \tag{26}$$

The measurement update in time:

$$\hat{r}_i(n) = \sum_{n=0}^{2} w_n\, Y_{n,i}(n|n-1) \tag{27}$$

$$\hat{\epsilon}_i(n|n) = \hat{\epsilon}_i(n|n-1) + Re(K_i(n)\{r_i(n) - \hat{r}_i(n)\}) \tag{28}$$

5. SIMULATION

The SCM-OFDM system is simulated with 12 subcarriers,

N_g =128(the number of a cyclic prefix). Carrier frequency f_c=2.5GHz and channel bandwidth B=10MHz. For simulation normalized CFO due to Doppler Effect and user movement is considered. For simulation Rayleigh channel is considered, BPSK is used to modulate the information bits in each of the SCM layers. We assume the CFO and channel estimation have been performed on the preamble. The sigma pointed Kalman filter aims at estimating the CFOs for the SCM-OFDM symbol. We carried out the testing of the SCM-OFDM system for estimating CFO with SPKF and EKF.

The results showed in figure 3, Sigma pointed Kalman filter is tracking the CFO better than the Extended Kalman filter. According to figure 1. Sigma pointed Kalman filter is accurately estimating the CFO than the Extended Kalman filter.

The percentage of error in estimating CFO using EKF is 11.01% whereas using SPKF is 0.24%. There is a significant reduction in error. Hence proposed novel scheme proves to be a powerful scheme to increase throughput by reducing BER.

The CFO effect on OFDM and SCM-OFDM without Kalman filters estimation is shown in figure 2. The result shows that the performance of the OFDM system is degraded. At the SNR 5dB using SPKF 4.76% of BER reduction is achieved on the SCM-OFDM scheme. Hence data rate is also increased by an amount of 1.14% compared to using EKF on the SCM-OFDM system.

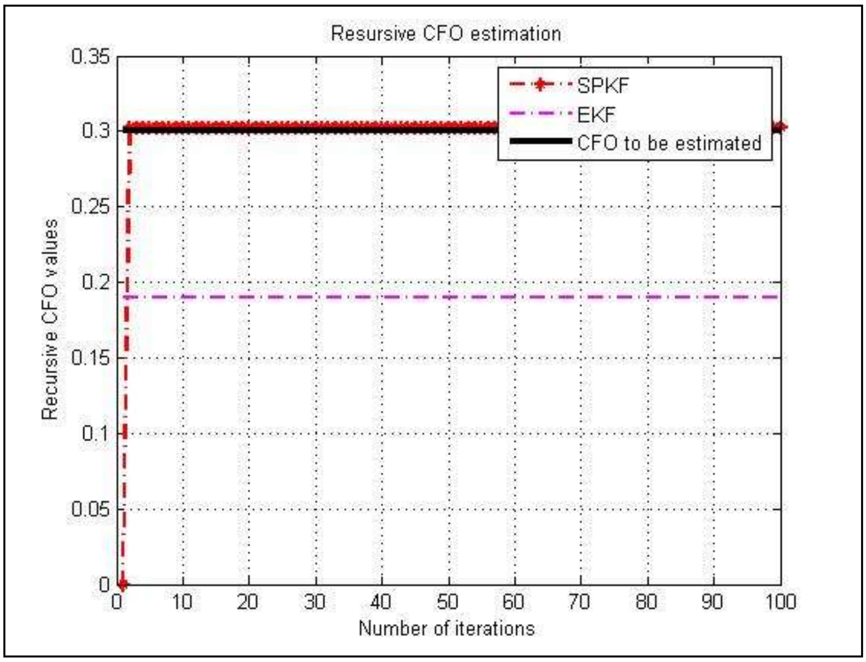


Fig. 1   Comparison of Sigma point Kalman Filter and Extended Kalman Filter for CFO estimation over normalized CFO =0.3 in the SCM-OFDM system.

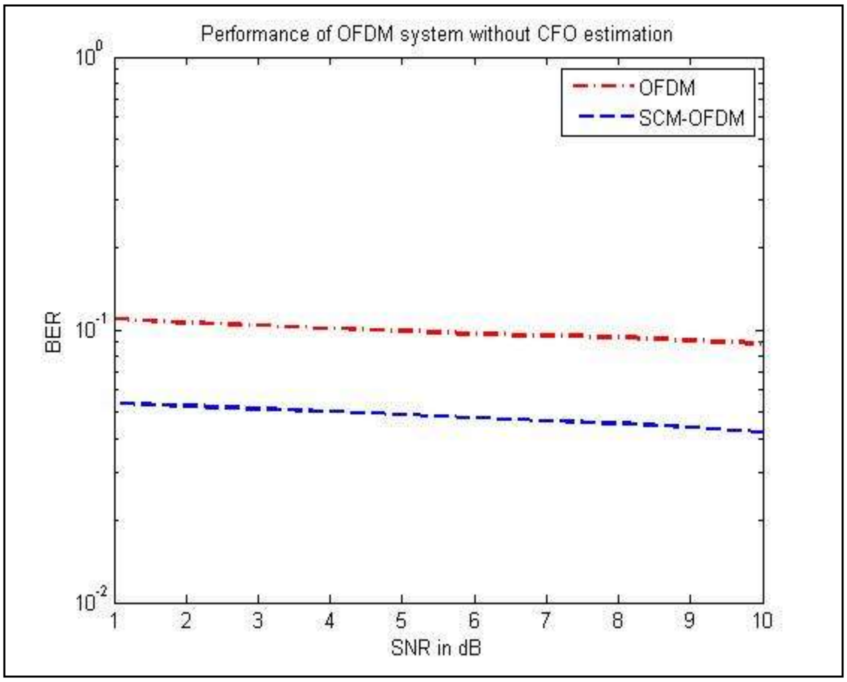


Fig. 2   Performance of OFDM and SCM-OFDM system without CFO estimation

Table I. Error variance for EKF and SPKF with SCM-OFDM for CFO estimation

E_b/N_o (dB)	Theoretical BER	BER (EKF)	BER (SPKF)	% A
0	1.464X 10^{-1}	0.769X10^{-1}	1.05628X 10^{-1}	40
5	6.418X10^{-2}	6.96X10^{-2}	6.23X10^{-2}	18.8
10	2.327X10^{-2}	5.49X10^{-2}	4.3X10^{-2}	200

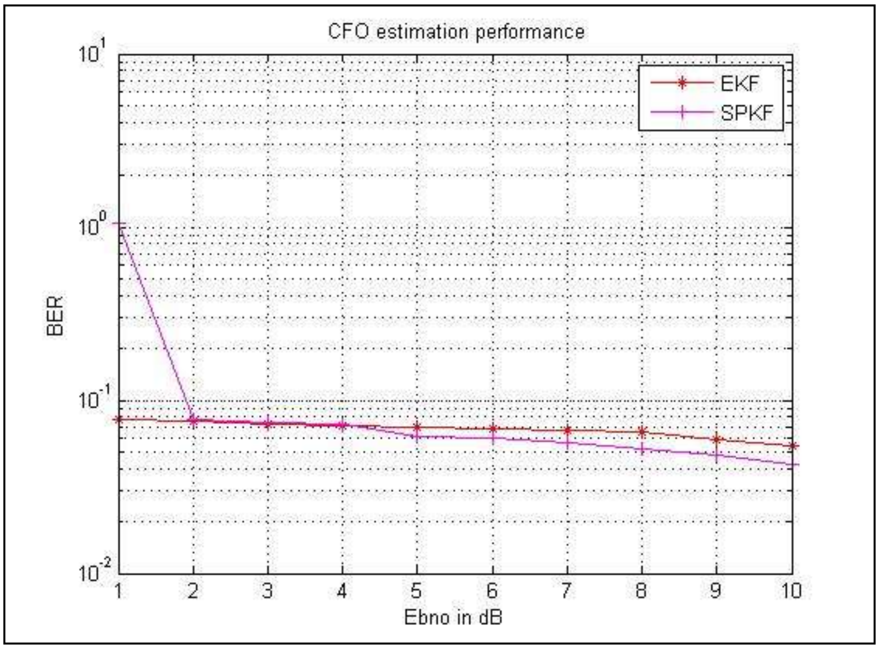


Fig. 3 BER performance using SPKF and EKF in the SCM-OFDM system

Figure 3 represents the SNR Vs BER performance of EKF and SPKF for normalized CFO estimation and correction for the SCM-OFDM system. The SPKF outperforms the EKF in BER reduction and thereby increasing the data rate and throughput by 1.14% compared to using EKF for estimating CFO with SCM-OFDM system

Table I presents the error variance of EKF and UKF, column **A** refers to the difference between theoretical BER and BER using SPKF and is defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(s_i - \hat{s})^2 \qquad (29)$$

S_i refers to the estimation at each step and $\hat{s}$ is the true value. From Table I it is evident that sigma point Kalman filter performance is superior with the smallest estimation error.

6. CONCLUSION

The Superposition Coded Modulation –OFDM system is simulated with sigma pointed Kalman filter for estimating normalized CFO the results show that compared to the Extended Kalman filter, the performance of SPKF improved under time-varying channel with Doppler Effect. SPKF is suitable for the nonlinear environment. There is an improvement in data rate due to SPKF compare to EKF. The proposed scheme estimates the CFO nonlinearities and corrects the same. CFO estimation for SCM-OFDM system is imperative for the next generation communication technologies. As the next generation technology demands high data rates with a better quality of service.

Future scope: The performance of the SCM-OFDM system will be compared with the MIMO-OFDM system.

REFERENCES

[1] Liu, L., Liang, C., Ma, J., & Ping, L. (2019). Capacity optimality of AMP in coded systems. arXiv preprint arXiv:1901.09559.
[2] Damrath, M., Hoeher, P. A., & Forkel, G. J. (2018). Piecewise linear detection for direct superposition modulation. Digital Communications and Networks, 4(2), 98-105.
[3] Osamura, A., Song, G., Cheng, J., & Cai, K. (2018, October). Sparse Multiple Access and Code Design with Near Channel Capacity Performance. In 2018 International Symposium on Information Theory and Its Applications (ISITA) (pp. 139-143). IEEE.
[4] Li, S., Bai, B., Zhou, J., He, Q., & Li, Q. (2018). Superposition coded modulation based faster-than-Nyquist signaling. Wireless Communications and Mobile Computing, 2018.
[5] Chi, Y., Liu, L., Guo, J., Song, G., Yuen, C., & Guan, Y. L. (2019). Variable-Rate Coding With Constant BER for NOMA via Multilevel IRA Coding. IEEE Transactions on Vehicular Technology, 68(5), 5149-5153.
[6] Cho, H., Nguyen, T. H., Nguyen, H. N., Choi, H. K., Choi, J. K., & Ro, S. (2018). A robust ICI suppression based on an adaptive equalizer for very fast time-varying channels in LTE-R systems. EURASIP Journal on Wireless Communications and Networking, 2018(1), 17.
[7] Vahidi, V., & Saberinia, E. (2018). ICI mitigation for high-speed OFDM communications in high-mobility vehicular channels. In Information Technology-New Generations (pp. 17-22). Springer, Cham.
[8] Lin, B., Tang, X., Ghassemlooy, Z., Li, Y., Zhang, M., Wu, Y., & Li, H. (2019). A NOMA scheme for visible light communications with single carrier transmission and frequency-domain successive interference cancellation. Optik, 183, 445-450.
[9] Wang, Y. Y., Huang, C. W., & Chen, W. W. (2018). Maximum likelihood carrier frequency offset estimation algorithm with adjustable frequency acquisition region. Journal of the Franklin Institute, 355(5), 2978-2985.
[10] Singh, S., Kumar, P., & Prasad, S. V. A. V. (2019). Improvement of BER in Wavelet Transform-based OFDM System with Kalman Filter. Journal of Communication Engineering & Systems, 4(1), 21-28.
[11] Zhang, X., Zhang, L., Xiao, P., & Wei, J. (2019). Fine Timing Synchronization Based on Modified Expectation Maximization Clustering Algorithm for OFDM Systems. IEEE Wireless Communications Letters.
[12] Malarvezhi, P., & Kumar, R. (2018). A Diversity Enhanced-Particle Filter based Multiple Carrier Frequency Offset Estimation Using Hadamard Matching in CA-OFDM Based LTE-A System. Wireless Personal Communications, 101(1), 87-99.
[13] Huu, N. N., Quoc, H. P., Van, L. P., Nguyen, T. H., Cho, H., Nguyen, V. D., & Jeon, B. (2018, January). An OFDM System Design with Efficient ICI Cancellation over Embedded LTE-R Channels Based on the USRP Platform. In Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication (p. 10). ACM.
[14] Liao, Y., Shen, X., Sun, G., Dai, X., & Wan, S. (2019). EKF/UKF-based channel estimation for robust and reliable communications in V2V and IIoT. EURASIP Journal on Wireless Communications and Networking, 2019(1), 144.
[15] Shen, Z., Xu, K., Wang, Y., & Xie, W. (2018, October). Angle-Domain Channel Tracking for High-Speed Railway Communications with Massive ULA. In 2018 IEEE 18th International Conference on Communication Technology (ICCT) (pp. 159-165). IEEE.
[16] Rashmi, N., & Sarvagya, M. (2016, March). Channel estimation using Extended Kalman filter for superposition coded modulation system. In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 3482-3486). IEEE.
[17] Rashmi, N., & Sarvagya, M. (2016). State Space Model-Based Channel Estimation using Extended Kalman Filter for Superposition Coded Modulation OFDM System. BVICA M's International Journal of Information Technology, 8(2), 1017.
[18] D J Costello and G D Forney "Channel coding: The Road to channel capacity," Proc. IEEE, vol.95, no.6, June 2007, pp.1150-77.
[19] L Duan, B. Rimoldi, and R. Urbanke, "Approaching the AWGN channel capacity without active shaping," in Proc. IEEE int. Symp.Information theory (ISIT'97), Ulm, Germany, Jun/Jul 1997, p.374.

Journal of Non-Crystalline Solids 544 (2020) 120196

Contents lists available at ScienceDirect

Journal of Non-Crystalline Solids

journal homepage: www.elsevier.com/locate/jnoncrysol

Electrical switching and crystalline peak studies on $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 7$) chalcogenide bulk glasses

Jagannatha K.B.[a], Diptoshi Roy[b], Chandasree Das[b,*]

[a] Dept. of Electronics and Communication, BMS Institute of Technology and Management, Bangalore-64, India
[b] Dept. of Electrical and Electronics, BMS College of Engineering, Bangalore-19, India

ARTICLE INFO

Keywords:
Chalcogenide glasses
Electrical switching
Crystalline peak study

ABSTRACT

The $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 7$) glasses are synthesized using melt quenching technique. Result of X-ray diffraction (XRD) exhibits the non-crystallinity of the prepared samples. Depending on applied on state current, the synthesized $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 7$) samples show both threshold and memory type of behavior. The composition dependence of threshold voltage (V_{th}) displays a decrease in the V_{th} values. Further the composition and morphological studies are carried out using Energy Dispersive X-ray analysis (EDAX) and Electron microscopy (SEM). The differential scanning calorimetry (DSC) is used for crystalline peak studies. Using DSC, the crystallization temperature (T_c) of representative glass samples are found. The crystalline peak studies are carried out by annealing representative samples at their respective T_c for two hours to find different phases present in the samples. The SET-RESET studies are performed to find the applicability of the given material in phase change memory (PCM) applications.

1. Introduction

At the beginning of 1968, Ovshinsky noticed a reversible electrical switching process in chalcogenide glasses (ChG) [1]. The ChG are suitable for many applications such as waveguides, energy, photonic, switches, thermoelectric, sensors, media and phase change memory (PCM) devices [2–9] . PCM has unique property which represents the logic '1' in crystalline phase and logic '0' in amorphous phase. The group 16 element of periodic table namely selenium, sulphur and tellurium (Te) are used for preparing ChG. The advantages of PCM over flash technologies are high durability, less read/write time, better scalability, non-volatile, high retention rate and less expensive [10]. The PCM materials shows high resistance when it is in amorphous state, and the material state can be transformed into low resistance by applying a suitable electric field. The logic '1' and logic '0' are considered as crystalline and amorphous phase respectively in PCM application [11]. From past three decades most of the researchers are working on Te-based or metal doped Ge-Te ChG such as Ge-Te, As-Te, Ga-Te, Si-Te, In-Te, Ge-Te [12–16] and Ge-Te-Sn, Ge-Te-Ag, Ge-Te-Cu and Ge-As-Te [17–20] respectively for their potential use in memory applications. Meanwhile, numerous research are carried out on binary and ternary silicon (Si) based ChG such as, Si-Te, Si-Te-Bi, Si-Te-Cu, Si-Te-Ge, As-Te-Si, Si-Te-Ge, Si-Te-In, Si-Te-Sb [21–28], which are found to be memory switching chalcogenides. Two types of switching are possible in ChG, (i) Threshold switching and (ii) Memory switching. Memory type of electrical switching is a reversible process, which comprises a morphological change (amorphous-crystalline); samples retain its original phase even after removing applied electric field; whereas, the threshold switching mechanism is irreversible and does not contain a morphological changeover even after the removing of the applied electric field [1, 29]. Although the electrical switching trials of PCM have been underway over past few decades, still flash technology is dominating the semiconductor industry [30]. The scientists are working to replace flash technology with PCM technology which is well-suitable with CMOS technology. The investigation of switching mechanisms has wide scope to understand and to identify the materials that might be suitable for PCM applications using switching studies [9]. In this current work, for the composition $Si_{20}Te_{80-x}Sn_x$, I-V characteristics, SET-RESET and structural studies are carried out over the range ($1 \leq x \leq 7$) to determine the materials suitability for PCM applications. Studies are also undertaken in thermally crystallized samples to interpret the effect of Sn doping on switching voltages and crystalline phases.

2. Experimental details

Conventional melt quenching method is used for preparing $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 7$) samples. 99.9999 percent pure Si, Sn and Te elements are measured in a suitable quantity via a balance meter and loaded in

* Corresponding author.
E-mail address: chandasreedas.eee@bmsce.ac.in (C. Das).

https://doi.org/10.1016/j.jnoncrysol.2020.120196
Received 27 January 2020; Received in revised form 15 May 2020; Accepted 27 May 2020

quartz ampoules those are washed with soap solution, water and acetone. The samples containing quartz ampoules are sealed at 10^{-6} mbar. The closed ampoules are positioned in the centre of a rotating furnace at the elevated temperature and the furnace temperature is augmented by 100˚C each 60 minutes until it hits 1100 ˚C. For absolute homogenization of melt, each ampoule is kept within the furnace in rotating condition at 1100 ˚C for 24 hours after reaching optimal temperature. In a bucket of water which is ice-cold (< 0 ˚C) and sodium hydroxide (NaOH) blend, the ampoules are quenched to obtain the bulk ingots of the samples. The XRD (PANalytical X'Pert3 Powder X-ray diffractometer) proves the amorphous nature of bulk ingots. The bulk ingots are treated with sandpaper to obtain a thickness of ~ 0.3 mm to carry out electrical characterization and the samples are placed between plane plate bottom electrode and a top electrode of a custom-built probe station capable of moving in X, Y, Z directions. Using a Keithley 2410[c] source measurement unit, which is programmable and controlled by LabVIEW 7, the current-voltage (I-V) characterization is executed. The source measurement unit is capable of supplying a peak voltage of 1100 V and current between 50 pA and 1.05 A. For each sample, I–V characteristics were reiterated for four times and characteristics are found to be reproducible within ± 5V. Typical error in the measurement of switching voltages is within ± 5 V. Errors are estimated using the method of standard deviation and the corresponding error bars are included in the figures. Using VEGA3 TESCAN, scanning electron microscopy (SEM) and Energy Dispersive X-ray analysis (EDAX) experiments are carried out to monitor morphological modifications in the samples during switching and elemental composition of materials. Mettler Toledo Differential Scanning Calorimeter is used to perform DSC studies on representative $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) bulk specimens. The crystallization peak studies are carried out on representative $Si_{20}Te_{79}Sn_1$, $Si_{20}Te_{77}Sn_3$ and $Si_{20}Te_{75}Sn_5$ samples. These samples are sealed under vacuum and annealed for two hours at their respective T_c. Then the XRD studies are performed on the annealed samples to comprehend the crystalline phases present in these samples.

3. Results

3.1. XRD and EDAX analysis of the $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) glasses

The XRD of the $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) bulk glasses are shown in Fig. 1. It is observed from Fig. 1 that there are no sharp peaks visible and hence, it confirms the amorphous nature of as prepared samples.

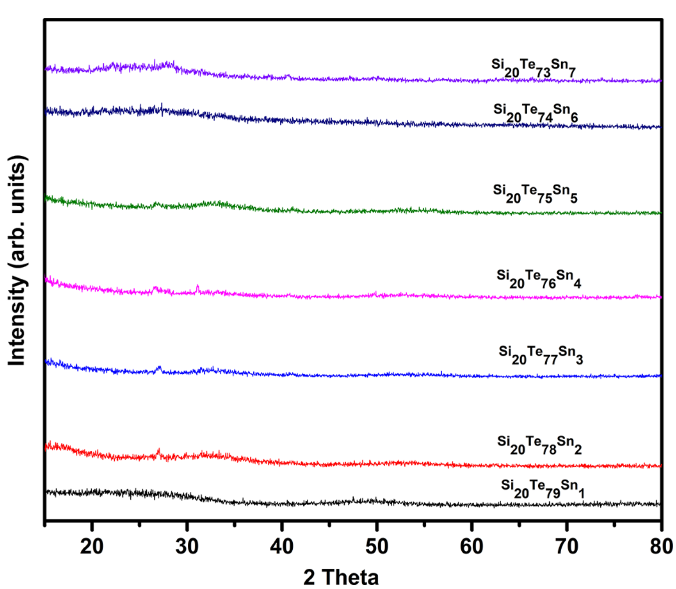


Fig. 1. X-ray diffraction spectrum of $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) glasses

EDAX analysis is carried out to find the stochiometric ratio of the quenched samples which confirms the existence of all constituent elements and, they closely match with their stoichiometric values. Fig. 2 and Table 1 shows the EDAX pattern and composition of the different constituent elements in as-prepared bulk samples. It can be seen that the sample composition closely matches to the composition of the original material, $Si_{20}Te_{79}Sn_1$ bulk sample.

3.2. Switching studies in $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) bulk samples

The mechanism of memory and threshold switching depends on local atomic structures, thermal stability and elemental composition of materials [31]. ChG memory switching involves local crystallization of material between two electrodes. Using the Warren's electrothermal mechanism, the memory switching mechanism is explained [32] and it is proposed that the current rises abruptly during memory switching when electric field crosses threshold voltage (V_{th}), suggesting filling of defects in the mobility gap of the glasses. Long Te-Te chains in structure and rapid atomic reorganization accelerate the memory switching [33]. The higher electrical conductivity causes the power dissipation during memory switching and the materials remains in low resistance state even after electric field withdrawal, owing to crystallization of the materials [34]. When charged defect states in material are filled by field injected charge carriers, the threshold switch is achieved. The threshold switching has higher structural cross-linking, difficulty in structural reorganization and a greater energy barrier for crystallization [35, 36]. In this case the high resistive off state is retained even after the withdrawal of electric field. The electrical switching studies are carried out on the $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) glasses. Fig. 3 displays memory and threshold switching characteristics of representative $Si_{20}Te_{79}Sn_1$, $Si_{20}Te_{77}Sn_3$ and $Si_{20}Te_{74}Sn_6$ bulk samples. As seen from Fig. 3(a) samples are originally at amorphous state (high-resistance OFF state), but after crossing V_{th}, the samples switch sharply to crystalline state (low-resistance ON state). The samples do not return to high resistance phase even after purging of applied electric field showing memory type of electrical switching behavior. Fig. 3(b) displays the return of the sample to the initial high resistance phase after removal of the electric field, which reflects threshold switching behavior by the samples. The switching experiments are conducted on four to five specimens of each composition of $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) bulk glasses showing both memory and threshold type switching with different current values. In every case switching voltages are reproducible within a voltage difference of ± 4%. The samples are found to display memory type switching with the currents of the order of 4 mA and above, and threshold type of switching is observed with less than 4 mA of current. Based on the percentage of Sn in Si-Te composition, the threshold voltage range is found to vary from 40 - 155 V in the I-V characteristics of $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) bulk samples for memory switching. An interesting observation can be made here is that, all the samples exhibited threshold tendency at lower currents (< 4 mA) for few cycles, but as the on-state current is increased the sample latches on to the memory state. Similar memory type switching characteristics are reported in Ge-Te-Si [24], As-Te-Si [37] and Si-Te-Sb [38] also. The easy crystallization and greater electrical conductance are the reasons for good memory switching in Si-Te glasses [39]. The $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) bulk samples contain a large Te chain with Si which is favorable for memory switching and therefore the obtained memory switching behavior of these glasses matches with other Si-Te glasses like Si-Te-Sb, and Si-Te-Bi [19 22, 28].

3.3. Compositional dependence of switching voltages of $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) glasses

Fig. 4 displays the composition dependence of V_{th} for $Si_{20}Te_{80\text{-}x}Sn_x$ (1 ≤ x ≤ 7) glasses. It has been seen that the V_{th} (for memory switching) and the OFF state electrical resistance of the samples

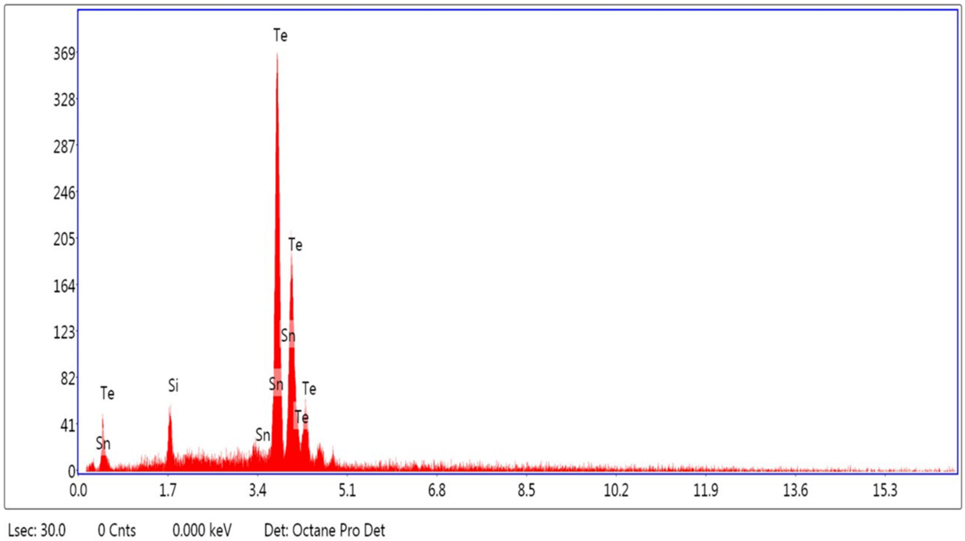


Fig. 2. EDAX pattern of representative $Si_{20}Te_{79}Sn_1$ bulk sample

Table 1
EDAX report of $Si_{20}Te_{79}Sn_1$ representative bulk sample

Element	Weight %	Atomic %
SiK	4.81	18.65
SnL	1.34	1.23
TeL	93.85	80.12

decreases with increase in Sn concentration [29]. The composition dependence of V_{th} in ChG depends on three factors, namely metallicity [40], chemical ordering [19], network connectivity and rigidity [41–43]. The inclusion of metallic dopants contributes to the decrease in V_{th} of ChG, as Sn is more metallic compared to Te ($\rho Te = 10^4$ μΩ-cm and $\rho Sn = 11.5$ μΩ-cm) [44] and replacing Te with Sn results in a decrease in the V_{th} values. A similar trend in compositional dependence of switching voltage and off state resistance has been seen in Ge-Te-Sn, Ge-Te-Bi, Si-Te-Bi, Si-Te-Sb [22, 38, 44–46] glasses. Further Fig. 5 demonstrates the dependence of threshold voltage with composition for threshold switching. From Fig. 5 a decrease of V_{th} with increasing Sn concentration is clearly been observed. Normally, the addition of metallic dopants decreases the V_{th} to a greater extent as seen in As–Te–Cu, Ge–Se–Tl, As–Te–Tl, and Ge–Se–Sb [34, 35, 38, 47] glasses. The addition of Sn to Si-Te matrix, results in a decrease in the value of V_{th} as Sn resistivity is lower than Te [17]. For a ChG, adding metallic dopants with four-fold or higher coordination elements like In, Bi etc. increases the connectivity and rigidity of the network, resulting in an increase in V_{th} value [19, 48]. Similarly as Sn is a metallic dopant and it is having higher coordination number compared to Te (coordination number of Sn = 4 and coordination number of Te = 2) [49], is likely to increase the rigidity and network connectivity when doped in Si-Te matrix, and subsequently increase the V_{th} in $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 7$) bulk glasses. However, the Fig. 4 shows a decrease in V_{th} and OFF state resistance with addition of Sn. Further the conductivity of a sample increases as resistance decreases, and as the current flowing through the sample increases, temperature across the channel increases, favouring memory switching phenomenon. Based on these findings, it can be inferred that composition dependence of V_{th} and off state resistance of $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 7$) bulk samples are largely due to the metallicity of Sn dopant and not for network connectivity. The decrease in V_{th} and Off state resistance are also observed in Si-Te-Bi glasses [39, 45]. Typical error in the measurement of compositional dependence on switching voltages is within ± 5V.

3.4. Crystalline peak studies

The DSC has widely been utilized in enterprises and research for the determination of glass transition and crystallization temperature, phase changes, heat capacity, and melting point etc. [50]. DSC is one of the

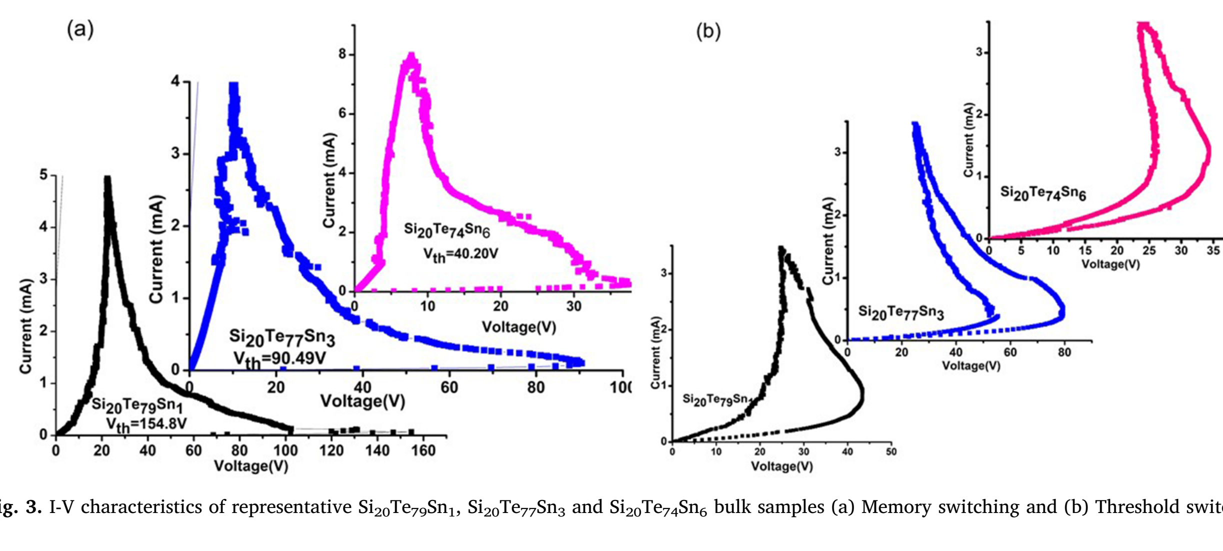


Fig. 3. I-V characteristics of representative $Si_{20}Te_{79}Sn_1$, $Si_{20}Te_{77}Sn_3$ and $Si_{20}Te_{74}Sn_6$ bulk samples (a) Memory switching and (b) Threshold switching

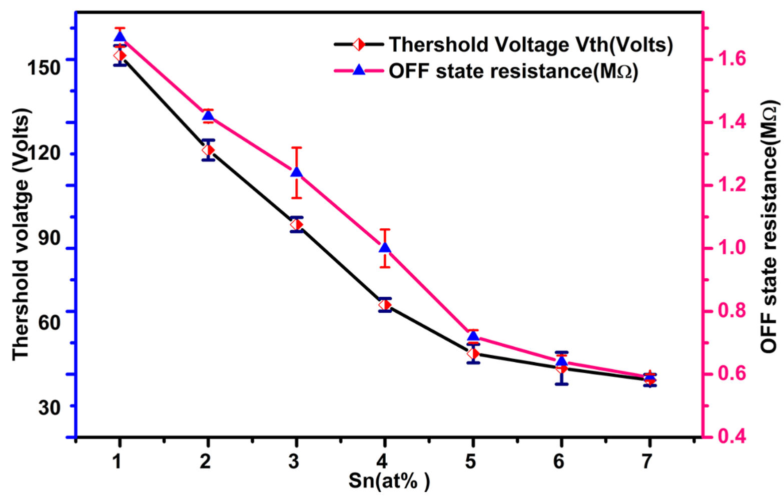


Fig. 4. Composition dependence of threshold voltage (V_{th}) and OFF State electrical resistance of $Si_{20}Te_{80\text{-}x}Sn_x$ ($1 \leq x \leq 7$) bulk samples for memory switching.

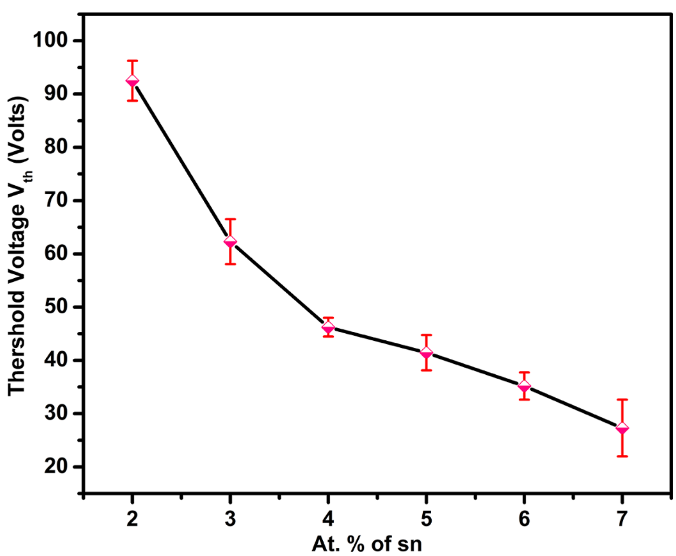


Fig. 5. Composition dependence of threshold voltage (V_{th}) of $Si_{20}Te_{80\text{-}x}Sn_x$ ($2 \leq x \leq 7$) bulk samples for threshold switching.

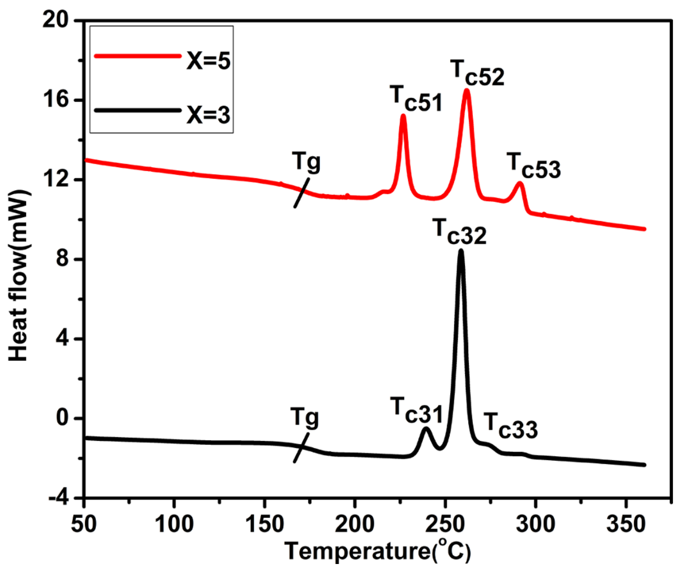


Fig. 6. The DSC thermogram of representative of $Si_{20}Te_{80\text{-}x}Sn_x$ bulk samples with different Sn concentration(x).

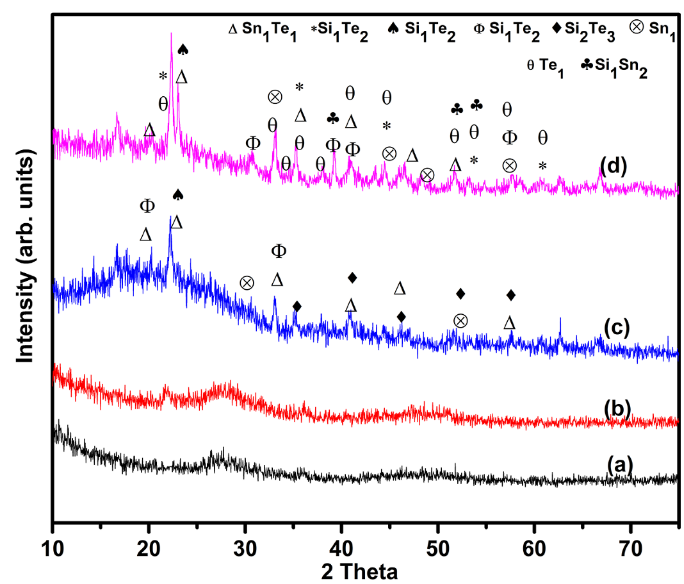


Fig. 7. XRD of representative (a) as-quenched $Si_{20}Te_{77}Sn_3$ (b) as-quenched $Si_{20}Te_{75}Sn_5$ samples showing amorphous nature (c) $Si_{20}Te_{77}Sn_3$ at 260 °C (T_{c32}) and (d) $Si_{20}Te_{75}Sn_5$ at 227 °C (T_{c51}) annealed for two hours.

calorimetric methods in which the warmth stream rate to the sample is checked while the temperature of the sample is varied in a predefined environment at a given rate. Fig. 6 shows the DSC thermogram of representative $Si_{20}Te_{80\text{-}x}Sn_x$ (x = 3,5) bulk samples. From Fig. 6 it is observed that DSC thermograms have one endothermic glass transition temperature (T_g) and three exothermic crystallization temperature T_c (T_{c51}, T_{c52} and T_{c53} for x = 5), there is no well-defined T_{c33} for x = 3. Similar results were observed in Si-Te-Ag glasses also [26]. To find out the crystalline phases present in $Si_{20}Te_{80\text{-}x}Sn_x$ (x = 3,5) glasses, the samples are annealed at their respective crystalline temperatures T_{c32} (260 °C) and T_{c51} (227 °C) for two hours in ampoules which are sealed at 10^{-6} mbar. Fig. 7 display the XRD images of as prepared $Si_{20}Te_{77}Sn_3$ and $Si_{20}Te_{75}Sn_5$ glasses which confirm the amorphous nature of the samples and representative annealed $Si_{20}Te_{77}Sn_3$ and $Si_{20}Te_{75}Sn_5$ bulk samples.

XRD studies on annealed $Si_{20}Te_{77}Sn_3$ and $Si_{20}Te_{75}Sn_5$ bulk samples divulge the presence of hexagonal Si_1Te_2, hexagonal Si_2Te_3, hexagonal Te_1 and cubic Sn_1Te_2. There is no finding of ternary phases of Sn. Majority of the binary phase of Si-Te in Si-Te-Sn samples indicates that Sn is not actively taking part in network connectivity to a greater extent, similar results are seen in Si Te Bi [51] Samples.

Sn can form a structural unit in the system as Sn-Sn matrix, if it is appended to Si-Te network and Sn can remain in parent Si-Te glass matrix as micro-inclusion. However, Sn can interact with the host matrix Si-Te and enhance the glass forming ability and network connectivity [52]. As Si and Te are 4-fold and 2-fold coordinated elements respectively, in the host matrix ($Si_{20}Te_{80}$) hence Si atoms bonds well with Te atoms and chalcogenide glassy network is formed primarily by the Si–Te linkages [17, 39, 53] and the residual Te atoms form a network among themselves (Te-Te). Though adding Sn atoms to Si-Te networks is expected to change the network, as Sn can bond with Te or Si as well as with the Si–Te, but Fig. 7 illustrates that Sn atoms are not bonded with the Si–Te network. This confirms that there is no covalent bonding of Sn with the Si-Te host matrix; hence network connectivity is not enhanced due to Sn addition. Similar trend in bonding is seen in Ge-Te-Sn glasses [44]. For the representative $Si_{20}Te_{79}Sn_1$, $Si_{20}Te_{77}Sn_3$ and $Si_{20}Te_{75}Sn_5$ samples, Sn-Sn, Te-Te, Si-Si, Sn-Te, and Si–Te bonds are formed according to bond energy considerations [54]. The bond energies (D, kJ/mol) are in the following order; D (Sn–Sn) = 187.1 ± 0.3; D (Te–Te) = 257.6 ± 4.1; D (Si-Si) = 325 ± 7; D (Sn–Te) = 359.8; D (Si–Te) = 429.2. From Fig. 6, which represents the DSC thermogram it is evident that, as doping of Sn increases the crystallization temperature decreases. This result, corelates with the composition dependence of switching voltages in which, as the doping concentration increases the threshold switching voltage decreases.

3.5. Set-reset studies in $Si_{20}Te_{80-x}Sn_x$ ($1 \le x \le 7$) bulk samples

The chalcogenide glasses switch quickly and easily between non-crystalline to crystalline state and vice versa under the submission of suitable current pulse. When the applied current pulse provides the required amount of heat for crystallization, then the phase transformation takes place from amorphous to crystalline which is denoted as SET process. The SET process is usually attained by applying a long-term current pulse with less amplitude whereas a higher and sharper current pulse is used to achieve the RESET process [38]. A triangular current pulse with a 2.5 mA pulse is applied for SET operation and a rectangular pulse 10 mA with 10 ms duration to the RESET operation for $Si_{20}Te_{80-x}Sn_x$ ($1 \le x \le 7$) glasses. This process is carried out for 5-20 cycles, and all samples of the series have been found to exhibit SET-RESET repeatedly. The SET–RESET cycles of a representative $Si_{20}Te_{77}Sn_3$ sample are shown in Fig. 8. During set operation the controlled slow cooling of materials take place due to the gradual reduction in current with negative ramp pulse [55]. The RESET operation is achieved using 10 mA rectangular pulses for a short duration of 10 ms.

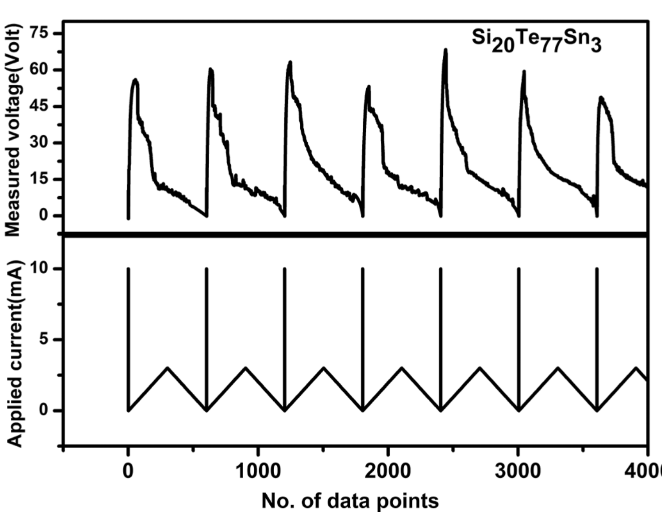


Fig. 8. SET-RESET operations of representative $Si_{20}Te_{77}Sn_3$ bulk sample.

During RESET operation current dissipates heat quickly to the surrounding areas of the conducting channel of the sample which results in re-amorphization of the samples [56]

3.6. SEM analysis of $Si_{20}Te_{80-x}Sn_x$ ($1 \le x \le 7$) glasses

To observe the structural changes during switching, SEM analysis is performed on a representative $Si_{20}Te_{78}Sn_2$ sample. The memory switching which has thermal origin is performed by passing current through the sample in the range of 1mA-10mA. The applied current provides an energy for crystallization and a conducting channel is formed between top and bottom electrode, which leads material to switch from non-crystalline to crystalline phase upon application of an electric field [57]. Morphological studies on memory switching (before and after switching) are performed on a representative sample to confirm structural changes occurred during switching. Fig. 9(a) and (b) displays the SEM images of unswitched and switched region of $Si_{20}Te_{78}Sn_2$ sample respectively. It is clearly seen that the unswitched $Si_{20}Te_{78}Sn_2$ film is smooth with no structural changes. On the other hand, a morphological variation in SEM picture of switched sample indicates the phase transformation in the electrode region. The image distinction in SEM can be attributed to any of the subsequent factors [58].

i) In ChG, particularly in memory switching, the phenomenon is related with the local melting and solidification into crystalline state. This process can result in a relief of the surface that create an image contrast.
ii) Amorphous to crystalline transition is correlated with spatial topography changes and compaction in the electrode area as the crystalline state becomes denser than non-crystalline state. The difference in density of the switched area may subsidize to a distinct image.
iii) In the electrode region there is a wide variation in the resistance of the material. Changes in the material conductivity can also be reason for SEM image contrast.

For $Si_{20}Te_{78}Sn_2$ switched sample, detailed analysis is necessary to understand the factors accountable for the changes in SEM images. However, similar results have been obtained earlier during electrical switching process [39, 58, 59]. It is evident from the pictures that, the sample has undergone structural phase transformation during switching. The SEM analysis is further extended by performing switching at the edge of a representative sample to evident the creation of crystalline channel like structure in the sample. The channel like formation has been observed in the SEM image of the sample in which switching has been carried out at the edge of the sample. Fig. 9(c) displays formation of channel like structure between two ends of the electrodes after conducting switching operation at the edge of the sample. Detail analysis is not done on the formation of crystalline channel in the representative sample.

4. Conclusions

The conventional melt quenching technique is used for the preparation of $Si_{20}Te_{80-x}Sn_x$ ($1 \le x \le 7$) bulk glasses. The prepared samples have revealed memory type of switching with the current of the order of 4mA and above, on the other hand threshold switching is observed with less than 4 mA of current. The decrease in V_{th} and off state resistance with increase in Sn concentration is explained on the basis of metallicity factor. Crystalline peak studies are performed to understand the phases present in the sample after memory switching and the study of annealed samples at respective crystallization temperature reveals the presence of only Si-Te, Sn-Si, and Te-Sn phases and no ternary phases of Si-Te-Sn is observed. This confirms that there is no covalent bonding of tin with the Si-Te host matrix; hence network

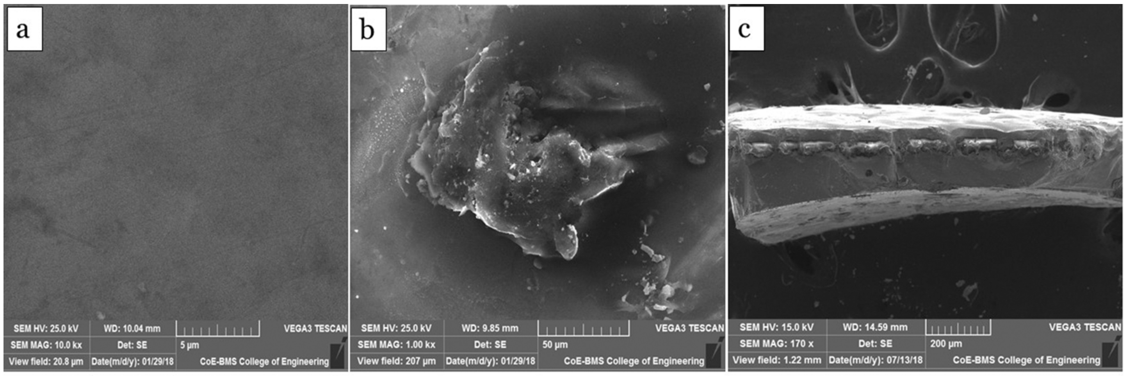


Fig. 9. SEM images of representative $Si_{20}Te_{78}Sn_2$ glassy sample as (a) unswitched bulk sample (b) **s**witched bulk sample (c) switching undertaken at the edge of the sample.

connectivity is not enhanced due to Sn addition. Switched and unswitched images of SEM, clearly demonstrate the structural variations at the place of switching. The change in phase from amorphous to crystalline and from crystalline to amorphous has been observed in SET-RESET process for few cycles on $Si_{20}Te_{80-x}Sn_x$ ($1 \le x \le 7$) bulk glasses which could lead to its application as PCM device.

Credit author statement

All authors have contributed equally in the research work from which the results are consolidated in this paper. It is agreed by all authors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Author Chandasree Das acknowledges Science and Engineering Research Board (SERB) India for the financial support to carry out this research work.

References

[1] S.R. Ovshinsky, Reversible electrical switching phenomena in disordered structures, Phys. Rev. Lett (1968).
[2] V.S. Shiryaev, M.F. Churbanov, E.M. Dianov, V.G. Plotnichenko, J.L. Adam, J. Lucas, Recent progress in preparation of chalcogenide As-Se-Te glasses with low impurity content, J. Optoelectron. Adv. Mater. (2005).
[3] Y. Huang, F. Chen, B. Qiao, S. Dai, Q. Nie, X. Zhang, Improved nonlinear optical properties of chalcogenide glasses in Ge-Sn-Se ternary system by thermal treatment, Opt. Mater. Express 6 (2016) 1644.
[4] A. Sharma, N. Mehta, Observation of switching behavior in some multi-component glasses of Se-Te-Sn-Pb system, Mater. Lett 178 (2016) 178–180.
[5] M. and minam Tohge, Electrical properties of n-Type semiconducting chalcogenide glasses in the system Pb-Ge-Se, J. Non. Cryst. Solids (1987) 96.
[6] A.K. Sharma, G.J. Mohr, Plasmonic optical sensor for determination of refractive index of human skin tissues, sensors actuators, B Chem 226 (2016) 312–317.
[7] A. Maignan, E. Guilmeau, F. Gascoin, Y. Bréard, V. Hardy, Revisiting some chalcogenides for thermoelectricity, Sci. Technol. Adv. Mater (2012) 13.
[8] Z. Han, P. Lin, V. Singh, L. Kimerling, J. Hu, K. Richardson, A. Agarwal, D.T.H. Tan, On-chip mid-infrared gas detection using chalcogenide glass waveguide, Appl. Phys. Lett. 108 (2016) 1–4.
[9] M. Wuttig, Phase-change materials: towards a universal memory? Nat. Mater 4 (2005) 265–266.
[10] S. Raoux, W. Wełnic, D. Lelmini, Phase change materials and their application to nonvolatile memories, Chem. Rev. 110 (2010) 240–267.
[11] A. Pirovano, A.L. Lacaita, A. Benvenuti, F. Pellizzer, R. Bez, Electronic Switching in phase-change memories, IEEE Trans. Electron Devices 51 (2004) 452–459.
[12] H. Fritzsche, Amorphous semiconductors for switching, memory, and imaging applications, IEEE Trans. Electron Devices ED-20 (1973) 91–105.
[13] N.F. Mott, Conduction in non-crystalline systems, Philos. Mag 17 (1968) 1259–1268.
[14] C.L. Dargan, P. Burton, S.V. Phillips, A.S. Bloor, P. Nesvadba, Comparison of memory switching operation in a number of amorphous chalcogenide systems, J, Mater. Sci. 9 (1974) 1595–1601.
[15] S. Murugavel, S. Asokan, "Composition tunable memory and threshold switching in $Al_{20}As_xTe_{80-x}$ semiconducting glasses, J. Mater. Res. 13.10 (1998) 2982–2987.
[16] G. Jones, R.A. Collins, Threshold and memory switching in amorphous selenium thin films, Phys. Status Solidi 53 (1979) 339–350.
[17] C. Das, M.G. Mahesha, G.M. Rao, S. Asokan, Electrical switching studies on amorphous Ge-Te-Sn thin films, AIP Conf. Proc. 1349 (2011) 633–634.
[18] K. Ramesh, S. Asokan, K.S. Sangunni, E.S.R. Gopal, Effect of topological thresholds on thermal behaviour of germanium telluride glasses containing metallic additive, Eur. Phys. J. B. 6 (1998) 207–212.
[19] C.N. Murthy, V. Ganesan, S. Asokan, Electrical switching and topological thresholds in Ge-Te and Si-Te glasses, Appl. Phys. A Mater. Sci. Process 81 (2005) 939–942.
[20] B.J. Madhu, H.S. Jayanna, S. Asokan, The composition dependence of electrical switching behavior of $Ge_7Se_{93-x}Sb_x$ glasses, J. Non. Cryst. Solids. 355 (2009) 2630–2633.
[21] K. Ramesh, S. Asokan, K.S. Sangunni, E.S.R. Gopal, Glass formation in germanium telluride glasses containing metallic additives, J. Phys. Chem. Solids. 61 (2000) 95–101.
[22] B.J. Fernandes, K. Ramesh, N.K. Udayashankar, Electrical switching in $Si_{20}Te_{80-x}Bi_x$($0 \le x \le 3$) chalcogenide glassy alloys, J. Non. Cryst. Solids. 483 (2018) 86–93.
[23] D. Roy, C.H. S Nadig, A. Krishnan, A. Karanam, J.K. B, C. Das, C.H. Nadig, Electrical switching studies on Si15Te85-xCux bulk ($1 \le x \le 5$) glasses, (1966) 20033.
[24] C. Das, G.M. Rao, S. Asokan, A comparative study of electrical switching behavior of certain tellurium based chalcogenide thin films for phase change memory (PCM) applications, 3rd Int, Conf. Multi-Funct. Mater. Struct. MFMS 123–125 (2010) (2010) 1207–1210.
[25] M. Anbarasu, S. Asokan, Electrical switching behavior of bulk As–Te–Si glasses: composition dependence and topological effects, Appl. Phys. A 80 (2005) 249–252.
[26] S. Rao Gunti, S. Asokan, Thermodynamic, Raman and electrical switching studies on $Si_{15}Te_{85-x}Ag_x$($4 \le x \le 20$) glasses, J. Appl. Phys (2012) 111.
[27] S.R. Gunti, A. Ayiriveetil, A. Sundarrajan, Thermodynamic, kinetic and electrical switching studies on $Si_{15}Te_{85-x}In_x$glasses: observation of boolchand intermediate phase, J. Solid State Chem 184 (2011) 3345–3352.
[28] R. Lokesh, N.K. Udayashankar, S. Asokan, Electrical switching behavior of bulk $Si_{15}Te_{85-x}Sb_x$ chalcogenide glasses - a study of compositional dependence, J. Non. Cryst. Solids. 356 (2010) 321–325.
[29] B.J. Fernandes, K. Sridharan, P. Munga, K. Ramesh, N.K. Udayashankar, Memory type switching behavior of ternary $Ge_{20}Te_{80-x}Sn_x$(0×4) chalcogenide compounds, J. Phys. D. Appl. Phys. 49 (2016) 295104.
[30] O. Brien, (12) United States patent, 2 (2004) 1–6.
[31] A.E. Owen, J.M. Robertson, Electronic conduction and switching in chalcogenide glasses, IEEE Trans. Electron Devices ED-20 (1973) 105–122.
[32] A.C. Warren, Thermal switching in semiconducting glasses, J. Non. Cryst. Solids. 4 (1970) 613–616 https://doi.org/10.1016/0022-3093(70)90098-0.
[33] D. Adlert, B.B. Schwartz, Science and technology institute for amorphous studies series series editors, 1991, ISBN : 978-1-4684-8747-3.
[34] S. Murugavel, S. Asokan, Local structure and electrical switching in chalcogenide glasses, Phys. Rev. B - Condens. Matter Mater. Phys 58 (1998) 3022–3025.
[35] S.R. Ovshinsky, Localized states in the gap of amorphous semiconductors, 36 (1976) 1469–1472.
[36] D. Adler, M.S. Shur, M. Silver, S.R. Ovshinsky, Threshold switching in chalcogenide-glass thin films, J. Appl. Phys. 51 (1980) 3289–3309.
[37] M. Anbarasu, S. Asokan, Electrical switching behavior of bulk As-Te-Si glasses: composition dependence and topological effects, Appl. Phys. A Mater. Sci. ProcessAppl. Phys. A Mater. Sci. Process 80 (2005) 249–252.
[38] R. Lokesh, N.K. Udayashankar, S. Asokan, Eelectrical switching behavior of antimony doped bulk Si-Te chalcogenide glasses, 3 (2011) 55–60.
[39] B.J. Fernandes, P. Munga, K. Ramesh, N.K. Udayashankar, Thermal stability and crystallization kinetics of Bi doped $Si_{15}Te_{85-x}Bi_x$ ($0 \le x \le 2$) chalcogenide glassy

alloys, Mater. Today Proc. 5 (2018) 16237–16245.
[40] S.B. Bhanu Prashanth, S. Asokan, A composition dependent thermal behavior of GexSe35-xTe65 glasses, J, Non. Cryst. Solids. 355 (2009) 1227–1230.
[41] R.A. Narayanan, S. Asokan, Evidence concerning the effect of topology on electrical switching in chalcogenide network glasses, Phys. Rev. B - Condens. Matter Mater. Phys 54 (1996) 4413–4415.
[42] R. Bez, A. Pirovano, Non-volatile memory technologies: emerging concepts and new materials, Mater. Sci. Semicond. Process. 7 (2004) 349–355.
[43] S. Asokan, K.P. Lakshmi, Electrical switching and other properties of chalcogenide glasses, J. Indian Inst. Sci. 91 (2011) 319–330.
[44] C. Das, G.M. Rao, S. Asokan, The electrical switching and thermal behavior of bulk $Ge_{15}Te_{85-X}Sn_x$ and $Ge_{17}Te_{83-X}$Snxglasses, J. Non. Cryst. Solids. 358 (2012) 224–228.
[45] B.J.F. and K.S. and P.M. and K.R, N.K. Udayashankar, Memory type switching behavior of ternary $Ge_{20}Te_{80-x}Sn_x$ ($0 \leqslant x \leqslant 4$) chalcogenide compounds, J. Phys. D. Appl. Phys. 49 (2016) 295104.
[46] C.N. Murthy, V. Ganesan, S. Asokan, Electrical switching and topological thresholds in Ge-Te and Si-Te glasses, Appl. Phys. A Mater. Sci. Process (2005).
[47] M. Anbarasu, S. Asokan, The influence of network rigidity on the electrical switching behaviour of Ge-Te-Si glasses suitable for phase change memory applications, J. Phys. D. Appl. Phys (2007).
[48] S. Murugavel, S. Asokan, Composition tunable memory and threshold switching in $Al_{20}As_xTe_{80-x}$semiconducting glasses, J, Mater. Res. 13 (1998) 2982–2987.
[49] B.J. Fernandes, N. Naresh, K. Ramesh, K. Sridharan, N.K. Udayashankar, Crystallization kinetics of Sn doped $Ge_{20}Te_{80-x}Sn_x$($0 \leq x \leq 4$) chalcogenide glassy alloys, J, Alloys Compd 721 (2017) 674–682.
[50] D.P. Olander, instrumental methods of analysis, sixth edition (Willard, Hobart H.; Merritt, Lynne L.; Dean, John A.; Settle, Frank A., Jr.), J. Chem. Educ. 61 (1984) A222.
[51] B.J. Fernandes, K. Ramesh, N.K. Udayashankar, Crystallization kinetics of $Si_{20}Te_{80-x}Bi_x$ ($0 \leq x \leq 3$) chalcogenide glasses, Mater. Sci. Eng. B Solid-State Mater. Adv. Technol 246 (2019) 34–41.
[52] A. Alegría, A. Arruabarrena, F. Sanz, Switching in Al-As-Te glass system, J. Non. Cryst. Solids. 58 (1983) 17–25.
[53] J.K. Babanna, D. Roy, S.G. Varma, S. Asokan, C. Das, Effect of Sn addition on glassy Si-Te bulk sample, (2018) 20034.
[54] D.R. Lide, G. Baysinger, CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data, Choice Rev. Online 41 (2004) 41-4368-41–4368.
[55] M. Anbarasu, S. Asokan, S. Prusty, A.K. Sood, Electrical switching and in situ Raman scattering studies on the set-reset processes in Ge-Te-Si glass, Appl. Phys. Lett 91 (2007) 0–3.
[56] M. Anbarasu, Thesis, D o c t o r o f P h i l o s o p h y, 2007.
[57] C. Das, G.M. Rao, S. Asokan, Electrical switching and thermal studies on bulk Ge-Te-Bi glasses, J. Non. Cryst. Solids. 357 (2011) 165–169.
[58] C. Das, R. Lokesh, G.M. Rao, S. Asokan, Electrical switching behavior of amorphous $Al_{23}Te_{77}$ thin film sample, J. Non. Cryst. Solids. 356 (2010) 2203–2206.
[59] A.K. Singh, Surface morphology and crystallization kinetics of multicomponent chalcogenide glasses, Mater. Focus 1 (2013) 50–56.

Journal of Non-Crystalline Solids 531 (2020) 119863

Contents lists available at ScienceDirect

Journal of Non-Crystalline Solids

journal homepage: www.elsevier.com/locate/jnoncrysol

Manifestation of intermediate phase in Cu doped Si-Te glasses

Diptoshi Roy[a], B. Tanujit[b], Jagannatha K. B[c], G. Sreevidya Varma[d], S. Asokan[b], Chandasree Das[a,*]

[a] Department of EEE, BMS College of Engineering, Bangalore, Karnataka 560019, India
[b] Department of Instrumentation & Applied Physics, Indian Institute of Science, Bangalore, Karnataka 560012, India
[c] Department of ECE, BMS Institute of Technology, Bangalore, Karnataka 560064, India
[d] Department of Physics, Presidency University, Bangalore, Karnataka 560064, India

ARTICLE INFO

Keywords:
Chalcogenide glass
Boolchand intermediate phase
Chemical threshold

ABSTRACT

Alternating differential scanning calorimetry (ADSC) analysis has been carried out on bulk $Si_{15}Te_{85\text{-}x}Cu_x$ ($1 \leq x \leq 10$) glasses to examine the thermal properties of the glassy samples in detail. The investigations on the compositional dependence of ΔH_{NR} (non- reversing enthalpy) manifests a trough between the composition $2 \leq x \leq 6$ which stands as the basis of the presence of Boolchand's Intermediate phase in that range. Additionally, anomalous behavior has been observed at the compositional dependence of various thermal parameters at $x = 9$, which stipulates the appearance of chemical threshold at the stated composition. Annealed bulk samples have been exposed to XRD studies to discern the type of formed crystalline phases. The study has reported the manifestation of Te, Si_1Te_2, Si_2Te_3, Cu_3Te_2 phases.

1. Introduction

Chalcogenide glasses being amorphous in nature displays various fascinating properties of which some are sizably employed in photonic and electronic devices [1] and stands as a reason for the intense research work on the amorphous glasses [2–6]. A considerable change can be incorporated in the structures of the amorphous material by extrinsic factors like light, heat, radiation of electron or the electric field, because of the presence of metastable states. The change in the structure of the material appreciably affects its various thermal properties like glass transition temperature (T_g), crystallization temperature (T_c) and its rate, optical constants, thermal and electrical conductivity, chemical activity, etc. Chalcogenides also manifests unusual deviations in their physical properties at rigidity percolation threshold where the floppy polymeric glass percolates to rigid amorphous solid, and chemical threshold. Investigation of the variation of various properties of chalcogenides with respect to the composition helps in identifying the network thresholds on these glasses.

The measurement of latent heats and transformation temperatures of diverse thermal transformations has been made easy by alternating or modulated Differential Scanning Calorimetry (ADSC or MDSC). Through ADSC the total heat flow curve (THF) can be deconvoluted into two parts - one that trails the variation in temperature and hence is familiar as reversing heat flow rate (RHF) and the other that is independent of the modulation in temperature (referred to as non-reversing heat flow rate (NHF)). The variation of various thermal parameters with composition such as non-reversing enthalpy change (ΔH_{NR}), change in specific heat (ΔC_p) can be quantified simultaneously with enhanced sensitivity.

Chalcogenides have enticed appreciable attention because of its potentiality in PCM application [7]. The PCM works on the basis of electrical switching [8, 9] which is electro-thermal in nature. Accordingly the thermal attributes like glass formability, thermal stability etc. are salient for determining the glass for PCM application.

Silicon telluride glasses have proved to be competent acoustic optic material and are utilized as a memory type switching diode [2, 3]. Incorporation of metal atoms impressively alters the connectivity, rigidity and structure of the network which inevitably alters the electrical properties namely the switching behavior of the glass [10–12]. The physical attributes of the glass are affected by the network rigidity and chemical composition [13]. Different metallic dopant in Si-Te glasses exhibits various properties and hence its study has gained particular interest. In this paper the thermal analysis of various variables such as T_g, thermal stability (ΔT), ΔH_{NR}, ΔC_p, with respect to composition for the Cu doped Si-Te glasses has been presented.

2. Experimental approach

Bulk $Si_{15}Te_{85\text{-}x}Cu_x$ ($1 \leq x \leq 10$) glasses have been prepared by quenching the melt from high temperature. Integrant elements of high

* Corresponding author.
E-mail address: chandasreedas.eee@bmsce.ac.in (C. Das).

https://doi.org/10.1016/j.jnoncrysol.2019.119863
Received 4 October 2019; Received in revised form 6 December 2019; Accepted 14 December 2019

purity have been taken in quartz ampoules which are evacuated and sealed at 10^{-6} mbar. The horizontal rotary furnace houses the sealed ampoules and has been programmed to heat the ampoule up to 1100 ^{0}C at the rate of 100 ^{0}C/h. The ampoules have been kept in continuous rotation in the furnace for 24 h to maintain the homogeneity of the melt and have been eventually quenched in ice-water mixed with NaOH. To perceive the amorphous nature of as-quenched bulk samples X-ray diffraction has been performed using a PANalytical X'Pert3 Powder X-ray diffractometer with CuK_α radiation in the 2θ range from 10° to 80° at a rate of 3°/min.

To effectuate thermal studies on $Si_{15}Te_{85-x}Cu_x$ ($1 \le x \le 10$) glasses, ADSC experiment using a Mettler Toledo 822^e ADSC instrument has been performed. The ADSC is graded for tau lag, temperature and heat flow using high purity indium. Two 40 ml aluminum pans are taken wherein one is kept empty and is considered as reference whereas the other is sealed with samples of identical thickness. Thermal scans are consummated in the temperature span of 40 °C to 300 °C at a 3 °C/min with 1 °C scan amplitude. A modulation period of 1 min is taken for ADSC scan. Argon with a flow rate of 75 ml/min is employed as the purge gas. THF can be separated into two parts, the RHF and the NHF; RHF is analogous to the variations in thermodynamic specific heat on heating whereas NHF corresponds to the kinetic processes and enthalpy variations that accompanies the structuring in the glassy network [14]. Typical error in the measurement of thermal parameters in within ± 2 °C. Errors are estimated using the method of standard deviation and the corresponding error bars are included in the figures.

To discern the thermally crystallized phases that would have appeared in the as-prepared sample, the later has been kept in sealed ampoules and annealed at their respective crystallization temperature for two hours followed by the diffraction study. Experimental analysis like EXAFS structural approach, isotopic substitution in neutron diffraction etc. have divulged that the metal impurities are primarily trihedral coordinated in glassy selenides like Ag-As-Se [15], Ge-Se-In [16] whereas it has been seen to coordinate tetrahedrally in glassy tellurides like Ag-Ge-Te [17].

3. Results and discussion

The THF curve of the representative $Si_{15}Te_{85-x}Cu_x$ ($1 \le x \le 10$) glasses showing the regions of T_g and T_c is shown in Fig. 1. The THF curves of $Si_{15}Te_{85-x}Cu_x$ where x is 2, 4, 6 display one endothermic glass transition peak and three prominent exothermic crystalline peaks stipulates that percolation of stable structural phases of Si-Te-Cu glasses occurs at different temperatures. However $Si_{15}Te_{77}Cu_8$ and

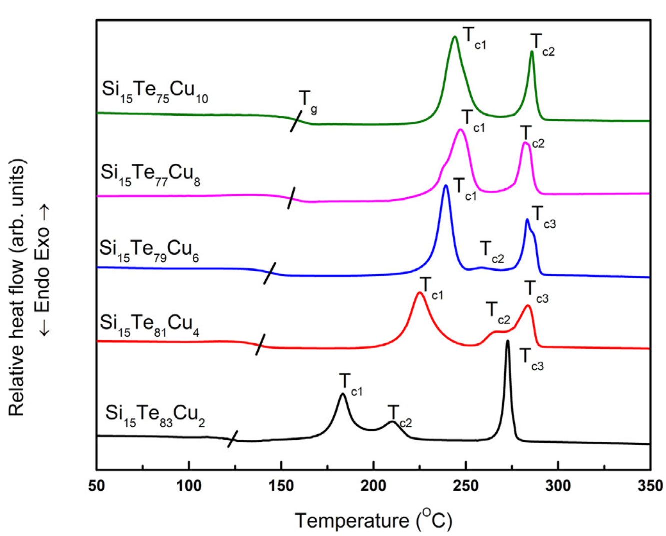


Fig. 1. ADSC THF curves of selected $Si_{15}Te_{85-x}Cu_x$ ($1 \le x \le 10$) glasses.

$Si_{15}Te_{75}Cu_{10}$ show only two crystalline peaks.

In almost all the chalcogenide systems, it has been seen that the coordination numbers of Si and Te follows the Mott's 8-N theory, where N is the number of electrons in the valence shell. Accordingly in the present Si-Te-Cu system the coordination number of Si and Te is presumed to be 4 and 2 respectively. Hypothetically, it is difficult to design amorphous semiconductors because of the interatomic associations that are in turn accountable for short range orders. Authors on the basis of RDF (Radial Atomic Distribution) studies propose that copper is tetrahedrally coordinated in chalcogenide glasses. The analysis of the EXAFS spectrum reveals the existence of four coordination numbers in copper [18], a number which has been observed in the other systems of copper such as CuFeS [19], Cu_3AsS_4 [19], and CuAsI [20]. For the current work, tetra coordinated copper has been presumed, gaining the requisite electrons for the sp^3 hybridation of certain tellurium atoms. The rigidity percolation model also indicates that the coordination number of Cu is 4 to explain the increasing glass forming ability [21]. Using coordination number of 4, 2, 4 for Si, Te and Cu respectively, the average coordination number ($<r>$) for a specific composition can be deduced using the following formula:

$$\langle r \rangle = \frac{\text{r1Si}(15) + \text{r2Te}(85 - \text{x}) + \text{r3Cu}(\text{x})}{100} \tag{1}$$

In case of $Si_{15}Te_{85-x}Cu_x$ ($1 \le x \le 10$) samples the probable bonds emerged in the system as per the bond energy consideration are Si-Te, Si-Si, Te-Te, Te-Cu and Cu-Cu. The bond energies (D) in KJ/mol in decreasing order are D(Si-Te) = 429.2; D(Si-Si) = 310; D(Te-Te) = 257.6 ± 4.1; D(Te-Cu) = 230.5 ± 14.6; D(Cu-Cu) = 201 [20]. The illustration by Bicerano and Ovshinsky [21] on chemical bond approach explains the formation of structural lattice in chalcogenides. Thence, in a chalcogenide network the building of bonds follows the order of declining bond energies before all the unoccupied valencies are filled. Additionally atoms tend to bind more emphatically with atoms of dissimilar one rather than of the same type. In $Si_{15}Te_{85}$ glass, the Si being 4- fold coordinated pairs with Te atoms which is 2- fold coordinated as the bond energy between Si-Te is highest and thus the glassy network is embraced by the Si-Te links. After the formation of Si-Te bonds, the remaining Te atoms combine among themselves to form a network. A crystallization study on $Si_{20}Te_{80}$ glasses [22] affirms this conjecture, where the formation of crystalline phases: Te and $SiTe_2$ on thermal devitrification has been revealed. Furthermore, a modification to the existing network can be expected by the addition of copper atoms which is 4- fold coordinated to $Si_{15}Te_{85}$ matrix by replacing Te atoms as copper atom can pair with Te atoms and also with Si-Te backbone. The X-ray diffraction (XRD) of crystallized Si-Te-Cu sample shown in Fig. 2 divulges that the phases which have crystallized out at T_{c1}, T_{c2} and T_{c3} are: Hexagonal Te with the unit cell defined by $a = b = 4.454$ Å, $c = 5.924$ Å; tetragonal Cu_3Te_2 with unit cell of $a = b = 3.98$ Å and $c = 6.55$ Å; hexagonal Si_2Te_3 with unit cell specified by $a = b = 7.429$ Å, $c = 13.471$ Å and Si_1Te_2. Thus Cu atoms bond with Te atoms leading to an intensification in network connectivity in the system.

Network entropy which is analogous to network connectivity, rigidity etc. is the factor which helps in determining the reason behind the variation of thermal variables with composition of chalcogenide glasses. Additionally, the abrupt change in the characteristic of thermal variables with respect to composition can be related with topological thresholds.

Fig. 3 shows the variation of T_g with composition for $Si_{15}Te_{85-x}Cu_x$ ($1 \le x \le 10$) glasses. As seen, T_g decreases initially up to the composition range of $x \le 2$ and further followed by continuous increase. The compositional dependence of T_g of glasses is used to ascertain the network connectivity. An increment in T_g normally signifies strengthening of network connection [23] whereas a saturation or decrease in T_g signals to the phase segmentation because of the separation of homopolar bonds [24]. The initial decline in T_g of $Si_{15}Te_{85-x}Cu_x$ glasses

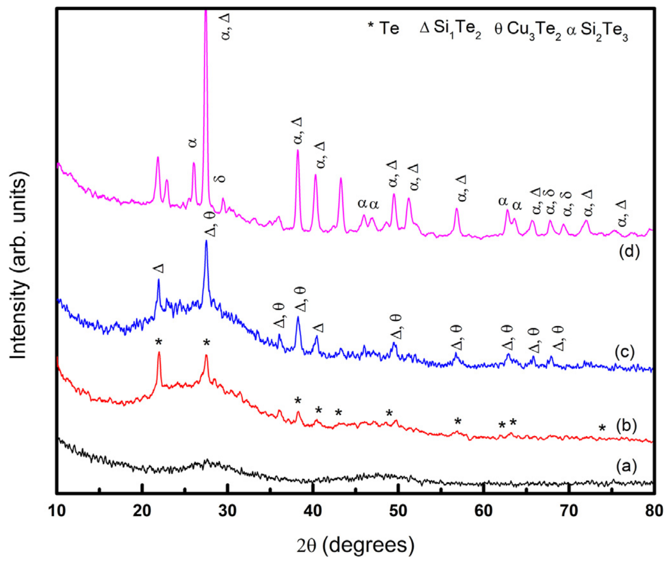


Fig. 2. XRD pattern of (a) as-prepared $Si_{15}Te_{84}Cu_1$ glass showing amorphous nature; (b) $Si_{15}Te_{83}Cu_2$ glass at T_{c1}; (c) $Si_{15}Te_{83}Cu_2$ glass at T_{c2}; (d) $Si_{15}Te_{79}Cu_6$ glass at T_{c3}.

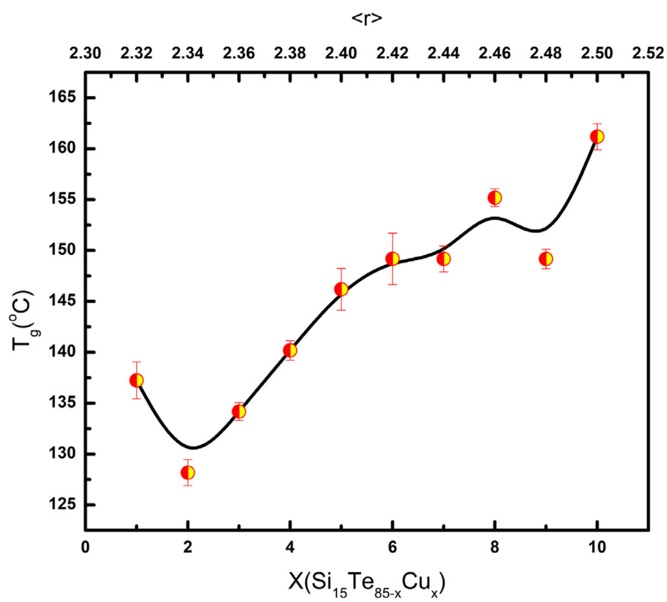


Fig. 3. The variation of glass transition temperature (T_g) of $Si_{15}Te_{85\text{-}x}Cu_x$ ($1 \le x \le 10$) samples with Cu composition and average coordination number. The line connecting the points, is used as a guide to the eye.

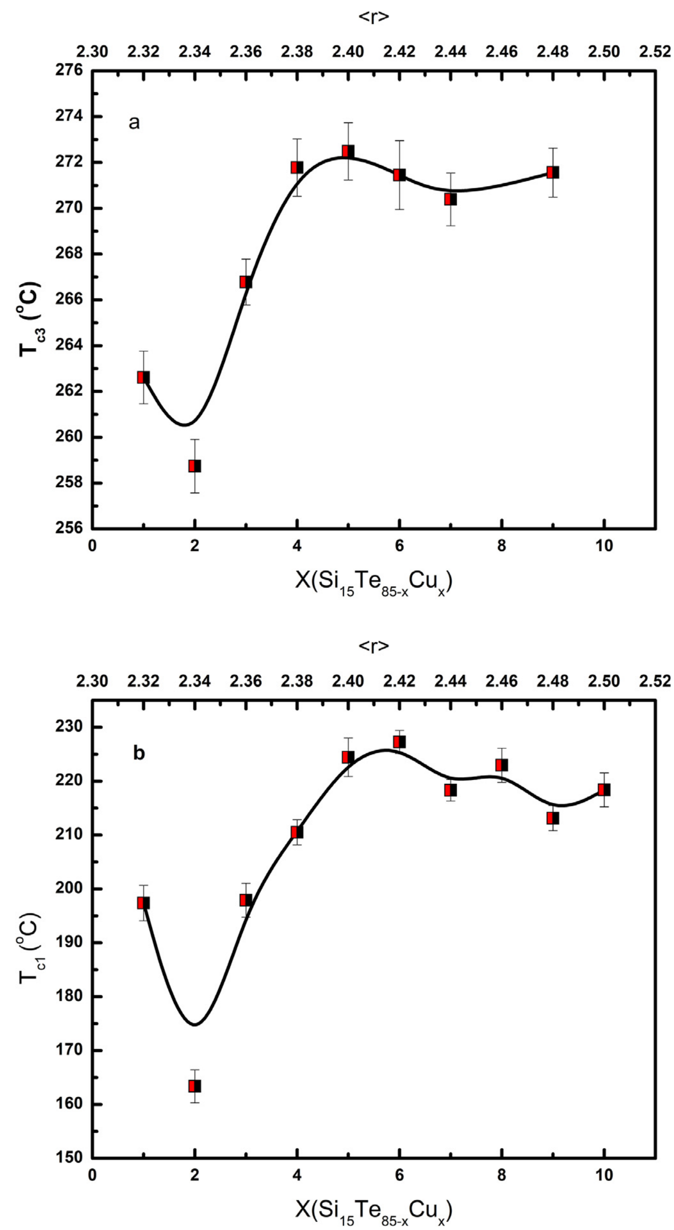


Fig. 4. The compositional dependence of (a) T_{c3} and (b) T_{c1} of $Si_{15}Te_{85\text{-}x}Cu_x$ ($1 \le x \le 10$) glasses. The line connecting the points, is used as a guide to the eye.

in the composition radius of $1 \le x \le 2$ can be attributed to the separation of homopolar Te-Te bonds in Si-Te-Cu network. An increase in T_g is marked from $x = 2$ which indicates that the copper atoms combines with the Te-Te matrix in a particular fashion resulting in the development of Cu-Te network.

An anomaly in different properties normally points out the percolation threshold of the glass howbeit the precise feature of variation rests on the property under analysis and equally differs with system. Glasses containing tellurium such as Ge–Te [25], Si–Te [26], etc., exhibits a minimal value in T_g at the stiffness threshold, while the amorphous selenides such as Ge–In–Se [27], Ge–Ga–Se [28], etc., exhibits a change in slope. It can be noticed from the compositional dependence of T_g (Fig. 3) in Si-Te-Cu system that the stiffness threshold corresponds to a minimum value thus manifesting a congruent behavior as in other telluride. Covalency of the constituents, size of the atom, metallicity of the additive, etc. are some of the factors responsible for determining the specific nature of peculiarity perceived in a certain property at the mechanical threshold.

A similar trend is discerned in the compositional dependence of first crystallization temperature (T_{c1}) of $Si_{15}Te_{85\text{-}x}Cu_x$ ($1 \le x \le 10$) system, as shown in Fig. 4. As can be seen the curve marks a steep increase in T_{c1} in the range of $2 \le x \le 6$ after which it saturates. T_{c3} has the same pattern as T_{c1} with insertion of copper. The crystallization of binary and ternary telluride glasses at T_{c1} typically marks the precipitation of Te, resulting in the disintegration of Te-Te chains. As a result of this procedure a glassy residue containing elements of higher coordination is obtained which is categorized by the higher T_g than the parent glass [26, 29]. The reduction in the Te-Te bonds of lower energy (257.6 ± 4.1 KJ/mol) and the increase in the Si-Te bonds of higher energy (429.2 KJ/mol) makes the precipitation of Te more adverse and inevitably leads to an augmentation in T_{c1} with the insertion of Cu.

The contrast between T_c and T_g denoted by "ΔT" highlights the

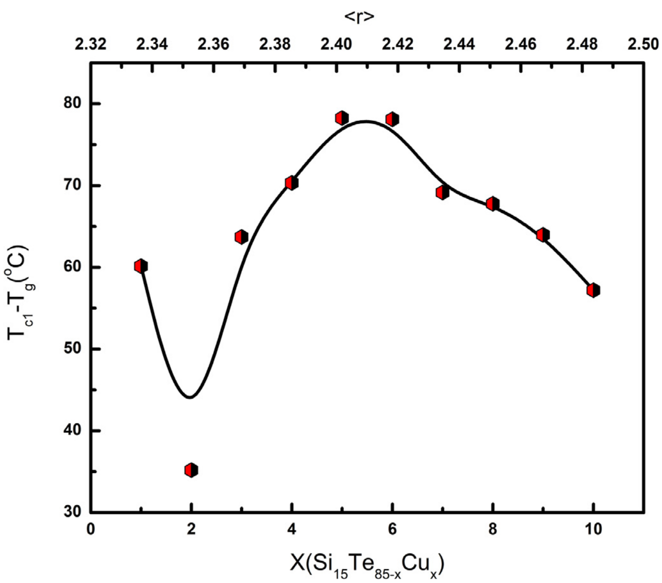


Fig. 5. The compositional dependence of $\Delta T = T_{c1}$-T_g of $Si_{15}Te_{85\text{-}x}Cu_x$ ($1 \leq x \leq 10$) glasses. The line connecting the points, is used as a guide to the eye.

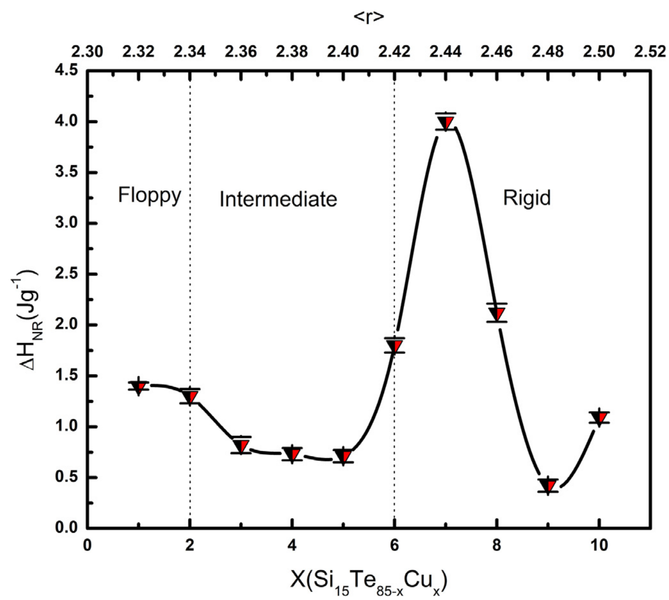


Fig. 6. Variation of non-reversing heat flow (ΔH_{NR}) of $Si_{15}Te_{85\text{-}x}Cu_x$ ($1 \leq x \leq 10$) glasses with composition (x) and average coordination number. The line connecting the points, is used as a guide to the eye.

propensity of the system towards crystallization [30]. More the dissimilarity less is the propensity towards crystallization. Moreover, the glass forming ability (GFA) in chalcogenide systems has been found to have direct proportional relationship with the difference between T_c and T_g. Fig. 5 shows the glass forming ability also known as thermal stability of the $Si_{15}Te_{85\text{-}x}Cu_x$ ($1 \leq x \leq 10$) samples. As seen from the figure, ΔT increases with copper insertion in the compass of $2 \leq x \leq 6$, which stipulates that the network connectivity and accordingly the network rigidity of $Si_{15}Te_{85\text{-}x}Cu_x$ increases in the range of $2 \leq x \leq 6$.

The glassy systems which are covalently bonded are seen to be restrained by bond stretching and bond bending forces. At an average coordination number of 2.40, the glassy system attains a critical point wherein the constraints per atom and degrees of freedom of the network become equal [31, 32]. Rigidity of the network percolates and the material experiences a change from floppy polymeric glass to rigid amorphous solid at the stiffness threshold. The stiffness threshold is also synonym as rigidity percolation or mechanical threshold. The stiffness threshold at $\langle r\rangle = 2.40$ is seen in the following binary systems like Si-Te [33], Ge-Se [34] and ternary systems like Al-As-Te [35], Ge-Se-Te [36] etc. Further, Tanaka reported that the systems having medium range interactions the stiffness threshold shifts to different average coordination number [37]. Prudent research and analysis have revealed that in some glassy systems the stiffness threshold compasses over a broad composition span [38–40]. These systems are found to transit from floppy polymeric phase to an isostatically rigid phase and from an isostatically rigid to a stressed rigid phase or in other words these types of systems display two stiffness transitions. The intermediate isostatically rigid phase are seen to be comprised of the glass composition which are self-organized, near-ideal and stress-free [38, 39]. Theoretical estimations using techniques such as graph theory, constraint counting, cluster approximations show the presence of "intermediate phase – IP" (familiar as Boolchand IP) in many chalcogenide systems [41]. ΔH_{NR} abbreviated for non-reversing enthalpy measures the latent heat between the glass and its corresponding melt and also quantifies in the glass transition area, the enthalpy required for composure of the glass structure. Moreover, the configurational contrast of the glass and the liquid is evident from the compositional dependence of ΔH_{NR}. In the IP it is interesting to find that the ΔH_{NR} of the glasses almost vanishes [38, 39], leading to the conclusion that in the IP both the liquid and the glass structures look similar and hence are stress free. This study reveals a fascinating result about the compositional dependence of ΔH_{NR} as shown in Fig. 6, which stipulates that ΔH_{NR} does not vary much in the composition range of $1 \leq x \leq 2$. Howbeit a decline is noticed in ΔH_{NR} at $x \geq 2$ followed by the formation of a trough in the compositional span of $2 \leq x \leq 6$, and this clearly stipulates the presence of IP in the span of $2 \leq x \leq 6$. Specific group of phase change glasses for instance Ge-Te-Si, Si-Te-In, etc. [42, 43] exhibits reversibility window in certain composition. Especially $Si_{15}Te_{85\text{-}x}In_x$ glasses are noticed to have IP in the average coordination span $\langle r\rangle = 2.32$ to 2.50, wherein the width (Δr) and the centroid of the IP are 0.18 and 2.41 respectively [43]; whereas in $Si_{15}Te_{85\text{-}x}Ag_x$ system the IP is not seen [44]. The structure of the glass is responsible for shaping the compositional width and centroid of the rigidity gradation; the breadth of the IP practically vanishes in certain networks showcasing a perfect accordance with the extended constraint theory [45]. IP region, IP width and the centroid of the present ternary system ($Si_{15}Te_{85\text{-}x}Cu_x$) has been found to be 2.32 to 2.50, 0.18 and 2.41 respectively which is similar to $Si_{15}Te_{85\text{-}x}In_x$ glasses.

The variation of density with the addition of Cu and also with average coordination number is displayed in Fig. 7. In spite the atomic weight of copper (63.546 amu) is less than Te (127.6 amu), the substitution of Te with copper intensifies the density primarily. Howbeit, the density increases with the Cu composition along with the plateau seen in the composition range of $3 \leq x \leq 6$. Literature specifies that molar volume manifests a minimum or a maximum is exhibited by molar density in the region related to Boolchand's IP [46]. The wide maximum exhibited by the density versus composition graph of $Si_{15}Te_{85-x}Cu_x$ glasses highlights the thermally reversing window in this chalcogenide glass.

The change in specific heat (ΔC_p) at glass transition with the insertion of copper is shown in Fig. 8. As seen ΔC_p jumps and exhibits a minimum at $x = 2$ ($<r> = 2.34$). Some glasses when subjected to ADSC have shown a minimum value of ΔC_p at the stiffness threshold [47]. The resilience of the glass structure can be defined in terms of the specific heat of the glass; a system which shows strong reluctance for network reformation at T_g can have least value for ΔC_p at the rigidity percolation threshold [30, 31].

Chemically Ordered Random Network (COCRN) model [48] proposes that at a certain critical composition (referred as chemical threshold) of chalcogenide system, the hetero-polar bonds are only seen

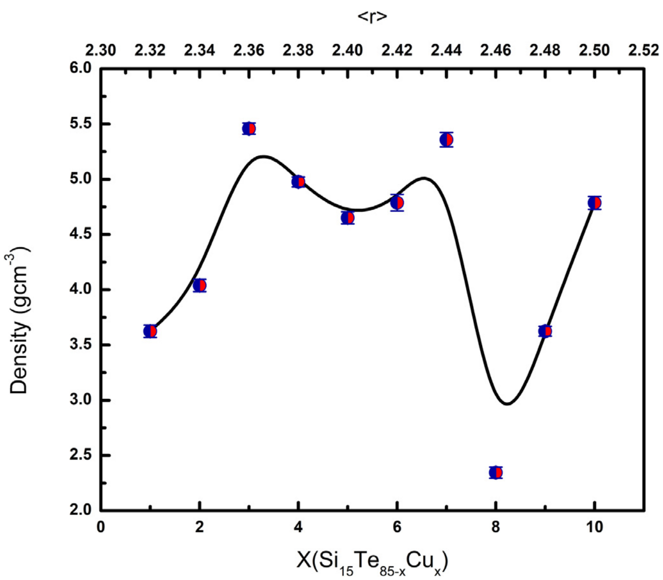


Fig. 7. Variation of density with Cu composition for $Si_{15}Te_{85-x}Cu_x$ ($1 \le x \le 10$) glasses. The line connecting the points, is used as a guide to the eye.

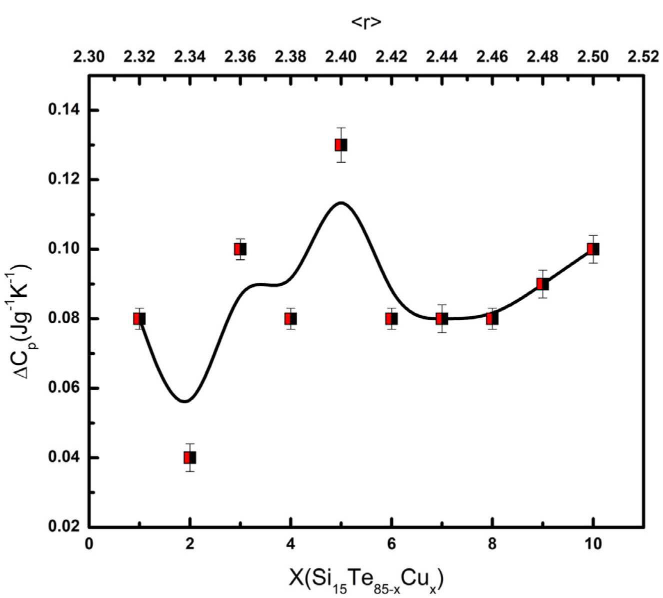


Fig. 8. Compositional dependence of specific heat (ΔC_p) measured at T_g for $Si_{15}Te_{85-x}Cu_x$ ($1 \le x \le 10$) glasses. The line connecting the points, is used as a guide to the eye.

to exist. At chemical threshold the glass is anticipated to show maximum chemical organization and is also marked by incongruous variations [34]. Although the chemical threshold is supposed to ensue at the mean coordination number of 2.67 however it is seen that chemical threshold of distinct chalcogenide systems varies and eventuates at different coordination number [49, 50]. In Si-Te-Cu glasses, a dip in T_g and T_{c1} has been observed at the composition (x) = 9, (Figs. 3 and 4) which can be credibly categorized as chemical threshold of this glasses. The structural network of the glass at the chemical threshold is known to be of paramount order, hence close to crystalline state. A minimal value is seen for molar density (Fig. 7) at the chemical threshold which implicates the existence of the network in least restrained state favorable for structural reorganization. Consequently, the driving forces and energy barrier essential for crystallization of the ordered glass are low. Hence the dip in T_g and T_c indicates the manifestation of chemical threshold at $x = 9$. It also supports the theoretical predictions which propose a minimum in GFA at the chemical threshold.

An intriguing fact that can be highlighted is that the appearance of IP in $Si_{15}Te_{85-x}Cu_x$ glass is noticed in the composition versus ΔH_{NR} characteristic while the variance of T_c and ΔC_p with composition manifests only the effect of rigidity percolation threshold. Customarily, in glasses the appearance of IP is noticed in the variation of attributes like thermal diffusivity [51], ΔH_{NR} [52, 39], photo response [53] with composition, howbeit attributes like electrical switching voltage [54], T_c [55], etc., manifests solely the effect of rigidity percolation threshold.

4. Conclusion

The ADSC studies on $Si_{15}Te_{85-x}Cu_x$ ($1 \le x \le 10$) glasses stipulate that the glasses constitute of distinct phases which are stable and percolate at different crystallization temperature. The variation of various thermal parameters like T_g, T_{c1}, T_{c3} with Cu composition indicates an initial decline in the composition till $x \le 2$ followed by a sustained increment in the range of $2 \le x \le 6$ indicating a better network connectivity in this region. The ΔC_p at T_g with composition manifests the presence of rigidity percolation threshold at $x = 2$. Also the compositional dependence of thermal stability reveals that the glasses in the span of $2 \le x \le 6$ are thermally more stable. The ΔH_{NR} manifests a trough between the composition $2 \le x \le 6$, which distinctly stipulates the existence of intermediate phase in the composition range. The variation of density with composition shows the signature of mean field threshold. The anomalous behavior seen in thermal parameters at $x = 9$ denotes the existence of chemical threshold for these glasses. In short, the study has helped in revealing the existence of intermediate phase in the range which is the most stable composition and also the existence of chemical threshold. These studies will help to correlate and analyze the results of other studies like electrical switching studies for the other series of glasses.

Author contribution statement

All the authors have contributed for the research work associated with this paper in terms of experimentation, documentation, result analysis, paper writing and communication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Author Chandasree Das and Diptoshi Roy acknowledge Science and Engineering Research Board (SERB) India and Technical Education Quality Improvement Program - II (TEQIP-II) for the financial support to carry out this research work.

References

[1] H.S. Nalwa, Handbook of Advanced Electronic and Photonic Materials and devices, Chalcogenide Glasses and Sol–Gel Materials, Academic Press, San Diego, 2001.
[2] Y. Huang, F. Chen, B. Qiao, S. Dai, Q. Nie, X. Zhang, Improved nonlinear optical properties of chalcogenide glasses in Ge-Sn-Se ternary system by thermal treatment, Opt. Mater. Express 6 (2016) 1644–1652.
[3] A. Sharma, N. Mehta, Observation of switching behavior in some multi-component glasses of Se-Te-Sn-Pb system, Mater. Lett. 178 (2016) 178–180.
[4] A. Maignan, E. Guilmeau, F. Gascoin, Y. Bréard, V. Hardy, Revisiting some chalcogenides for thermoelectricity, Sci. Technol. Adv. Mater. 13 (2012) 0530039pp.
[5] S. Balasubramanian, K.J. Rao, A molecular dynamics study of atomic correlations in glassy B2S3, J. Phys. Chem. 98 (1994) 9216–9221.
[6] Z. Han, P. Lin, V. Singh, et al., On-chip mid-infrared gas detection using chalcogenide glass waveguide, Appl. Phys. Lett. 108 (2016) 1–4.
[7] M.H.R. Lankhorst, B.W.M.M. Ketelaars, R.A.M. Wolters, Low-cost and nanoscale

non-volatile memory concept for future silicon chips, Nat. Mat. 4 (2005) 347–352.
[8] D. Roy, K.B Jagannatha, C. Das, Electrical switching and morphological studies on $Si_{15}Te_{84}Cu_1$ thin film, Mater. Today Proceed. 5 (2018) 3052–3055.
[9] D. Roy, C.H.S. Nadig, A. Krishnan, A. Karanam, R. Abhilash, K.B. Jagannatha, C. Das, Electrical switching studies on $Si_{15}Te_{85\text{-}x}Cu_x$ bulk ($1 \leq x \leq 5$) glasses, AIP Conf. Proc. 1966 (2018) 020033.
[10] K. Ramesh, S. Asokan, E.S.R. Gopal, Chemical ordering and fragility minimum in Cu–As–Se glasses, J. Non-Cryst. Solids 352 (2006) 2905–2912.
[11] K. Ramesh, N. Ramesh Rao, K.S. Sangunni, E.S.R. Gopal, Electrical resistivity of Cu doped As–Se glasses at high pressure, Phys. Status Solidi (B) 235 (2003) 536–541.
[12] S. Murugavel, S. Asokan, Local structure and electrical switching in chalcogenide glasses, Phys. Rev. B 58 (1998) 3022–3025.
[13] U. Senapati, A.K. Varshneya, Configurational arrangements in chalcogenide glasses: a new perspective on Phillips' constraint theory, J. Non-Cryst. Solids 185 (1995) 289–296.
[14] D.M. Price, Modulated-temperature thermo mechanical analysis, Thermochim. Acta 357 (2000) 23–29.
[15] C.J. Senmore, P.S. Salmon, Structure of fast-ion conducting chalcogenide glasses: the Ag-As-Se system, J. Non-Cryst. Solids 156–158 (1993) 720–724.
[16] G. Saffarini, Atomic density versus average coordination number in Ge–In–Se glasses, Phys. Status Solidi B213(1999) 261.
[17] A. Ferhat, R. Ollitrault-Fichet, V. Mastelaro, S. Benazeth, E.J. Rivet, Etude des verres du système Ag-Ge-Te, J. Phys. IV (Colloque) (1992) C2; J. Phys. III (Suppl. 2) (1992) C2-201.
[18] S.H. Hunter, A. Biennenstack, T.M. Hayes, in The Structure of Non-Crystalline Materials, in: P.H Gaskell (Ed.), 1977 Taylor and Francis, London.
[19] L. Pauling, Uniones Quimicas, Kapelusz, Buenos Aires, 1969.
[20] R.B. Heslop, K. Jones, Inorganic Chemistry, Elsevier, Amsterdam, 1976.
[21] J. Bicerano, S.R. Ovshinsky, Proceedings of the Noble Laureates Symposium on Applied Quantum Chemistry, D, Reidel Publishing Co., Honolulu, Hawaii, 1984.
[22] S. Asokan, G. Parthasarathy, E.S.R. Gopal, Double glass transition and double stage crystallization of bulk $Si_{20}Te_{80}$ glass, J. Mater. Sci. Lett. 4 (1985) 502–504.
[23] P. Boolchand, D.G. Georgiev, M. Micoulaut, Nature of glass transition in chalcogenides, J. Optoelectron. Adv. Mater. 4 (2002) 823–836.
[24] P. Boolchand, D.G. Georgiev, T. Qu, F. Wang, L. Cai, S. Chakravarthy, Nanoscale phase separation effects near r = 2.4 and 2.67, and rigidity transitions in chalcogenide glasses, C.R. Chim. 5 (2002) 713–724.
[25] K. Ramesh, S. Asokan, K.S. Sangunni, E.S.R. Gopal, Effect of topological threshold on thermal behavior of germanium telluride glasses containing metallic additive, Eur. Phys. J. B 6 (1998) 207–212.
[26] S. Asokan, G. Parthasarathy, E.S.R. Gopal, Crystallization studies on bulk Si_xTe_{100-x} glasses, J. Non-Cryst. Solids 86 (1986) 48–64.
[27] A. Giridhar, S. Mahadevan, The Tg versus Z dependence of glasses of the Ge-In-Se system, J. Non-Cryst. Solids 151 (1992) 245–252.
[28] S. Mahadevan, A. Giridhar, Coexistence of topological and chemical ordering effects in Ge-Ga-Se glasses, J. Non-Cryst. Solids 152 (1993) 42–49.
[29] B. Norban, D. Pershing, R.N. Enzweiler, P. Boolchand, J.E. Griffiths, J.C. Phillips, Coordination-number-induced morphological structural transition in a network glass, Phys. Rev. B 36 (1987) 8109–8114.
[30] A. Hruby, Evaluation of glass-forming tendency by means of DTA, Czech. J. Phys. B 22 (1972) 1187–1193.
[31] J.C. Phillips, Chemical bonding, internal surfaces, and the topology of non-crystalline solids, Phys. Stat. Sol. (b) 101 (1980) 473–479.
[32] M.F. Thorpe, Continuous deformations in random networks, J. Non-Cryst. Solids 57 (1983) 355–370.
[33] S. Asokan, G. Parthasarathy, E.S.R. Gopal, Evidence for a critical composition in group-IV–VI chalcogenide glasses, Phys. Rev. B 35 (1987) 8269.
[34] S. Asokan, M.V.N. Prasad, G. Parthasarathy, E.S.R. Gopal, Mechanical and chemical thresholds in IV-VI chalcogenide glasses, Phys. Rev. Lett. 62 (1989) 808–810.
[35] S. Murugavel, K.V. Acharya, S. Asokan, Pressure-induced metallization in Al-As-Te glasses, J. Non-Cryst. Solids 191 (1995) 327.
[36] R. Ganesan, A. Srinivasan, K.N. Madhusoodanan, K.S. Sangunni, E.S.R. Gopal, Composition dependence of the glass transition in Ge–Se–Te glasses, Phys. Status Solidi B 190 (1995) K23.
[37] K. Tanaka, Structural phase transitions in chalcogenide glasses, Phys. Rev. B 39 (1989) 1270–1279.
[38] D. Selvanathan, W.J. Bresser, P. Boolchand, Stiffness transitions in Si_xSe_{1-x} glasses from Raman scattering and temperature-modulated differential scanning calorimetry, Phys. Rev. B 61 (2000) 15061.
[39] D.G. Georgiev, P. Boolchand, M. Micoulaut, Rigidity transitions and molecular structure of As_xSe_{1-x} glasses, Phys. Rev. B 62 (2000) R9228.
[40] Diptoshi Roy, G.Sreevidya Varma, S. Asokan, Chandasree Das, Electrical switching and optical bandgap studies on quaternary Ag-doped Ge–Te–In thin films, IEEE Trans. Electron Devices 66 (2019) 1881.
[41] S.A. Altunyan, V.S. Minaev, M.S. Minazhdinov, B.K. Skachkov, Formation of glasses in Si-Te system and memory-type switching diodes made of Si-Te semiconducting glasses, Sov. Phys. Semicond. 4 (1971) 1906.
[42] M. Anbarasu, K.K. Singh, S. Asokan, Evidence for a thermally reversing window in bulk Ge–Te–Si glasses revealed by alternating differential scanning calorimetry, Phil. Mag. 88 (2008) 599–605.
[43] S.R. Gunti, A. Ayiriveetil, S. Asokan, Thermodynamic, kinetic and electrical switching studies on $Si_{15}Te_{85\text{-}x}In_x$ glasses: observation of Boolchand intermediate phase, J. Solid State Chem. 184 (2011) 3345–3352.
[44] S.R. Gunti, S. Asokan, Thermodynamic, Raman and electrical switching studies on $Si_{15}Te_{85\text{-}x}Ag_x$ ($4 \leq x \leq 20$) glasses, J. Appl. Phys. 111 (2012) 033518.
[45] N.F. Mott, Conduction in non-crystalline materials, Phil. Mag. 19 (1969) 835–852.
[46] S. Manohar, S. Murugavel, S. Asokan, Thermal diffusivities and molar volumes of ternary $Al_{20}As_xTe_{80-x}$ alloy glasses: evidence of self-organization, Solid State Commun. 135 (2005) 323–326.
[47] X. Feng, W.J. Bresser, P. Boolchand, Direct evidence for stiffness threshold in chalcogenide glasses, Phys. Rev. Lett. 78 (1997) 4422–4425.
[48] G. Lucovsky, F.L. Gleener, R.H. Geils, R.C. Keezer, The Structure of Non-Cryst, in: P.H. Gaskel (Ed.), Materials, Taylor and Francis, London, 1977, p. 127.
[49] B.H. Sharmila, J.T. Devaraju, S. Asokan, A modulated differential scanning calorimetric study of As–Te–In glasses, J. Non-Cryst. Solids 326–327 (2003) 154–158.
[50] P. Zaheerudeen Saheb, B.H. Sharmila, S. Asokan, K. Appaji Gowda, Extended rigidity percolation and chemical thresholds in Ge–Te–Pb glasses as revealed by MDSC, Solid State Commun. 129 (2004) 765–768.
[51] P. Pattanayak, N. Manikandan, M. Paulraj, S. Asokan, Photo-thermal deflection and electrical switching studies on Ge–Te–I chalcohalide glasses, J. Phys. Cond. Matter. 19 (2007) 036224.
[52] P. Boolchand, D. Georgiev, B. Goodman, Discovery of the intermediate phase in chalcogenide glasses, J. Optoelectron. Adv. Mater. 3 (2001) 703–720.
[53] N. Manikandan, S. Asokan, Signatures of an extended rigidity percolation in the photo-degradation behavior and the composition dependence of photo-response of Ge–Te–In glasses, J. Non-Cryst. Solids 354 (2008) 3732–3734.
[54] C.N. Murthy, V. Ganesan, S. Asokan, Electrical switching and topological thresholds in Ge-Te and Si-Te glasses, Appl. Phys. A 81 (2005) 939–942.
[55] R. Aravinda Narayanan, S. Asokan, A. Kumar, Influence of chemical disorder on electrical switching in chalcogenide glasses, Phys. Rev. B 63 (2001) 09223.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 07 | July 2019 www.irjet.net p-ISSN: 2395-0072

SMART REAL TIME MANHOLE MONITORING SYSTEM

Prof. Chandraprabha R[1], Ashwini C.V[2], Dharani M[3], Harshitha G[4], Kruti Mohan[5]

[1]*Assistant Professor, Dept. of ECE, BMSIT & M, Karnataka. India*
[2,3,4,5]*Students, Dept. of ECE, BMSIT & M, Karnataka, India*

---***---

Abstract – *A smart city is the future goal to have cleaner and better amenities for the society. Smart underground infrastructure is an important feature to be considered while implementing a smart city. Drainage system monitoring plays a vital role in keeping the city clean and healthy. Since manual monitoring is incompetent, this leads to slow handling of problems in drainage and consumes more time to solve. To mitigate all these issues, the system using a wireless sensor network, consisting of sensor nodes is designed. The proposed system is low cost, low maintenance, IoT based real time which alerts the managing station through an email when any manhole crosses its threshold values. This system reduces the death risk of manual scavengers who clean the underground drainage and also benefits the public.*

Key Words- Smart underground, Drainage system monitoring, wireless sensor network, low cost.

1. INTRODUCTION

An integral part of any drainage system is the access points into it when it comes to cleaning, clearing, and inspection. Metropolitan cities have adopted underground drainage system and the city's municipal corporation must maintain its cleanliness. If the sewage maintenance is not proper, ground water gets contaminated causing infectious diseases. Blockages in drains during monsoon season, causes problems in the routine of the public. Hence, there should be a facility in the city's corporation, which alerts the officials about blockages in sewers, their exact location and also if the manhole lid is open automatically. Underground drainage consists of sewage system, gas pipeline network, water pipelines, and manholes. Temperature sensors are used to monitor electric power lines that are installed underground. Pressure sensors are deployed to avoid manhole explosions due to chemical release and electrical energy.

Paper [1] represents the implementation and design function of and Underground Drainage and Manhole Monitoring System (UDMS) with separate transmitter and receiver models. The vital considerations of this design are low cost, low maintenance, fast deployment, and a high number of sensors, long life-time and high quality of service. Paper [2] mainly acknowledges in the field of alerting the people about the gas explosion, increase in the water level and the opened lid. It uses IoT to make the drainage monitoring system in a highly automotive by using sensor for detecting and sending alerts through audible alarms with glowing of LED light and messages via Wi-Fi module to the authorities, storing the data in the cloud and displaying the details in the web browser.

This project overcomes the demerits of paper [1] by detecting drainage water blockage by installing water flow rate sensors at the intersection of nodes. When there is a blockage in a particular node, there is variation in the flow of drainage water which when cross the set value will display the alert in the managing station. Also demerits of paper [2] are solved by detecting temperature variations inside the manhole and alerting the same to the managing station through automatic mail. Also, flow rate sensors are used to detect the over flow of the drainage water and alerting the same to the managing station through automatic mail.

So, the main focus of this project is to provide a system which monitors water level, atmospheric temperature, water flow and toxic gases. If drainage gets blocked and sewage water overflows, manhole lid opens, it is sensed by the sensors and this data is sent to the corresponding managing station via transmitter located in that area. Maintenance of manholes manually is tedious and dangerous due to the poor environmental conditions inside. It is, therefore dangerous to go inside the manholes for inspection of its current state. To solve all the problems related to underground sanitation, a remote alarm system is necessary for transmitting data collected by the sensors set inside the manhole to the managing station. This project uses Wireless Sensor Networks (WSN) to implement this system. These nodes are composed of controller, memory, transceiver and battery to supply power.

Fig 1a: Manual scavenger clearing blockage in a manhole

Fig 1b: Sensors embedded

2. METHODOLOGY

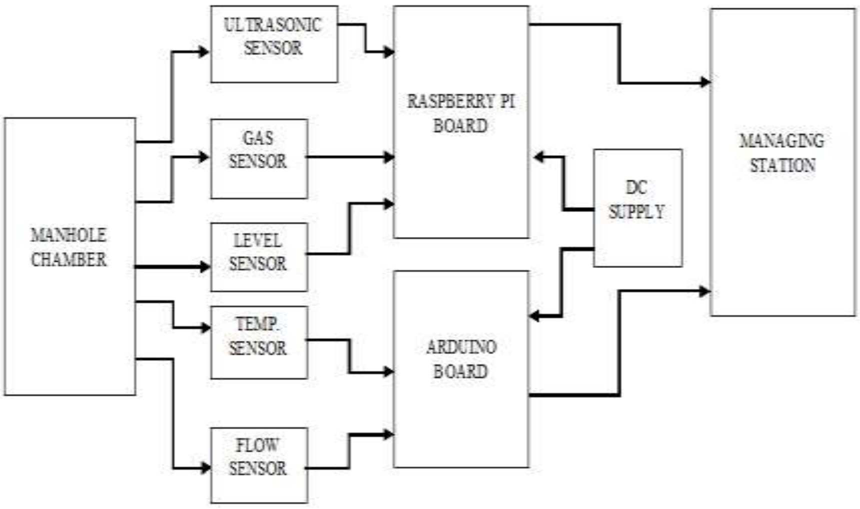


Fig 2. Functional block diagram

The functional block diagram describes the monitoring of manhole in underground drainage system. Any blockages, rise in temperature, explosion due to toxic gases, overflow, manhole lid left open is detected by the sensors. The signals from the sensors are fed to the controller, which is programmed to generate alerts.

In this we use sensors to detect blockage, floods, and gases. The sensors will identify the clogging inside the drainage system and will give information about the location and further actions will be taken care by the municipal. This system consists of

2.1 Sensor network

2. 2 Transmission station

2.1 Sensor Network

Sensor Network consists of:

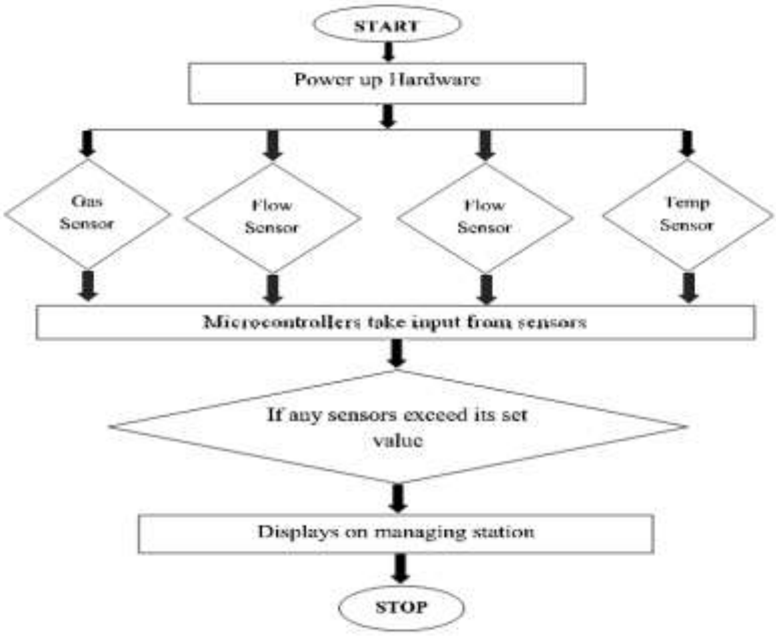
i. Flow Sensor- YF-S201: Flow sensors are mainly used to measure the quantity or the rate of flow of liquids or gases. We are using it to detect overflow.
ii. Float horizontal sensor- FSH-01: Float sensor is used to detect the level of water in the system. This can turn on to be as a pump, alarm and indicator. We are using it to detect blockages in drainage.
iii. Temperature sensor-LM35: Temperature sensor is a device used to measure the hotness or coldness of an object. We are using this sensor to obtain the temperature underground.
iv. Ultrasonic sensor-HC-SR04: Ultrasonic detection is most commonly used in industrial applications to detect hidden tracks etc. We are using it to detect the opening and closing of lid of manhole.
v. Gas Sensor-MQ2: MQ2 Gas sensor detects combustible gasses and smoke.

2.2 Transmission Station

This station helps to send signals from sensors. This station consists of Raspberry Pi model B and Arduino Uno which are the two microcontrollers that are interfaced. The signals received by the Arduino from the sensors is converted from analog signals to digital signals with the help of ADC located in the Arduino board and further is processed and sent to the cloud and Raspberry Pi ,takes this as input data. An alert is displayed in the managing station and an email is sent to the respective authority.

3. FLOW CHART

The working of the whole unit can be explained by using the following flowchart as shown below.



4. PROPOSED SYSTEM

In the proposed method, development of IoT based drainage and manhole monitoring system is designed. This system monitors atmospheric temperature, release of toxic gases, blockages, overflow in drains and manhole lid position. Maximum levels are set and sensors keep monitoring the changing conditions. As the levels reach a maximum set point the sensors detect and send the signal to controller, where it commands the IoT network to generate alerts to the municipal corporation.

Fig 4a. Model of smart city

Fig 4b. Experimental setup

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 07 | July 2019 www.irjet.net p-ISSN: 2395-0072

5. RESULT

This system detects the blockages and water level in the manhole. It also monitors the continuous water flow rate. With the help of sensors temperature, humidity and gas leakage can be identified. The system also informs whether the manhole lid is open or closed by using the ultrasonic sensor.

When a particular sensor reaches the respective threshold level, then that respective value of the sensor will be sent to the microcontroller. Microcontroller updates the live values of all the sensors using IoT. If any problem arises in the manhole, sensor senses it and sends that information to the microcontroller. Furthermore, the microcontroller sends the signal and the exact location of the manhole through IoT to the managing station. Then, an automatic mail is sent by Raspberry Pi. This alerts the person-in-charge to take the required actions regarding the problem occurring inside the manhole.

Fig 5a. Alerts displayed in the managing station

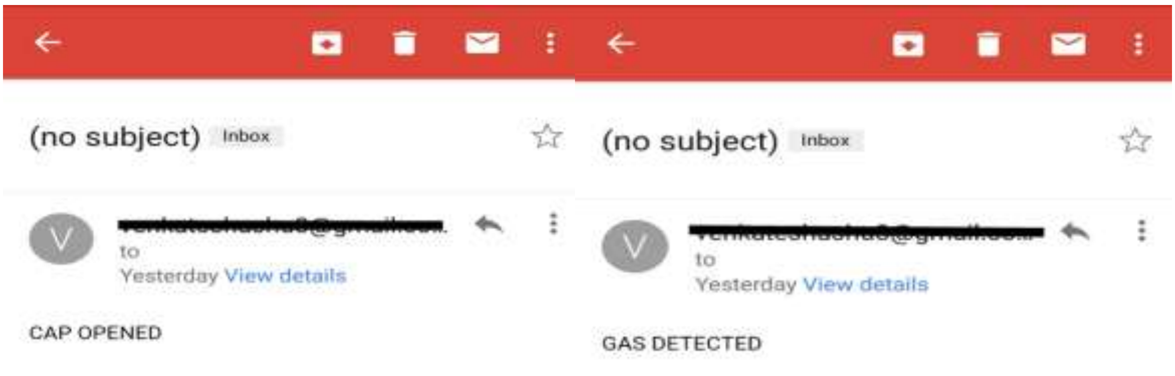


Fig 5b. Alerts sent through email

6. CONCLUSION

Sensor unit automatically senses and updates the live values of the physical parameters like temperature, humidity, water level and flow rate, blockages, and manhole cap is open or closed through IoT. This makes the system smart and automated. The deployment of Wireless Sensor Networks (WSN), helps in the implementation of the Smart cities in a developing countries. This WSN can also be useful in designing of environmental monitoring systems, which helps in monitoring of volcanic activities, flood detectors and other system. By a small modification in the implementation, this project can be used in agriculture fields or other environmental fields to monitor and control the systems.

In future, Smart cities infrastructure could be modified for intelligent communication and management of traffic signals, street lights, transit applications, active lanes, and so on. With the integration of smart devices in a city infrastructure can makes life in a city a lot easier. Also further by using PLC controller and SCADA systems, drainage water can be controlled, monitored and also this water can be used to irrigate plants, clean toilets, etc. This PLC and SCADA systems can be used as a treatment system for drainage water. Primarily, PLC controls the process of sewage treatment plant and SCADA is a remote terminal unit, which monitors and controls the entire area.

7. REFERENCES

[1] Prof Muragesh SK, Santhosha Rao, "Automated Internet of Things For Underground Drainage and Manhole Monitoring Systems For Metropolitan Cities." International Journal of Information & Computation Technology, ISSN 0974-2239 Vol. 4, 2014.

[2] Dhanalakshmi.G, Akhil.S, Francisca Little Flower.M, Haribalambika.R, "Explosion detection and drainage monitoring system by Automation System" International Journal of Innovative research in computer and communication engineering, vol. 6, issue 2, February 2018.

[3] Gaurang Sonawane, Chetan Mahajan, Anuja Nikale, Yogita dalvi, "Smart Real-Time Drainage Monitoring System USing IoT" May 2018, IRE Journals, Vol. 1 issue 11, ISSN: 2456-8880.

[4] Lazarescu, M.T., "Design of a WSN Platform for Long-Term Environmental Monitoring for IoT Applications," Emerging and Selected Topics in Circuits And Systems, IEEE Journal on, vol.3, no.1, pp.45, 54, March 2013.

[5] G.Gowtham, K.Hari Haran, G.Keerthee Rajan, A.Sweeto Jeison, "Sewage level maintenance using IoT" International Journal of Mechanical Engineering and Technical, vol. 9, Issue 2, February 2018.

[6] Timofte, R.; Van Gool, L., "Multi-view manhole detection, recognition, and 3D localisation," Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference, vol., no., pp.188.195, 6-13 Nov. 2011

Superlattices and Microstructures 145 (2020) 106589

Contents lists available at ScienceDirect

Superlattices and Microstructures

journal homepage: www.elsevier.com/locate/superlattices

Effect of copper concentration on CTS thin films for solar cell absorber layer and photocatalysis applications

Sabina Rahaman [a,b], Monoj Kumar Singha [c,*], M. Anantha Sunil [d], Kaustab Ghosh [a]

[a] School of Electronics Engineering, Vellore Institute of Technology, Chennai Campus, Chennai, 600127, India
[b] Department of Electronics & Communication Engineering, BMS Institute of Technology, Bangalore, 560064, India
[c] Department of Instrumentation and Applied Physics, Indian Institute of Science, Bangalore, 560012, India
[d] Department of Electronics & Communication Engineering, BMS College of Engineering, Bangalore, 560019, India

ARTICLE INFO

Keywords:
Cu_2SnS_3 thin films
Ultrasonic spray pyrolysis
Copper chloride concentration
Crystal structure
Absorber layer
Photocatalyst

ABSTRACT

Earth abundant, nontoxic Cu_2SnS_3 (CTS) thin films are prepared using low-cost ultrasonic spray pyrolysis (USP) technique by varying different copper (Cu) precursor (copper chloride) concentration. Influence of Cu concentration on the electrical, optical and structural properties of CTS thin films are studied. XRD, TEM results confirm the formation of tetragonal structure of CTS films with orientation along (112) plane. Raman analysis reconfirms the tetragonal structure of CTS thin film and Cu_2S secondary phase in films with higher Cu concentration samples. Stoichiometric film is found for 0.025M Cu concentration. XPS and PL studies are also performed. The absorption coefficient of CTS films is found to be $\sim 10^5$ cm^{-1} and optical bandgap is varying from 1.32 eV to 1.49 eV. P-type semiconductor nature of the film is found from electrical characteristics with carrier concentration of $\sim 10^{21}$/cm^3. The resistivity of the film decreases from 4.6×10^{-1} to 9.5×10^{-4} Ω-cm as Cu concentration increases. These properties are suitable for solar cell absorber layer. The optimized CTS thin film is used as a photocatalyst and photocatalysis activity of methylene blue (MB) dye is examined under visible light irradiation. A significant photodegradation (~90%) of MB achieved under 3 h visible light irradiation.

1. Introduction

Increasing demands on energy and depletion of non-renewable energy sources at a faster rate give attention to find non-renewable energy source. Solar cell is one of the prominent non-renewable energy source. Researchers are given significant attention across the world to develop thin film solar cell due to its high absorbance and better photovoltaic performance. Materials like CdTe, $CuIn(Se,S)_2$ and $CuInGaSe_2$ (CIGS) are used as a solar absorber materials for fabrication of thin film solar cells. A record efficiency of 23.3% is achieved using Cd–free CIGS solar cells [1]. However, presence of toxic element and fabrication cost of CIGS solar cells leads to search for alternative absorber materials. Among other absorber materials, Cu_2ZnSnS_4 (CZTS) is a suitable solar absorber and photocatalytic material, which can be substitute by existing popular CIGS and CdTe materials [2]. Elements present in CZTS are inexpensive because indium (In) is replaced by low cost Zn and Sn, and it is also safe, and made of earth abundant materials. It has high absorption coefficient ($>10^4$ cm^{-1}), direct band gap (1.4–1.6eV) and forms a p type semiconductor. Highest solar efficiency of 10% is achieved for CZTS thin films obtained from vacuum techniques [3]. The major drawback with CZTS solar cells are formation of single phase

* Corresponding author.
E-mail addresses: monojsingha@yahoo.co.in, monojsingha@iisc.ac.in (M.K. Singha), sunil.ece@bmsce.ac.in (M.A. Sunil).

https://doi.org/10.1016/j.spmi.2020.106589
Received 2 January 2020; Received in revised form 28 April 2020; Accepted 24 May 2020
Available online 6 June 2020

crystalline films, composition control of elements and preventing the possible occurrence of binary and ternary secondary phase such as Cu_2S, SnS_2, Sn_2S_3, Cu_2SnS_3, Cu_3SnS_4 and Cu_4SnS_4 [4]. To overcome these challenges, Chalcogenide based solar cells have been a focus of research due to their suitable chemical and physical properties.

The I-IV-VI group ternary material Cu_2SnS_3 (CTS) exhibits p type semiconductor, high absorption coefficient ($\sim 10^4$ cm^{-1}), direct band gap and earth abundant, makes CTS as a suitable absorber material for solar cell fabrication [5,6]. At higher temperature it exhibits isomorphic phase (cubic) and at lower temperatures it shows polymorphic phase (tetragonal, monoclinic, triclinic) [7–9]. CTS film's band gap can be easily varied in the range of 0.9–1.7 eV based on the different crystal formation [10]. Few research groups explored new semiconductor material which can be used for degradation of environmental pollutants [11–13]. Mostly, textile, chemical, pharmaceutical industries are the main source of these pollutants. Metal-based oxide semiconductors, such as ZnO, TiO_2 as a photocatalyst, are used to remove organic pollutants [14–19]. The drawback of these materials is wide band gap (more than 3.3eV) and sensible under UV light irradiation. But nearly 46% of total solar energy is visible light which cannot be absorbed by these materials [20]. CTS has shown a tunable energy band gap that makes it an appropriate candidate for visible light driven photocatalysis [21–23].

Researchers have explored both Physical and chemical method to prepare CTS thin films. R Chierchia et al. fabricated solar device using sputtering method and reported 3.05% solar efficiency at 520 °C sulfurization temperature [24]. In this work, sulfurization is achieved at 520 °C and 560 °C. All samples are showed monoclinic phase with band gap of 0.92eV and they have reported that sulfurization temperature has no effect to change the bandgap. Robles et al. made CTS films using co-evaporation method and studies the properties of CTS thin films with the influence of Cu/Sn ratio (changes from 0.8 to 1.7) and substrate temperature maintained at 350 °C and 400 °C [25]. Fernandes et al. studies the effect of sulfurization temperature on the properties of CTS films using DC magnetron sputtered techniques. In Sulphur atmosphere films are sulfurized for 10 min and temperature is varied from 350 to 520 °C. Tetragonal, cubic and orthogonal crystal structure of CTS films are found at temperature 350 °C, 400 °C and 520 °C respectively and Cu_2SnS_4 phase is formed due to the evaporation of tin at higher temperature (520 °C) [26]. Tiwari et al. have fabricated CTS based solar cell using dip coating techniques and reported 2% solar efficiency [27]. D.M Berg el al. reported a technique to produce a P type semiconductor CTS material via the annealing of an electrodeposited precursor [28]. Chalapathi et al. studies the effect of temperature on the properties of CTS thin film using spray pyrolysis for solar cell application. They have annealed the as-deposited sample in Sulphur atmosphere to improve the Sulphur deficiency [29]. J.J. Chaudhari have fabricated the solar device (FTO coated glass/-ZnO/CTS/Al contact) using sol-gel techniques followed by sulfurization process for 1 h and reported 1.12% of solar efficiency [30]. Few research groups have used chemical spray techniques for deposition of thin films, studied about the effect of vacuum, sulphur annealing, different precursor on the properties of thin film [31–33].

The present work focused on growth of stoichiometric Cu_2SnS_3 thin films by varying Cu concentration. We have used the automated indigenously developed ultrasonic spray pyrolysis (USP) system. The advantages of USP over conventional spray pyrolysis is that it has narrow distribution size, uniform films, less deposition time and flow rate of gas is not dependent on aerosol flow rate. Ruan et al. have prepared a potassium doped Cu_2SnS_3 films using ultrasonic spray pyrolysis method and they have reported that there is an improvement in solar efficiency by 2 at % [34]. Ignacio et al. have studied the properties of sample which is prepared by ultrasonic spray pyrolysis and then thermally annealed at 380 °C for 50 min in vacuum environment. Mixture of tetragonal and monoclinic phase of Cu_2SnS_3 films are achieved from as-deposited and annealed sample [35]. The automated USP set up was previously used for successful growth of $Cu_2(Zn,Sn)S_4$ and $AgInS_2$ thin films [36,37]. The present work focuses on effect of various Cu precursor concentration on the structural, morphological, optical band gap and electrical properties of CTS thin films. The optimum CTS thin film is also used as a photocatalyst and studied the photodegradation behavior of methylene blue (MB) dye under visible light irradiation.

2. Experimental details

CTS thin films are prepared on glass substrates by automated ultrasonic spray pyrolysis techniques. This automated USP consists of ultrasonic nebulizer (frequency 1.7 MHz), heater, software interface, thermocouple which control the temperature of the heater and spray head which can move in X–Y directions. The distance between ultrasonic nebulizer and substrate heater is 3 cm, this distance is kept constant throughout the experimentation. 60 ml of total precursor solution is taken for deposition, spray flow rate is 2 ml/min and total 30 min of time is required to prepare each sample of CTS thin film.

Copper chloride ($CuCl_2$. H_2O), Tin chloride ($SnCl_2$. $2H_2O$), Thiourea [$CS(NH_2)_2$] are chosen as precursors for Copper, Tin and Sulphur respectively. The glass substrates are thoroughly washed in clean water and soap solution followed by ultrasonication in deionized water, acetone and iso-propanol alcohol for 20min to get high-quality cleaning. Cleaned glass substrates are immediately dried using nitrogen flush and dried in hot air oven at 80 °C for 2hr.

In the present work, the concentration of copper chloride ($CuCl_2$. H_2O) solution is varied from 0.015M to 0.03M in 0.005 steps and samples are labeled as CTS1, CTS2, CTS3 and CTS4. Based on our previous work, we have fixed the substrate temperature at 500 °C [38]. Substrate temperature, molar concentrations of tin chloride (0.01 M) and thiourea (0.2M) are kept constant throughout the experiment. Thiourea concentration is taken in excess due to volatile nature of Sulphur. Salts are measured at required molarities and then they were dissolved in deionized water to form the homogeneous CTS solution. These prepared solutions were finally sprayed on pre heated substrates at 500 °C.

Bruker D8 (Cu Kα line with 1.5405 Å as a radiation source.), operated at 40 mA and 40 kV, X-ray diffraction (XRD) was used to determine crystal structure of CTS films from 10° to 80° in θ-2θ configuration. Horiba LABRAM HR with 532 nm wavelength Argon laser as excitation source was used for Raman analysis on the CTS thin film. Surface morphology and elemental composition of samples are determined using FEI ESEM ZEISS-50 for Scanning electron microscopy (SEM) and Energy Dispersive X-ray analysis (EDX). Bandgap is measured from absorption spectra which is obtained from the UV–Vis spectrometer (SPECORD 5600 UV–vis). Four-point probe method (Four Dimensions probe metal model 280) is used to study of Electrical properties of CTS thin film.

Photocatalytic test was performed in a double jacket photoreactor with 60 ml methylene blue (MB) solution (1×10^{-5} M) under 300 W halogen lamp. Three pieces of CTS thin films (1.5×2.5 cm^2) are kept in the MB solution. Prior to irradiation of light, films with MB solution were kept on magnetic stirrer for 30min in the dark environment to achieve absorption desorption equilibrium between MB and CTS photocatalyst. Then the solution is exposed under visible light which is obtained from 300W halogen lamp. After certain interval, UV–Vis spectrometer is used to measure the absorption spectra of exposed solution. During the experiment process, there is a possibility to evaporate the MB solution due to produce of heat which can be reduced by circulating cold water through the double jacket reactor.

3. Results and discussion

Fig. 1 shows the XRD patterns of the Cu_2SnS_3 films deposited for different Cu precursor concertation. All the samples showed peaks at 28.47°, 33.03°, 47.14° and 56.36° which correspond to (112), (200), (204) and (312) planes and is matched with tetragonal crystal pattern (JCPDS 89-4715) of CTS [39]. Prepared samples are polycrystalline nature due to presence of three planes. Sample CTS4 exhibited additional peaks at 46.26° which corresponds to secondary phases of Cu_2S (JCPDS 12-0227) [40]. The secondary phase Cu_2S formed in CTS4 is due to high concentration of Cu present in the precursor solution. With increase in Cu concentration (Cu > 0.015M), intensity of (112) plane become more prominent which indicates there is an improvement in crystallinity. Further increase in Cu concentration (Cu > 0.025M), intensity of the (112) plane is found to reduce due to the presence of secondary phase Cu_2S. So, we have not carried out further variation in Cu concentration. XRD result confirm that sample CTS3 is highly crystalline in nature without any secondary Cu_2S phase.

Using Debye Scherrer's formula (equation (1)) the average Crystalline size (x) of the thin film was found.

$$x = \frac{0.9\lambda}{\beta cos\theta} \tag{1}$$

Where β is full width at half maximum (FWHM) of the peak and λ is the wavelength of X-ray source (1.5405 Å), θ is the diffraction angle and both β and θ are measured in radians. Crystallite size varies from 11 to 43 nm with increasing of copper molarity and it is reported in Table 1. Higher crystallite size allows less recombination of charge carriers and it will increase the transport of charge carriers which, in turn, increases the efficiency of solar cell.

The TEM analysis of the optimized sample CTS3 is carried out using Talos F200S instrument. TEM and the high resolution TEM (HRTEM) image of CTS3 sample is shown in the supporting information Fig. 1 (a,b). From supporting Fig. 1(b), it is observed the lattice fringes and the corresponding interplanar distance is found to be 3.13 Å which corresponds to the (112) plane of tetragonal structure of the CTS. The SAED pattern of the sample is shown in the supporting Fig. 1(c). From the supporting Fig. 1(c), we can observe that our prepared material is polycrystalline in nature. It shows various plane which indexed to (112), (204) and (312) plane of the tetragonal structure of Cu_2SnS_3. SAED pattern is matching with the diffraction peaks of XRD pattern.

Raman spectroscopy is the technique to find out the crystaline structure and secondary phase present in the thin films. Fig. 2 depicts the Raman images of the CTS film for different Cu concentration. All films exhibited intense peak at 332 cm^{-1} which contributes to A1 vibration mode of tetragonal phase of CTS film [41]. As Cu concentration (Cu > 0.025M) increases, the peak intensity of 332 cm^{-1} is decrease which may happen due to the non-stoichiometric film. All the samples show weak peak at 287 cm^{-1} and 343 cm^{-1} which are attributed to tetragonal phase of CTS film [42]. From Fig. 2(d), we can clearly observe a visible peak at 470 cm^{-1}, which corresponds to Cu_2S phase. As CTS4 has higher concetration of Cu which leads to formation of Cu_2S impurity [43,44]. Both Raman and XRD results confirmed the presense of Cu_2S impurities for CTS4 sample. Small variation in peak values are due to imperfection in structure,

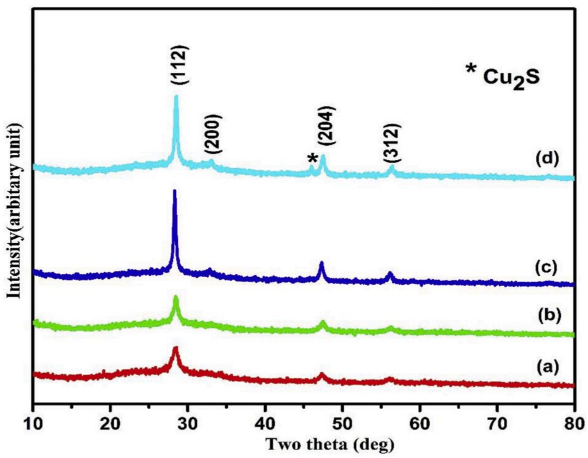


Fig. 1. XRD pattern of films deposited at different Cu concentration, (a) CTS1, (b) CTS2, (c) CTS3 and (d) CTS4.

Table 1
Atomic percentage of elements in CTS Films deposited at different Cu concentrations.

Sample Name	Atomic % of elemental composition			Grain size(nm)	Bandgap(eV)
	Cu (%)	Sn (%)	S(%)		
CTS1	30.11	21.62	48.27	11	1.49
CTS2	33.06	18.92	48.02	21	1.44
CTS3	34.63	16.64	49.73	43	1.35
CTS4	41.28	14.88	43.84	37	1.32

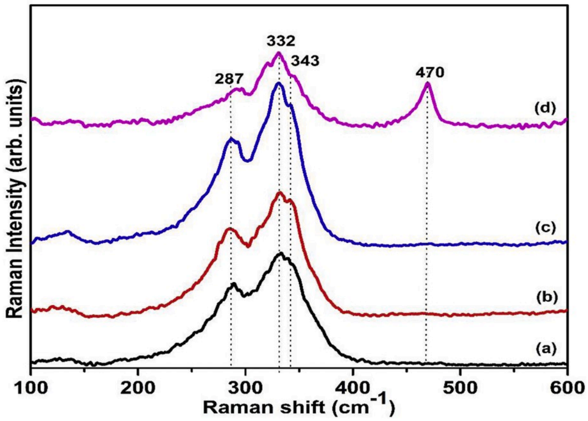


Fig. 2. Raman signature of the CTS films deposited with different Cu precursor concentrations, (a) CTS1, (b) CTS2, (c) CTS3 and (d) CTS4.

composition and growth condition of the CTS films. Also existence of a secondary phase creates extra strain in the film which effects the vibration mode. As a result a slight shift in peak can be obsereved [45].

We have also performed the photoluminescence (PL) spectroscopy of the CTS films. More details on the PL are found in the supporting documents.

EDX analyses are carried out to know composition of the sample and atomic percent of elements are measured within the accuracy of ±2% to determine the atomic percentage composition of CTS thin films. Table 1 shows the atomic percentage of elemental composition value of CTS thin films which are prepared at different Cu concentrations. CTS1 is found to have poor atomic percentage

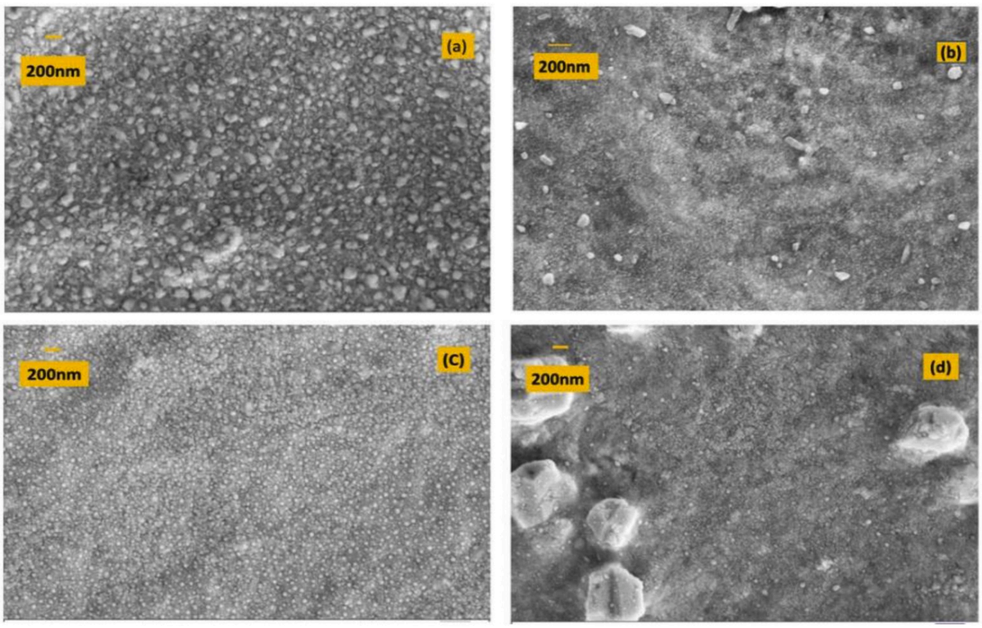


Fig. 3. SEM images of CTS thin films (a) CTS1, (b) CTS2, (c) CTS3, (d) CTS4.

of Cu and having rich atomic percentage of Sn. Sample CTS3 is found to be stoichiometric in nature. As Cu concentration is increased to 0.03M, there is considerable change observed in Cu composition which may be due to the presence of binary phase Cu_2S, which is also noticed in XRD and Raman spectrum. For solar cell application, Cu deficient film is preferable as it contribute to increase the shunt resistance value [46]. If the prepared CTS thin films are not in stoichiometric nature, then electrical, structural and optical properties of films will be affected. Therefore it is very important to control the elemental composition of the film. Moreover, solar efficiency also significantly depends on electrical and optical properties of thin film.

Fig. 3 shows the SEM micrograph of CTS films at different Cu concentrations. As Cu molarity increases, morphology of CTS films is become more uniform, dense and compact structure. Cu concentration at 0.025M (CTS3) shows smooth and compact surface which is highly preferable to collect the minority charge carrier. For Cu at 0.03M (CTS4), some standing wedges are found on the surface. These structures are found due to presence of secondary phase Cu_2S. Chierchia et al. reported solar cell efficiency decreases for Cu rich and non-uniform surface [24]. CTS4 sample is Cu rich (based on EDX data) and non-compact surface which is not apt for solar cell application. All the samples are found to be crack free without any pinholes.

Resistivity, carrier concentration and mobility of CTS thin films are measured using four probe method and all parameter values are listed in Table 2. All films are having p type conductivity. The carrier concentration of CTS films is in the order of $10^{19}/cm^3$ to $10^{21}/cm^3$. Mobility of CTS is found to increase from 0.48 to 2.64 $cm^2/V.s$ with increase in copper concentration. The highest mobility is obtained in CTS4 could be due to the presence of secondary phase Cu_2S which is highly conductive in nature [40]. It is observed that the resistivity of film is highly dependent on concentration of Cu. Cu percentage is maximum at sample CTS4 and it showed least resistivity value. As concentration of Cu increases, resistivity is reduced which is may be due to the higher grain size of the films. As a result it improves the mobility of the film as shown in Table 2 [47].

Optical analysis of CTS thin films is carried out using UV–Vis spectrometer. Optical absorption spectra of CTS films are recorded in the wavelength range from 300 to 1100 nm. Band gap of CTS films are calculated using Tauc formula (equation (2)) [48,49].

$$\alpha h\nu = A\left(h\nu - E_g\right)^n \tag{2}$$

Where Eg is the band gap of material, A is proportionality factor, α is absorption coefficient of material, h is planck's constant, ν is the frequency, n is a parameter value which can be derived from density of state of any material and we have taken 'n' value as 1/2 for direct band gap material. Fig. 4 shows the graph plotted between photon energy ($h\nu$) versus $(\alpha h\nu)^2$. A straight line is drawn at the linear portion of $(\alpha h\nu)^2$ graph and it is extrapolated up to X ($h\nu$) axis. The intersection point at X axis gives the bandgap of CTS thin film which is summarized at Table 1. The band gap energy of CTS thin film varies from 1.32 eV to 1.49eV. It is clearly observed that band gap of CTS films decreases while Cu concentration increases,. Several factors are responsible for reduction of band gap such as crystallite size, quantum confinement effect [30,50]. Crystallite size increases with increase in Cn concentration and possibilities of electron scattering at boundaries of grain reduced. Therefore, electron moves easily from valence to conduction band; suggesting a reduction of band gap. There is also antibonding formation happened in the valence band and conduction band of the films which reduce the bandgap of the films. In the valence band d states in Cu and p states in S participate for antibonding whereas in the conduction band s state of both sulphur and tin participate in the antibonding formation [47].

Study of X-ray photoelectron spectroscopy (XPS) is conducted using the Kratos-Axis DLD photoelectron spectrometer at a pressure of 2.6×10^{-7} Pa. At 15 kV and 10 mA, the monochromatic Al $K_{(alpha)}$ line with an energy of 1.486 KeV is used as the X ray source. XPS analysis of sample CTS3 is carried out to determine the oxidation states of the material and results confirm the presence of single-phase CTS thin film. Fig. 5 (a), (b) and (c) represents the binding energy of core level of Cu 2p, Sn 3d and S 2p. The binding energy is calibrated at the value of C1s that is observed at 284.8eV. The binding energy values for Cu $2p_{3/2}$ and Cu $2p_{1/2}$ respectively have been noticed at 931.52 and 951.27eV, which corresponds to Cu^{+1} valence state of Cu. Absence of characteristic peak at 943eV indicates that the Cu^{+2} oxidation state is not present. The binding energies at 486.12 eV and 498.82 eV correspond to Sn $3d_{5/2}$ and Sn $3d_{3/2}$ states. The obtained energies indicate that Sn is present in +4 oxidation state. The 161.11 eV and 162.02 eV binding energies corresponds to core level of S $p_{3/2}$ and S $p_{1/2}$, which indicate that the sulphur is present in S^{-2} ionic state. From the above analysis, Cu, Sn and S of prepared CTS films present are present in +1, +4 and −2 oxidation state respectively. Obtained results are well matching with the literature [9].

Based on the above discussion it is found that CTS3 film shows best structural, crystallinity, electrical, optical properties which are suitable for solar cell absorber layer.

3.1. Photocatalysis effect

The film deposited at 0.025M Cu concentration showed the optimal properties in terms of crystallinity, elemental composition,

Table 2
Conductivity, resistivity, carrier concentration and mobility of CTS thin films.

Sample name	Type of conductivity	Resistivity (Ω cm)	Carrier Concentration(ions/cm3)	Mobility (cm2/V.s)
CTS1	P	4.6×10^{-1}	3.31×10^{19}	0.48
CTS2	P	3.2×10^{-3}	2.04×10^{21}	0.98
CTS3	P	1.5×10^{-3}	2.86×10^{21}	1.4
CTS4	P	9.5×10^{-4}	3.18×10^{21}	2.64

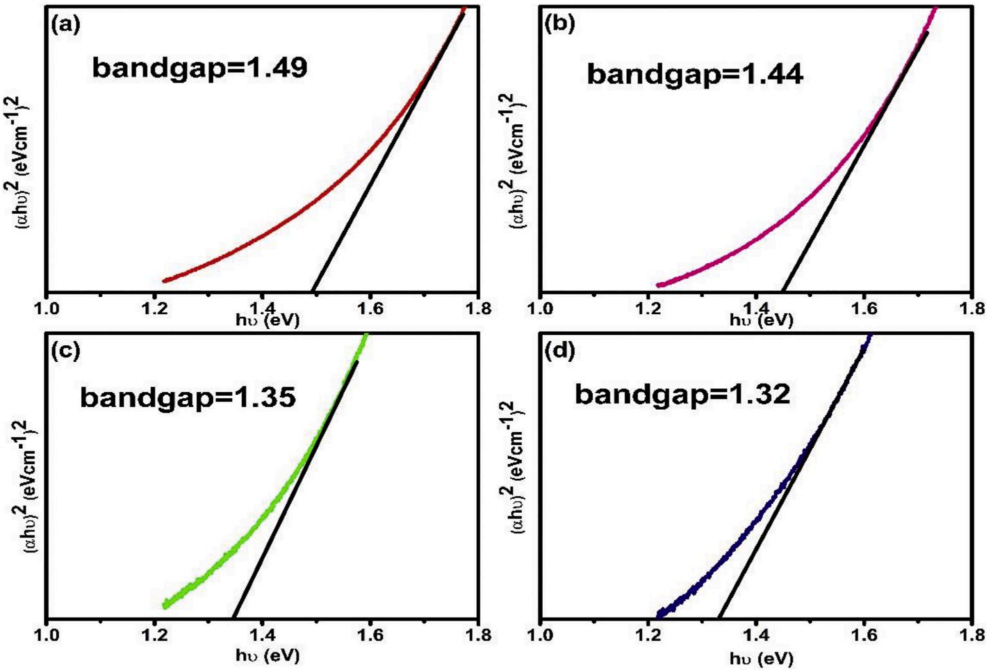


Fig. 4. Band gap of the CTS thin film at different Cu concentration.

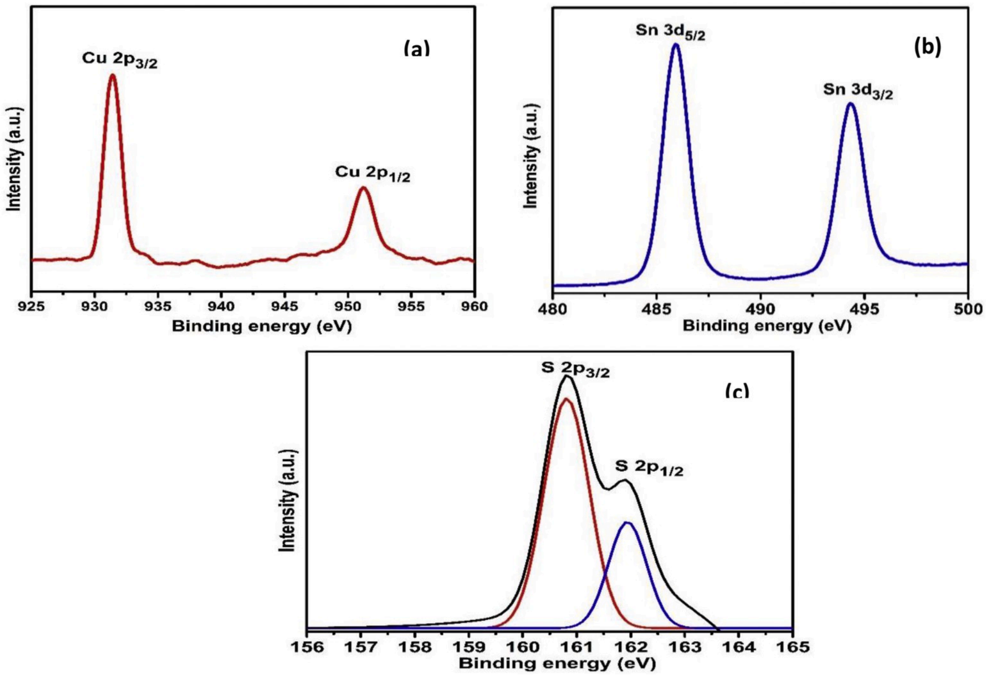


Fig. 5. XPS spectra of CTS3 thin film (a) core level of Cu 2p, (b) core level of Sn 3d (c) core level of S 2p.

optical and electrical analysis. This optimized CTS3 film is further used as a photocatalyst to degrade methylene blue (MB) dye. Here we have degraded 1×10^{-5} M MB (60 ml) dye solution in the presence of three pieces of CTS thin film (1.5×2.5 cm^2) under visible light irradiation. The absorption spectra of methylene blue dye at different time is shown in Fig. 6. Characteristics peak (λmax = 664 nm) is chosen for the reference. Without photocatalyst there is no degradation observed in MB dye. But during the adsorption process it was found the dye was slightly degraded (below 5%). In the presence of CTS catalyst, we have observed 90% degradation in MB dye solution within 3-h time under visible light irradiation. Intensity of absorbance peak reduces gradually as reaction time increases but

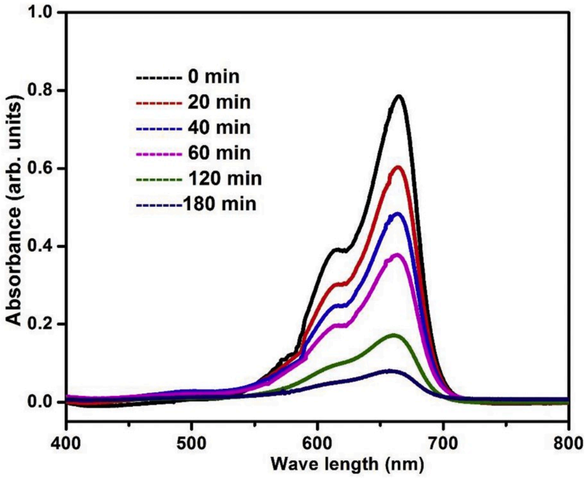


Fig. 6. Absorbance spectra of MB degradation at different time.

there is no shift observed in absorbance wavelength.

Photocatalytic degradation efficiency of a catalyst is given by equation (3) [51,52]:

$$\text{photodegradation}\% = \frac{A_0 - A_t}{A_0} \times 100\% = \frac{C_0 - C_t}{A_0} \times 100\% \tag{3}$$

Where A_0 is the absorbance of MB dye at time 0 and A_t is the absorbance value after addition of CTS film at interval time 't'. Again this is proportional to concentration of dye where C_0 is the concentration of MB dye before catalysis process (at time $t = 0$) and C_t is the concentration of dye after time t under the presence of photocatalysis.

Fig. 7 shows the photodegradation percentage of MB dye at different time. It is observed from Fig. 7 that initially the degradation of MB dye was fast under illumination but as reaction time increases the dye solution degrades exponentially. When halogen light falls on MB dye solution in the presence of CTS photocatalyst, valence band electrons in CTS film acquire light energy and it goes to excited state. These excited state electrons then move from valence to conduction band and it produces hole and electron in valence and conduction band respectively. Afterwards, hole reacts with hydroxyl ion which generates free OH radicals and conduction band electrons react with oxygen molecules which produce superoxide radical anion that again reacts with water molecules and produced free OH radicals. These free OH radicals only mainly control the degradation rate of MB dye.

Fig. 8 shows the first order reaction of MB degradation. Based on Langmuir Hinshelwood model, photocatalysis reaction follows pseudo first order reaction kinetic equation. If the initial concentration of the pollutant is low, then this model is used for photocatalysis degradation process and it is given by Eq. (4) [53]:

$$\ln\left(\frac{C_t}{C_0}\right) = -Kt \tag{4}$$

Where C_t, C_0 and K are initial concentration, final concentration and degradation rate constant respectively. Fig. 8 shows the degradation rate of MB and it is found to be 0.01296 min^{-1} with high adj. R square value of 0.9996.

4. Conclusion

Nontoxic, earth abundant, chalcogenide CTS thin films were successfully deposited on glass substrate. The effects of different Cu concentration on the properties of films were examined. From XRD and Raman analysis, it is conformed that the CTS thin films structures are tetragonal in nature with orientation along (112) plane. Secondary phase, Cu_2S, is observed in the films when higher Cu concentration was used. As copper concentration increases, the band gap of CTS films are decreases from 1.49 to 1.32 eV which is suitable bandgap for absorber layer for solar cell application. P-type conductivity of all samples were confirmed using Hall measurement. Resistivity decrease from 4.6×10^{-1} to 9.5×10^{-4} Ω cm with increase in Cu concentration. When MB dye solution with CTS film is exposed to visible light irradiation, the MB dye degrades to ~90% within 3-h. The rate of degradation of MB dye solution is 0.01296 min^{-1}. Sample CTS3 thin film has stoichiometric elemental composition with high crystallinity. Based on above result we can conclude that CTS3 film is a suitable material for absorber layer for solar cell application and it also helps to remove the organic pollutants from water using visible light irradiation.

Authors agreements

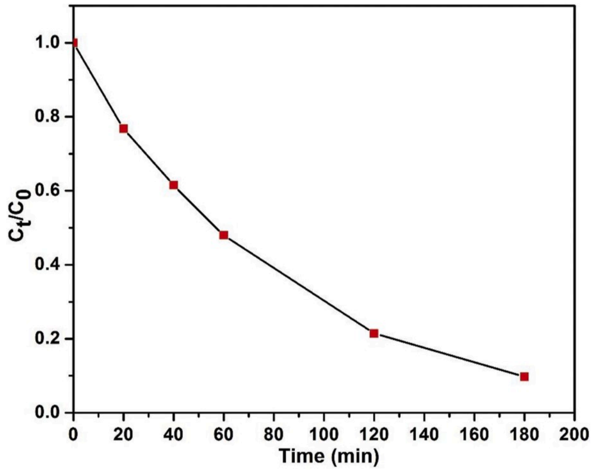


Fig. 7. MB degradation at different time.

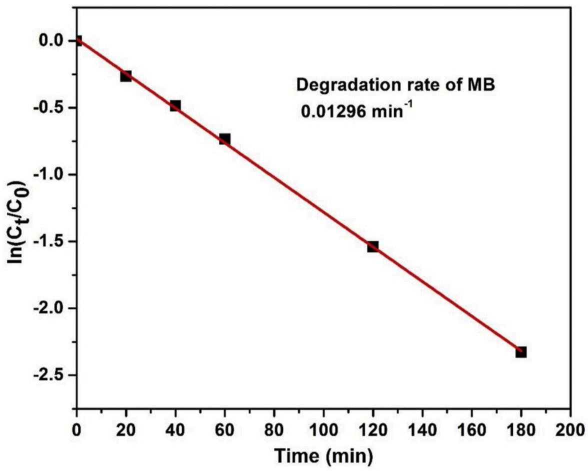


Fig. 8. MB degradation rate.

- Share their article for Personal Use, Internal Institutional Use and Scholarly Sharing purposes, with a DOI link to the version of record on ScienceDirect (and with the Creative Commons CC-BY-NC- ND license for author manuscript versions)
- Retain patent, trademark and other intellectual property rights (including research data).
- Proper attribution and credit for the published work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge "Micro Nano Characterization Facility (MNCF)" at Centre for Nano Science & Engineering (CeNSE), IISc for providing characterization facility. We would also like to thank the Centre for Nano and Soft Matter Sciences, Bangalore, for providing the facility to carry out TEM and PL characterizations.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.spmi.2020.106589.

References

[1] M.A. Green, E.D. Dunlop, D.H. Levi, J. Hohl-Ebinger, M. Yoshita, A.W.Y. Ho-Baillie, Solar cell efficiency tables (version 54), Prog. Photovoltaics Res. Appl. 27 (2019) 565–575, https://doi.org/10.1002/pip.3171.
[2] K.G. Deepaa, P.C. Ramamurthy, M.K. Singha, Mesoporous Cu_2ZnSnS_4 nanoparticle film as a flexible and reusable visible light photocatalyst, Opt. Mater. 98 (2019) 109492.
[3] M.A. Green, Y. Hishikawa, E.D. Dunlop, D.H. Levi, J. Hohl-Ebinger, A.W.Y. Ho-Baillie, Solar cell efficiency tables (version 52), Prog. Photovoltaics Res. Appl. 26 (2018) 427–436.
[4] T. Srinivasa Reddy, R. Amiruddin, M.C. Santhosh Kumar, Deposition and characterization of Cu_2SnS_3 thin films by co-evaporation for photovoltaic application, Sol. Energy Mater. Sol. Cells 143 (2015) 128–134, https://doi.org/10.1016/j.solmat.2015.06.049.
[5] S. Thiruvenkadam, S. Prabhakaran, S. Chakravarty, V. Ganesan, V. Sathe, M.C. Santhosh Kumar, A. Leo Rajesh, Effect of Zn/Sn molar ratio on the microstructural and optical properties of Cu2Zn1-xSnxS4thin films prepared by spray pyrolysis technique, Phys. B Condens. Matter 533 (2018) 22–27, https://doi.org/10.1016/j.physb.2017.12.065.
[6] M. Adelifard, M.M. Bagheri Mohagheghi, H. Eshghi, Preparation and characterization of Cu $_2$ SnS $_3$ ternary semiconductor nanostructures via the spray pyrolysis technique for photovoltaic applications, Phys. Scripta 85 (2012), 035603, https://doi.org/10.1088/0031-8949/85/03/035603.
[7] H. Dahman, S. Rabaoui, A. Alyamani, L. El Mir, Structural, morphological and optical properties of Cu_2SnS_3thin film synthesized by spin coating technique, Vacuum 101 (2014) 208–211, https://doi.org/10.1016/j.vacuum.2013.08.019.
[8] M. Onoda, X. Chen, A. Sato, H. Wada, Crystal structure and twinning of monoclinic Cu2SnS3, Mater. Res. Bull. 35 (2000) 1563–1570.
[9] J. Han, Y. Zhou, Y. Tian, Z. Huang, X. Wang, J. Zhong, Z. Xia, B. Yang, H. Song, J. Tang, Hydrazine processed Cu2SnS3 thin film and their application for photovoltaic devices, Front. Optoelectron. 7 (2014) 37–45, https://doi.org/10.1007/s12200-014-0389-3.
[10] M. Nakashima, T. Yamaguchi, H. Itani, J. Sasano, M. Izaki, Cu2SnS3 thin film solar cells prepared by thermal crystallization of evaporated Cu/Sn precursors in sulfur and tin atmosphere, Phys. Status Solidi 12 (2015) 761–764.
[11] A.B. Djurišić, Y.H. Leung, A.M.C. Ng, Strategies for improving the efficiency of semiconductor metal oxide photocatalysis, Mater. Horizons 1 (2014) 400–410.
[12] N. Shaham-Waldmann, Y. Paz, Away from TiO2: a critical minireview on the developing of new photocatalysts for degradation of contaminants in water, Mater. Sci. Semicond. Process. 42 (2016) 72–80, https://doi.org/10.1016/j.mssp.2015.06.068.
[13] A. Mills, S. Le Hunte, An overview of semiconductor photocatalysis, J. Photochem. Photobiol. Chem. 108 (1997) 1–35.
[14] I.A. Perales-Martínez, V. Rodríguez-González, S.W. Lee, S. Obregón, Facile synthesis of InVO4/TiO2 heterojunction photocatalysts with enhanced photocatalytic properties under UV-vis irradiation, J. Photochem. Photobiol. Chem. 299 (2015) 152–158, https://doi.org/10.1016/j.jphotochem.2014.11.021.
[15] H. Han, R. Bai, Highly effective buoyant photocatalyst prepared with a novel layered-TiO2 configuration on polypropylene fabric and the degradation performance for methyl orange dye under UV–Vis and Vis lights, Separ. Purif. Technol. 73 (2010) 142–150.
[16] Y. Li, J. Yan, Q. Su, E. Xie, W. Lan, Preparation of graphene–TiO2 nanotubes/nanofibers composites as an enhanced visible light photocatalyst using a hybrid synthetic strategy, Mater. Sci. Semicond. Process. 27 (2014) 695–701.
[17] A. Di Mauro, M. Cantarella, G. Nicotra, V. Privitera, G. Impellizzeri, Low temperature atomic layer deposition of ZnO: applications in photocatalysis, Appl. Catal. B Environ. 196 (2016) 68–76.
[18] A. Samad, M. Furukawa, H. Katsumata, T. Suzuki, S. Kaneco, Photocatalytic oxidation and simultaneous removal of arsenite with CuO/ZnO photocatalyst, J. Photochem. Photobiol. Chem. 325 (2016) 97–103.
[19] M.K. Singha, A. Patra, V. Rojwal, D. Kumar, K.G. Deepa, Ultrasonic Spray Pyrolysis Deposited ZnO Thin Film for Photocatalytic Activity" 3rd International Conference on Optoelectronic and Nano Materials for Advanced Technology (icONMAT 2019), 2019. Kochi, India, 3-5th January.
[20] Y. Tan, Z. Lin, W. Ren, W. Long, Y. Wang, X. Ouyang, Facile solvothermal synthesis of Cu 2SnS 3 architectures and their visible-light-driven photocatalytic properties, Mater. Lett. 89 (2012) 240–242, https://doi.org/10.1016/j.matlet.2012.08.117.
[21] S. Vadivel, D. Maruthamani, B. Paul, S.S. Dhar, A. Habibi-Yangjeh, S. Balachandran, B. Saravanakumar, A. Selvakumar, K. Selvam, Biomolecule-assisted solvothermal synthesis of Cu2SnS3 flower/RGO nanocomposites and their visible-light-driven photocatalytic activites, RSC Adv. 6 (2016) 74177–74185, https://doi.org/10.1039/c6ral2068g.
[22] S. Dias, S.B. Krupanidhi, Solution processed Cu2SnS3 thin films for visible and infrared photodetector applications, AIP Adv. 6 (2016), 025217, https://doi.org/10.1063/1.4942775.
[23] Q. Chen, D. Ma, Preparation of nanostructured Cu2SnS3 photocatalysts by solvothermal method, Int. J. Photoenergy 2013 (2013) 2012–2014, https://doi.org/10.1155/2013/593420.
[24] R. Chierchia, F. Pigna, M. Valentini, C. Malerba, E. Salza, P. Mangiapane, T. Polichetti, A. Mittiga, Cu2SnS3 based solar cell with 3% efficiency, Phys. Status Solidi 13 (2016) 35–39.
[25] N. Aihara, A. Kanai, K. Kimura, M. Yamada, K. Toyonaga, H. Araki, A. Takeuchi, H. Katagiri, Sulfurization temperature dependences of photovoltaic properties in Cu 2SnS3-based thin-film solar cells, Jpn. J. Appl. Phys. 53 (2014) 1–6, https://doi.org/10.7567/JJAP.53.05FW13.
[26] P.A. Fernandes, P.M.P. Salomé, A.F. Da Cunha, A study of ternary Cu2SnS3 and Cu3SnS4 thin films prepared by sulfurizing stacked metal precursors, J. Phys. D Appl. Phys. 43 (2010) 215403.
[27] D. Tiwari, T.K. Chaudhuri, T. Shripathi, U. Deshpande, R. Rawat, Non-toxic, earth-abundant 2% efficient Cu2SnS3 solar cell based on tetragonal films direct-coated from single metal-organic precursor solution, Sol. Energy Mater. Sol. Cells 113 (2013) 165–170.
[28] D.M. Berg, R. Djemour, L. Gütay, G. Zoppi, S. Siebentritt, P.J. Dale, Thin film solar cells based on the ternary compound Cu2SnS3, Thin Solid Films 520 (2012) 6291–6294.
[29] U. Chalapathi, Y. Jayasree, S. Uthanna, V. Sundara Raja, Effect of annealing temperature on the properties of spray deposited Cu2 S n S 3 thin films, Phys. Status Solidi 210 (2013) 2384–2390.
[30] J.J. Chaudhari, U.S. Joshi, Fabrication of high quality Cu2SnS3 thin film solar cell with 1.12% power conversion efficiency obtain by low cost environment friendly sol-gel technique, Mater. Res. Express 5 (2018) 36203.
[31] B. Patel, R.K. Pati, I. Mukhopadhyay, A. Ray, Effect of vacuum and sulphur annealing on the structural properties of spray deposited Cu2SnS3 thin films, Vacuum 158 (2018) 263–270.
[32] M. Bouaziz, M. Amlouk, S. Belgacem, Structural and optical properties of Cu2SnS3sprayed thin films, Thin Solid Films 517 (2009) 2527–2530, https://doi.org/10.1016/j.tsf.2008.11.039.
[33] G. Sunny, T. Thomas, D.R. Deepu, C.S. Kartha, K.P. Vijayakumar, Thin film solar cell using earth abundant Cu2SnS3 (CTS) fabricated through spray pyrolysis: influence of precursors, Optik 144 (2017) 263–270.
[34] C. Ruan, J. Tao, C. Zhu, C. Chen, Effect of potassium doping for ultrasonic sprayed Cu2SnS3 thin films for solar cell application, J. Mater. Sci. Mater. Electron. 29 (2018) 12824–12829, https://doi.org/10.1007/s10854-018-9401-9.
[35] I.E. Espinoza, Y. Matsumoto Kuwabara, M.O. Lopez, J. Octavio Estevez Espinoza, Phase composition of Cu 2 SnS 3 thin films prepared by Ultrasonic Spray Pyrolysis as potential photovoltaic material, in: 2018 15th Int. Conf. Electr. Eng. Comput. Sci. Autom. Control. CCE 2018, Institute of Electrical and Electronics Engineers Inc., 2018, https://doi.org/10.1109/ICEEE.2018.8533900.
[36] T. Prabhakar, J. Nagaraju, Ultrasonic spray pyrolysis of CZTS solar cell absorber layers and characterization studies, in: Photovolt. Spec. Conf. (PVSC), 2010 35th IEEE, 2010, pp. 1964–1969.
[37] M.A. Sunil, K.G. Deepa, J. Nagaraju, Growth of AgInS2 thin films by ultrasonic spray pyrolysis technique, Thin Solid Films 550 (2014) 71–75.
[38] S. Rahaman, M.A. Sunil, M.K. Singha, K. Ghosh, Temperature dependent growth of Cu_2SnS_3 thin films using Ultrasonic Spray Pyrolysis for solar cell absorber layer and photocatalytic application", Mater. Res. Express 6 (2019) 106417.
[39] International Centre for Diffraction Data—Reference Code, 01-089-4714, (n.d.).

[40] K.G. Deepa, T.H. Sajeesh, N. Jampana, Opto-electronic properties of Cu 2 ZnSnS 4 thin films grown by ultrasonic spray pyrolysis, J. Electron. Mater. 47 (2018) 530–535.
[41] B. Patel, M. Waldiya, A. Ray, Highly phase-pure spray-pyrolysed Cu2SnS3 thin films prepared by hybrid thermal treatment for photovoltaic applications, J. Alloys Compd. 745 (2018) 347–354.
[42] V. V Brus, I.S. Babichuk, I.G. Orletskyi, P.D. Maryanchuk, V.O. Yukhymchuk, V.M. Dzhagan, I.B. Yanchuk, M.M. Solovan, I. V Babichuk, Raman spectroscopy of Cu-Sn-S ternary compound thin films prepared by the low-cost spray-pyrolysis technique, Appl. Optic. 55 (2016) B158–B162, https://doi.org/10.1364/AO.55.00B158.
[43] H. Yoo, J. Kim, Growth of Cu2ZnSnS4 thin films using sulfurization of stacked metallic films, Thin Solid Films 518 (2010) 6567–6572.
[44] D. Dumcenco, Y.-S. Huang, The vibrational properties study of kesterite Cu2ZnSnS4 single crystals by using polarization dependent Raman spectroscopy, Opt. Mater. 35 (2013) 419–425.
[45] B. Patel, M. Waldiya, A. Ray, Highly phase-pure spray-pyrolysed Cu 2 SnS 3 thin films prepared by hybrid thermal treatment for photovoltaic applications, J. Alloys Compd. 745 (2018) 347–354, https://doi.org/10.1016/j.jallcom.2018.02.220.
[46] A.C. Lokhande, R.B. V Chalapathy, M. He, E. Jo, M. Gang, S.A. Pawar, C.D. Lokhande, J.H. Kim, Development of Cu2SnS3 (CTS) thin film solar cells by physical techniques: a status review, Sol. Energy Mater. Sol. Cells 153 (2016) 84–107.
[47] Y.X. Guo, W.J. Cheng, J.C. Jiang, J.H. Chu, The effect of substrate temperature, Cu/Sn ratio and post-annealing on the phase-change and properties of Cu2SnS3 film deposited by ultrasonic spray pyrolysis, J. Mater. Sci. Mater. Electron. 27 (2016) 4636–4646, https://doi.org/10.1007/s10854-016-4341-8.
[48] T.S. Moss, Optical Properties of Semicoundutors, Acad. Press., NY, 1959, pp. 345–346.
[49] M. Calixto-Rodriguez, H. Martínez, M.E. Calixto, Y. Peña, D. Martínez-Escobar, A. Tiburcio-Silver, A. Sanchez-Juarez, Optical and electrical characterization of AgInS2thin films deposited by spray pyrolysis, Mater. Sci. Eng. B Solid-State Mater. Adv. Technol. 174 (2010) 253–256, https://doi.org/10.1093/nar/gks140.
[50] V. Maheskumar, B. Vidhya, Investigation on the morphology and photocatalytic activity of Cu3SnS4 synthesized by ball milling and solvothermal method, J. Photochem. Photobiol. Chem. 356 (2018) 521–529, https://doi.org/10.1016/j.jphotochem.2017.12.026.
[51] G. Poongodi, P. Anandan, R.M. Kumar, R. Jayavel, Studies on visible light photocatalytic and antibacterial activities of nanostructured cobalt doped ZnO thin films prepared by sol–gel spin coating method, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 148 (2015) 237–243.
[52] X. Zhang, X. Liu, C. Fan, Y. Wang, Y. Wang, Z. Liang, A novel BiOCl thin film prepared by electrochemical method and its application in photocatalysis, Appl. Catal. B Environ. 132 (2013) 332–341.
[53] Monoj Kumar Singha, Aniket Patra, Highly efficient and Reusable ZnO microflower photocatalyst on stainless steel mesh under UV–Vis and natural sunlight, Opt. Mater. (2020), https://doi.org/10.1016/j.optmat.2020.110000. In press.

Materials Research Express

ACCEPTED MANUSCRIPT

Temperature dependent growth of Cu₂SnS₃ thin films using Ultrasonic Spray Pyrolysis for solar cell absorber layer and photocatalytic application

To cite this article before publication: Sabina Rahaman *et al* 2019 *Mater. Res. Express* in press https://doi.org/10.1088/2053-1591/ab3928

Temperature dependent growth of Cu_2SnS_3 thin films using Ultrasonic Spray Pyrolysis for solar cell absorber layer and photocatalytic application

Sabina Rahaman[a,b] , M. Anantha Sunil[c], Monoj Kumar Singha*[d], Kaustab Ghosh[a]

[a] School of Electronics Engineering, Vellore Institute of Technology, Chennai Campus, Chennai-600127.

[b] Department of Electronics & Communication Engineering, BMS Institute of Technology, Bangalore, INDIA-560064.

[c] Department of Electronics & Communication Engineering, BMS college of Engineering, Bangalore, INDIA-560019.

[d] Department of Instrumentation and Applied Physics, Indian Institute of Science, Bangalore-560012.

***Corresponding authors**: sunil.ece@bmsce.ac.in, monojsingha@iisc.ac.in

Abstract:

Ternary chalcogenide Cu_2SnS_3 (CTS) thin films are deposited on glass substrates using low cost ultrasonic spray pyrolysis (USP) technique by varying substrate temperatures 300°C to 550°C. Crystal structure of CTS thin film is found to change its phase from amorphous to tetragonal phase with increase in substrate temperature. Properties of as-deposited Cu_2SnS_3 thin films are studied using various characterization techniques. X-ray powder diffraction (XRD) studies confirm the formation of Cu_2SnS_3 tetragonal phase with orientation along (112) plane with increase in substrate temperature. Raman analysis revealed the formation of binary phase at 550°C substrate temperature. Increase in substrate temperature resulted in stoichiometric nature of Cu_2SnS_3 films at 500°C. The absorption coefficient of CTS films is found to be ~ 10^5 cm^{-1} with band gap ranging from 1.35eV to 1.48 eV. Electrical properties of CTS films exhibited p type conductivity with carrier concentration of the order of $10^{21}/cm^3$. The resistivity of the CTS films is found to vary from 1.5×10^{-3} to 3.2×10^{-3} Ω-cm. Photocatalytic effect of optimized Cu_2SnS_3 thin film under visible light (300W) is examined with methylene blue (MB) dye and ~90% MB is

degraded under 3 hours visible light irradiation. The above properties indicate that Cu_2SnS_3 is a potential candidate to be used for solar cell absorber layer and photocatalytic activity.

Keywords: Cu_2SnS_3 (CTS), Ultrasonic spray pyrolysis, absorber layers for solar cell, crystal structure, photocatalysis, MB degradation

Introduction:

In recent time, solar energy is one of the important sources of the renewable energy due to lack of non-renewable energy sources. Many researchers are using solar absorber materials such as CdTe, $CuInSe_2$, $CuInGaSe_2$ (CIGS) extensively for the fabrication of thin films solar cells. Till date, CIGS solar cells prepared by two-step process holds record efficiency of 22.9% conversion efficiency [1]. Due to its toxicity, presence of rare earth elements and high cost, kesterite materials such as $Cu_2ZnSnSe_4$ (CZTSe), Cu_2ZnSnS_4 (CZTS), $Cu_2ZnSn(SSe)_4$ (CZTSSe) are alternatively used for solar cell application. Cu_2ZnSnS_4 (CZTS) is a potential absorber layer material, which can be replaced with the existing CIGS and CdTe materials. Among the kesterite materials, CZTS absorber material consists of earth abundant nontoxic elements which has direct band gap energy in the range of 1.4 to 1.6 eV, exhibit p-type conductivity and high absorption coefficient (> 10^4 cm^{-1}). CZTS thin film solar cells are fabricated using vacuum method produced a record efficiency of 10.0%[1]. The main drawback of this semiconductor is controlling the elemental composition of CZTS thin films during its synthesis and obtaining single kesterite CZTS phase due to the formation of binary and ternary phases like Cu_2S, SnS_2, Sn_2S_3, Cu_2SnS_3, Cu_3SnS_4 and Cu_4SnS_4 [2]. Ternary materials are introduced to overcome this problem whose opto-electrical properties and crystal structure matches well with the CZTS. The I-IV-VI ternary chalcogenide compound semiconductor Cu_2SnS_3 (CTS) exhibit p type conductivity, direct band gap material, high absorption co-efficient (> $10^4 cm^{-1}$). Elements used in CTS are nontoxic and earth abundant which makes it a potential candidate for absorber material in solar cell. The absorber layer plays a key role in determining the efficiency of solar cell. The presence of defects, secondary phases and non-stoichiometric composition in the absorber compound affects the photovoltaic parameters of the solar cell. The absorber layer should be single phase in crystal structure and free from defects. Phase control of absorber layer can be obtained by controlling the composition of metallic precursors. In addition to these factors, thickness of absorber layer also plays a key role in affecting the conversion efficiency. With the phase transition temperature of CTS being at 780°C, it has the ability to exhibit isomorphic phase (cubic) at higher temperatures and polymorphic phase (tetragonal, monoclinic, triclinic) at lower temperatures[3–5]. Based on the formation of crystal phase, CTS has a unique feature of tuning the band gap in the range of 0.9 eV to 1.7 eV[6]. Presently various research groups are working on photocatalysis to remove the organic pollutants from the water. Source of this pollutants are mainly textile, chemical, pharmaceutical industries etc. Mostly metal oxides like TiO_2, ZnO are used to remove these pollutant[7,8]. Due to band gap of these metal oxides higher than 3.3eV, they are used in UV region. The unique property of tunable band gap helps the CTS films to be used as suitable candidate for visible light (sun light) driven photocatalysis.

Various deposition techniques have been used for growth of CTS thin films using vacuum and non-vacuum methods. Recently, R Chierchia et. al. [9] have grown CTS films on glass substrate using sputtering techniques and reported an solar efficiency of 3%. Katagiri et.al.[10] have got 4% solar efficiency using CTS films and technique used for this work is co-evaporation. Kuku et al.[11] discuss about the characteristics of photovoltaic and optical absorption of CTS thin film prepared by the direct evaporation technique. M. Bouaziz et al.[12]

effectively prepared cubic CTS film by the superposition of SnS_2 and Cu_xS thin films on Pyrex glass substrate using a sulfur annealing process at 550°C. A. Fernandes et al.[13] Prepared sample using sulfurization of DC magnetron sputtered metals and reported the tetragonal and cubic crystal structure of CTS films with the band gap of 1.35eV and 0.96 eV.

Tiwari et. al [14] have used non vacuum dip coating techniques for fabrication of CTS solar cell and they have reported 2% efficiency. D.M Berg el al.[15] have prepared monoclinic CTS film with band gap of 0.93 eV through the annealing of an electrodeposited precursor in a S and SnS environment. Chalapathi et al[16] discussed the annealing effect on the properties of CTS thin film solar cell and film annealed at 400°C, 450°C and 500°C in Sulphur atmosphere. J chaudhari[17] have reported 1.12% of solar power conversation efficiency using sol-gel techniques. Few research groups have prepared CTS thin films using chemical spray technique and reported influence of vacuum annealing and change in precursor concentration [18,19]. Mohamed H. Sayed et al have prepared Cu_2SnS_3 layers by chemical spray pyrolysis and studied the impact of annealing on the device characteristics of prepared absorber layers [20].

In the present work, ultrasonic spray pyrolysis (USP) method is used for the growth of CTS films. The advantages of USP over conventional spray pyrolysis is like narrow distribution size, lesser deposition time, uniform surface of the film, flow rate of the gas is not dependent on flow rate of aerosol [21]. M. W. Alam, et al. have successfully synthesized Cu-SnO_2 nanoparticles using novel and economical ultrasonic spray pyrolysis method. The prepared nanoparticles were used for detection of H_2S sensor applications [22]. Ruan et al have improved the conversion efficiency of ultrasonically sprayed Cu_2SnS_3 films by doping with potassium[23]. Ignacio et al. have studied the phase composition of Cu_2SnS_3 films prepared by ultrasonic spray pyrolysis [24]. The automated set up is used successfully for growth of $Cu_2(Zn,Sn)S_4$ and $AgInS_2$ thin films

[25,26]. In this paper, CTS films are deposited on glass substrates by varying the substrate temperature from 300 to 550°C and the properties of as-deposited films are discussed. Finally, photocatalytic effect of optimized CTS thin film is studied.

Experimental Details

Automated ultrasonic spray pyrolysis system is used for deposition of the CTS films on soda lime glass substrate. The developed automated ultrasonic spray pyrolysis system consists of ultrasonic nebulizer, x-y movement of spray head, substrate heater, temperature controller and software interface. The distance between the substrate and ultrasonic nebulizer is 3 cm. The total volume of precursor solution is kept constant to 60 ml. The spray flow rate and duration of deposition is 2ml/min and 30 min respectively. The above spray conditions are kept constant throughout the experiment. Prior to deposition, glass substrates are thoroughly washed in soap solution followed by ultrasonication in deionized water, acetone and IPA for 20min respectively. The cleaned glass substrates are dried using nitrogen flush and dried in hot air oven at 80°C for 30min. $CuCl_2$. H_2O (0.02M), $SnCl_2$. $2H_2O$ (0.01M), CS $(NH_2)_2$ (0.2M) are taken as precursors and dissolved separately in deionized water to form the CTS precursor solution. CTS solution is atomized using ultrasonic nebulizer with frequency of 1.7 MHz and is sprayed onto hot glass substrates resulting in growth of Cu_2SnS_3 thin films. Due to volatile nature of sulfur, higher molar concentration of sulfur is taken. The substrate temperature of prepared CTS films is varied from 300°C to 550°C in steps of 50°C and samples are named as T300, T350, T400, T450, T500 and T550 respectively.

Crystal structures of CTS films are determined using Bruker D8 X-ray diffractometer with Cu Kα line of wavelength 1.5405Ǻ is used as radiation source. X-ray pattern is recorded from 10° to 80° with operating voltage and current of 40 kV and 40mA respectively. The scan mode used is θ-2θ configuration. Raman analysis is carried out on as-deposited films using Horiba LABRAM HR with 532nm wavelength Argon laser as excitation source. The surface morphology of CTS films is observed using FESEM Quanta-200 Scanning electron microscopy (SEM). The elemental composition of CTS films is measured using Energy Dispersive X-ray analysis (EDX) with higher acceleration potential of 20 kV. To estimate the bandgap of CTS, absorption spectrum is recorded in the range from 300 to 1000nm using UV-Vis spectrometer (SPECORD 5600 UV-vis). Electrical properties of CTS thin film are determined using Hall Effect method (Ecopia HMS5000).

Photocatalytic experiment is carried out in a double jacket photoreactor with 60 ml MB solution (1×10^{-5}M) under 300 W halogen lamp to produce the sunlight environment. Three CTS samples of area 1.5×2.5 cm^2 are placed in the MB solution. Before irradiating the light, films are kept in the solution for 30 minutes under the dark environment for achieving adsorption-desorption equilibrium. Post to light irradiation, solutions are taken at certain intervals to measure the absorption spectra using UV-vis spectrometer. During the light irradiation, cold water is circulated through the double jacket reactor to prevent the evaporation of MB solution.

Results and discussions

Fig.1. depicts the X-ray diffraction pattern of the CTS films prepared at different substrate temperature (Ts) ranging from 300°C to 550°C. Except T300, all the films exhibited tetragonal Cu_2SnS_3 phase with preferential orientation along (112) plane. The other planes (204) and (312) correspond to the tetragonal CTS phase and matches well with JCPDS 89-4715. The three planes observed in the XRD pattern confirm that as-deposited films are polycrystalline in nature.

At lower substrate temperature (T300), there is no prominent peak observed in the XRD pattern resulting in amorphous nature of CTS phase. It also indicated that the substrate temperature set during the deposition is not enough for proper growth of CTS thin films. Hence further characterization of T300 is not carried out except SEM. With increase in Ts (Ts>300°C), the intensity of (112) plane is found to increase resulting in significant improvement of crystallinity of CTS films. Further increase in Ts (Ts>500°C), intensity of (112) plane is found to decrease resulting in poor crystalline films. Among all the samples, T500 is found be highly crystalline in nature.

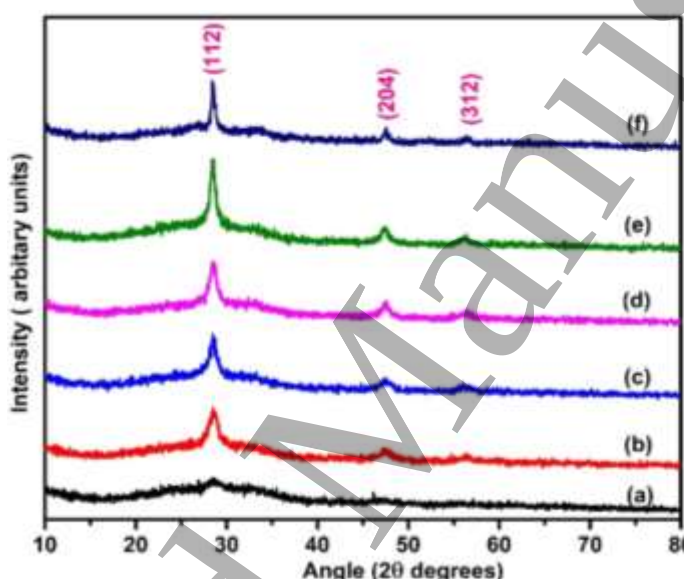


Fig. 1. Pattern of X-ray diffraction of the CTS films, (a) T300, (b) T350, (c) T400 (d) T450, (e) T500 and (f) T550

The average crystallite size (L) of the thin film was found using Debye Scherrer's formula [17].

$$L = \frac{0.9\lambda}{\beta \cos\theta}$$

Where λ is the wavelength of x-ray source (1.5405Å), β is full width at half maximum (FWHM) of the (112) peak and θ is the diffraction angle

Table 1. Crystallite size of all CTS thin films

Sample name	FWHM in radian	Ɵ in Radians	Crystallite size(nm)
T350	0.041	0.25	33.15
T400	0.043	0.25	31.04
T450	0.040	0.25	33.33
T500	0.034	0.25	39.40
T550	0.016	0.25	85.53

The unit of β, θ is measured in radians. With increase in crystallinity, crystallite size of CTS films is found to increase from 31.04 to 85.53 nm. The estimated value of FWHM, theta, crystallite size of the CTS samples derived from XRD is shown in Table 1.

Raman Analysis

Raman analysis is carried out to reconfirm the formation of tetragonal CTS phase and existence of binary phases. Fig.2. depicts the Raman spectra of CTS thin films contain modes at 287 cm^{-1}, 310 cm^{-1}, 333 cm^{-1}, and 343 cm^{-1}. The prominent peak at 333 cm^{-1} corresponds to A1 vibration mode of tetragonal Cu_2SnS_3 structure [27]. With increase in substrate temperature, the intensity of peak 333 cm^{-1} is found to decrease and found to be less dominant for sample T550. The other peaks at 287 cm^{-1} and 343 cm^{-1} also corresponds to tetragonal phase of Cu_2SnS_3. However, peak at 343 cm^{-1} is dominant for sample T550. The peak at 310 cm^{-1} observed for sample T550 is due to formation of secondary phase SnS_2.

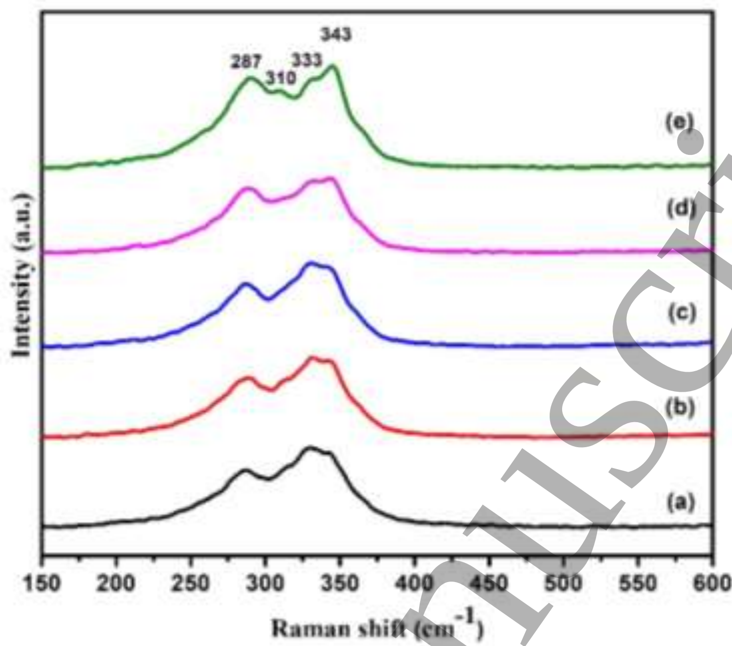


Fig. 2. Raman spectra of the CTS films, (a) T300, (b) T350, (c) T400, (d) T450, (e) T500 and (f) T550

The formation of this secondary phase can be due to deposition carried out at very high temperature [28]. The slight shift in the peak position can be due to defects present in the films. From both the structural analysis, sample T500 is found to be highly crystalline in nature.

EDX analysis is carried out to know about the atomic percentage of chalcogenide CTS thin films prepared at different temperature and summarized at table 2. Atomic percentage of each composition of each sample is measured with the accuracy of ±2%. From the table it is observed that sample T500 is stoichiometric in composition. Due to volatile nature of sulfur, all the samples are found to be sulfur deficient. As temperature increases to 550°C, there is a drastic change in Sn composition which is due to the formation of secondary phase of SnS_2. It is necessary to control the elemental composition of CTS thin films because the electrical and optical properties of the CTS films are greatly affected if films are not stoichiometric in nature.

Table 2. Elemental Composition bandgap of CTS Films deposited at different temperature

Sample name	Atomic percentage of elements			Bandgap
	Cu (%)	Sn (%)	S(%)	(eV)
T350	36.26	19.10	44.64	1.48
T400	34.49	18.53	46.97	1.46
T450	35.82	17.98	46.20	1.45
T500	35.02	17.50	47.48	1.35
T550	31.66	25.42	42.92	1.44

Fig.3. shows the SEM images of the CTS thin films which are grown on glass substrate at temperature 300, 350, 400, 450, 500 and 550℃. It is observed from the image that film deposited at 300°C has adhesion issues with substrate and it peel off from the surface. There is a change in surface morphology observed in samples T350 and T400. In sample T350, porous structured is formed. Sample T400 showed flake like structure which is reported in previous literature[29]. When temperature increased to 450°C, there is noticeable change in structure which is changed to spherical in shape. Also, it is observed that grain size increases with the increase in temperature, it is due to increase in crystallinity of CTS films. Sample prepared at 500℃ shows full dense and smooth microstructure but again at 550℃, porosity is observed in the film which can be due to formation of secondary phase as observed in Raman spectrum[6]. All samples are crack and pin holes free. As seen in figure 4, increase in substrate temperature results in shape transformation from uneven grains to spherical in structure. It can be implied that CTS formation significantly depends on substrate temperature because the binary phases such as CuS and SnS phases reacts with each other to form CTS at specific temperature. Similar phenomenon is

observed growth of Cu_2SnS_3 thin films using vacuum evaporation methods[30]. From SEM images, it is confirmed that sample prepared at 500℃ shows the smooth and compact surface compare to all other sample and this result is a good agreement with the high crystallinity formed by XRD and no secondary phase found by Raman analysis.

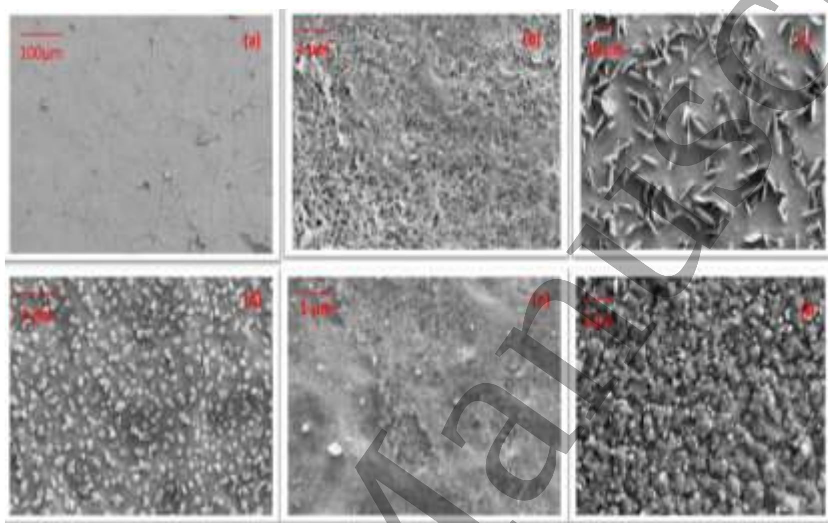


Fig.3. SEM images of CTS thin films (a) T300, (b) T350, (c) T400, (d) T450 (e) T500 and (f) T550

Average thickness of CTS films is found to be 650.4nm. The thickness of CTS films is measured by taking cross section of SEM image as shown in fig.4. Since constant volume (60ml) of solution is sprayed all through the experiment, it is observed that there is a negligible change in the thickness of the film. As shown in figure 4, thickness is measured at three different points and the average of measurement is considered.

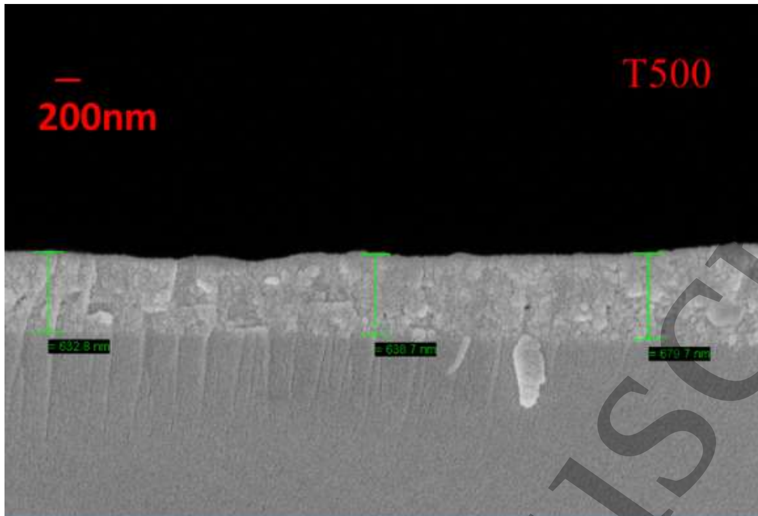


Fig. 4. Cross-sectional image for sample T500

Optical and electrical properties:

At different substrate temperature, the optical bandgap of CTS thin films was calculated using UV-Vis spectroscopy. From the absorption spectra, the bandgap of the CTS film is measured using Tauc formula[31–33] which is given by

$$\alpha h\nu = A\left(h\nu - E_g\right)^n \tag{1}$$

Where Eg is the value of band gap, A is proportionality constant, α is absorption coefficient of material, h is planck's constant, ν is the frequency of excitation, n=1/2 for allowed direct band gap [34]. A graphical representation between photon energy (hν) versus $(\alpha h\nu)^2$ gives the Tauc plot as shown in fig. 5. A slope is drawn at the linear portion of $(\alpha h\nu)^2$ which is extended up to X axis and intersect at point X. It gives the band gap value of CTS thin film.

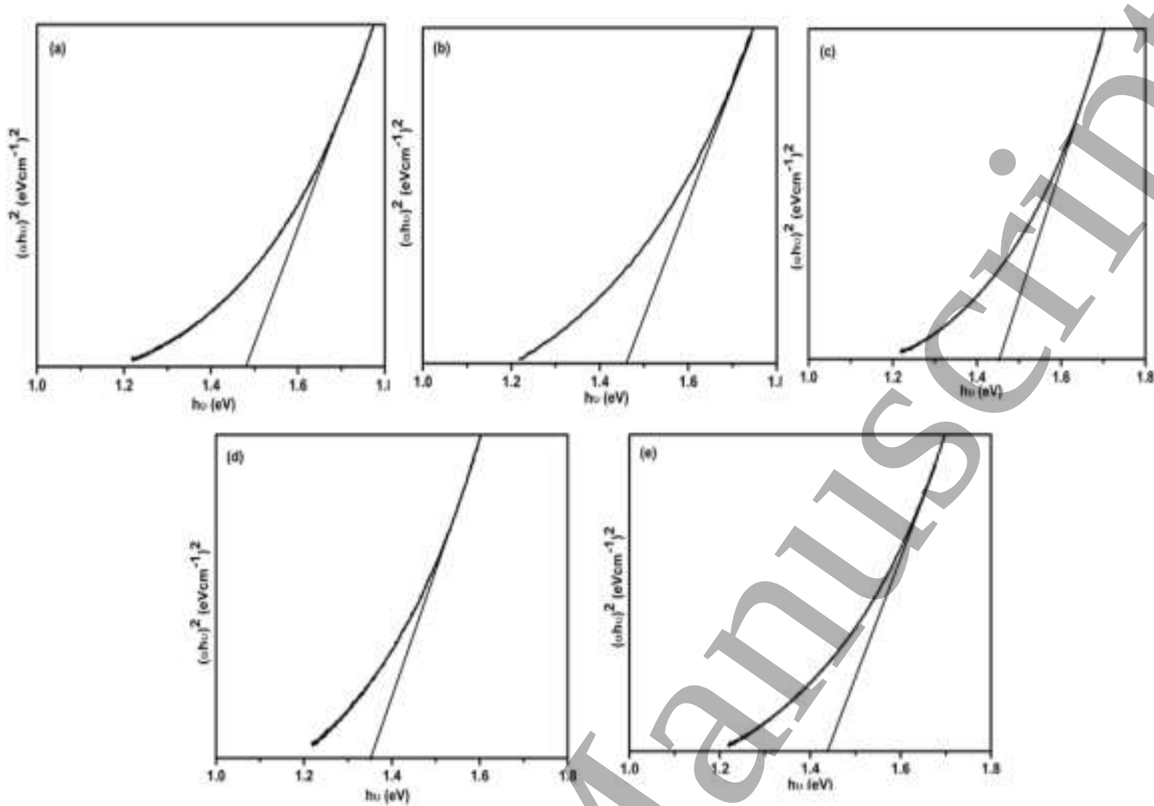


Fig. 5. Band gap of the CTS films, (a) T350, (b) T400, (c) T450, (d) T500 and (e) T550

Increase in substrate temperature resulted in decrease of band gap of the CTS films except for sample T550. The band gap reduces due to antibonding formation between d states in Cu and p states in S in the valence band and antibonding formation between s states of both S and Sn in conduction band [35,36]. The other possible reasons for band gap reduction can be due to improvement in crystallinity as observed from XRD pattern.

Reported value of band gap of tetragonal CTS film is 1.35 [35] which is matching with our result. Band gap of T550 is increased to 1.44 eV, the reason for this increase can be due to the existence of SnS_2 binary phase as observed in Raman spectrum. The band gap value of the CTS thin films is changed from 1.35 to 1.48 eV and the obtained results are suitable for optical applications.

Electrical measurement of the CTS thin films is carries out on the selected sample (T450, T500) using Hall Effect method at room temperature. It is observed that all samples exhibit p- type conductivity. Electrical properties of CTS films exhibited p type conductivity with carrier concentration of the order of $10^{21}/cm^3$. The resistivity of the CTS films is found to decrease from 3.2×10^{-3} Ω-cm to 1.5×10^{-3} Ω-cm. Reduced resistivity can be due to the increase of crystallite size which was observed by other research groups [35].

Photocatalysis effect

Based on above results, it is confirmed that CTS films prepared at temperature 500°C exhibited optimum properties in terms of crystallinity, composition and conductivity. Hence sample T500 is taken as a photocatalyst for degradation of methylene blue (MB) dye under visible light irradiation. Methylene blue is selected as a dye and degradation behavior of MB dye with respect to different time in the presence of the catalyst has been studied. With the help of UV-Vis spectrometer, absorbance spectra of the MB solution have been measured at different time. Fig.6. indicates the photocatalytic effect of CTS thin film at different time interval under the visible light irradiation. Characteristics peak (λmax = 664 nm) is used for reference absorbance peaks. It is observed that there is no degradation of MB solution under the visible light without any catalyst. In the presence of a catalyst in MB dye solution under visible light, it is observed that there is ~90% reduction in absorbance within a three hours duration of time. Photodegradation rate is calculated using following relation [37,38]

$$\text{photodegradation}\% = \frac{A_0 - A_t}{A_0} \times 100\% \qquad (2)$$

Where, A_0 absorbance of dye without photocatalyst, A_t absorbance of dye with the presence of photocatalyst after a time t.

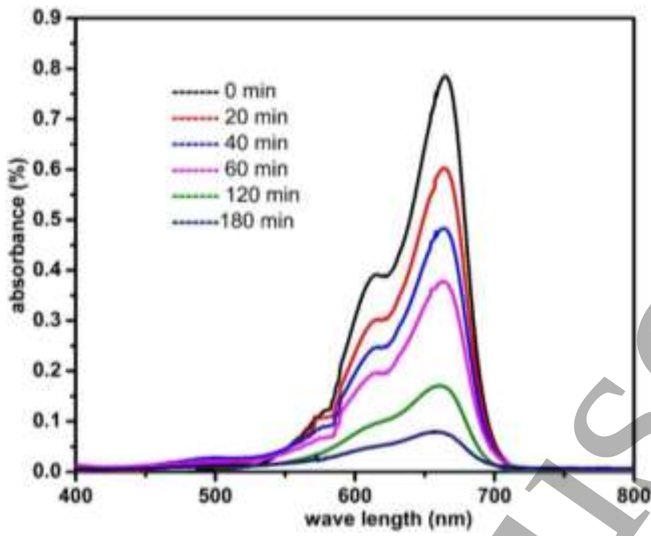


Fig.6: Absorbance spectra of MB photodegradation at different time

Fig.7. indicates photodegradation of MB dye solution as a function of irradiation time with CTS photocatalyst. It is calculated from the absorbance peak at 664 nm for different time interval. From this figure it is clearly observed that initially MB solution is degraded fast under illumination of light. With increase in time, solution degrades exponentially. When light illuminated to the MB dye solution in the presence of catalyst (CTS), it generates electron-hole pair. The generated electron and hole later produce superoxide and super hydroxyl radicals. The produced radicals are mainly used to degrade MB. Fig. 8. shows the first order reaction of MB degradation. First order kinetics is given by the following equation:

$$\ln\left(\frac{C_t}{C_0}\right] = -Kt \qquad (3)$$

Where K is the degradation rate constant and t is the time interval. C_t and C_0 are degradation at time t and 0 respectively. C_t and C_0 are proportional to the absorbance of the solution at time t and 0 respectively. Fig.8. is a straight line in which slope provides the information about degradation rate of MB dye solution. First order kinetic reaction curve is measured by taking logarithmic of Y axis of fig.7. Degradation rate constant is found to be 0.01296 min^{-1} with high adjacent R square value of 0.9994.

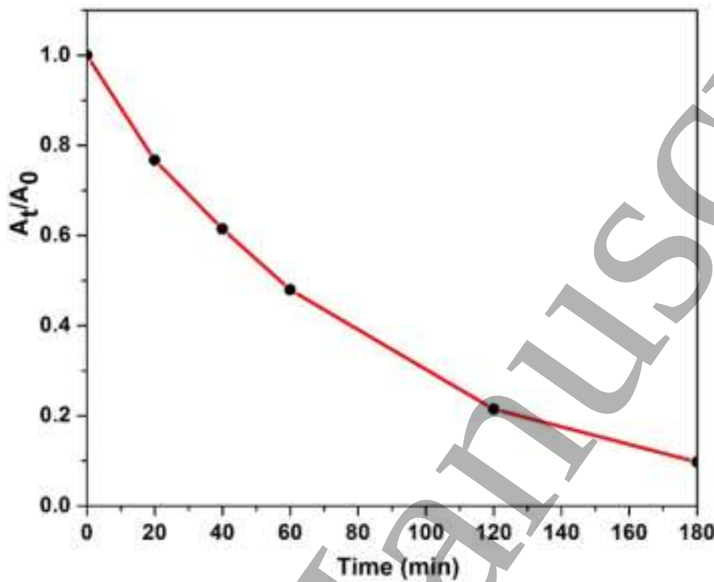


Fig.7. Photodegradation of MB dye solution at different time under the presence of CTS thin films

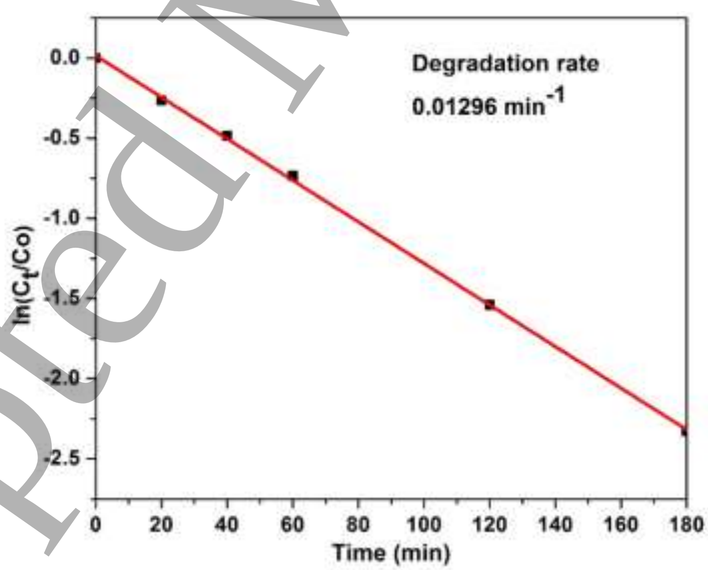


Fig.8. First order kinetic reaction curve of MB degradation

Conclusion

We have successfully deposited chalcogenide Cu_2SnS_3 thin film using low cost automated ultrasonic spray pyrolysis techniques at different substrate temperatures in the order of 300°C to 550°C in steps of 50°C. Highly crystalline films with tetragonal Cu_2SnS_3 phase is obtained for films prepared at substrate temperature of 500°C. Binary phase SnS_2 is found for films deposited at higher substrate temperature (Ts >500°C). The average thickness of the films is found to be 650.4nm. Band gap of optimized film (Ts=500°C) is found to be 1.35eV which closely matches with solar spectrum for maximum conversion efficiency. Electrical properties of CTS films exhibited p type conductivity with carrier concentration of the order of $10^{21}/cm^3$. The resistivity of the films varied from 1.5×10^{-3} to 3.2×10^{-3} Ω-cm. Optimized CTS film (Ts=500°C) degrades ~90% MB dye within three hours of exposure to visible light. Degradation rate of MB solution is found to be 0.01296 min^{-1}. Hence, the above properties indicate that CTS films can be used as potential absorber layer material for thin film solar cell application as well as degrade organic pollutants from water for photocatalyst applications using visible light.

Acknowledgement

The authors would like to thank Department of Instrumentation and Applied Physics, and Micro-nano Characterization facility, Indian Institute of Science (IISc) for providing the characterization facility.

References

[1] Green M A, Hishikawa Y, Dunlop E D, Levi D H, Hohl-Ebinger J and Ho-Baillie A W Y 2018 Solar cell efficiency tables (version 52) *Prog. Photovoltaics Res. Appl.* **26** 427–36

[2] Srinivasa Reddy T, Amiruddin R and Santhosh Kumar M C 2015 Deposition and characterization of Cu_2SnS_3 thin films by co-evaporation for photovoltaic application *Sol.*

Energy Mater. Sol. Cells **143** 128–34

[3] Dahman H, Rabaoui S, Alyamani A and El Mir L 2014 Structural, morphological and optical properties of Cu_2SnS_3 thin film synthesized by spin coating technique *Vacuum* **101** 208–11

[4] Onoda M, Chen X, Sato A and Wada H 2000 Crystal structure and twinning of monoclinic Cu_2SnS_3 *Mater. Res. Bull.* **35** 1563–70

[5] Han J, Zhou Y, Tian Y, Huang Z, Wang X, Zhong J, Xia Z, Yang B, Song H and Tang J 2014 Hydrazine processed Cu_2SnS_3 thin film and their application for photovoltaic devices *Front. Optoelectron.* **7** 37–45

[6] Nakashima M, Yamaguchi T, Itani H, Sasano J and Izaki M 2015 Cu_2SnS_3 thin film solar cells prepared by thermal crystallization of evaporated Cu/Sn precursors in sulfur and tin atmosphere *Phys. status solidi* **12** 761–4

[7] Singha M K, Patra A, Rojwal V, Deepa K G and Kumar D 2019 Ultrasonic spray pyrolysis deposited ZnO thin film for photocatalytic activity *AIP Conference Proceedings* vol 2082 p 30023

[8] Nakata K and Fujishima A 2012 TiO_2 photocatalysis: Design and applications *J. Photochem. Photobiol. C Photochem. Rev.* **13** 169–89

[9] Chierchia R, Pigna F, Valentini M, Malerba C, Salza E, Mangiapane P, Polichetti T and Mittiga A 2016 Cu_2SnS_3 based solar cell with 3% efficiency *Phys. status solidi* **13** 35–9

[10] Kanai A, Toyonaga K, Chino K, Katagiri H and Araki H 2015 Fabrication of Cu_2SnS_3 thin-film solar cells with power conversion efficiency of over 4% *Jpn. J. Appl. Phys.* **54** 08KC06

[11] Kuku T A and Fakolujo O A 1987 Photovoltaic characteristics of thin films of Cu_2SnS_3 *Sol. Energy Mater.* **16** 199–204

[12] Bouaziz M, Amlouk M and Belgacem S 2009 Structural and optical properties of Cu_2SnS_3 sprayed thin films *Thin Solid Films* **517** 2527–30

[13] Fernandes P A, Salomé P M P and Da Cunha A F 2010 A study of ternary Cu_2SnS_3 and Cu_2SnS_4 thin films prepared by sulfurizing stacked metal precursors *J. Phys. D. Appl. Phys.* **43** 215403

[14] Tiwari D, Chaudhuri T K, Shripathi T, Deshpande U and Rawat R 2013 Non-toxic, earth-abundant 2% efficient Cu_2SnS_3 solar cell based on tetragonal films direct-coated from

single metal-organic precursor solution *Sol. Energy Mater. Sol. Cells* **113** 165–70

[15] Berg D M, Djemour R, Gütay L, Zoppi G, Siebentritt S and Dale P J 2012 Thin film solar cells based on the ternary compound Cu_2SnS_3 *Thin Solid Films* **520** 6291–4

[16] Chalapathi U, Jayasree Y, Uthanna S and Sundara Raja V 2013 Effect of annealing temperature on the properties of spray deposited Cu_2SnS_3 thin films *Phys. Status Solidi Appl. Mater. Sci.* **210** 2384–90

[17] Chaudhari J J and Joshi U S 2018 Fabrication of high quality Cu_2SnS_3 thin film solar cell with 1.12% power conversion efficiency obtain by low cost environment friendly sol-gel technique *Mater. Res. Express* **5** 36203

[18] Patel B, Pati R K, Mukhopadhyay I and Ray A 2018 Effect of vacuum and sulphur annealing on the structural properties of spray deposited Cu_2SnS_3 thin films *Vacuum* **158** 263–70

[19] Sunny G, Thomas T, Deepu D R, Kartha C S and Vijayakumar K P 2017 Thin film solar cell using earth abundant Cu_2SnS_3 (CTS) fabricated through spray pyrolysis: Influence of precursors *Optik (Stuttg).* **144** 263–70

[20] Sayed M H, Robert E V C, Dale P J and Gütay L 2019 Cu_2SnS_3 based thin film solar cells from chemical spray pyrolysis *Thin Solid Films* **669** 436–9

[21] Patil P S 1999 Versatility of chemical spray pyrolysis technique *Mater. Chem. Phys.* **59** 185–98

[22] Waqas Alam M, Khatoon U and Qurashi A 2012 Synthesis and characterization of Cu-SnO_2 nanoparticles deposited on glass using ultrasonic spray pyrolysis and their H2S sensing properties *Curr. Nanosci.* **8** 919–24

[23] Ruan C, Tao J, Zhu C and Chen C 2018 Effect of potassium doping for ultrasonic sprayed Cu_2SnS_3 thin films for solar cell application *J. Mater. Sci. Mater. Electron.* **29** 12824–9

[24] Espinoza I E, Matsumoto Kuwabara Y, Lopez M O and Octavio Estevez Espinoza J 2018 Phase composition of Cu_2SnS_3 thin films prepared by Ultrasonic Spray Pyrolysis as potential photovoltaic material *2018 15th International Conference on Electrical Engineering, Computing Science and Automatic Control, CCE 2018* (Institute of Electrical and Electronics Engineers Inc.)

[25] Prabhakar T and Nagaraju J 2010 Ultrasonic spray pyrolysis of CZTS solar cell absorber layers and characterization studies *Photovoltaic Specialists Conference (PVSC), 2010 35th*

IEEE pp 1964–9

[26] Sunil M A, Deepa K G and Nagaraju J 2014 Growth of $AgInS_2$ thin films by ultrasonic spray pyrolysis technique *Thin Solid Films* **550** 71–5

[27] Patel B, Waldiya M and Ray A 2018 Highly phase-pure spray-pyrolysed Cu_2SnS_3 thin films prepared by hybrid thermal treatment for photovoltaic applications *J. Alloys Compd.* **745** 347–54

[28] Lokhande A C, Chalapathy R B V, He M, Jo E, Gang M, Pawar S A, Lokhande C D and Kim J H 2016 Development of Cu2SnS3(CTS) thin film solar cells by physical techniques: A status review *Sol. Energy Mater. Sol. Cells* **153** 84–107

[29] Robles V, Trigo J F, Guillén C and Herrero J 2015 Copper tin sulfide (CTS) absorber thin films obtained by co-evaporation: Influence of the ratio Cu/Sn *J. Alloys Compd.* **642** 40–4

[30] Lokhande A C, Chalapathy R B V, He M, Jo E, Gang M, Pawar S A, Lokhande C D and Kim J H 2016 Development of Cu_2SnS_3 (CTS) thin film solar cells by physical techniques: A status review *Sol. Energy Mater. Sol. Cells* **153** 84–107

[31] Patidar D, Rathore K S, Saxena N S, Sharma K and Sharma T P 2008 Energy band gap and conductivity measurement of CdSe thin films *Chalcogenide Lett.* **5** 21–5

[32] T.S.Moss 1959 Optical Properties of semicoundutors *Acad. Press. NY* 345–6

[33] Chopra K L and Das S R Thin Film Solar Cells (Plenum Press, New York, 1983).

[34] Calixto-Rodriguez M, Martínez H, Calixto M E, Peña Y, Martínez-Escobar D, Tiburcio-Silver A and Sanchez-Juarez A 2010 Optical and electrical characterization of $AgInS_2$ thin films deposited by spray pyrolysis *Mater. Sci. Eng. B Solid-State Mater. Adv. Technol.* **174** 253–6

[35] Guo Y X, Cheng W J, Jiang J C and Chu J H 2016 The effect of substrate temperature, Cu/Sn ratio and post-annealing on the phase-change and properties of Cu_2SnS_3 film deposited by ultrasonic spray pyrolysis *J. Mater. Sci. Mater. Electron.* **27** 4636–46

[36] Zhai Y-T, Chen S, Yang J-H, Xiang H-J, Gong X-G, Walsh A, Kang J and Wei S-H 2011 Structural diversity and electronic properties of Cu_2SnS_3 (X= S, Se): A first-principles investigation *Phys. Rev. B* **84** 75213

[37] Poongodi G, Anandan P, Kumar R M and Jayavel R 2015 Studies on visible light photocatalytic and antibacterial activities of nanostructured cobalt doped ZnO thin films prepared by sol--gel spin coating method *Spectrochim. Acta Part A Mol. Biomol.*

Spectrosc. **148** 237–43

[38] Zhang X, Liu X, Fan C, Wang Y, Wang Y and Liang Z 2013 A novel BiOCl thin film prepared by electrochemical method and its application in photocatalysis *Appl. Catal. B Environ.* **132** 332–41

A statistical approach to spectrum sensing using bayes factor and p-values

Deepa N. Reddy[1], Y. Ravinder[2]

[1]Department of Electronics and Telecommunication Engineering, JSPM's Rajarshri Shahu College of Engineering, Savitribai Phule Pune University, India

[1]Department of Electronics and Communication Engineering, B.M.S Institute of Technology and Management, India

[2]Department of Electronics and Telecommunication Engineering, Pune Institute of Computer Technology, Savitribai Phule Pune University, India

Article Info

Article history:

Received Sep 10, 2018
Revised Apr 4, 2019
Accepted Apr 9, 2019

Keywords:

Bayes factors
Diversity combiners
Meta-analytic approach
Probability of detection
p-value
Spectrum sensing

ABSTRACT

The sensing methods with multiple receive antennas in the Cognitive Radio (CR) device, provide a promising solution for reducing the error rates in the detection of the Primary User (PU) signal. The received Signal to Noise Ratio at the CR receiver is enhanced using the diversity combiners. This paper proposes a statistical approach based on minimum Bayes factors and p-Values as diversity combiners in the spectrum sensing scenario. The effect of these statistical measures in sensing the spectrum in a CR environment is investigated. Through extensive Monte Carlo simulations it is shown that this novel statistical approach based on Bayes factors provides a promising solution to combine the test statistics from multiple receiver antennas and can be used as an alternative to the conventional hypothesis testing methods for spectrum sensing. The Bayesian results provide more accurate results when measuring the strength of the evidence against the hypothesis.

Corresponding Author:

Deepa N. Reddy,
Department of Electronics and Telecommunication Engineering,
JSPM's Rajarshri Shahu College of Engineering,
Savitribai Phule Pune University,
Pune-411046, India.
Email: reddydeepa2680@gmail.com

1. INTRODUCTION

Spectrum sensing in Cognitive Radio (CR) is a binary hypothesis testing problem. It is the decision on the Signal absent, noise-only (null) hypothesis denoted by H_0 and the signal-present (alternate) hypothesis denoted by H_1. Energy Detection (ED) is a very simple technique that decides on hypothesis H_0 or H_1 using the average energy in the observations as the decision statistic. The threshold chosen for ED is dependent on the noise power. This makes the performance of the ED sensitive to uncertainty in the noise variance, especially at low Signal to Noise Ratio (SNR) [1]. Diversity techniques are employed to mitigate the effect of small scale fading and hence provide improvement in the received SNR to achieve higher probability of detection [2, 3].

From the studies, the maximal ratio combining (MRC), equal gain combining (EGC), and selection combining (SC) are the most commonly used diversity combiners. These diversity combining techniques no doubt provide improvement in the receive SNR but they demand the learning of the channel state information (CSI). Hence this increases the implementation complexity [3-7].

Several diversity combining techniques were proposed in the literature which would mitigate the impact of the channel estimation error on the performance of diversity receivers. The non-coherent combining schemes which do not need the CSI are investigated in [1, 7]. Under this category the square law combiner (SLC) and square law selection (SLS) are studied which produces the decision statistic using the outputs of the square-law devices available in each of the diversity branch. In [8] the use of evolutionary algorithms on receiver diversity based on the imperialistic competitive algorithm (ICA) is proposed. It is shown that this combiner does not demand the CSI, and it provides superior performance compared to MRC.

The Goodness of fit test based sensing (GoFT) using Anderson Darling (AD) and Jarque Bera (JB) tests check for the distribution of test statistic under null hypothesis and is independent of the noise distribution [9-11]. These statistical methods were primarily meant to measure the strength of evidence for drawing an accurate decision in hypothesis testing based on a sample. The use of these statistical methods in spectrum sensing is investigated in previous studies but the studies do not focus on receiver diversity using multiple antennas at the CR device.

In this paper two benchmarks coined as p-Value and Minimum Bayes factor (MBF) are used in decision making in hypothesis testing. The first metric, p-Value is viewed as an index of the "strength of evidence" against H_0, with small p indicating an unlikely hypothesis [12] and is widely used in medical research and decision-making. Using the p-Value the compatibility of the data with the null hypothesis is measured. The second metric is the Bayes factor and is often referred to as the "strength of evidence" or "weight of evidence". The Minimum Bayes factor provides the smallest amount of evidence that can be stated for the null hypothesis. The sound theoretical foundation of the MBF and its interpretation allows its usage in both inference and decision making. They have straightforward interpretation as the strength of the evidence in favor of H_1 relative to H_0 [13-16]. This work is the first of its kind in which the use of p-values and Minimum Bayes factors are proposed in the context of spectrum sensing. The paper is organized as follows: Section 2 gives the overview of blind sensing schemes for primary user detection, Section 3 discusses the proposed method, Section 4 discusses the results and Section 5 concludes the paper.

2. SYSTEM MODEL AND PROBLEM FORMULATION

Consider the scenario of Single Input Multiple Output (SIMO) system with one transmit antenna and multiple receiver antennas. Assume that each CR contains M antennas. The M diversity branches are assumed to be sufficiently far from each other. Hence this study takes full advantage of this assumption that the received signals are statistically independent and the correlation among them is considered to be negligible. Corresponding to the signal received in the i^{th} antenna of the CR device the hypotheses H_0 and H_1 are defined as

$$H_0 : x_i[k] = v_i[k]$$

$$H_1 : x_i[k] = h\, s[k] + v_i[k] \quad (1)$$

where, h is the amplitude gain of the channel, i is the antenna index (i=1,2,..M) at each CR, $s[k]$ is the transmitted signal by PU and $v_i[k]$ is the AWGN noise component.

2.1. Existing blind sensing methods

2.1.1. Square law detector

Energy detector (ED) or Square law detector is the most commonly used method for hypothesis testing in a CR environment. Each individual branch at the receiver is provided with an energy detector to provide the instantaneous individual branch energy measurements. The energy of the received signal at the i^{th} branch is Y_i and N is the sample size. The decision static Y_i is compared against a fixed threshold λ.

$$Y_i = \sum_{k=1}^{N} |x_i[k]|^2 \quad (2)$$

The simple hypothesis testing problem is formulated in Equation (3) as

$$Y_i \underset{H_0}{\overset{H_1}{\gtrless}} \lambda \quad (3)$$

The operation in Equation (2) is executed using a square law device provided at each diversity branch of the CR receiver. The following conventional square law combining techniques are used to form a better estimate of the primary user signal [7].

a. Square-law selection

The energy vectors from M diversity branches, Y_1, Y_2, · · · , Y_M are used in SLS. The branch with the highest energy is chosen. The test statistic is given as

$$Y_{sls} = max(Y_1, Y_2, \ldots .. Y_M) \tag{4}$$

b. Square law combining

The energy vectors from M diversity branches, Y_1, Y_2, · · · , Y_M are gathered and combined in SLC to make a combined decision. The test statistic is as

$$Y_{slc} = \sum_{i=1}^{M} Y_i \tag{5}$$

2.1.2. Goodness of fit tests based sensing

Another blind sensing method is the goodness of fit tests (GoFT). These tests are blind non-parametric hypothesis testing method, which decides on the null hypothesis if the received samples follow the noise Cumulative Distribution Function (CDF) denoted as F_0. Let $x[k]$ denote the set of N discrete time vector observations $k=1, 2....N$. Let the i^{th} component of $x[k]$ be denoted as $x_i[k]$, $i=1, 2...M$. The signal detection in noise is therefore given as a simple hypothesis testing problem in [9-11] and is expressed as

Decide on H_0: if $F_n(x) = F_0(x)$

Decide on H_1: if $F_n(x) \neq F_0(x)$ (6)

where, $F_n(x)$ is the empirical CDF of the received sample. The popular goodness of fit tests are:

a. Anderson darling (AD) test

To test the normality of a random sample x[k] the Anderson Darling test statistic formulated in [17] is given as:

$$\mathrm{A_n}^2 = -N - \frac{\sum_{k=1}^{N} (2\mathrm{k}-1)(\ln \mathrm{z_k} - \ln \mathrm{z_{(N+1-k)}})}{N} \tag{7}$$

when the mean and variance of the sample are unknown the adjusted AD statistic as given in [18] is

$$\mathrm{A} = \mathrm{A_n}^2 \left(1 + \frac{0.75}{N} + \frac{2.25}{N^2}\right) \tag{8}$$

where $z_k = F_0(y_k)$ is the assumed distribution, N denotes the sample size, ln is the natural logarithm with $y_k = (x_i - \breve{x})/S$ where $\breve{x} = \sum {x_k}/{N}$ and $S^2 = \sum (x_k - \breve{x})^2/(N-1)$. The spectrum sensing problem is expressed as:

H_0: $A \leq \lambda_{cv}$

H_1: $A > \lambda_{cv}$ (9)

where, λ_{cv} is a critical value. If A exceeds the critical value then H_0 is rejected. A table of thresholds for different values of P_f is given in [19].

b. Kolmogorov-smirnov (KS) test

In the KS test the distance between Fn (x) and F0 (x) is given by:

$$D_n = \max | F_n(x) - F_0(x) | \tag{10}$$

where $F_n(x)$ is the empirical distribution. If the samples under test are coming from F0 (x), then, D_n converges to 0. If the value of D_n exceeds the critical value then H_0 is rejected. A table of thresholds for different values of P_f is given in [20].

c. Jarque and bera test

The Jarque and Bera (JB) test is another goodness-of-fit test to check for normal distribution. It uses the skewness and kurtosis to determine whether the sample data is from a normal distribution [21]. The JB test statistic is the combination of the squares of normalized skewness and kurtosis and is given as follows:

$$J = \frac{N}{6}\left({\gamma_1}^2 + \frac{(\gamma_2 - 3)^2}{4}\right) \tag{11}$$

where, γ_1 is the skewness and γ_2 is the kurtosis and N is the number of samples. The critical values of the JB test for different sample sizes are given in [21]. The primary user signal is declared present if the Jarque Bera test statistic is greater than the critical value and is declared as noise otherwise. The spectrum sensing problem using JB test can be expressed as

$$H_0: J \leq \lambda_{cv}$$

$$H_1: J > \lambda_{cv} \tag{12}$$

2.2. Significance of statistical measures

2.2.1. p-value

Fisher justified that the p -Value can be viewed as an index of the "strength of evidence" against H_0, with small p indicating an unlikely hypothesis [12]. The test statistic is used to determine the p-Value using the formula mentioned in Table 1 as given in [18] and the interpretation of the test results are given in Table 2.

Table 1. The p-value formula for Anderson darling test

AD statistic	p-Value Formula
A >153.467	$p = 0$
0.6<A ≤ 153.467	$p = e^{(1.2937-5.709*A+0.0186A^2)}$
0.34 < A ≤ 0.60	$p = e^{(0.9177-4.279*A-1.38A^2)}$
0.20 < A ≤ 0.34	$p = 1 - e^{(-8.318+42.796*A-59.938A^2)}$
A≤ 0.20	$p = 1 - e^{(-13.436+101.14*A-223.73A^2)}$

Table 2. Decision table

Method	Condition	Decision
Classical test	if(test statistic > Critical value)	H_0 is rejected
Classical test	if (test statistic < Critical value)	H_0 cannot be rejected
p –Value	(p -Value < α)	H_0 is rejected
p-Value	(p -Value > α)	H_0 cannot be rejected

The steps involved in hypothesis testing using p-Values given in [12] are as follows:

a. Define the null and alternative hypotheses.
b. Compute the test statistic from the sample data.
c. Determine the p-Value using the test statistic obtained from step 2.
d. Fix the significance level α=0.05 and interpret the results using Table 2.

2.2.2. Bayes factor method

The term Bayes factor (BF) is also called as likelihood ratio. The Bayes factor is often referred to as the "strength of evidence" or "weight of evidence". Bayes factors show that p-Values greatly overstate the evidence against the null hypothesis**.** The Bayes factors have direct interpretation as the strength of the evidence in favor of H_1 relative to H_0. The use of Bayes factors can avoid the misinterpretations that arise from dependency on the p-Value in decisions [13, 14]. Minimum Bayes factors have the advantage that they do not depend on the prior probability. The proof of the minimum Bayes factor as function of the p-Value is given in [14]. Figure 3 shown Categorization of Bayes Factors BF<1 into levels of evidence against H_0.

$$minBF(p) = \begin{Bmatrix} -eplogp \; for \; p < \frac{1}{e} \\ 1 \quad otherwise \end{Bmatrix} \quad (13$$

Table 3. Categorization of bayes factors BF<1 into levels of evidence against H_0 as given in [14]

Strength of Evidence	Bayes factor
Weak	1 to 1/3
Moderate	1/3 to 1/10
Substantial	1/10 to 1/30
Strong	1/30 to 1/100
Very strong	1/100 to 1/300
Decisive	<1/300

3. PROPOSED METHOD

The block diagram of the proposed method is given in Figure 1. This paper adopts the following statistical methods to integrate the statistical measures from independent tests [22-25] to have an overall assessment on the detection of the primary user signal activity.

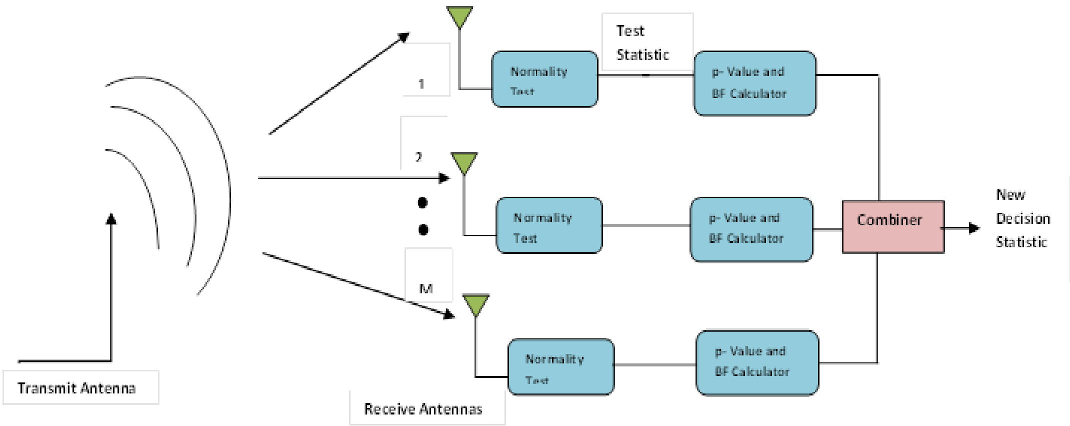


Figure 1. Block diagram of the proposed method

3.1. p-value based diversity combiner

The M independent samples $x_i[k]$ i=1, 2…M are received from M diversity branches of the CR receiver. The test statistics (A_1,A_2….A_M) and its corresponding p-Values (p_1,p_2…..p_M) are computed. This study adopts the following statistical methods to integrate the p -Values from independent tests [22, 23] to have an overall assessment on the detection of the primary user signal activity.

3.1.1. Fisher's test

A popular method of combining the p-Values is the Fisher's method [22]. Let p_1, p_2, …, p_M be the significance probabilities of the test statistic A or J from the i^{th} sample received from each diversity branch of the CR receiver. The joint assessment of the normality is based on the M values of the statistic. The different significance probabilities obtained from M diversity branches are combined using Fisher's method as given below.

$$F_T = -2ln(\sum_{i=1}^{M} p_i) \tag{14}$$

3.2. Bayes Factor based diversity combiner

From the values (p_1,p_2…..p_M) their corresponding Minimum Bayes factors are computed from (13). In the context of spectrum sensing the following method is proposed to integrate these statistical measures from independent tests [24] to have an overall assessment on the detection of the primary user signal activity. The method proposed for combining the data is by calculating the product of the Bayes factor calculated from M independent samples and is defined as the Group Bayes Factor (GBF) as given in [24].

$$GBF_{ij} = \prod_{n=1}^{M} BF_{i,j}^{(n)} \tag{15}$$

where, the subscripts i,j refer to the hypothesis models being compared, and the bracketed superscript refers to the M-th sample. Since the measured data is treated as conditionally independent samples, the probabilities are multiplied.

Algorithm 1. Statistical approach to spectrum sensing

1. Obtain *M* observation samples from each of the diversity branches of the CR node.
2. Let Z_i, (i=1…*M*) be the observation vector. Sort the observations from each branch in ascending order.
3. Calculate the AD test statistic using Equation (7) and (8). Let A_i (i=1…*M*) denote the test statistic obtained for *M* diversity branches.
4. Using the formula given in Table 1 calculate the p-Value p_1 p_2, …, p_M and their respective MBFs using Equation (13).
5. The MBFs and p-Values from *M* diversity branches are combined using Equation (14) and (15) to obtain the new decision statistic.
6 Reject null Hypothesis if the new decision statistic is less than the predefined significance level.

4. RESULTS AND ANALYSIS

4.1. Monte carlo simulations

The performance analysis of spectrum sensing using receiver diversity in a CR environment are carried out using 1) Conventional Hypothesis Testing and 2) Statistical Hypothesis testing. The statistical hypothesis testing is carried out using the following two methods 1) p-Values 2) Minimum Bayes Factor.

The detection probability is used as a standard of measurement to determine the sensing accuracy. The following assumptions are made in the simulations.

a. The system model has Single Input Multiple Output.
b. The primary transmitter signal is a sinusoidal pilot signal of known frequency.
c. Additive White Gaussian Noise with μ=0 and σ^2 =1.
d. For the Hypothesis H_1 to be declared true (signal is present)

Method 1: If the p-Values is less than the significance level α =0.05
Method 2: If the minimum Bayes factor is less than 1/100

Figure 2 provides a comparison of the Goodness of fit tests in the context of primary user signal detection in cognitive radio. The Conventional method of hypothesis testing in a CR environment is compared with the statistical method of hypothesis testing. The number of samples in the test is taken as 100. The tests include the the conventional energy detector, Anderson Darling test, Kolmogorov-Smirnov Test and Jarque and Bera Test. Energy detector shows better performance compared to the other tests. But AD test provides better detection performance compared to the other normality tests. The statistical measure coined as p-Value and Bayes Factor are used as statistical measures in hypothesis testing in Figures 3-6.

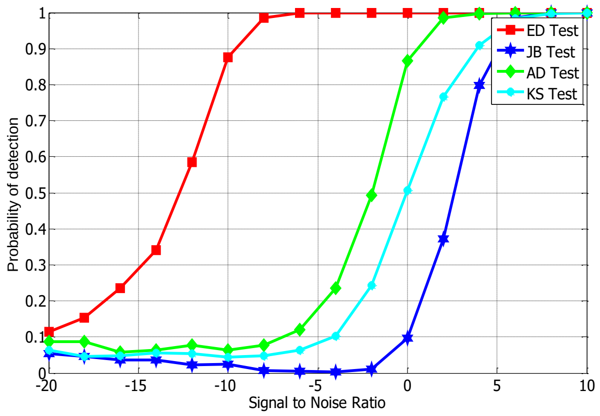


Figure 2. Spectrum sensing using Goodness of fit tests when the primary user is present

The following observations are made:

a. It is observed that the p-Value method overstates the evidence against the null hypothesis than the MBF method.
b. The diversity improves the detection probability in the low SNR regimes using both the methods.
c. The statistical methods provide detection probability close to the conventional square law methods of combining.

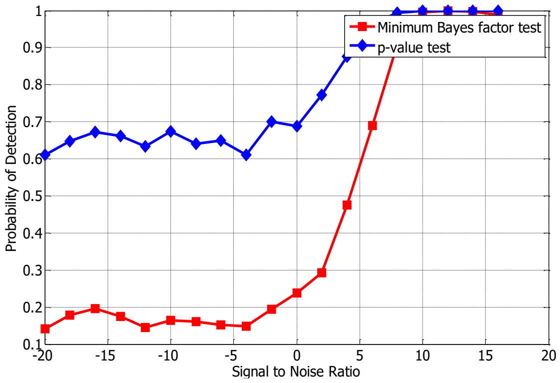


Figure 3. Primary User detection of a Sinusoidal Pilot signal with M=1 ,α=0.05 and N=100 Samples using Anderson Darling Statistic

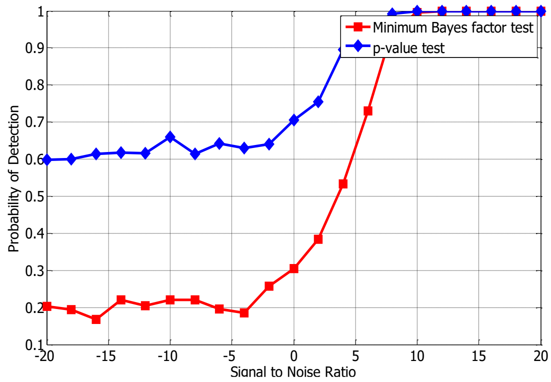


Figure 4. Primary User detection of a Sinusoidal Pilot signal with M=1 ,α=0.05 and N=100 Samples using Jarque Bera Statistic

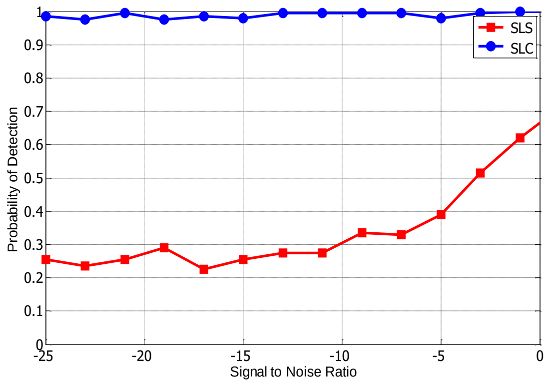


Figure 5. Primary User detection of a Sinusoidal Pilot signal with M=3 , P_f =0.05 and N=100 Samples using Conventional Square Law diversity combining techniques

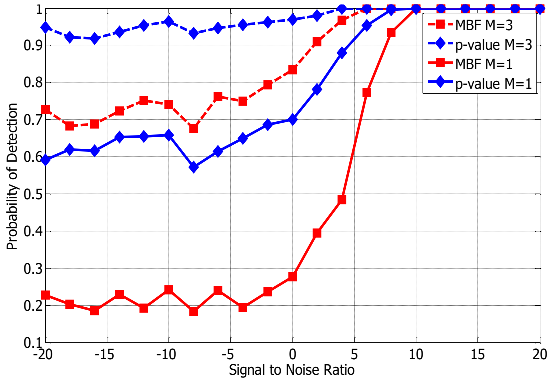


Figure 6. Primary User detection of a Sinusoidal Pilot signal with M=3 , α =0.05 and N=100 Samples using statistical approach a)Minimum Bayes Factor method b) p- Value Method

Given below is a detailed illustration to support point 1 of the observation:

If the MBF=0.262788=1/3.8 with p-value=0.070552 then based on the observed evidence, the result is that the H_1 (the alternative hypothesis) is 3.8 times as likely as H_0 (the null hypothesis). Since MBF/p-Value =0.262788/0.070552 it can be seen that Fisher's p-Value states the amount of evidence against H_0 as 3.7 times as much as the MBF does. This means that the exaggeration of the statistical significance by Fisher' p-Value is almost 4 times as much as that of MBF. Therefore it can be stated that Fisher's p-Value provides less accurate results as a measure of the strength of evidence against H_0 [14].

5. CONCLUSION

In this paper a novel Statistical approach for the hypothesis testing problem in a spectrum sensing environment is proposed. The Minimum Bayes factors and p-Values are proposed for combining the data received from a secondary user equipped with multiple antennas. The effect of these statistical measures in sensing the spectrum in a Cognitive Radio environment is investigated. A ballpark figure of the merits of these diversity combining methods are provided in this study. Results show that p-Values magnifies the evidence against the null hypothesis. The Bayes factor has a straightforward interpretation as the strength of the evidence in favor of H_1 relative to H_0. Also the algorithm proposed improves the detection of the PU in low SNR regimes. Through extensive Monte Carlo simulations it is shown that Bayes Factors provides a promising solution to combine the test statistics from multiple receiver antennas and can be used as an alternative to the conventional hypothesis testing methods for spectrum sensing. Hence this novel statistical approach using bayes factors provide more accurate and relevant test results when measuring the strength of the evidence against the hypothesis.

REFERENCES

[1] S. Atapattu, *et al*., "Energy detection based cooperative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 10, pp. 1232-1241, 2011.
[2] Yucek T., *et al.*, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE communications surveys & tutorials*, vol. 11, pp. 116-30, 2009.
[3] Mietzner J., *et al.*, "Multiple-antenna techniques for wireless communications-a comprehensive literature survey," *IEEE communications surveys & tutorials*, vol. 11, pp. 87-105, 2009.
[4] Wang P., *et al.*, "Multiantenna-assisted spectrum sensing for cognitive radio," *IEEE Transactions on Vehicular Technology*, vol/issue: 59(4), pp. 1791-800, 2010.
[5] Digham F. F., *et al.*, "On the energy detection of unknown signals over fading channels," *IEEE transactions on communications*, vol. 55, pp. 21-24, 2007.
[6] Annavajjala R., *et al.*, "Performance analysis of linear diversity-combining schemes on Rayleigh fading channels with binary signaling and Gaussian weighting errors," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2267-78, 2005.
[7] Herath S. P., *et al.*, "On the energy detection of unknown deterministic signal over Nakagami channels with selection combining," *IEEE Canadian conference on the Electrical and Computer Engineering*, pp. 745-749, 2009.
[8] Akbari M., *et al.*, "Receiver diversity combining using evolutionary algorithms in rayleigh fading channel," *The Scientific World Journal*, 2014.
[9] Arshad K., *et al.*, "Robust spectrum sensing based on statistical tests," *IET Communications*, vol. 7, pp. 808-17, 2013.
[10] Carvalho F. B., *et al.*, "Cognitive spectrum sensing based on statistical tests in fading channels," *Communications (LATINCOM), 7th IEEE Latin-American Conference*, pp. 1-6, 2015.
[11] H. Wang, *et al.*, "Spectrum sensing in cognitive radio using goodness of fit testing," *IEEE Transactions on Wireless Communications*, vol/issue: 8(11), 2009.
[12] Berger J. O., "Could Fisher, Jeffreys and Neyman have agreed on testing?" *Statistical Science*, vol. 18, pp. 1-32, 2003.
[13] Katki H. A., "Invited commentary: evidence-based evaluation of p values and Bayes factors," *American Journal of Epidemiology*, vol. 168, pp. 384-388, 2008.
[14] Held L., *et al.*, "How the maximal evidence of p-values against point null hypotheses depends on sample size," *The American Statistician*, vol. 70, pp. 335-41, 2016.
[15] Goodman S. N., "Toward evidence-based medical statistics. 1: The P value fallacy," *Annals of internal medicine*, vol. 130, pp. 995-1004, 1999.
[16] S. N. Goodman, "Toward evidence-based medical statistics. 2: The Bayes factor," *Annals of internal medicine*, vol. 130, pp. 1005-1013, 1999.
[17] Romeu J. L., "Anderson-Darling: a goodness of fit test for small samples assumptions," *RAC START*, 2003.
[18] R. B. D'Agostino and Stephens M. A., "Goodness-of-fit techniques," Statistics: Text books and monograph, Boca Raton, Florida: CRC Press, 1986.
[19] Stephens M. A., "EDF statistics for goodness of fit and some comparisons," *Journal of the American statistical Association*, vol. 69, pp. 730-7, 1974.
[20] S. Facchinetti, "A procedure to find exact critical values of Kolmogorov-Smirnov test," *Statistica Applicata*, vol. 21, pp. 337-359, 2009.
[21] T. Thadewald and H. Büning, "Jarque–Bera test and its competitors for testing normality–a power comparison," *Journal of Applied Statistics*, vol. 34, pp. 87-105, 2007.
[22] Bhandary M., *et al.*, "Comparison of several tests for combining several independent tests," *Journal of Modern Applied Statistical Methods*, vol. 10, 2011.
[23] Whitlock M. C., "Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach," *Journal of evolutionary biology*, vol. 18, pp. 1368-73, 2005.
[24] Stephan K. E., *et al.*, "Bayesian model selection for group studies. Neuroimage," vol. 46, pp. 1004-17, 2009.
[25] Pettitt A. N., "Testing the normality of several independent samples using the Anderson-Darling statistic," *Applied Statistics*, pp. 156-61, 1977.

A Comparative Analysis of Performance of Several Wavelet Based ECG Data Compression Methodologies

Shanti Chandra, Ambalika Sharma, Girish Kumar Singh

PII: S1959-0318(20)30100-7
DOI: https://doi.org/10.1016/j.irbm.2020.05.004
Reference: IRBM 612

To appear in: *IRBM*

Received date: 12 June 2019
Revised date: 9 February 2020
Accepted date: 15 May 2020

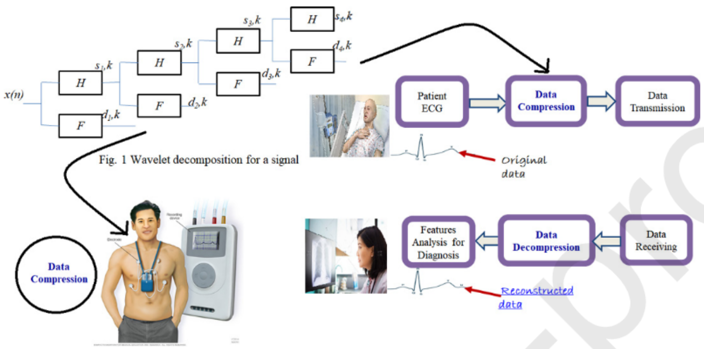
Graphical abstract

IRBM ●●●●, ●●●, ●●●

A Comparative Analysis of Performance of Several Wavelet Based ECG Data Compression Methodologies

Shanti Chandra[a], Ambalika Sharma[b], Girish Kumar Singh[b]

[a] *Department of Electronics and Communication Engineering, BMSIT & M Bangalore, India*
[b] *Department of Electrical Engineering, Indian Institute of Technology, Roorkee, India*



Highlights

- Compression performance comparative analysis.
- Compression ratio.
- Data reconstruction without diagnostic information.

A Comparative Analysis of Performance of Several Wavelet Based ECG Data Compression Methodologies

Shanti Chandra[1], Ambalika Sharma[2] and Girish Kumar Singh[3]
[1]Department of Electronics and Communication Engineering
BMSIT & M Bangalore, India
[2,3]Department of Electrical Engineering
Indian Institute of Technology
Roorkee, India

chndra.shanti@gmail.com[1], bchs38@yahoo.com[2] and gksngfee@gmail.com[3]

Abstract: Compression of an electrocardiogram (ECG) signal has given much consideration to the researchers since the computer-aided analysis of ECG has come into being. In some critical cases, *viz*., astronauts, a person under cardiac surveillance, ambulatory patients and in Holter monitoring system, continuous ECG data recording and transmitting from one location to other location is required. However, the size of the recorded data becomes so voluminous, that its transmission of data becomes practically impossible. In this paper, ECG data compression using wavelet-based techniques are presented, such that: a) wavelet packet transform with run-length encoding (RLE), b) wavelet transform with Huffman encoding, c) wavelet transform with RLE and d) wavelet transform and Lempel ZivWelch (LZW). The results have been tested using MIT-BIH (Massachusetts Institute of Technology/Beth Israel Hospital) arrhythmia databases. The performances of these methodologies are examined in the quantitative and qualitative manner. From Tabular results, it can be observed that the methodology based on WT and LZW provides efficient results in terms of compression ratio (*CR ~=20 to 30*) and peak root mean square difference (*PRD~=0.01 to 1.8*) both, hence the overall QS value is improved.
Keywords: Electrocardiogram, Data compression, Wavelet, Telemedicine, Encoding, Vector quantization.

1. Introduction

The electrocardiogram (ECG) is one of the most important physiological signals that illustrates the electromechanical activity of human heart. In this study, the essential aspects of biomedical signal compression techniques are described [1]. Several methodologies have been developed for data compression of ECG rhythms, which are reported by Jalaleddine *et. al*., [4]. From this unified approach it can be observed that Data compression is one of the impertinent fields of research from the last 5 decades [45]. Two most important applications of data compression are storage requirement reduction and transmission cost reduction. In general, data compression is mainly classified into two types: lossless data compression (LLDC) and lossy data compression (LDC). In the LLDC schemes, exactly original data can be recovered, but low compression ratio (CR) is achieved [2]. Huffman and run-length encoding (RLE) are the famous examples of LLDC. In contrast, the high compression ratio can be achieved using lossy data compression (LDC) methods by eliminating irrelevant information from the data [3]. In most of the ECG data compression applications, *viz*., telemedicine and e-healthcare, Holter monitor systems and in reducing the storage requirement, the lossless data compression does not provide sufficient value compression ratio (CR). Therefore, LDC techniques are more preferred for data compression of ECG rhythms over the LLDC techniques. However, algorithms used for ECG signal compression must be provided the acceptable fidelity. ECG data compression methods are classified into three types: a) direct data compression method, b) transformational data compression methods and c) parameter extraction based data compression methods [4]. Among these three categories, transform based data compression is always preferred, because these techniques provide efficient results in terms of CR and signal reconstruction [37,40]. In these categories, signal is first transformed from one domain (mostly time domain) to the other domain and then rejection of irrelevant coefficients is performed. The examples of transformed based technique are: sine transform, Fourier transform (FT), cosine transform, wavelet transform (WT), *etc*. Wavelet packet (WP) algorithm is a powerfull tool for signal processing and analysis. In this technique the signal is decompose into number of frequency bands, in a chain process. In a study, single lead ECG data compression is performed using WP [5]. WP is used here, because of its high efficiency and flexibility. In another study, detection of R peaks, R-R interval and ECG data compression is done using discrete wavelet transform (DWT). Here, researchers have developed a methodology based on orthogonal WT and adaptive quantization. This work provides a high compression ratio and low implementation complexity [6]. WT is used to compress ECG data along with vector quantization (VQ). VQ is applied to wavelet coefficients to achieve the compression by choosing scales of long durations and low dynamic range retains features integrity of the ECG with a very low bit per sample rate [7].

In the transformed based techniques, the use of encoding is employed to achieve higher value of CR. Different types of coding is used to compress the data, *viz*., run-length encoding (RLE), Huffman encoding, Lempel ZivWelch (LZW) encoding, *etc*. An efficient coding has been proposed by Rajoub [8]. Here, first, the author has used WT to decompose the signal into a number of frequency bands and then applied a new efficient coding. This method provides better performance than other methods in terms of CR and PRD. Other transform methods are also used to perform data compression, such as; fast Fourier transform (FFT) [9], set partitioning in hierarchical trees (LSPIHT) [10], discrete sinc interpolation (DSI) technique [11], DCT and Laplacian pyramid based compression method [12], *etc*. A novel coding strategy is proposed to enhance the coding efficiency of SPIHT at the less significant bit representation of the WT coefficients [13]. Subsequently, in another work, 2-D wavelet technique is used to compress the ECG signal. In this work, the researchers have used modified SPIHT. The methodology is tested by using several records of the MIT-BIH arrhythmia database [14]. Other researchers have also used SPIHT and subband energy compression methodology for data compression of ECG signal. They have utilized 2-D ECG data, in which each row of the array indicates one or more period and amplitude normalized ECG beats. The overall method is superior than 1D ECG signal compression [15]. A novel method for compression of ECG rhythms has been reported by Manikandn & Dandapat using discrete sinc interpolation (DSI) technique. Here, the realization of DSI is done by an efficient discrete Fourier transform (DFT). A new wavelet-based quality measure is introduced by Al-Fahoum. This approach is based on the division of the interest segment in the repeated band, where a score is given based on dynamic range and diagnostic significance. Quantitative and qualitative measurement of the performance is done, which shows that given approach is insensitive to error variation and provides an accurate comparison in original and reconstructed diagnostic features [16]. Another novel approach is used to compress the ECG data is reported by Ku *et al.*, This methodology uses full wavelet coefficients with mean value in the termination. Here, the reversible round-off non-recursive 1-D discrete periodized wavelet transform (NRDPWT) is utilized to resist truncation error propagation. Further, quantization is used to compress the data using octave coefficients [17]. A hybrid approach based on wavelet, uniform scalar zero zone quantizer (USZZQ) and Huffman coding on differencing significance map (DSM) is done for compression of ECG signal. Here, the thresholding value selection is done using the energy packing efficiency (EPE) of each subband after this, the quantization process is utilized. Then after storing process of the indices of significant coefficients is done by creating a significance map. Finally, the Huffman coding is used to encode the map [18]. Another approach of ECG signal compression is done based on the modified embedded zero-tree wavelet

(MEZW). Two different thresholding values are applied to the decomposed coefficients to improve the values of CR and PRD [19]. A methodology for the compression of biomedical signals, such as ECG, electromyogram, electroencephalogram, *etc*., is done with mother wavelet optimization and best-basis WP selection. Here, the researchers have used discreet WP for decomposing the ECG waveform. The optimization process is involved to acquire the optimum wavelet decomposition [20].

A constrained ECG compression algorithm using the block-based DCT is given by Benzid *et al*., This algorithm controlled quality criterion in terms of PRD [21]. SPIHT algorithm is modified to provide better performance of compression called enhanced set partitioning in hierarchical trees (ESPIHT). This method is faster than conventional SPIHT. This method also reduces the number of bits in a bit stream. The ESPIHT is applied to the multichannel ECG data [22]. Another wavelet-based method is used for ECG signal compression, in which first, the DWT is applied to the digitized ECG signal. And then uniform scalar dead-zone quantizer is applied to the DWT coefficients, after this these coefficients are decomposed into four symbol streams, representing a binary significance stream, the signs, the positions of the most significant bits, and the residual bits. An adaptive arithmetic coder with several different context models is employed for the entropy coding of these symbol streams [23]. Lee & Lee have introduced an algorithm to compress ECG signal for the holter monitoring. This algorithm involves different steps, such as; ECG signal differentiation, the R-R interval division and classification, DCT application, window filtering, signal assembling, and Huffman coding. This algorithm provides a good value of CR and PRD [24]. Another work has been reported, in which a new 2-D wavelet-based ECG signal compression algorithm is introduced. An ECG signal (1-D) is first segmented and aligned to a 2-D signal (array), consequently, the two types of correlation of heartbeat signals can be fully utilized. Then after, 2-D WT is applied to the created 2-D ECG data array. After this, a modified VQ is applied to the resultant coefficients of previous stage [25]. DCT and Laplacian pyramid based compression method for ECG signal is proposed in [26], where first, the transformation is done using DCT algorithm, then after a thresholding is applied based on bisection algorithm. Then after, the binary lookup table is formulated for storing the position map. In other work, a new modified SPIHT algorithm is used to code the coefficients of ECG after WT decomposition [27]. An ECG compression algorithm is proposed based on wavelet and linear quality control scheme. In this work, the 1-D reversible round-off non-recursive discrete periodic WT is used to reduce the word-length-growth (WLG) effect [28].

In other work, ECG signal compression and classification is done using the quad level vector for ECG Holter system [29]. In a comparative study, transform based methods, *viz*., FFT, DCT, and WT are analyzed for ECG data compression. Researchers have observed that the DCT based method increases CR by 58.97% than FFT. And WT methods further improve the value of CR by 31% than DCT with low PRD value [30]. Bendifallah *et al*., have introduced an algorithm in which improvement in the DCT method is done, first. And then improved DCT algorithm is applied to the ECG signal. DCT coefficient then quantized using a uniform scalar dead zone quantizer. After this, quantized coefficients are coded using arithmetic encoding method [31]. A new approach based on compressive sensing (CS) is formulated for data compression of ECG signal [32]. Lee *et al*., have developed a real-time signal compression technique for e-healthcare terminals. This methodology includes five procedures, such as; downsampling, classification of one rhythmic period, DCT, filtering using the windowing method and coding using Huffman coding [33]. Another work is presented by Sharma *et al*., for data compression of multichannel ECG signal. Here, researchers have used wavelet transform to decompose all channel coefficients, and then multichannel principle component analysis is used. Multichannel compression is implemented using uniform quantizer and entropy coding of principal component analysis (PCA) coefficients [34]. Kumar *et al*., have used optimized wavelet filter bank based methodology for the compression of ECG rhythms. Authors first, designed wavelet filter bank using a Kaiser window function, which is optimized using the linear iterative algorithm. Then applied it to ECG signal for frequency division. Further, they have used RLE technique to encode the wavelet coefficients [35]. An improved wavelet-based technique is proposed by Huang *et al*., for data compression of ECG signal. Here, first, DWT is applied to the ECG signal, which provides tree structure decomposition. Then, the quantization of the decomposed coefficients is done using vector-scalar quantizer. And then, the context modeling arithmetic coding is applied to quantized coefficients [36]. Another approach of ECG data compression is given by Kumar *et al*., by using WT. Here, researchers have used beta wavelet and lossless coding. Here, the threshold value estimation is done in such a way, the low PRD value (for MIT-BIH Rec. 112, *PRD = 0.6*) can be achieved [37]. Recently, several studies have been done to compress ECG signal, in which researchers have used MIT-BIH dataset [38- 43]. In this work also, MIT-BIH dataset is used to examine the performance of the proposed data compression method.

From the above, literature survey, the following points are observed : a) ECG data compression are always needed in telemedicine and e-healthcare, Holter monitor system, and in reducing the storage requirement for the digital ECG signal, b) lossy data compression is always preferred because it provides high compression performance, c) the decompression performance of an biological signal must be evaluated by comparing the diagnostic features of the signal, d) transform based methods are chosen over other methods because of their abilities of providing suitable CR and PRD, and e) among all transformed technique wavelet based methods are always preferred to process a non-stationary signal. Therefore in this paper, an attempt has been made to examine the performance of different combinations of wavelet and encoding methods.

The rest of the paper is constructed as follows: a brief introduction of the methods, *viz*., wavelet transform, RLE, Huffman coding, LZW coding, *etc*., is presented in Section 2. Section 3 involves the steps of data compression using different combinations of wavelet transform and encoding methodology. In Section 4, experimental results of the methods given in Section 3 are presented. Section 5 has discussions of the obtained results of Section 3 and 4. And in the last conclusions of this study is presented in section 6.

2. Methods

2.1. Wavelet Transform

The wavelet transform (WT) has emerged as an excessive mathematical tool in the area of non-stationary signal analysis because it examines signal structures at different scales. The fundamental principle of WT is: it divides a signal into different functions by using the property of translation and dilation of the mother wavelet ($\varphi(t)$), which acts as a prototype function and can be defined as [6]:

$$\varphi_{ab}(t) = |a|^{-1/2}\,\varphi\left(\frac{t-b}{a}\right) \qquad a,b \in R \tag{1}$$

where, a and b are the dilation and translation function, respectively. When the parameters a and b are restricted to discrete values such as: $a = 2^{m/2}$, $b = n2^{-m}$, then, a new family of discrete wavelets is formulated, which is expressed as [7]:

$$\varphi_{mn}(t) = 2^{m/2}\,\varphi(2^{mt-n}), \tag{2}$$

in this expression, m=$n = \pm 0, \pm 1, \ldots, \pm\infty$ and φ must satisfy the condition given below:

$$\int R^{\varphi(t)}\,dt = 0 \tag{3}$$

WT is classified into two types: continuous wavelet transforms (CWT) and discrete wavelet transform (DWT). CWT can be expressed as [18]:

$$W_f(b,a) = |a|^{-1/2} \int_{-\infty}^{\infty} f(t)\varphi^*\left(\frac{t-b}{a}\right)dt \tag{4}$$

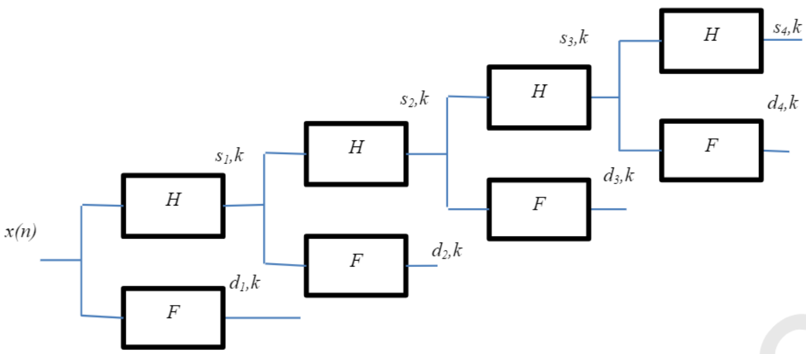


Fig. 1 Wavelet decomposition for a signal

In the above expression, the symbol '*' represents a complex conjugate and multiplication of $|a|^{-1/2}$ and it is used for stabilizing the transformed signal with the same energy at every scale. Adaptive nature and multiresolution property of the WT, makes analysis easy and effective. In the case of CWT, the values of *a* and *b* are continuous over the real number (*R*), which increases the computational complexity. Therefore, DWT is used as a more efficient transform to analyze a signal. DWT is represented by Eqn. (5) [10].

$$Wx(m,n) = \int_{-\infty}^{\infty} x(t)\varphi_{m,n}(t)dt = a_0^{-m/2}\int_{-\infty}^{\infty} x(t)\varphi(a_0^{-m}t - nb_0)dt \tag{5}$$

This expression is achieved by choosing the values: $a = a_0^m$ and $b = nb_0a_0^m$ for different values of *m* and *n*, (*i.e.*, $m{=}n = \pm 0, \pm 1, \ldots, \pm\infty$). For both types of wavelet transforms, condition given in Eqn. (6) should be satisfied.

$$\int \varphi(t)dt \tag{6}$$

When, $a_0 = 2$ and $b_0 = 1$, a new discrete wavelet can be constructed, which is given in Eqn. (7). This wavelet consists of an orthogonal basis for $L^2(R)$.

$$\varphi_{m,n}(t) = 2^{-m/2}\varphi(2^{-m}t - n) \tag{7}$$

In DWT, the input signal is decomposed into different scales by following expression given below:

$$x(t) = \sum_{j=1}^{K}\sum_{k=-\infty}^{\infty} d_j(k)\varphi_{j,k}(t) + \sum_{-\infty}^{\infty} a_k(k)\psi_{K,k}(t) \tag{8}$$

where, $d_j(k)$, $\varphi_{j,k}(t)$ and $\psi_{K,k}(t)$ are detailed signals, discrete analysis wavelets, and discrete scaling functions, respectively. Wavelet decomposition tree is depicted in Fig. 1, and the implementation of DW filter bank can be done using Eqns. (9) and (10).

$$h(n) = \frac{1}{2}\langle\psi(t), \psi(2t-n)\rangle \tag{9}$$

$$g(n) = \frac{1}{\sqrt{2}}\langle \varphi(t), \psi(2t-n)\rangle = \int_{-\infty}^{\infty} (-1)^n h(1-n) \quad (10)$$

The value of the function $d_j(k)$ at the level j can be estimated by using the mathematical function, *i.e.,* convolution function with the signal $f(n)$ at $j-1$ [25]. In wavelet transform only lower frequency bands are decomposed as depicted in Fig. 1, however in wavelet packet decomposition, the detail signals are also decomposed. Fig. 2 shows the decomposing of a signal using wavelet packet algorithm.

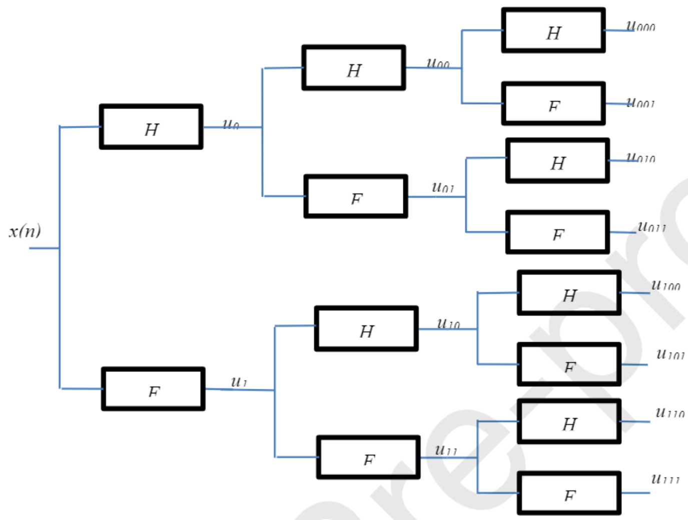


Fig. 2 Wavelet packet decomposition for a signal

2.2. Encoding Methodologies

Encoding methodologies are always needed to improve the data compression efficiency in the reduction of an ECG signal size. Three types of encoding methodologies are utilized in this work such as: Run-length, Huffman and LZW encoding. These are discussed below.

2.2.1. Run-Length Encoding (RLE)

Run-length encoding schemes later employed in the transmission of television signals. It is a simple and lossless data compression technique, in which the data is compressed by representing the consecutive runs of the same value in the data as the value followed by the count. For example: aaaabbccccccrrrrrrffff can be represented as 4a2b6c6r4f. It is a fast scheme, however, the compression efficiency of this method depends on the data type. It can also be expressed in multiple ways to put up data properties and additional compression methods. RLE is also very useful in image compression [4]. The following steps are used to apply the RLE;

Step1. Find the length of the data/vector and accordingly start a loop stretch to string length.

Step2. Here, examination of key/coefficient *i.e.,*

Step3 (a) If the key is new, then the count will be equal to zero there will be addition in the key and value to the dictionary is done.

Step3 (b) Else if the key is not new, then there will be increment in the counter value.

Step4. Reparation of Step2 and 3 will be done continuously until the loop is termination.

Step5. Print the dictionary key/coefficient and value together.

These steps are given in Fig. 3

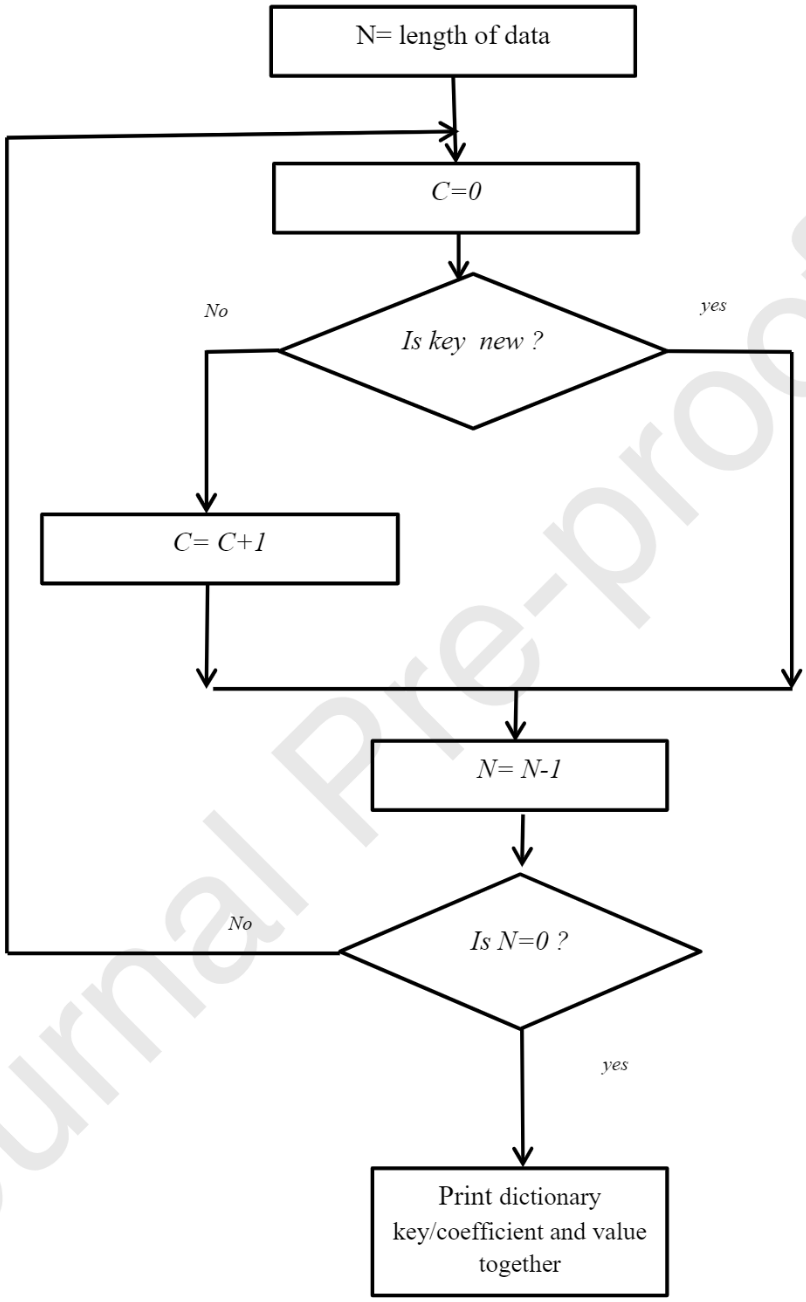


Fig. 3 Run-length encoding flow chart

2.2.2. Huffman Coding

Huffman coding allocates variable length code words to a fixed length input data/characters, based on their frequencies. The allocation is done in such manner so that the more repeated characters are allocated by smaller code words and the less repeated characters are allocated by larger code words. Here, code digits are assigned to all edges along the path to a character. A tree structure is formed in this compression scheme, in that zero is used for the left side and one for the right side. The leaves contain a letter and its frequency count. In all the other nodes, instead of one character, a zero, and the count of its frequency and its descendants is depicted. The tree structure of the Huffman encoding is depicted in Fig. 4.

For example, consider the string *ABRACADABRAADAABRACDADCC*. There are 24 characters in the sequence. The string is having *A, B, C, D* and *R* characters with frequency of *10, 3, 4, 4* and *3*, respectively. *R* and *B* have smallest frequency, *i.e.*, 3. Tree can be started from lowest frequencies, two nodes; *R,3* and *B,3*. The root will become the sum of the frequencies of these character; *i.e., 3+3=6*. And then next lowest frequency character is either *C* or *D*. Here, *C* will be the next node and root node will be φ =10. This will be repeatedly done until all characters are included. The top node should have φ =24 [18].

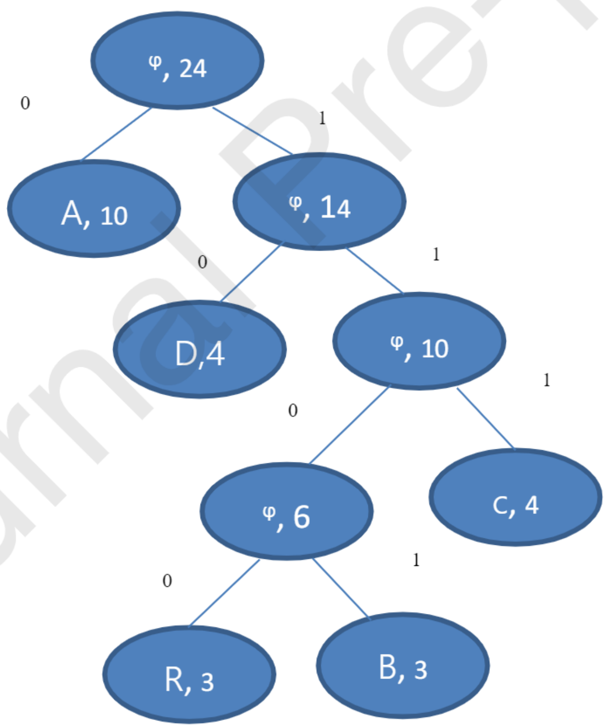


Fig. 4 Huffman tree

2.2.3. LZW Encoding

Lempel ZivWelch (LZW) is also a lossless compression method. This method is a general purpose method, which has some advantages over other coding methodologies such as simplicity and versatility. Generally, this method can compress data using coding way, in which the resultant file becomes one-half of the original size. The LZW method provides very good results in terms of the compression ratio in some applications such as: such as tabulated numbers, computer source code, and signal.

LZW compression uses a code table, as illustrated in Fig. 5. A common choice is to provide 4096 entries in the table. In this case, the LZW encoded data consists entirely of 12-bit codes, each referring to one of the entries in the code table. The decompression is achieved by taking each code from the compressed file and translates it through the code table to find what character or characters it represents. Codes 0-255 in the code table are always assigned to represent single bytes from the input file. For example, if only the first 256 codes were used, each byte in the original file would be converted into 12 bits in the LZW encoded file, resulting in a 50% larger file size. During the decompression, each 12-bit code would be translated via the code table back into the single bytes [38].

The following steps are used to perform the LZW encoding:

Step1. Initialization of the dictionary is done.

Step2. Estimation of the longest string is done in the dictionary that matches the current input.

Step3. Reduction of the dictionary is done, for W to output and remove W from the input.

Step4. Addition of "W" is done and followed by the up-coming coefficient/symbol in the input.

Step 5. Step2 is carried out.

A dictionary is initialized for containing the single-character strings corresponding to all the probable i/p characters. This technique works by scanning through the input string for consecutively lengthier sub-strings till the new string entry. If suitable string is extracted, the index for the string deprived of the preceding character is recovered from the dictionary and directed to o/p, and the new string is added to the dictionary with the succeeding accessible code. After this, the last i/p character is used for the next beginning point to scan for substrings. The same procedure will be follow; consecutively lengthier strings are registered in the dictionary and accessible aimed at succeeding encoding as single o/p values. This method works preeminently on data with the repetitive configurations; therefore, the initial parts of any information see little compression. As the information increases in size, though, the CR tends asymptotically to the maximum.

Decoding

Decoding of LZW includes, reading a value of the encoded data and outputting the consistent string from the initialized dictionary [39]. To reconstruct the dictionary in the same way as it was constructed during encoding, it also finds the subsequent value from the i/p and adds to the dictionary the concatenation of the current string and the first character of the string obtained by decoding the next input value, or the first character of the string just output if the next value cannot be decoded (If the next value is unknown to the decoder, then it must be the value added to the dictionary this iteration, and so its first character must be the same as the first character of the current string being sent to decoded output) [39]. The decoder then proceeds to the next input value (which was already read in as the "next value" in the previous pass) and repeats the process until there is no more input, at which point the final input value is decoded without any more additions to the dictionary [46].

Thus, the decoder constructs a dictionary which is duplication of the dictionary which was utilized by the encoder, and utilizes to decode succeeding i/p coefficients. In this way, the full dictionary does not required to be directed with the encoded coefficients. Only the preliminary dictionary which contains the single-character strings/coefficient is sufficient [47].

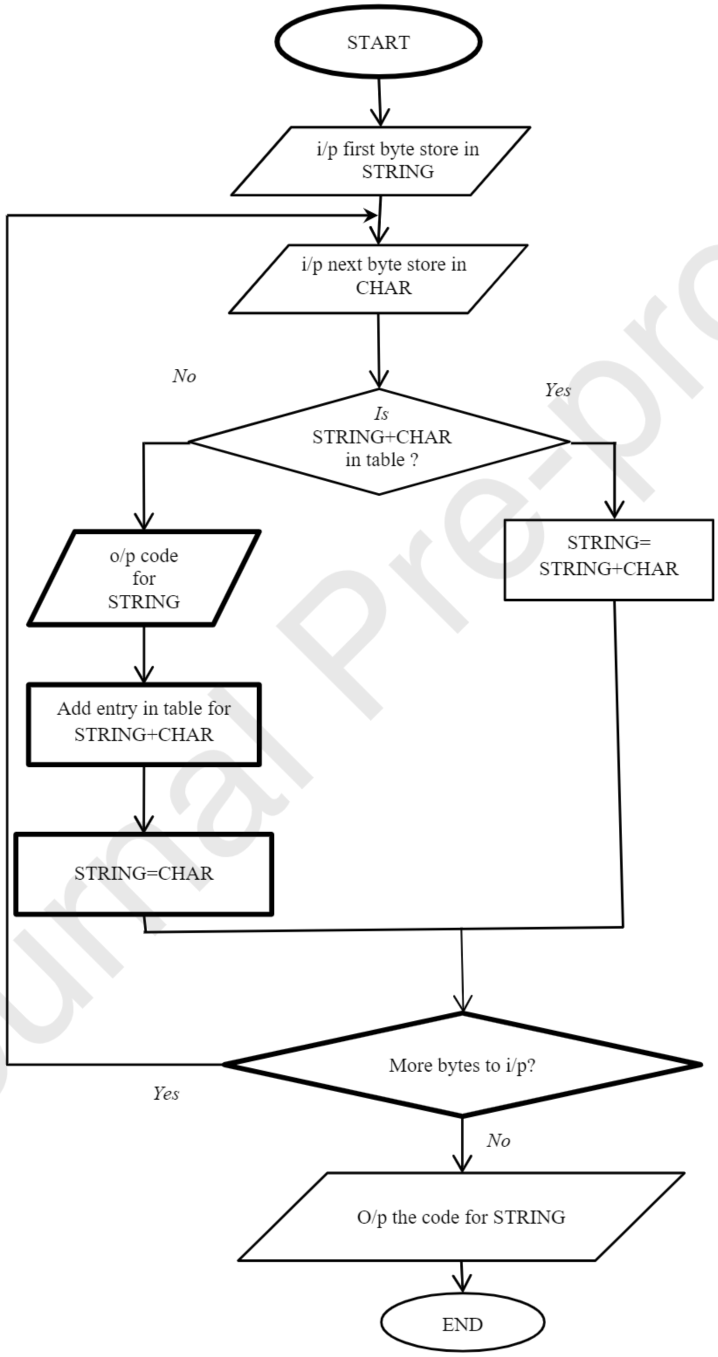


Fig. 5 LZW encoding algorithm

3. Performance Parameters

This section includes the parameters which are used for estimating the performance data compression and reconstruction [37].

- Compression ratio (*CR*)

$$CR = \frac{N_x}{H + N_s + N_v} \tag{11}$$

In this equation, N_x is the number of bits in the *x(n)*, N_v is the number of bits used to code the significant coefficients, N_s represents the number of bits in the compressed implication map and *H* is a 64-bit header [32]. This parameter is used to measure the how much amount of data is reduced.

- Percent root mean square difference (*PRD*)

$$PRD = \left(\frac{Reconstrutcted\ noise\ energy}{Origional\ signal\ energy}\right)^{1/2} \times 100 \tag{12}$$

This parameter as quality measurement can mask the real performance of an algorithm since the PRD depends a lot on the mean value of the original signal. Reconstructed noise energy, is the energy of the noisy signal energy achieved after reconstruction. The value of this parameter should be as less as possible.

- Percent root mean square difference normalized (PRDN),

$$PRDN(\%) = 100 \times \sqrt{\frac{\sum_{n=1}^{N}(x(n)-y(n))^2}{\sum_{n=1}^{N}(x(n)-\bar{x}(n))^2}} \tag{13}$$

where, *x(n)* is input signal, *y(n)* is reconstructed signal and $\bar{x}(n)$ is the mean value of input signal. It is normalized version of PRD, which is independent of the signal mean value input signal.

- Correlation coefficient (γ_{xy}).

$$\gamma_{xy} = \frac{\sum_{n=1}^{N}(x(n)-\bar{x}(n))(y(n)-\bar{y}(n))}{\sqrt{\sum_{n=1}^{N}(x(n)-\bar{x}(n))^2(y(n)-\bar{y}(n))^2}} \tag{14}$$

where, $\bar{y}(n)$ is the mean value of reconstructed signal. It is arithmetical quantity of the degree to which deviations to the value of one variable predict variation to the value of another. In positively correlated variables, the value increases or decreases in tandem. The magnitude of CC is negative, then the value of one increases as the value of the other decreases.

- Signal to noise ratio *(SNR)*

$$SNR = 100log_{10}\left\{\frac{\sum x^2(n)}{\sum |x(n)-y(n)|^2}\right\} \tag{15}$$

SNR is defined as the ratio of signal power to the noise power, often expressed in decibels. A ratio higher than 1:1 (greater than 0 dB) indicates more signal than noise.

- Mean Square error (*MSE*)

$$MSE = \frac{1}{2}\sum |x(n) - y(n)|^2 \tag{16}$$

MSE measures the average of the squares of the errors that is, the average squared difference between the estimated values and the actual value. MSE is a risk function, corresponding to the expected value of the squared error loss. The fact that MSE is almost always strictly positive (and not zero) is because of randomness or because the estimator does not account for information that could produce a more accurate estimate.

- Maximum error (*ME*)

$$ME = \max_{n} |x(n) - y(n)| \quad (17)$$

It is the maximum difference between the point estimate and the actual parameter, which is 1/2 the width of the confidence interval for means and proportions.

- Quality Score (*QS*)

$$QS = \frac{CR}{PRD} \quad (18)$$

It represents the overall performance of data compression and decompression. CR shows how much data is reduced and PRD shows data reconstruction quantity, and QS measures the trade-off between CR and PRD.

4. Methodologies

4.1. Data Compression using Wavelet Packet Decomposition and RLE

In this section, the wavelet packet (WP) based compression methodology is presented to compress the ECG signals. This mythology employs the WP, level thresholding along with RLE by executing the steps as follows;
Step 1: Acquisition of ECG data.
In this work, MIT-BIH arrhythmia database has been chosen for examining the performance of the proposed method.
Step 2: Signal decomposition
The original signal should be decomposed in such a way so that, it can successively decompose into components of lower resolution, while the high frequency components can be analysed any further. An ECG signal has PQRST components of frequency 0 to 100 Hz. Wavelet packet decomposition is done to divide the ECG signal into the number of frequency bands. In [37], researchers have done data compression using Beta2, db10, Beta3 and coif4. Here, authors have decomposed signal with different level decomposing (1 to 5) and compared results in terms of CR, SNR and PRD. Comparison results show that value of CR is increasing by increasing number of level of decomposition. Value of SNR and PRD are improving by increasing level of decomposition in some cases, while in other cases performance of SNR and PRD are diminishing (for 4 and 5 level of decomposition) [37]. Therefore in this study, level 3/4/5 decomposition is done, to check resolution quality of the decomposed signal.
For example, if the total frequency of a signal is 100Hz, by applying 3 level decomposition using WP, produce 8 signals, *viz*., *Aaa* (0-12.5Hz), *Aad* (above 12.50-25Hz), *Ada* (above 25-37.5Hz), *Add* (above 37.5-50Hz), *Daa* (above 50-62.5Hz), *Dad* (above 62.5-75Hz), *Dda* (above 75-87.5Hz), and *Ddd* (above 87.5-100Hz). The lowest frequency signal (*Add*) is also known as approximation subband and reaming signals are known as detail subbands.
Step 3: Application of thresholding

Level thresholding (*i.e.*, different thersholding values are used for their respective level of decomposed signals) is applied to all decomposed signal, except the lowest frequency signal (*Aaa*). Here, all the coefficients which have a magnitude less than the threshold value are replaced by zero.

Table 1 Signals naming before and after thresholding

Before thresholding	*Aaa*	*Aad*	*Ada*	*Add*	*Daa*	*Dad*	*Dda*	*Ddd*
After thresholding	-	thr_7	thr_6	thr_5	thr_4	thr_3	thr_2	thr_1

Step 4: Vector '*v*' formation

After obtaining the thresholded coefficients (all coefficients *i.e.*, zero valued coefficients obtained after applying thresholding and actual magnitude coefficients that have magnitude larger than threshold value) of all bands, these are placed in a vector (*v*), which is illustrated in Eqn. (19).

$$v= [Aaa\ thr_7\ thr_6\ thr_5\ thr_4\ thr_3\ thr_2\ thr_1] \tag{19}$$

where, thr_1 to thr_7 are signals obtained after applying thresholding, as depicted in Table 5.1.

Step 5: Application of modified run-length encoding.

In this step, RLE is applied to *v*. RLE is a lossless compression method, therefore it increases the performance of the compression without loss of information. In this work, modified RLE is deployed in which, two-stage RLE is used, and therefore the output of modified RLE is in the form of three vectors (v_{10}, $v_{110,}$ and v_{111}). The modified RLE is presented in Fig. 6.

Reconstruction of data is done by following steps given below:

Step a: Application of run-length decoding.

Run-length decoding (RLD) is applied two times *i.e.,* firstly to the vectors v_{110} and v_{111} which gives one output vector "v_{11}", again RLD is applied to v_{11} and v_{10}, and it provides reconstructed coefficients of vector *v*.

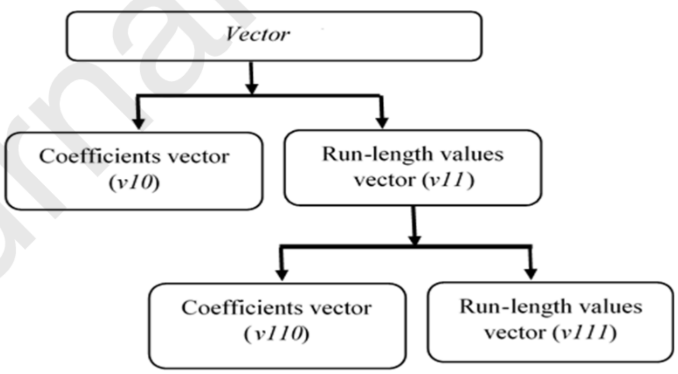


Fig. 6 Modified RLE

For example: MIT-BIHrecord no. 101 is taken, obtained the coefficient of vector

v_{110}=	[78 1	5	1	4	1	3	1	1	1	5	1	7
	1	7	1	1	1	1	1	2	1	3	1	2
	1	2	1	2	1	3	1	4	1	3	1	4
	1	4	1	4	1	4	1]					

v_{111}=	[8	1	4	1	7	1	2	1	3	1	6	1
	3	1	4	1	7	1	11	1	17	1	17	1
	18	1	19	1	19	1	39	1	35	1	37	1
	36	1	38	1	39	1	1562]					

v_{11}'' = [1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	8	1	1	1	1	1	4
	1	1	1	1	7	1	1	1	2	1	3	1
	1	1	1	1	6	1	1	1	1	1	1	1
	3	1	1	1	1	1	1	1	4	1	7	1
	11	1	1	17	1	1	1	17	1	1	18	1
	1	19	1	1	19	1	1	1	39	1	1	1
	1	35	1	1	1	37	1	1	1	1	36	1
	1	1	1	38	1	1	1	1	39	1	1	1
	1	1562].										

Step b: Coefficient distribution of *v*.

The coefficients of *v* are divided into eight groups to reconstruct signals, such as: *Aaa*, thr_7, thr_6, thr_5, thr_4, thr_3, thr_2 and thr_1.

Step c: Inverse transform.

The inverse transform is used to reconstruct the signal. These steps are depicted in the block diagram given in Fig. 7.

In this section, experimental results of WP and RLE based compression method are described. Here, parameters for evaluating the performance are considered which are given in Section 3. The experiments are done by using different mother wavelets, such as: bi-orthogonal (bior), Daubechies (db), symlet and Fejér-Korovkin (FK) and symlets. Here, the different class of wavelet filters have been used for the camparasion of the performce. Here, the 3/4 level of decomposition is taken. The tabular results of these experiments are presented in Tabls 2 (a), (b), (c), (d) and (e). The original and reconstructed signal of this method is given in Fig. 8. Here, MIT-BIH record no.109 has been taken. In Fig. 8(a) noise free signal is shown before appliying data compression (or decomposition), and in Fig. 8(b), (MIT-BIH record no.109) is presented after reconsonstruction by following the reconstructed steps. It can be seen that both signals are identical to each other.

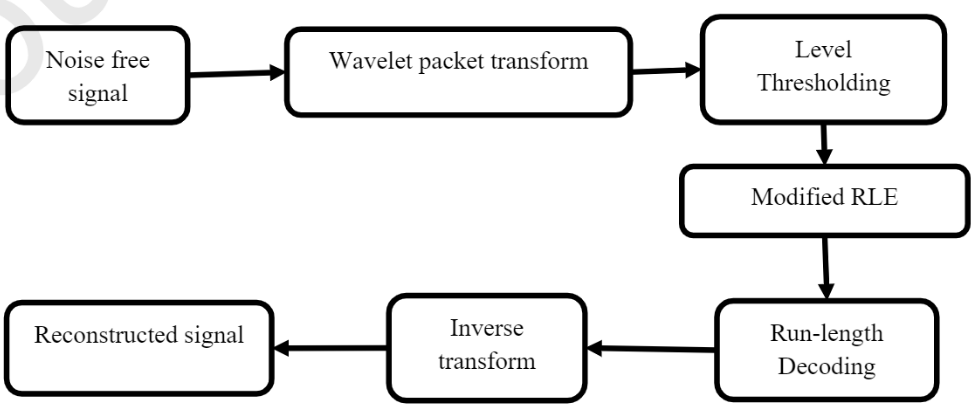


Fig. 7 ECG data compression methodology using WP

In this study, different wavelets have been taken to test whether these can be utilized for the designed methodologies. And if these wavelets will be use, it is also analysed, that provided resultant parameter (detailed is given in Section 3) are in the desire range or not. Also comparison of the possible combinations, *i.e.*, at different level of decomposition (*L=3*, *L=4* and *L=5*,), and truncating all coefficients of decomposed band (*D1=0 and Ad=0*) is presented. These results are depicted in Table 2 (a), (b), (c) (d) and (e). Best values of these parameters are highlighted. From these tables, it can be observed that for highest value of compression ratio, Haar wavelet is suitable. However, the PRD value in this case is not best. The best value of QS is obtained in case of bior3.5.

Table 2 (a) Fidelity parameters of proposed WP based data compression technique (using bior2.8)

Level	Signal	%CR	PRD	SNR	CC	QS
L=3 *D1=0*	MIT-BIH-100	5.6774	0.3621	48.8246	0.9942	15.6790
	MIT-BIH-101	**6.2301**	0.3214	49.8579	0.9965	19.3842
	MIT-BIH-102	4.3909	0.2989	50.4907	**0.9999**	14.6902
	MIT-BIH-103	5.6021	0.4245	47.4431	0.9979	13.1969
	MIT-BIH-104	5.2932	0.2079	53.6426	0.9989	25.4603
	MIT-BIH-105	5.0286	0.1605	55.8910	0.9997	31.3308
	MIT-BIH-106	5.8830	0.4715	46.5295	0.9979	12.4772
	MIT-BIH-107	4.6829	0.3824	48.3486	0.9997	12.2460
	MIT-BIH-108	3.4010	**0.1085**	**59.2892**	0.9994	**31.3456**
	MIT-BIH-109	4.4277	0.1938	54.2527	0.9997	22.8467
L=4 *D1=0, Ad=0*	MIT-BIH-100	**7.0000**	0.4815	46.3484	0.9897	14.5379
	MIT-BIH-101	**7.0000**	0.3804	48.3950	0.9951	18.4016
	MIT-BIH-102	5.4768	0.4446	47.0414	0.9998	12.3184
	MIT-BIH-103	6.6469	0.6784	43.3699	0.9945	09.7979
	MIT-BIH-104	**7.0000**	0.2653	51.5252	0.9982	26.3852
	MIT-BIH-105	6.4000	0.2992	50.4813	0.9989	21.3903
	MIT-BIH-106	6.1202	0.5517	45.1657	0.9971	11.0933
	MIT-BIH-107	5.7143	0.5349	45.4348	**0.9995**	10.6829
	MIT-BIH-108	5.0564	**0.1801**	**54.8895**	0.9984	**28.0755**
	MIT-BIH-109	5.5309	0.4079	47.7883	0.9987	13.5594
L=4 *D1=0*	MIT-BIH-100	**8.4528**	0.5117	45.8200	0.9883	16.5190
	MIT-BIH-101	8.0866	0.4133	47.6742	0.9943	19.5659
	MIT-BIH-102	6.4928	0.6935	43.1787	**0.9996**	09.3623
	MIT-BIH-103	8.3895	0.7136	42.9313	0.9939	11.7565
	MIT-BIH-104	7.1565	0.4116	47.7097	0.9955	17.3870
	MIT-BIH-105	6.9565	0.4081	47.7838	0.9980	17.0460
	MIT-BIH-106	7.0662	0.5796	44.7375	0.9968	12.1915
	MIT-BIH-107	6.8085	0.6372	43.9149	0.9949	10.6850
	MIT-BIH-108	5.5309	**0.2675**	**51.4540**	0.9965	**20.6762**
	MIT-BIH-109	7.0000	0.5641	44.9721	0.9975	12.4091

Table 2 (b) Fidelity parameters of proposed WP based data compression technique (using bior1.5)

Level	Signal	%CR	PRD	SNR	CC	QS
L=3 D1=0	MIT-BIH-100	5.5568	0.8460	41.4531	0.9691	06.5687
	MIT-BIH-101	**5.8743**	0.7846	42.1069	0.9797	07.4869
	MIT-BIH-102	3.8216	1.0871	39.2743	0.9989	03.5153
	MIT-BIH-103	4.5286	1.2883	37.7994	0.9807	03.5151
	MIT-BIH-104	4.0314	**0.4452**	47.0286	0.9948	09.0549
	MIT-BIH-105	4.5689	0.7517	42.4793	0.9934	06.0782
	MIT-BIH-106	3.5633	1.0191	39.8359	0.9903	03.4966
	MIT-BIH-107	3.5085	1.2267	38.2249	**0.9974**	02.8600
	MIT-BIH-108	3.3215	0.2663	**51.4923**	0.9966	**12.4723**
	MIT-BIH-109	3.2480	0.5909	44.5691	0.9973	05.4963
L=4 D1=0, Ad=0	MIT-BIH-100	**6.2041**	2.6543	31.5209	0.6975	02.3373
	MIT-BIH-101	5.7826	2.4567	32.1929	0.8016	02.3538
	MIT-BIH-102	4.4895	4.5388	26.8611	**0.9815**	00.9891
	MIT-BIH-103	5.2673	4.3616	27.2071	0.7795	01.2077
	MIT-BIH-104	6.1503	1.3528	37.3756	0.9523	**04.5465**
	MIT-BIH-105	5.4845	2.6798	31.4378	0.9161	02.0466
	MIT-BIH-106	4.7713	4.3176	27.2951	0.8275	01.1051
	MIT-BIH-107	5.1401	4.4471	27.0384	0.9660	01.1558
	MIT-BIH-108	5.1776	**0.9673**	**40.2889**	0.9548	05.3527
	MIT-BIH-109	5.3200	2.2966	32.7782	0.9596	02.3164
L=4 D1=0	MIT-BIH-100	**7.5730**	2.6818	31.4316	0.6924	02.8239
	MIT-BIH-101	6.2222	2.4605	32.1795	0.8012	02.5288
	MIT-BIH-102	4.9145	4.5566	26.8272	**0.9814**	01.0786
	MIT-BIH-103	6.0974	4.3658	27.1987	0.7795	01.3966
	MIT-BIH-104	7.1171	1.4414	36.8244	0.9463	04.9377
	MIT-BIH-105	5.6446	2.6867	31.4156	0.9157	02.1009
	MIT-BIH-106	5.6000	4.3519	27.2265	0.8254	01.2868
	MIT-BIH-107	5.5707	4.4892	26.9567	0.9654	01.2409
	MIT-BIH-108	5.8947	**1.0095**	**39.9178**	0.9510	**05.839**2
	MIT-BIH-109	5.8301	2.4420	32.2451	0.9543	02.3874

Table 2 (c) Fidelity parameters of proposed WP based data compression technique (using bior3.5)

Level	Signal	%CR	PRD	SNR	CC	QS
L=3 D1=0	MIT-BIH-100	5.8039	0.2659	51.5054	0.9969	21.8273
	MIT-BIH-101	**6.6624**	0.2119	53.4758	0.9985	31.4412
	MIT-BIH-102	5.7080	0.2393	52.4223	**0.9999**	23.8529
	MIT-BIH-103	6.4750	0.2507	52.0183	0.9993	25.8276
	MIT-BIH-104	6.1667	0.2290	52.8043	0.9986	26.9288
	MIT-BIH-105	5.1160	**0.1134**	58.9098	0.9998	45.1146
	MIT-BIH-106	6.6410	0.3324	49.5678	0.9990	19.9789
	MIT-BIH-107	5.8039	0.3310	49.6028	0.9998	17.5344
	MIT-BIH-108	3.7000	0.1138	**58.8769**	0.9994	**32.5131**
	MIT-BIH-109	4.8411	0.1829	54.7540	0.9997	26.4685
L=4 D1=0, Ad=0 yes	MIT-BIH-100	8.8163	0.5418	45.3232	0.9869	16.2722
	MIT-BIH-101	8.5714	0.4788	46.3976	0.9923	17.9018
	MIT-BIH-102	7.8261	0.5611	45.0199	**0.9997**	13.9477
	MIT-BIH-103	**9.1525**	0.8756	41.1538	0.9909	10.4528
	MIT-BIH-104	8.5714	0.4730	46.5032	0.9941	18.1213
	MIT-BIH-105	8.1509	0.3074	50.2449	0.9989	26.5156
	MIT-BIH-106	8.4375	0.8490	41.4222	0.9932	09.9381
	MIT-BIH-107	8.2443	0.9272	40.6566	0.9985	08.8916
	MIT-BIH-108	6.9903	**0.2413**	**52.3499**	0.9972	**28.9693**
	MIT-BIH-109	8.8889	0.5605	45.0289	0.9976	15.8588
L=4 D1=0 yes	MIT-BIH-100	9.0011	0.6644	43.5510	0.9810	13.5460
	MIT-BIH-101	8.9256	0.5241	45.6110	0.9907	17.0303
	MIT-BIH-102	8.3077	1.1152	39.0533	**0.9989**	07.4495
	MIT-BIH-103	9.2704	0.8915	40.9971	0.9905	10.3986
	MIT-BIH-104	**9.9083**	0.7536	42.4569	0.9850	13.1479
	MIT-BIH-105	8.5039	0.3826	48.3442	0.9983	22.2266
	MIT-BIH-106	9.1915	0.9075	40.8427	0.9922	10.1283
	MIT-BIH-107	8.3398	1.0231	39.8020	0.9982	08.1515
	MIT-BIH-108	8.0297	**0.3547**	**49.0034**	0.9939	**22.6380**
	MIT-BIH-109	9.7297	0.8164	41.7625	0.9949	11.9178

Table 2 (d) Fidelity parameters of proposed WP based data compression technique (using bior6.8)

Level	Signal	%CR	PRD	SNR	CC	QS
L=4 D1=0	MIT-BIH-100	**9.2562**	0.7149	42.9154	0.9770	12.9475
	MIT-BIH-101	8.7843	0.5851	44.6559	0.9884	15.0133
	MIT-BIH-102	7.6976	0.5924	44.5481	**0.9997**	12.9939
	MIT-BIH-103	8.5824	0.7271	42.7680	0.9937	11.8036
	MIT-BIH-104	7.6190	0.4276	47.3785	0.9952	17.8180
	MIT-BIH-105	9.2181	0.7501	42.4976	0.9933	12.2891
	MIT-BIH-106	8.4528	0.4696	46.5655	0.9979	18.0000
	MIT-BIH-107	7.2492	0.9921	40.0686	0.9983	07.3069
	MIT-BIH-108	6.4368	**0.2641**	**51.5646**	0.9966	24.3725
	MIT-BIH-109	7.9715	0.5807	44.7213	0.9974	13.7274
L=4 D1=0 yes	MIT-BIH-100	9.2946	0.7551	42.4398	0.9744	12.3091
	MIT-BIH-101	8.9600	0.9569	40.3831	0.9687	09.3635
	MIT-BIH-102	8.4528	0.9126	40.7941	**0.9993**	09.2623
	MIT-BIH-103	5.6281	0.7440	42.5687	0.9934	7.56465
	MIT-BIH-104	**9.4515**	0.7015	43.0789	0.9870	13.4732
	MIT-BIH-105	9.4118	1.0000	40.0002	0.9881	09.4118
	MIT-BIH-106	8.9960	0.5751	44.8055	0.9969	15.6425
	MIT-BIH-107	8.1752	1.3453	37.4238	0.9969	06.0768
	MIT-BIH-108	7.2258	**0.4372**	**47.1861**	0.9907	**16.5274**
	MIT-BIH-109	8.6154	0.8892	41.0196	0.9939	09.6889

Table 2 (e) Fidelity parameters of proposed WP based data compression technique

Level	Signal	%CR	PRD	SNR	CC	QS
Sym2 L=4 D1=0 yes	MIT-BIH-100	08.3967	0.7039	43.0493	0.9777	11.9281
	MIT-BIH-101	**08.4667**	0.6279	44.0420	0.9867	13.4838
	MIT-BIH-102	05.7892	0.9102	40.8174	0.9993	06.3604
	MIT-BIH-103	06.7508	0.9079	40.8390	0.9901	07.4354
	MIT-BIH-104	06.5338	0.4516	46.9046	0.9946	14.4675
	MIT-BIH-105	07.0311	0.7173	42.8855	0.9939	09.8017
	MIT-BIH-106	06.5128	0.7388	42.6295	0.9948	08.8154
	MIT-BIH-107	06.2331	0.9062	40.8556	**0.9986**	06.8784
	MIT-BIH-108	05.6602	**0.3217**	**49.8507**	0.9950	**17.5941**
	MIT-BIH-109	05.7401	0.5443	45.2838	0.9977	10.5465
Fk4 L=4 D1=0 yes	MIT-BIH-100	09.0131	2.1807	33.2283	0.7979	04.1332
	MIT-BIH-101	09.6449	2.1446	33.3731	0.8466	04.4973
	MIT-BIH-102	05.5335	3.3561	29.4830	**0.9898**	01.6488
	MIT-BIH-103	06.8344	3.2072	29.8775	0.8808	02.1310
	MIT-BIH-104	07.7594	1.2705	37.9206	0.9579	06.1074
	MIT-BIH-105	**09.3818**	2.1601	33.3106	0.9444	04.3433
	MIT-BIH-106	08.1905	3.2210	29.8401	0.9035	02.5428
	MIT-BIH-107	04.8112	3.3675	29.4538	0.9809	01.4287
	MIT-BIH-108	09.0526	**0.8991**	**40.9242**	0.9606	**10.0690**
	MIT-BIH-109	07.6729	1.8788	34.5224	0.9731	04.0839
Db10 L=4 D1=0 yes	MIT-BIH-100	07.4492	0.6154	44.2168	0.9830	12.1045
	MIT-BIH-101	**07.9164**	0.5339	45.4509	0.9904	14.8276
	MIT-BIH-102	06.7219	0.4458	47.0179	0.9998	15.0796
	MIT-BIH-103	06.2590	0.4030	47.8939	**0.9981**	15.5309
	MIT-BIH-104	06.2762	0.3488	49.1473	0.9968	17.9914
	MIT-BIH-105	07.2821	0.5685	44.9056	0.9962	12.8096
	MIT-BIH-106	08.1434	0.5622	45.0021	0.9970	14.4847
	MIT-BIH-107	07.0779	0.8569	41.3416	0.9987	08.2601
	MIT-BIH-108	05.0265	**0.2073**	**53.6673**	0.9979	**24.2458**
	MIT-BIH-109	05.3585	0.4765	46.4386	0.9983	11.2453
Haar, L=4 D1=0 yes	MIT-BIH-100	12.5379	6.5632	23.6577	0.6047	03.1842
	MIT-BIH-101	**12.7949**	3.7268	28.5733	0.5498	03.4332
	MIT-BIH-102	09.8164	7.0182	23.0755	0.9554	01.3987
	MIT-BIH-103	11.5759	6.5195	23.7157	0.5077	01.7756
	MIT-BIH-104	12.2424	2.1655	33.2886	0.8788	05.6533
	MIT-BIH-105	11.3483	4.0703	27.8074	0.8052	02.7880
	MIT-BIH-106	12.5370	6.5632	23.6577	**0.6047**	01.9103
Coif1, L=4 D1=0 yes	MIT-BIH-100	07.8479	**1.1243**	**38.9825**	0.9429	**06.9804**
	MIT-BIH-101	**09.2556**	0.9549	40.4013	0.9690	09.6933
	MIT-BIH-102	06.2927	1.6729	35.5308	**0.9975**	03.7616
	MIT-BIH-103	05.8971	1.6241	35.7876	0.9684	03.6310
	MIT-BIH-104	06.6154	**0.5893**	**44.5929**	0.9909	**11.2253**
	MIT-BIH-105	06.6367	0.9563	40.3882	0.9891	06.9400
	MIT-BIH-106	07.4783	1.6216	35.8012	0.9750	04.6117
	MIT-BIH-107	06.8344	1.7434	35.1723	0.9948	03.9203

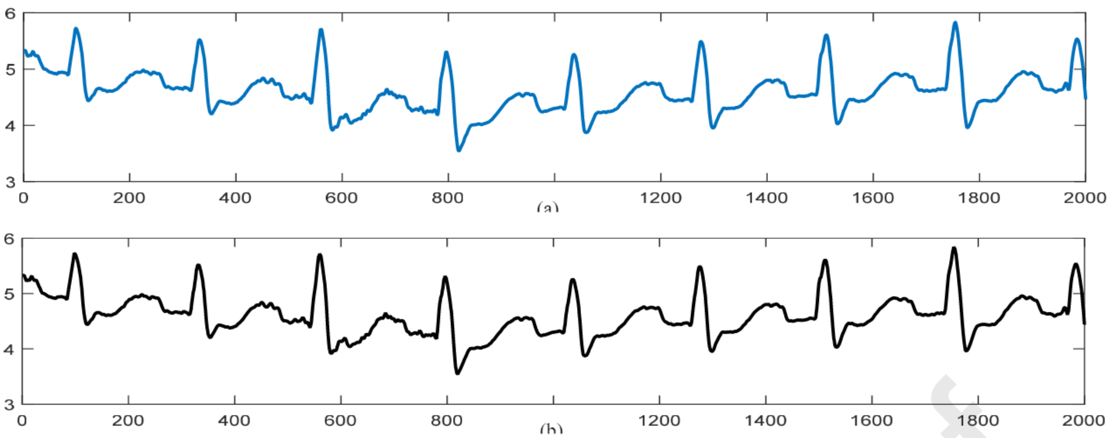


Fig. 8 a) Original ECG signal (MIT-BIH record no. 109) and b) reconstructed signal using 4-level decomposition by WP algorithm (sym2)

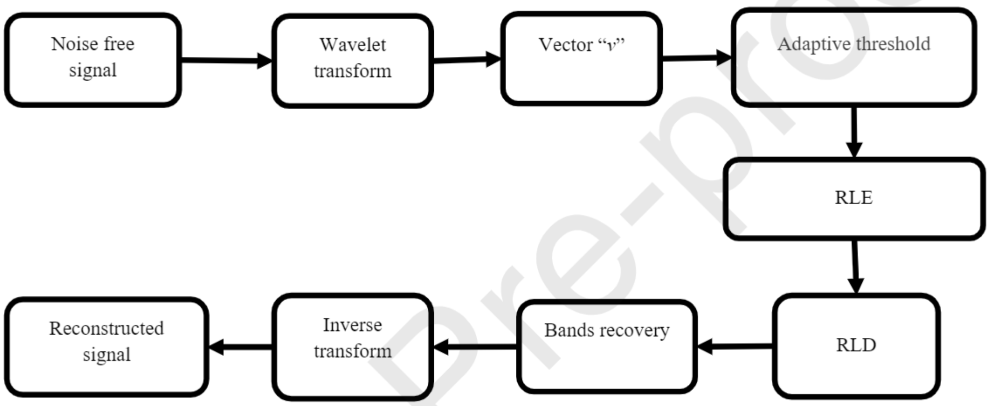


Fig. 9 ECG data compression using wavelet and RLE

4.2. ECG Data Compression using Wavelet Transform and RLE

In this section, ECG compression based on wavelet transform and RLE is presented. The steps used to compress the signal are depicted in Fig. 9. Here, first the wavelet transform decomposition is done, that splits the signal into different frequency bands. In this case, (level 4/5) is used to decompose the signal into approximation (*a*) and detail bands (*d*)). Allocation of decomposed bands is presented in Table 3, and is done on the basis of research given in [3]. Then after the resultant coefficients are subjected by a threshold (Adaptive threshold) using the principle of Stein's unbiased risk estimation. The coefficients below the threshold value are replaced by zero while remaining coefficients are kept as the same magnitude as before. And then these coefficients are kept in vector "*v*". Next, the RLE is applied. The performance of this method is given in Table 4 using Sym2, Bior5.5 and Db10 wavelet. Here also, performance of different wavelets at 3 level of decomposing is analysed. The best results are highlighted for every wavelet. In this case also, Haar wavelet performs better in terms of compression ratio and Db10 provides best results in terms of PRD and QS. The original and reconstructed signals were obtained by using the methodology given in Fig. 10. In this figure, 2000 samples of original noise free (*i.e.*, before applying compression) MIT-BIH record no. 105 has been taken, given in Fig. 10 (a), and signal is reconstructed by following the steps 6 to 9), which is given in Fig 10 (b). The comparison of these figure shows that both signals have identical wave.

Table 3 WT decomposition

Level 3 decomposition	Level 4 decmposition	Level 5 decomposition
a_3	a_4	a_5
d_3	d_4	d_5
d_2	d_3	d_4
d_1	d_2	d_3
-	d_1	d_2
-	-	d_1

Table 4 Performance of compression ratio using wavelet transform and RLE

Level	Signal	%CR	PRD	SNR	CC	QS
Sym2	MIT-BIH-100	11.0385	1.3839	37.1779	0.9110	07.9764
L=4	MIT-BIH-101	**11.4800**	1.2916	37.7776	0.9422	08.8884
	MIT-BIH-102	09.9950	1.4794	36.5986	**0.9980**	06.7564
	MIT-BIH-103	08.9289	2.1442	33.3745	0.9437	04.1641
	MIT-BIH-104	10.4635	0.7161	42.9010	0.9865	14.6126
	MIT-BIH-105	10.6862	1.1601	38.7104	0.9839	09.2118
	MIT-BIH-106	08.4059	1.9279	34.2981	0.9642	04.3600
	MIT-BIH-107	09.3442	1.7280	35.2491	0.9949	05.4075
	MIT-BIH-108	10.0450	**0.4728**	**46.5066**	0.9891	**21.2460**
	MIT-BIH-109	09.7053	0.8682	41.2281	0.9942	11.1792
Bior5.5	MIT-BIH-100	11.0978	1.4563	36.7352	0.9009	07.6208
D1=0 yes	MIT-BIH-101	**11.1585**	1.0190	39.8368	0.9645	10.9507
	MIT-BIH-102	09.7238	0.7237	42.8089	**0.9995**	13.4363
	MIT-BIH-103	10.1089	2.0955	33.5741	0.9464	04.8240
	MIT-BIH-104	10.1089	0.5875	44.6198	0.9909	17.2066
	MIT-BIH-105	10.3655	0.9412	40.5263	0.9894	11.0130
	MIT-BIH-106	10.2613	1.4954	36.5049	0.9786	06.8619
	MIT-BIH-107	09.6777	1.5067	36.4396	0.9961	06.4232
	MIT-BIH-108	09.5421	**0.3472**	**49.1887**	0.9941	**27.4839**
	MIT-BIH-109	09.2398	0.5858	44.6440	0.9970	15.7728
Db10 L=4	MIT-BIH-100	10.1122	1.3842	37.1758	0.9110	07.3052
D1=0 yes	MIT-BIH-101	**10.5765**	1.0481	39.5920	0.9623	10.0912
	MIT-BIH-102	08.9740	0.7281	42.7567	0.9995	12.3252
	MIT-BIH-103	08.8970	1.9778	34.0763	0.9523	04.4984
	MIT-BIH-104	08.6375	0.5858	44.645	0.9910	14.7450
	MIT-BIH-105	09.3801	0.8197	41.7265	0.9920	11.4427
	MIT-BIH-106	10.3650	1.6260	35.7775	0.9747	06.3745
	MIT-BIH-107	09.8714	1.4617	36.7029	0.9963	06.7534
	MIT-BIH-108	09.0130	**0.2979**	**50.5187**	0.9957	**30.2558**
	MIT-BIH-109	08.9740	0.6277	44.0443	**0.9970**	14.2956
Haar,	MIT-BIH-100	07.2727	1.8886	34.4771	0.9568	03.8508
L=4	MIT-BIH-101	**11.7647**	1.7152	35.3139	0.8956	06.8593
D1=0 yes	MIT-BIH-102	08.6207	3.1439	30.0506	0.9911	02.7420
	MIT-BIH-103	08.2305	3.0409	30.3398	0.8830	02.7065
	MIT-BIH-104	10.0503	0.9490	40.4542	0.9761	10.5898
	MIT-BIH-105	07.2727	1.8886	34.4771	0.9568	03.8508
	MIT-BIH-106	08.6580	2.4454	32.2331	0.9418	03.5406
	MIT-BIH-107	08.0972	3.0116	30.4242	0.9843	02.6887
	MIT-BIH-108	10.2041	**0.7142**	42.9237	0.9749	14.2875
	MIT-BIH-100	10.2386	**0.4354**	**47.2229**	**0.9908**	**23.5171**
Coif1,	MIT-BIH-101	**11.4602**	1.5487	36.2006	0.9158	07.3999
L=4	MIT-BIH-102	09.9360	1.3431	37.4381	0.9984	07.3980
D1=0 yes	MIT-BIH-103	07.3345	2.2516	32.9500	0.9378	03.2574
	MIT-BIH-104	09.3814	0.7579	42.4082	0.9848	12.3780
	MIT-BIH-105	10.0850	1.1187	39.0257	0.9851	09.0149
	MIT-BIH-106	08.6940	1.5083	36.4305	0.9783	05.7643
	MIT-BIH-107	10.2908	1.6295	35.7587	0.9954	06.3152
	MIT-BIH-108	09.6971	0.8622	41.2874	**0.9943**	11.2464
db3, L=4	MIT-BIH-100	11.1436	1.5542	36.1697	0.8863	07.1699
D1=0 yes	MIT-BIH-101	**11.5920**	1.3841	37.1766	0.9334	08.3751
	MIT-BIH-102	09.5142	0.9009	40.9062	**0.9993**	10.5604
	MIT-BIH-103	08.8855	2.2911	32.7991	0.9355	03.8782
	MIT-BIH-104	10.2386	0.6753	43.4104	0.9880	15.1621
	MIT-BIH-105	10.2386	1.0083	39.9283	0.9879	10.1545
	MIT-BIH-106	09.2100	1.5939	35.9508	0.9757	05.7783
	MIT-BIH-107	10.1869	1.4073	37.0321	0.9966	07.2385
	MIT-BIH-108	09.7440	**0.3742**	**48.5374**	0.9932	**26.0379**

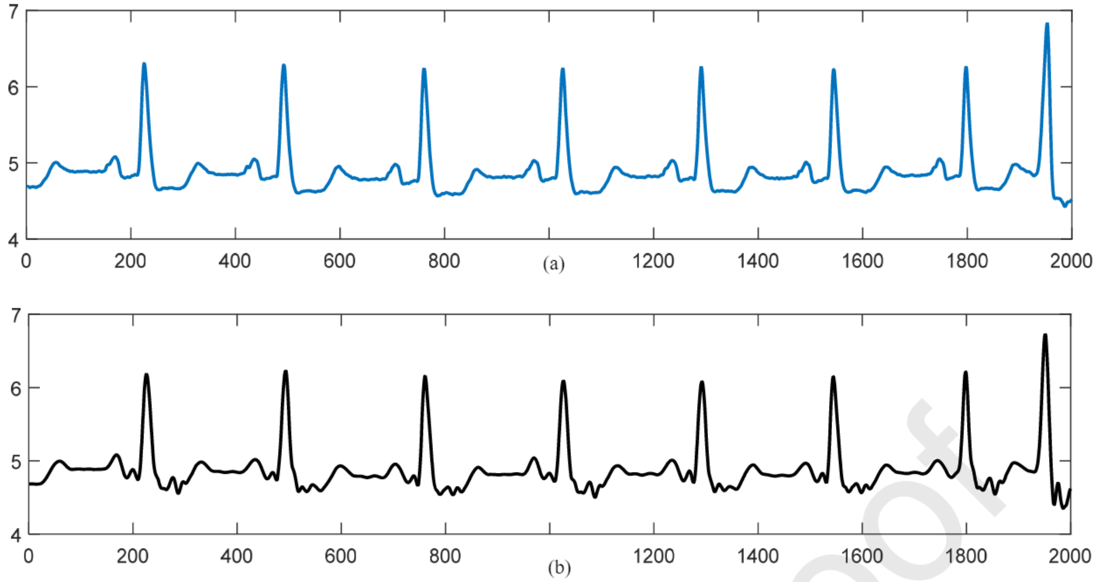


Fig. 10 a) Original ECG signal (MIT-BIH record no. 105) and b) Reconstructed signal using 4-level decomposition by WT (db10)

4.3. ECG Data Compression using Wavelet Transform and Huffman Coding

In this section, ECG compression based on wavelet transform and Huffman's coding is presented. The steps used to compress the signal are depicted in Fig. 11. Here, also the first step is the decomposition of the signal into different frequency bands (level 4). Then after the by vector quantization is performed to the coefficient of vector "*v*". The steps of the quantization are evaluated using expression given below:

$$step = floor\left(N\frac{\log(v/v\min)}{\log(vMax/vMin)}\right) \tag{20}$$

Next, the Huffman's coding is applied. The performance of this method is given in Table 5 and Fig. 12. On comparing results, it is found that the methodology given in this section provide better results in comparison to performance given in Table 2 and 3 in terms of both CR and PRD. Here also, best result for each wavelet is highlighted for all the parameters. From this Table, the best value of compression ratio is given by Fk4 wavelet. Similar to the previous algorithm the best results in terms of PRD and QS is given when db10 wavelet is used.

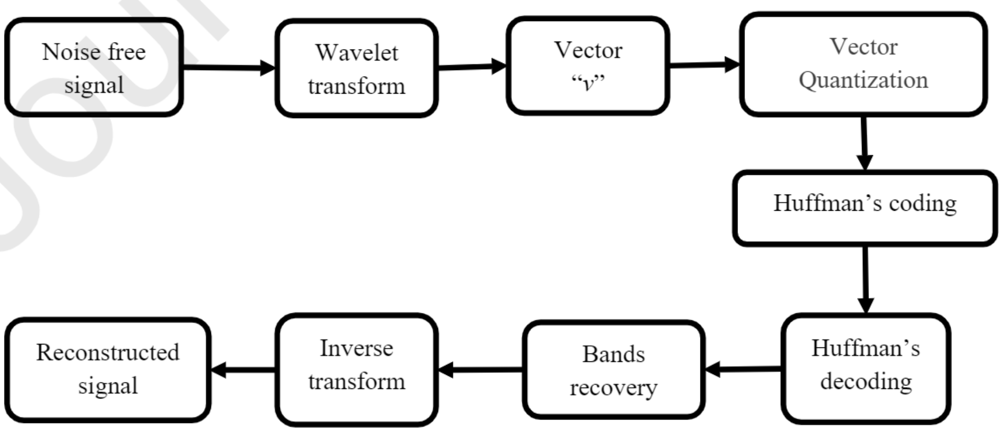


Fig. 11 ECG compression using wavelet and Huffman's coding.

Table 5 Performance of compression ratio using wavelet transform and Huffman's coding

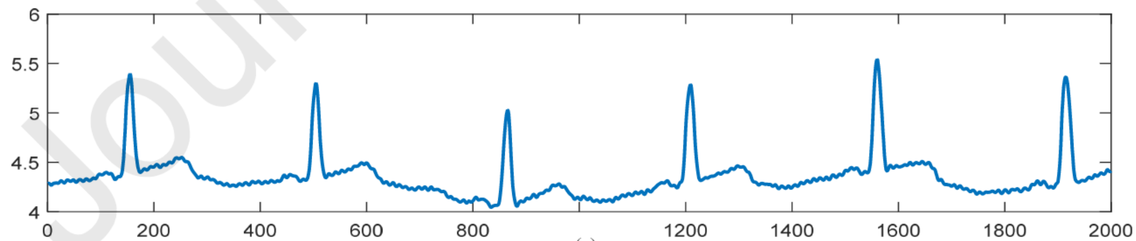
Level	Signal	CR	PRD	SNR	CC	QS
Sym2 L=4 D1=0 yes	MIT-BIH-100	13.1062	0.0765	56.3084	0.9990	171.3666
	MIT-BIH-101	12.8970	0.0665	57.5235	0.9994	193.9524
	MIT-BIH-102	10.9058	0.0575	58.7797	**1.0000**	189.5283
	MIT-BIH-103	11.7271	0.1063	53.4491	0.9995	110.0065
	MIT-BIH-104	12.4788	0.0432	61.2679	0.9998	288.798
	MIT-BIH-105	11.9350	0.0429	61.3291	0.9999	278.1698
	MIT-BIH-106	12.8662	0.1084	53.2791	0.9996	118.0277
	MIT-BIH-107	09.8297	0.0805	55.8680	**1.0000**	122.1716
	MIT-BIH-108	**13.3223**	**0.0220**	**67.1402**	0.9999	**606.2051**
	MIT-BIH-109	11.0609	0.0352	63.0368	**1.0000**	313.8049
Fk4 L=4 D1=0 yes	MIT-BIH-100	12.9695	0.1683	49.4595	0.9950	77.0776
	MIT-BIH-101	12.8356	0.1532	50.2764	0.9968	83.8043
	MIT-BIH-102	10.5208	0.2598	51.4491	**0.9998**	40.4959
	MIT-BIH-103	11.4854	0.2655	45.4972	0.9966	43.2553
	MIT-BIH-104	12.5857	0.0882	55.0661	0.9992	142.6309
	MIT-BIH-105	11.6510	0.1544	50.2044	0.9989	75.4427
	MIT-BIH-106	13.0980	0.2643	45.5361	0.9974	49.5492
	MIT-BIH-107	09.2607	0.2592	45.7068	0.9995	35.7286
	MIT-BIH-108	**13.2242**	**0.0580**	**58.7163**	0.9993	**228.1475**
	MIT-BIH-109	10.9115	0.1312	51.6198	0.9995	83.1587
Db10 L=4 D1=0 yes	MIT-BIH-100	**11.3563**	0.0020	88.0265	**1.0000**	**5.7226e+03**
	MIT-BIH-101	11.3007	0.0018	88.6740	**1.0000**	**6.1353e+03**
	MIT-BIH-102	10.4724	0.0034	83.4032	**1.0000**	**3.0991e+03**
	MIT-BIH-103	09.9020	0.0019	88.1814	**1.0000**	**5.0795e+03**
	MIT-BIH-104	11.4529	0.0034	83.4378	**1.0000**	**3.4028e+03**
	MIT-BIH-105	10.3457	**0.0016**	**89.7840**	**1.0000**	**6.3825e+03**
	MIT-BIH-106	11.6315	0.0039	82.1375	**1.0000**	**2.9754e+03**
	MIT-BIH-107	08.6742	0.0040	81.9599	**1.0000**	**2.1705e+03**
	MIT-BIH-108	08.6742	0.0019	88.4014	**1.0000**	**6.2610e+03**
	MIT-BIH-109	10.3723	0.0046	80.7925	**1.0000**	**2.2726e+03**
Haar ,L=4 D1=0 yes	MIT-BIH-100	13.2459	0.2220	47.0508	0.9912	59.6553
	MIT-BIH-101	12.6448	0.2031	47.8257	0.9944	62.2623
	MIT-BIH-102	09.8058	0.3570	42.9255	**0.9995**	27.4656
	MIT-BIH-103	11.5429	0.3547	42.9818	0.9940	32.5413
	MIT-BIH-104	12.5369	**0.1154**	**52.7355**	0.9986	**108.6408**
	MIT-BIH-105	11.6427	0.2107	47.5042	0.9979	55.2447
	MIT-BIH-106	**13.1169**	0.3524	43.0386	0.9953	37.2214
	MIT-BIH-107	08.9529	0.3520	43.0487	0.9991	25.4350
Coif1, L=4 D1=0 yes	MIT-BIH-100	12.6845	0.4486	40.9428	0.9642	28.2776
	MIT-BIH-101	12.4212	0.4100	41.7246	0.9774	30.2989
	MIT-BIH-102	10.8667	0.7155	36.8873	**0.9982**	15.1877
	MIT-BIH-103	11.6175	0.7153	36.8895	0.9756	16.2413
	MIT-BIH-104	12.3265	**0.2329**	**46.6347**	0.9943	**52.9180**
	MIT-BIH-105	11.4529	0.4232	41.4488	0.9915	27.0635
	MIT-BIH-106	**12.8662**	0.7107	36.9455	0.9809	18.1033
	MIT-BIH-107	09.8657	0.7069	36.9928	0.9966	13.9571

4.4. ECG Data Compression using Wavelet Transform and LZW

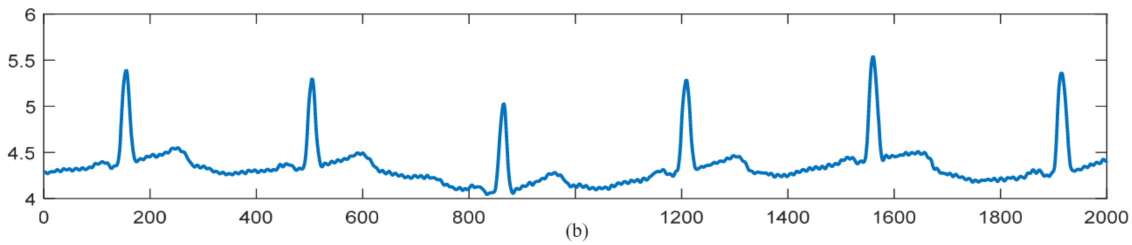
In this section, ECG data compression is done using WT and LZW coding. LZW is lossless "dictionary-based" compression algorithm [38]. LZW encoding is chosen as the entropy encoder for simplicity. The steps taken to compress ECG signal are given in Fig. 12. These steps are same as in the previous methodologies except coding method. The quantitative results of this methodology are presented in Table 6. Here, symlet, bi-orthogonal, Daubechies and Fejér-Korovkin wavelet filters are used and 5-level decomposition is done. Here, Db22 provides best result in terms of *CR* (**30.8060**) and Bior3.5 provides beat results in terms of PRD and QS.

Table 6 Performance of compression ratio using wavelet transform and LZW encoding

Level	Signal	CR	PRD	SNR	CC	QS
Sym6 L=5 D1=0 yes	MIT-BIH-100	**27.8974**	0.6331	43.9705	0.9838	44.0646
	MIT-BIH-101	25.9048	0.7368	42.6535	0.9826	35.1606
	MIT-BIH-102	20.5283	0.7657	42.3192	**0.9995**	26.8109
	MIT-BIH-103	21.9798	0.5572	45.0799	0.9965	39.4476
	MIT-BIH-104	23.6522	0.7141	42.9247	0.9869	33.1211
	MIT-BIH-105	23.3978	0.7231	42.8160	0.9939	32.3576
	MIT-BIH-106	23.6522	0.8150	41.7767	0.9940	29.0204
	MIT-BIH-107	18.2857	1.0776	39.3510	0.9981	16.9693
	MIT-BIH-108	25.9048	**0.4919**	**46.1624**	0.9884	**52.6625**
	MIT-BIH-109	18.9217	0.6491	43.7533	0.9968	29.1493
Fk4 L=5 D1=0 yes	MIT-BIH-100	**30.8060**	1.7270	29.2335	0.9036	17.8378
	MIT-BIH-101	25.1707	1.5992	29.9015	0.9692	15.7398
	MIT-BIH-102	19.6571	2.8617	24.8470	**0.9704**	06.8691
	MIT-BIH-103	22.4348	2.7785	25.1032	0.8461	08.0744
	MIT-BIH-104	23.4545	0.9250	34.6490	0.9110	25.3341
	MIT-BIH-105	22.9333	1.6685	29.5328	0.8691	13.7448
	MIT-BIH-106	22.4348	2.7697	25.1306	0.8187	08.1000
	MIT-BIH-107	18.2655	2.8114	25.0008	0.9467	06.4969
	MIT-BIH-108	25.1707	**0.6542**	**37.6648**	0.9177	**38.4740**
Db22 L=5 D1=0 yes	MIT-BIH-100	25.0566	0.7750	42.2144	0.9806	32.0304
	MIT-BIH-101	**32.0304**	0.7750	42.2144	0.9806	32.0304
	MIT-BIH-102	19.6741	0.8170	41.7551	**0.9994**	24.0795
	MIT-BIH-103	20.7500	1.2514	38.0519	0.9820	16.5810
	MIT-BIH-104	23.5044	0.7499	42.4995	0.9855	31.3418
	MIT-BIH-105	24.1455	0.7214	42.8364	0.9939	33.4700
	MIT-BIH-106	22.3193	0.7557	42.4335	0.9948	29.5363
	MIT-BIH-107	16.9172	0.9964	40.0316	0.9983	16.9789
	MIT-BIH-108	23.7143	**0.4756**	**46.4560**	0.9891	**49.8666**
	MIT-BIH-109	19.9699	0.5797	44.7361	0.9975	34.4493
L=5 Bior3.5 D1=0 yes	MIT-BIH-100	24.4494	0.5490	45.2080	0.9877	44.5314
	MIT-BIH-101	24.1778	0.5043	45.9460	0.9919	47.9419
	MIT-BIH-102	19.0877	0.7403	42.6117	**0.9995**	25.7833
	MIT-BIH-103	21.3333	0.6648	43.5459	0.9950	32.0888
	MIT-BIH-104	23.3978	0.7020	43.0729	0.9873	33.3287
	MIT-BIH-105	22.2041	0.5148	45.7671	0.9969	43.1310
	MIT-BIH-106	22.4330	0.8178	41.7470	0.9939	27.4307
	MIT-BIH-107	17.5484	0.8073	41.8598	0.9989	21.7385
	MIT-BIH-108	**24.7273**	**0.4079**	**47.7891**	0.9921	**60.6218**
	MIT-BIH-109	18.4407	0.5285	45.5386	0.9979	34.8905



(a)



(b)

Fig. 12 a) Original ECG signal (MIT-BIH record no. 121) and b) Reconstructed signal using 4-level decomposition by WT (bior3.5)

The original and reconstructed signal of this method is given in Fig. 11 using methodology given in Fig. 12. In this figure, 2000 samples of MIT-BIH record no. 121 has been taken, given in Fig. 12 (a). Signal is reconstructed by following the steps 6 to 9 (Fig 13), which is given in Fig. 11 (b). The comparison of these figure shows that both signals have identical wave.

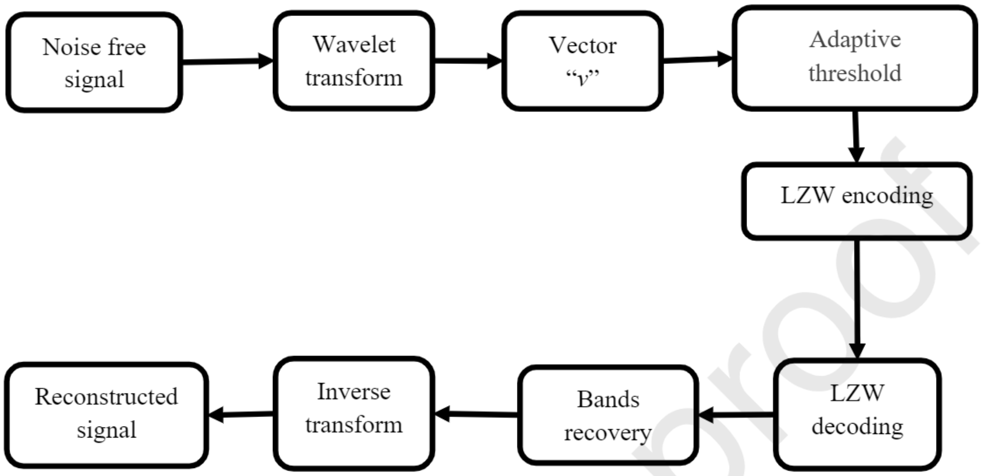


Fig. 13 ECG compression using wavelet and LZW encoding

In Fig. 13, steps of data compression and decompression are given. Here also MIT-BIH dataset is used to evaluate the performance of the ECG data compression. In this methodology, WT, adaptive threshholding and LZW are used to perform data reduction. Here, noise free signal is taken and then signal decomposition in done using WT into 6 signals. The coefficients of the decomposed signals are placed in one vector (*i.e.,* vector "*v*") and then adaptive thresholding is applied on the vector. After this LZW encoding is applied the vector coefficients. In this stage, compressed data is obtained. Compression ratio is measured using equation (11). For data reconstruction the first step is: LZW decoding is applied to the compressed signal. After this bands are formed by coefficient distribution. Finally, data is reconstructed using inverse WT.

5. Discussion

In this work, transformation-based methodologies are used to compress the ECG rhythms. The methodologies such as: a) WP algorithm and b) wavelet transform are employed for ECG data compression. In these methods, thresholding/ quantization and coding is also used after decomposition of the signal. The level (different threshold value used for different level decomposed signal) and for obtaining threshold value adaptive method is applied. Thresholding is done to truncate irrelevant coefficients of the signal. And then coding, such as; modified RLE/Huffman's/ LZW coding are applied. The performance comparison is examined by observing the compression ratio and signal reconstruction quality (using different fidelity parameters and the visual depiction of comparison of beats present in both the signals). From the tabular and graphical results, it can be perceived that modified RLE based methods are easy to implement. The performance of WP algorithm with thresholding and modified RLE is given in Tables 2 (a), (b), (c), (d) and (e) using different mother wavelet and level of decomposition. The average value of the performance indices are; *CR = 6.15*, PRD= 0.411857, *SNR= 48.50619,* *CC= 0.996853 and QS= 17.0833*. In WP algorithm, more decomposition are used in comparison to WT therefore, number of multiplications and additions are more. In WT, higher band decomposition is not necessarily required, therefore, in this work WT is also used for data compression. The performance indices of WT, thresholding and modified RLE based data compression methodology is presented in Table 4. *CR = 9.8725, PRD= 1.149583, SNR= 39.8572, CC= 0.974563* and *QS= 10.73856*. Performance of RLE based methodology is dependent on data type (if data has not similar coefficient values continuously than this method provide does not provide good result). It provides efficient results if repeatedly same sample values are presented in the data. Therefore, in thresholding based methodologies RLE algorithms are preferred, however in case of quantization based methodologies, entropy based encoding methods are preferred. In this work, WT, quantization and Huffman's coding and WT, quantization and LZW coding are used to compress the ECG data. The performances of these methodologies are given in Table 5 and 6. The average value of the performance indices of WT, quantization and Huffman's coding are; *CR = 11.4284, PRD= 0.082323, SNR= 64.87706 CC= 0.999303* and *QS= 1558.421*. And average value of the performance indices of WT, quantization and LZW coding are; *CR = 11.4284, PRD= 0.082323, SNR= 64.87706 CC= 0.999303* and *QS= 1558.421*

Table 6 shows that proposed WT and LZW based method gives high CR (~24.1778) and methodology using WT and Huffman's coding provides low PRD (0.0020), and overall QS value WT and Huffman's coding is superior to other. Therefore it can be observed that the limited bandwidth problem can be overcome by minimizing the number of bits in transmission of compressed ECG data. The applications of these methodologies include the real-time hospital environment that is estimated to give proficient solutions to ECG data storage and transmission difficulties.

The performance of this work is also compared with several existing methods (Table 7), which shows that methodology based on WT and LZW provides better results in terms of CR and PRD. The comparison of performances of data compression using WT and RLE, WT and HC and WT and LZW is given in Fig. 14 and 15. From these figures, it is observed that PRD value, which determines the data reconstruction quality, is best in case of WT and HC based method.

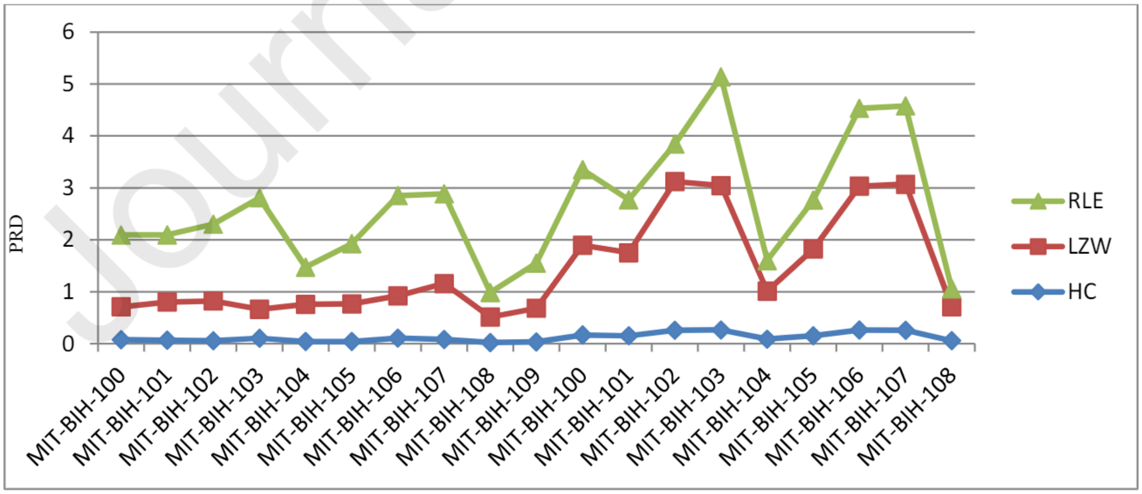
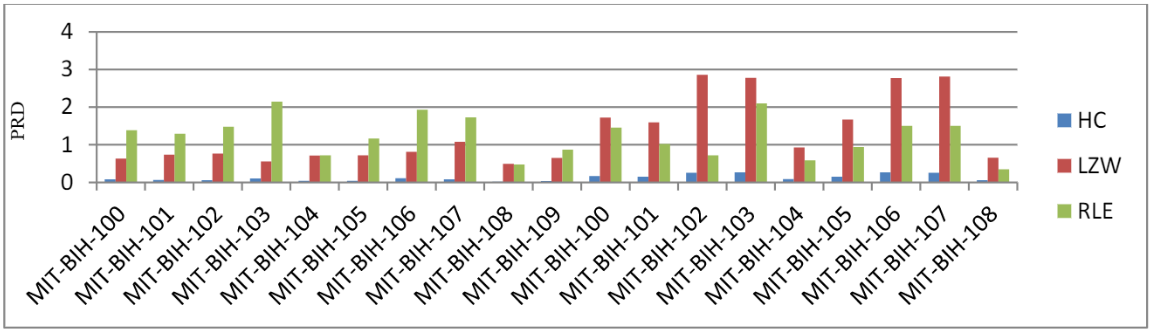


Fig. 14 PRD values by using RLE, HC and LZW



Signal record no.

Fig. 15 PRD values by using RLE, HC and LZW

Table 7 Performance comparisons with several other methods

	Level	Signal	CR	PRD	SNR	CC	QS
WT and LZW	**L=5 D1=0 yes**	**MIT-BIH-100**	**27.8974**	**0.6331**	**43.9705**	**0.9838**	**44.0646**
		MIT-BIH-101	**25.9048**	**0.7368**	**42.6535**	**0.9826**	**35.1606**
		MIT-BIH-108	**25.9048**	**0.4919**	**46.1624**	**0.9884**	**52.6625**
	Fk4 L=5 D1=0 yes	**MIT-BIH-100**	**30.8060**	**1.7270**	**29.2335**	**0.9036**	**17.8378**
		MIT-BIH-101	**25.1707**	**1.5992**	**29.9015**	**0.9692**	**15.7398**
		MIT-BIH-108	**25.1707**	**0.6542**	**37.6648**	**0.9177**	**38.4740**
	Db22 L=5 D1=0 yes	**MIT-BIH-100**	**25.0566**	**0.7750**	**42.2144**	**0.9806**	**32.0304**
		MIT-BIH-101	**32.0304**	**0.7750**	**42.2144**	**0.9806**	**32.0304**
	L=5 Bior3.5	**MIT-BIH-100**	**24.4494**	**0.5490**	**45.2080**	**0.9877**	**44.5314**
		MIT-BIH-101	**24.1778**	**0.5043**	**45.9460**	**0.9919**	**47.9419**
		MIT-BIH-108	**24.7273**	**0.4079**	**47.7891**	**0.9921**	**60.6218**
wavelet transform and Huffman's coding	**Sym2 L=4 D1=0 yes**	**MIT-BIH-100**	**13.1062**	**0.0765**	**56.3084**	**0.9990**	**171.3666**
		MIT-BIH-101	**12.8970**	**0.0665**	**57.5235**	**0.9994**	**193.9524**
		MIT-BIH-108	**13.3223**	**0.0220**	**67.1402**	**0.9999**	**606.2051**
	Fk4 L=4 D1=0 yes	**MIT-BIH-108**	**13.2242**	**0.0580**	**58.7163**	**0.9993**	**228.1475**
		MIT-BIH-106	**13.0980**	**0.2643**	**45.5361**	**0.9974**	**49.5492**
	Db10 L=4 D1=0 yes	**MIT-BIH-100**	**11.3563**	**0.0020**	**88.0265**	**1.0000**	**5.7226e+03**
		MIT-BIH-101	**11.3007**	**0.0018**	**88.6740**	**1.0000**	**6.1353e+03**
[17]	Wavelet Transform and Modified SPIHT	MIT-BIH record 107	8:1	0.035	-	-	-
This work	**WT and LZW**	**MIT-BIH record 107**	**17.5484**	**0.8073**	**41.8598**	**0.9989**	**17.5484**
[18]	WT using USZZQ and Huffman coding	MIT-BIH 119	21.81:1	3.716	-	-	-
This work	**WT and LZW**	**MIT-BIH record 119**	**20.7525**	**0.7268**	**42.7716**	**0.9977**	**28.5529**
[23]	uniform scalar dead-zone quantizatio	MIT-BIH 101	14.29	4.13	-	-	-
		MIT-BIH 111	21.48	5.99	-	-	-
This work	**WT and LZW**	**MIT-BIH record 111**	**23.0330**	**0.4536**	**46.8667**	**0.9924**	**50.7788**
[24]	DCT and Huffman Coding	MIT-BIH avg	8.82:1	1.82	-	-	-
This work	**WT and LZW**	**MIT-BIH avg**	**11.4284**	**0.082323**	**64.8770**	**0.99930**	**1558.421**
[25]	2-D ECG wavelet transform quantization	MIT-BIH 100.	30	3.16	-	-	-
[31]	DCT	MIT-BIH	13.02	2.72	-	-	-
			19.32	4.15		-	
[33]	Real-Time ECG Data Compression	MIT-BIH 121.	25.85	0.54	22.23	-	48.0
		MIT-BIH 207	24.45	0.61	21.8	-	41.31
This work	**WT and LZW**	**MIT-BIH record 121**	**24.9524**	**0.3539**	**49.0235**	**0.9976**	**70.5155**
This work	**WT and LZW**	**MIT-BIH record 207**	**24.413**	**0.2361**	**47.8667**	**0.9917**	**103.40**
[34]	wavelet Filter Bank based on Kaiser Window	MIT-BIH 124	5.37	0.96	40.30	-	-
		MIT-BIH 112	7.11	1.29	37.77	-	-
This work	**WT and LZW**	**MIT-BIH record 112**	**21.1717**	**0.4931**	**46.1419**	**0.9940**	**42.9389**
[37]	Beta wavelet with lossless coding	MIT-BIH Rec. 112	5.33	0.61	34.19	-	8.73
[41]	WT 'db3'	MIT-BIH Rec. 205	7.7953	0.3995	-	-	-
	WT 'haar'	MIT-BIH Rec. 205	10.8167	0.4320	-	-	-
	WT 'sym2'	MIT-BIH Rec. 205	7.8587	0.3798	-	-	-
	WT 'coif1'	MIT-BIH Rec. 205	7.8416	0.3511	-	-	-
	WT 'db10'	MIT-BIH Rec. 205	7.4060	0.4105	-	-	-
This work	**WT and LZW**	**MIT-BIH Rec. 205**	**42.1587**	**0.4072**	**47.8032**	**0.9262**	**103.5257**

6. Conclusions

In this work, transform based data compression of ECG signal is presented. Presented data compression algorithms provide higher compression performance with low signal quality degradation at the decomposition than several other algorithms. Therefore, these methodologies can be used at the transmission. These are protected because transmitted data are encoded with decomposed coefficients. Hence the method is applicable to the 1-D signal compression with more security. Several signals are used to obtain the results of the methodologies taken from the MIT-BIH database. From the tabular and graphical results, it can be observed RLE based methods are easy to implement. However, Huffman and LZW based methods provide better results in terms of CR and PRD both. In comparison to Huffman encoding method, LZW based method is superior. It is also observed that these methods can use to compress all kinds of ECG rhythms (single lead) in an efficient way. Using these methodologies, truncation of a bit stream at any point can be done also a good quality of reconstruction can be obtained. In comparison to other class of wavelet, Db10, Fk4 and Db22 provides better results in terms of CR and PRD. The method present in this work is also superior to other compression method according to Table 7.

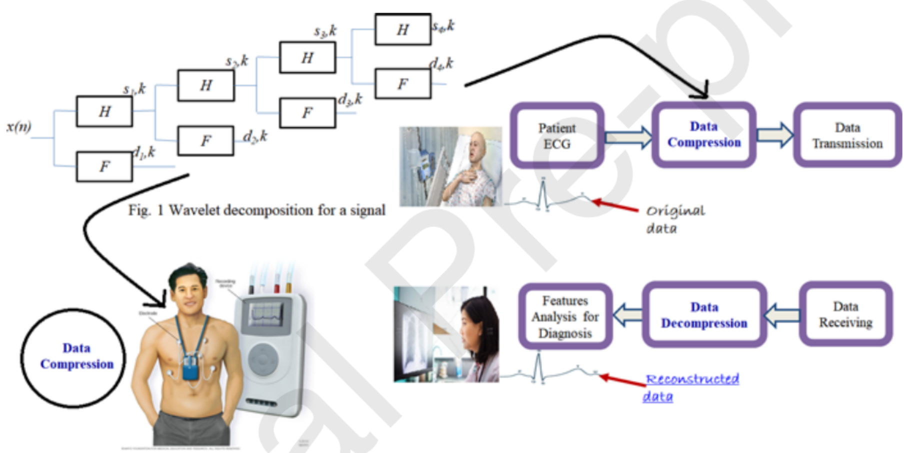
References

1. Yildirim, Özal. "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification." *Computers in biology and medicine* 96 (2018): 189-202.
2. Claggett, Steven, Sahar Azimi, and Martin Burtscher. "SPDP: An Automatically Synthesized Lossless Compression Algorithm for Floating-Point Data." *2018 Data Compression Conference*. IEEE, 2018.
3. Otero, E., et al. "The effect of lossy data compression in computational fluid dynamics applications: resilience and data postprocessing." *Direct and Large-Eddy Simulation XI*. Springer, Cham, 2019. 175-181.
4. Jalaleddine, Sateh MS, et al. "ECG data compression techniques-a unified approach." *IEEE transactions on Biomedical Engineering* 37.4 (1990): 329-343.
5. Bradie B., "Wavelet packet-based compression of single lead ECG", *IEEE transactions on Biomedical Engineering*, vol. 43, no. 5, pp. 493-501, May 1996
6. Chen J. and Itoh S., "A wavelet transform based ECG compression method guaranteeing desired signal quality", *IEEE transactions on Biomedical Engineering*, vol. 45, no. 12, pp. 1414-1419, December 1998.
7. Anant K, Dowla F. and Rodrigue G., "Vector quantization of ECG wavelet coefficients", *IEEE signal processing letters*, vol.2, no. 2, July 1995.
8. Rajoub B.A., "An efficient coded algorithms for the compression of ECG signals using the wavelet transform", *IEEE transactions on Biomedical Engineering*, vol. 49, no. 4, pp. 355-362, April 2002
9. Kulkami P.K., Kumar Vinod and Verma H.K., "Diagnostic acceptability of FFT- based ECG data compression", *Journal of Medical Engineering and Technology*, vol. 21, no. 5, pp. 185-189, Sept-Oct., 1997
10. Hwang WJ, Chine CF, Li KJ. Scalable medical data compression and transmission using wavelet transform for telemedicine applications. *IEEE transactions on information technology in biomedicine*. 2003 Mar;7(1):54-63.
11. Moazami-Goudarzi M, Moradi MH. Electrocardiogram signal compression using multiwavelet transform. *IEEE Signal processing*. 2005 Jun;4:12.
12. Aggarwal V, Patterh MS. Quality controlled ECG compression using Discrete Cosine transform (DCT) and Laplacian Pyramid (LP). *Multimedia, Signal Processing and Communication Technology*, IMPACT. 2009 Mar 14;2009:12-5.
13. Miaou SG, Chao SN. Wavelet-based lossy-to-lossless ECG compression in a unified vector quantization framework. *IEEE Transactions on Biomedical Engineering*. 2005 Mar;52(3):539-43.

14. Tai SC, Sun C, Yan WC. A 2-D ECG compression method based on wavelet transform and modified SPIHT. *IEEE Transactions on Biomedical Engineering*. 2005 Jun;52(6):999-1008.
15. Rezazadeh IM, Moradi MH, Nasrabadi AM. Implementing of SPIHT and sub-band energy compression (SEC) method on two-dimensional ECG compression: a novel approach. In Engineering in Medicine and Biology Society, 2005. *IEEE-EMBS 2005. 27th Annual International Conference of the 2006 Jan 17* (pp. 3763-3766). IEEE.
16. Al-Fahoum AS. Quality assessment of ECG compression techniques using a wavelet-based diagnostic measure. IEEE *Transactions on Information Technology in Biomedicine*. 2006 Jan;10(1):182-91.
17. Ku CT, Wang HS, Hung KC, Hung YS. A novel ECG data compression method based on nonrecursive discrete periodized wavelet transform. *IEEE Transactions on Biomedical Engineering*. 2006 Dec;53(12):2577-83.
18. Manikandan MS, Dandapat S. Wavelet threshold based ECG compression using USZZQ and Huffman coding of DSM. *Biomedical Signal Processing and Control*. 2006 Oct 1;1(4):261-70.
19. Tohumoglu G, Sezgin KE. ECG signal compression by multi-iteration EZW coding for different wavelets and thresholds. *Computers in Biology and Medicine*. 2007 Feb 1;37(2):173-82
20. Brechet L, Lucas MF, Doncarli C, Farina D. Compression of biomedical signals with mother wavelet optimization and best-basis wavelet packet selection. *IEEE Transactions on Biomedical Engineering*. 2007 Dec;54(12):2186-92.
21. Benzid R, Messaoudi A, Boussaad A. Constrained ECG compression algorithm using the block-based discrete cosine transform. *Digital Signal Processing*. 2008 Jan 1;18(1):56-64.
22. Sharifahmadian E. Wavelet compression of multichannel ECG data by enhanced set partitioning in hierarchical trees algorithm. *In Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE 2006 Aug 30* (pp. 5238-5243). IEEE.
23. Chen J, Wang F, Zhang Y, Shi X. ECG compression using uniform scalar dead-zone quantization and conditional entropy coding. *Medical Engineering & Physics*. 2008 May 1;30(4):523-30.
24. Lee S, Lee M. A real-time ECG data compression algorithm for a digital holter system. *In Engineering in Medicine and Biology Society, 2008. EMBS 2008.* 30th Annual International Conference of the IEEE 2008 Aug 20 (pp. 4736-4739). IEEE.
25. Wang X, Meng J. A 2-D ECG compression algorithm based on wavelet transform and vector quantization. *Digital Signal Processing.* 2008 Mar 1;18(2):179-88.
26. Aggarwal V, Patterh MS. Quality controlled ECG compression using Discrete Cosine transform (DCT) and Laplacian Pyramid (LP). *Multimedia, Signal Processing and Communication Technology, IMPACT.* 2009 Mar 14;2009:12-5.
27. Lee, Rong-Choi, and King-Chu Hung. "New Modified SPIHT Algorithm for Data Compression System." *Journal of Medical and Biological Engineering* 39.1 (2019): 18-26
28. Ku CT, Hung KC, Wu TC, Wang HS. Wavelet-based ECG data compression system with linear quality control scheme. *IEEE Transactions on Biomedical Engineering*. 2010 Jun;57(6):1399-409.
29. Kim H, Yazicioglu RF, Merken P, Van Hoof C, Yoo HJ. ECG signal compression and classification algorithm with quad level vector for ECG holter system. *IEEE Transactions on Information Technology in Biomedicine*. 2010 Jan;14(1):93-100.
30. Shinde AA, Kanjalkar P. The comparison of different transform based methods for ECG data compression. *In Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), 2011 International Conference on 2011* Jul 21 (pp. 332-335). IEEE.
31. 101. Bendifallah A, Benzid R, Boulemden M. Improved ECG compression method using discrete cosine transform. *Electronics letters*. 2011 Jan;47(2):87-9.
32. Polania LF, Carrillo RE, Blanco-Velasco M, Barner KE. Compressed sensing based method for ECG compression. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on 2011 May 22* (pp. 761-764). IEEE.

33. Lee S, Kim J, Lee M. A real-time ECG data compression and transmission algorithm for an e-health device. *IEEE Transactions on Biomedical Engineering*. 2011 Sep;58(9):2448-55.
34. Sharma LN, Dandapat S, Mahanta A. Multichannel ECG data compression based on multiscale principal component analysis. *IEEE Transactions on Information technology in Biomedicine.* 2012 Jul 1;16(4):730-6.
35. Ranjeet K, Kuamr A, Pandey RK. ECG signal compression using optimum wavelet filter bank based on Kaiser window. *Procedia engineering*. 2012 Jan 1;38:2889-902.
36. Huang B, Wang Y, Chen J. ECG compression using the context modeling arithmetic coding with dynamic learning vector–scalar quantization. *Biomedical Signal Processing and Control*. 2013 Jan 1;8(1):59-65.
37. Kumar R, Kumar A, Pandey RK. Beta wavelet based ECG signal compression using lossless encoding with modified thresholding. *Computers & Electrical Engineering*. 2013 Jan 1;39(1):130-40.
38. Moody GB, Mark RG. The MIT-BIH arrhythmia database on CD-ROM and software for use with it. In[1990] *Proceedings Computers in Cardiology* 1990 Sep 23 (pp. 185-188). IEEE.
39. Knieser, Michael J., et al. "A technique for high ratio LZW compression." *Proceedings of the conference on Design, Automation and Test in Europe-Volume 1*. IEEE Computer Society, 2003.
40. Lu Z, Kim DY, Pearlman WA. Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm. *IEEE transactions on Biomedical Engineering*. 2000 Jul;47(7):849-56.
41. Tun HM, Moe WK, Naing ZM. Analysis on ECG Data Compression Using Wavelet Transform Technique. *International Journal of Psychology*. Brain Sci. 2017 Dec 22; 2:127.
42. Parkale YV, Nalbalwar SL. Investigation on Daubechies Wavelet-Based Compressed Sensing Matrices for ECG Compression. *In Computing, Communication and Signal Processing 2019* (pp. 707-716). Springer, Singapore.
43. Rebollo-Neira L. Effective high compression of ECG signals at low level distortion. *Scientific reports*. 2019 Mar 14;9(1):4564.
44. Jha CK, Kolekar MH. Diagnostic quality assured ECG signal compression with selection of appropriate mother wavelet for minimal distortion. *IET Science, Measurement & Technology*. 2019 Jan 7;13(4):500-8.
45. Ahmed N, Milne PJ, Harris SG. Electrocardiographic data compression via orthogonal transforms. *IEEE Transactions on Biomedical Engineering*. 1975 Nov (6):484-7.
46. Lin CH, Xie Y, Wolf W. LZW-based code compression for VLIW embedded systems. *In Proceedings Design, Automation and Test in Europe Conference and Exhibition 2004 Feb 16* (Vol. 3, pp. 76-81). IEEE.
47. https://www.dspguide.com/ch27/5.htm

graphical abstract



Author responsibilities, integrity, ethics

Article title :

ELSEVIER

Human and animal rights

☑ The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans as well as in accordance with the EU Directive 2010/63/EU for animal experiments.

☐ Not applicable.

Informed consent and patient details

☑ The authors declare that this report does not contain any personal information that could lead to the identification of the patient(s).

☐ Not applicable

☑ The authors declare that they obtained a written informed consent from the patients and/or volunteers included in the article. The authors also confirm that the personal details of the patients and/or volunteers have been removed.

☐ Not applicable

Disclosure of interest

☐ The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

☐ The authors declare the following financial or personal relationships that could be viewed as influencing the work reported in this paper:

Funding

☑ This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

☐ This work has been supported by:

Author contributions

☑ All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Acknowledgements

Authors to thank Indian institute of Technology, for their support to carried out this work.

Author Statement

Title of Paper: A Comparative Analysis of Performance of Several Wavelet Based ECG Data Compression Methodologies

Name of Principal Author: Shanti Chandra

Contribution to the Paper: Performed analysis on all samples, interpreted data, wrote manuscript and acted as corresponding author.

Signature: *shanti* (Shanti Chandra)

Name of Co-Authors: 1) Ambalika Sharma

2) Girish Kumar Singh

Co-Author Contributions: Supervised development of work, helped in data interpretation and manuscript evaluation

Signature: 1) *asharma* (Ambalika Sharma) and **2)** *girish* (Girish Kumar Singh)

Energy-Delay Based Route Request Scheme for Load Balanced Routing in Wireless Mesh Networks

Keerthi D S
Deparment of Electronics &Communication Engineering
Malnad College of Engineering
Hassan,India
keerthikumards@gmail.com`

Shobha Rani A
Department of Electronics and Communication Engineering
BMSIT
Bengaluru, India
shobharani@bmsit.nic.in

T G Basavaraju
Department of Computer Science & Engineering
Govt. SKSJTI
Bengaluru, India
tg.braju@gmail.com

***Abstract*—The tremendous growth in the usage of internet technology has pawed way for advancement of diverse wireless networks. One of the domain of wireless networks that have gathered a lot of attention from researchers is Wireless Mesh Network (WMN). WMN is preferred because of its numerous advantages like better coverage area, communication with other networks, cost effectiveness, increased network capacity and is compatible with all IEEE 802.11 standards. Low Capacity, management of the traffic and gateway nodes and end to end fairness problem are the major issues in the employment of WMNs. In order to reduce the existing challenges in WMNs load balancing plays a vital role. In our work we have proposed a load balancing protocol to overcome the problem of congestion using residual energy and delay. The simulation results are derived and they prove that our proposed protocol enhances the performance and life time of WMN compared to standard protocol with reference to residual energy, throughput, packets received jitter and delay.**

Keywords—WMN, Load balancing, Energy, Delay

I. Introduction

A wireless mesh network is an infrastured network which is based on mesh topology. The architecture of a WMN [1] consists of three major components which are clients, routers and gateways. The clients are often laptops, desktops, tablets, and mobile phones etc. which are end devices. The router performs the task of traffic management through gateways. In a mesh network, even when a single node fails to operate, the communication is not hampered as the rest of the nodes are still able to establish communication with each other either directly or via intermediate nodes. Thus this uniqueness of WMNs has made them highly reliable, resilient and offers a higher degree of redundancy. The integration of WMN [2] technology can also be implemented along technologies such as 802.11, 802.16, 802.15 standards in cellular technologies.

A WMN can be visualized as a distinctive wireless Ad-hoc network. However a WMN has a systematic i.e. configuration and delivers a dynamic and cost effective deployment of network over a specific geographical area. An Ad-hoc network is a unpremeditated network where communication can be established only when the wireless devices come in each other's communication range. WMN [3] can also be used for wide variety of applications such as transport systems, wireless sensor network, surveillance system, broadband home networking, building automation, health and medical system, community and neighborhood networking.

A mobile node consumes its battery energy not only when it actively sends or receives packets but also when it stays idle listening to the wireless medium for any possible communication requests from other nodes. Thus, energy efficient routing protocols minimize either the active communication energy required to transmit and receive data packets or the energy during inactive periods. In some protocols the active communication energy can be reduced to reach the receiving node by adjusting the node's radio power just enough to reach the receiving node. The optimal routing path that reduces the total transmission energy needed to send the data packets to the destination can be determined by using the transmission power control method. When there is no data to send or receive each node can retain the inactivity energy changing its mode of operation into sleep power down mode or simply turn it off like in other protocols. Considerable energy saving can be achieved here, if the network conditions is specified with low duty cycle of communication activities. But for guaranteed data transmission when many nodes are sleep and not forwarding packets for the next nodes it may require much well designed routing protocols. Load distribution approach in another powerful approach for optimizing active communication energy. It maximizes the new lifetime when selecting a routing path by avoiding over utilized nodes and also balances the energy usage among nodes. Though each protocol will have advantages and disadvantages and is convenient for certain situation, it is not clear that any one algorithm or a clan of algorithms is ideally suited for all situations. However, it is possible to offer a better energy saving routing mechanism by combining and integrating the existing solutions. Many efforts have been directed to develop energy aware-MAC and transport protocols to tackle energy efficiently in other network layers. In layered network architecture each layer is supposed to operate in isolation. But to maximize the energy performance some recent studies have recommended the cross layer design.

The paper is arranged in the following manner. The existing technology related to proposed work is discussed in second Section. The proposed routing scheme EBL-AODV is completely described in third section. The simulation results, performance comparison and analysis are presented in fourth section. The inference of the paper is in fifth section.

II. Related Work

In this work we have proposed a modified AODV protocol called Energy-Delay Based Load Balancing Ad-hoc on Demand Protocol (EDBL-AODV), which considers residual energy and path delay as a routing metric during route discovery.

Aron et al. [4] have described a novel distributed topology control algorithm whose main objective is to reduce the consumption of energy by each node during transmission and extending lifetime of the network. It is a three phased topology control algorithm which involves establishing accessible neighborhood topology, constructing minimum energy local topology view and finally determining transmission power. Thus the above algorithm overcomes the topology control problem in hybrid WMNs equipped with heterogeneous wireless devices and altering transmission ranges by taking decisions based on information gathered locally and scaling it for large WMN.

Panagiotis et al. [5] have proposed a new energy efficient multi cost routing algorithm for WMN. According to the new routing algorithm, a vector of cost parameters are calculated at each link using appropriate operators for selecting an optimal path. The parameters are hop count, power required by the node to transmit and the remaining energy at each node. There are two models- Network evaluation model and dynamic one to one communication model under which the performance of the proposed protocol is evaluated. In the former evaluation model, the network aims to transmit maximum number of packets available preceding energy depletion. In the latter model, the data packets are generated uninterruptedly as the nodes are competent of periodically recharging their energy. This protocol is said to enhance the durability of the network.

Awad et al. 6] have proposed a cross layer optimization algorithm subject to a delay deadline constraint. The assumption is that all packets successfully received should have their end to end delay less than their corresponding delay deadline. The proposed algorithm optimizes network and physical layer by exploiting different control parameters viz. path selection parameter in the network layer, the transmission and modulation energy in the physical layer. The protocol is a little time consuming as for every path in the network, all cross layer parameters are calculated through all paths in the networks. The authors [7] have proposed a novel routing model called ERDB-AODV protocol in which remaining energy in the battery and energy draining rate are incorporated as additional metrics for an optimal path selection. Thus, the hop count, available energy at the node and the rate of energy consumption are considered as important factors of the path selection. The protocol avoids nodes with higher energy consumption and lesser energy levels to extend the lifetime of the network. However, this may affect the traffic to traverse a long path to reach destination despite successful transfer.

III. Energy-Delay Based Load Balancing AODV Protocol

We have proposed a novel routing protocol Energy-Delay based load balancing EDBL-AODV which is improved form of DBL-AODV and which considers both energy and delay for path selection. EDBL-AODV has a modified routing table entry, route discovery and selection process. The additional constituents included in EDBL-AODV are expiration time field, hop count field and advertised hop court for achieving maximum possible paths. The operation of EDBL-AODV routing protocol can be described in Figure 1.

The source node broadcasts the RREQ packet. For immediate nodes with the help of RREQ packet computes the minimum unused energy depending on the incurred energy consumption and delay. The energy consumed can be obtained by the equation:

EC = Power Received or Transmitted * (8 * Packet size / Bandwidth).

$$EC = P * \left(8 * \frac{PS}{B_w}\right) \tag{1}$$

where, EC = Energy consumption.
P = Power Received/Transmitted.
PS = Packet size.
B_W = Bandwidth.

The relay nodes enter the delay and calculated energy consumption in RREQ message. They also check sequence number for a new route and update the same in routing table. The destination node replies to the source node via RREP message. During route selection, the same node chooses an optimal path possessing lesser delay and maximum residual energy.

The delay is calculated in the algorithm is as given below

Delay= (time at which RREQ received at current node) - (time at which RREQ sent from previous node) (2)

The average delay (D_a) is calculated by

$$D_a = \frac{\sum(Tpr - Tps)}{T} \tag{3}$$

where, T_{pr} = Packet received time
T_{ps} = Packet sent time.
T = Total data packets received.

The path selection (PA) process is performed by using

$$P_s = \text{Max}\left[\sum\{(\frac{B1r}{B1f})\ (\frac{B2r}{B2f})\ \ldots\ldots\ldots\ldots(\frac{Bnr}{Bnf})\}\right] \text{ and Min}\left[\sum\{(D_1)\ (D_2)\ \ldots\ldots\ldots\ldots\ldots\ldots(D_n)\}\right] \tag{4}$$

where, n = advertised hop count
B_r = remaining energy in node
B_f = full energy in node
D = time delay

IV. Performance Evaluation

The effectiveness of EDBL-AODV algorithm is verified by comparing it with the AODV protocol algorithm using Qualnet simulator. The parameters for all simulations are shown in Table 1. We change the number of nodes froml6 to 120. For Each node density higher CBR is generated to create the congestion. We used five different quantitative measures to compare the performance of the proposed and existing protocol. In this comparative study, our proposed protocol ‘EDBL-AODV’ maintains network life time, higher throughput, and lesser delay unlike in AODV protocol.

We fix the packet size as 1024 bytes and change the number of node densities from 16 to 120 to observe the impact of load and node densities on the network performances. The performances of AODV and EDBL-AODV protocols are shown in Fig.1 to Fig.5.

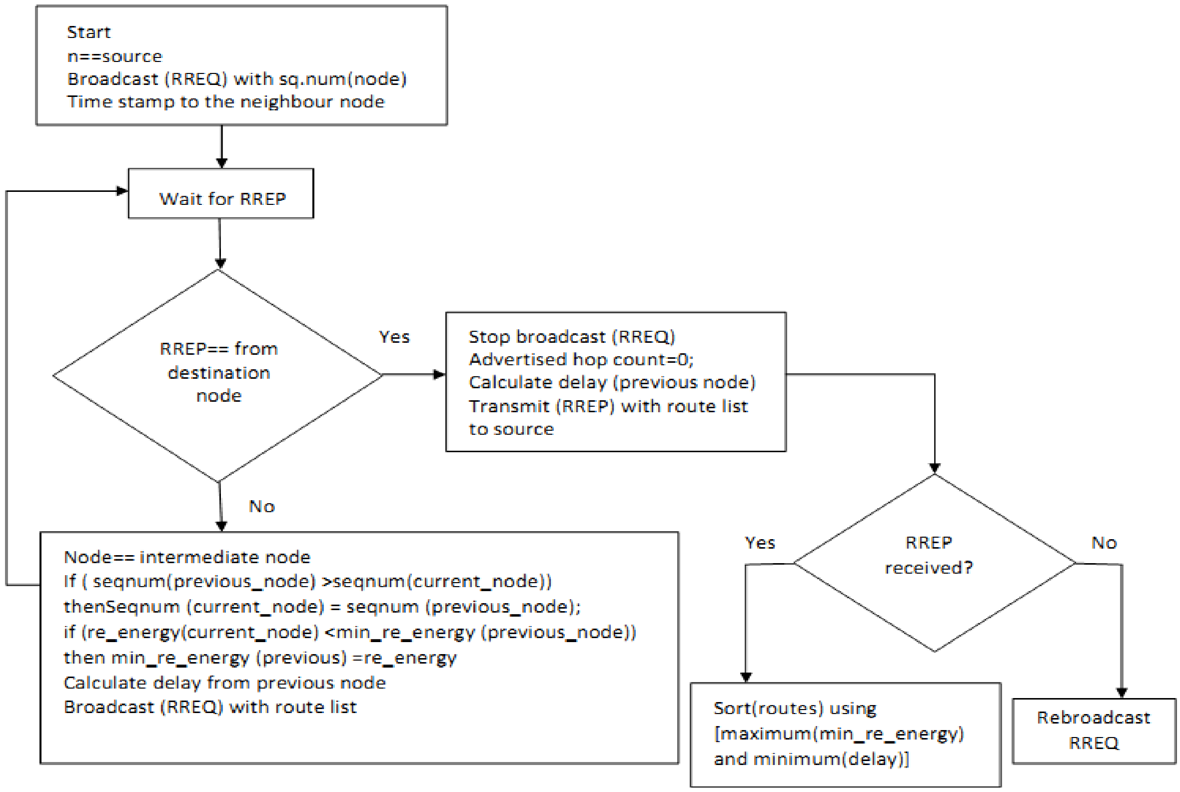


Figure 1: Flow chart of the algorithm

Table 1.Simulation Parameters

Parameter	Value
Protocols	AODV, EDBL-AODV
No. of Nodes	16, 25, 45, 70, 90, 120
Radio type	802.11b Radio
MAC Protocol	802.11s
Antenna Model	Omni-directional
Path Loss Model	Two ray propagation
Traffic Type	CBR

Fig.1 shows throughput comparison with respect to node density. As the number of nodes increases, the average network throughput of EDBL-AODV is always higher than AODV because data transmission is performed based on less delay, path having higher residual energy and minimum number in less time.

Fig. 2 shows that the total number of messages of nodes path. Due to less delay, more number of bits is transmitted received of EDBL-AODV remains higher than AODV with different number of node density since packet transmission is mainly based on less delay and minimum distance path. Due to this consideration, packets are transmitted evenly without any occurrence of node failure. Therefore, higher amount of packets is transmitted. As shown in Fig. 3, the average end-to-end delay of EBL-AODV is lower than that of AODV. Since AODV RREQ packet is modified by introducing delay parameter. Routing is performed using the delay parameter; less delay path is selected for packet transmission. Therefore, end-to-end delay is reduced. For the existing model, AODV packet transmission is performed based on minimum distance which increases congestion and leads to delay, which is even true for Jitter.

Fig. 5 shows that the residual energy of EDBL-AODV is higher than AODV for different node densities. This is because packets are transmitted based on higher residual energy, less delay and minimum distance path. Due to this consideration, packets are not queued at nodes which lead to less resource utilization and less energy consumption.

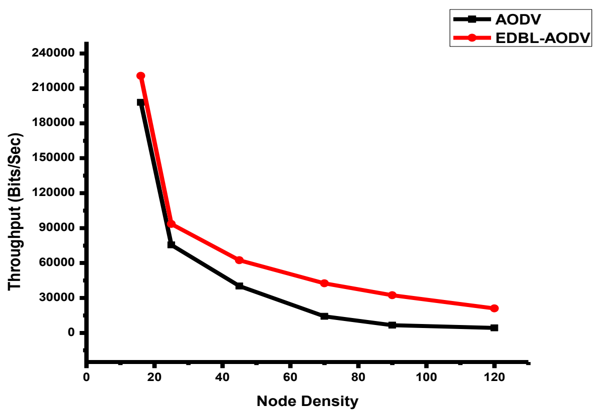


Figure 2: Variation of throughput v/s node densities

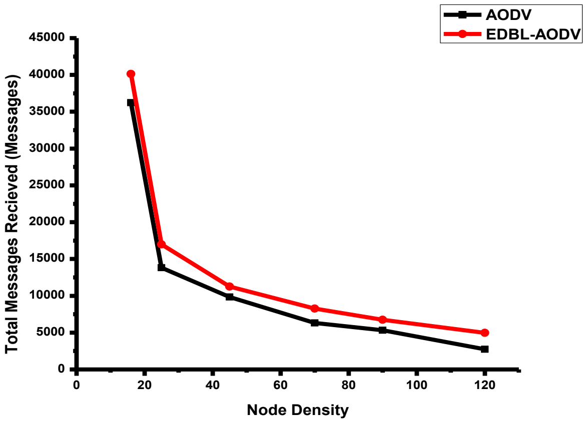


Figure3: Variation of total messages received v/s node densities

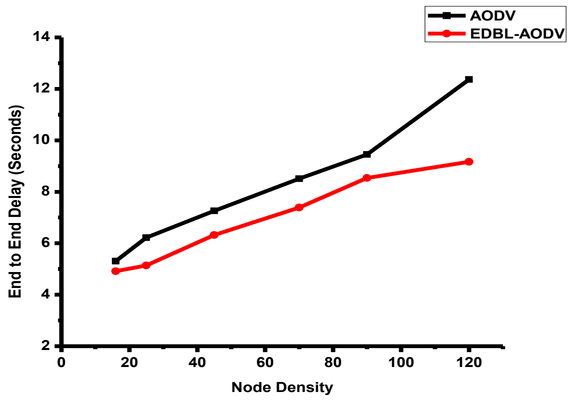


Figure 4: Variation of End to End Delay v/s node densities

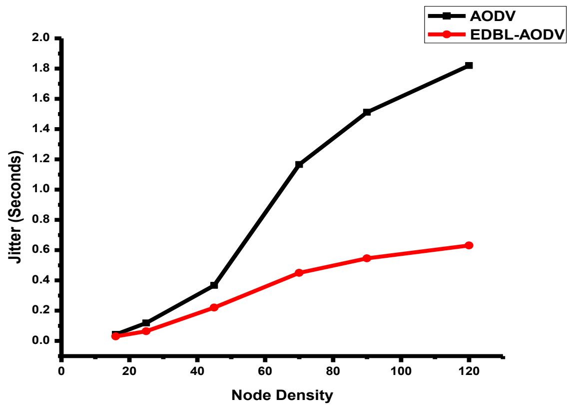


Figure 5: Variation of Jitter v/s node densities

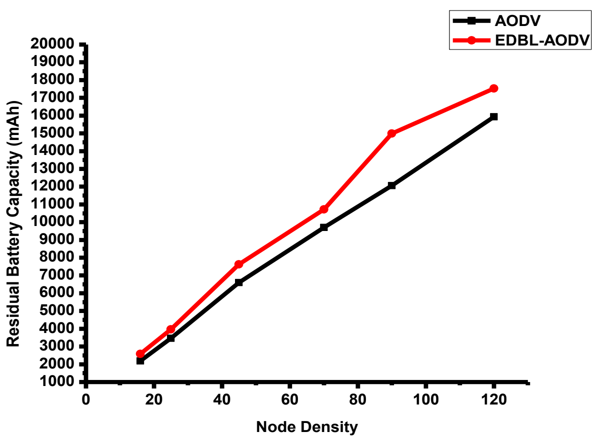


Figure 6: Residual Battery Power v/s node densities

V. CONCLUSION

In our proposed work energy efficiency is enhanced considerably by reducing routing overhead and delay. Comparative study has been carried out with existing routing protocol and the results are exhibited. Simulation results showed that the proposed protocol improves performance of the network and prolongs the lifetime of a WMN in comparison with the standard protocol. On an average, there is an overall increase in throughput of 12 to 14%, total number of messages received 15 to 18%, decrease in energy consumption of 4 to 6% and decrease in end-to-end delay of 6 to 8% when compared between proposed and existing algorithm.

ACKNOWLEDGMENT

This work is supported by the Technical Education Quality Improvement Program Phase III with the World Bank assistance. The authors would like to thank State Project Facilitation Unit, Directorate of Technical Education, Government of Karnataka, National Project Implementation Unit, MHRD, New Delhi for their support and encouragement.

REFERENCES

[1] F. Akyildiz, X. Wang, and W. Wang "Wireless mesh networks: a survey", in Computer Networks Journal, vol.47, 2005, pp.445-487.

[2] Eslami, M, Karimi, O, Khodadadi T "A survey on wireless mesh networks: Architecture, specifications and challenges" Control and System Graduate Research Colloquium (ICSGRC), 2014, pp. 219 – 222.

[3] M. L. Sichitiu "Wireless mesh networks: Opportunities and challenge," in Proc. of the Wireless World Congress, (Palo Alto, Ca), May 2005.

[4] F. O. Aron, T. O. Olwal, A. Kurien, and M. O. Ohanga, "A distributed topology control algorithm to conserve energy in heterogeneous wireless mesh networks," *World Acad. Sci. Eng. Technol.*, vol. 40, 2009.

[5] P. Kokkinos, C. Papageorgiou, and E. Varvarigos, "Multi-cost routing for energy and capacity constrained wireless mesh networks," *Wirel. Commun. Mob. Comput.*, vol. 13, no. 4, pp. 424–438, Mar. 2013.

[6] A. Awad, O. A. Nasr, and M. M. Khairy, "Energy-aware routing for delay-sensitive applications over wireless multihop mesh networks," in *2011 7th International Wireless Communications and Mobile Computing Conference*, 2011, pp. 1075–1080.

[7] N. TN, "Energy Draining Rate Based-AODV Routing in Wireless Mesh Network, In 12th FRUCT conference, ISSN2305-7254." [Online]. Available: https://fruct.org/publications/abstract12/files/Nag.pdf. [Accessed: 10-Sep-2016].

[8] Kamal Mistry; Sanjay Srivastava; R. B. Lenin "Buffer aware routing in interplanetary Ad Hoc Network", First International Communication Systems and Networks and Workshops, 2009, pp. 1-10.

[9] JaniLakkakorpi, MikkoPitkänen and JörgOtt "Using buffer space advertisements to avoid congestion in mobile opportunistic DTNs" Proceedings of the 9th IFIP TC 6 international conference on Wired/wireless internet communications, 2011, pp. 386-397.

[10] Maroufi B.E., Rahmani M.D., Rziza M. (2016) Congestion Avoidance in AODV Routing Protocol Using Buffer Queue Occupancy and Hop Count. In: El Oualkadi A., Choubani F.

[11] El Moussati A. (eds) Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015. Lecture Notes in Electrical Engineering, vol 381. Springer, Cham

[12] KrutiN. Kapadia. Dayanand D. Ambawade "Congestion aware load balancing for multi-radio Wireless Mesh Network" International Conference on Communication, Information & Computing Technology (ICCICT) 2015.

[13] Raffaele Bruno, Marco Conti, Antonio Pinizzotto "A Queuing Modeling Approach for Load-Aware Route Selection in Heterogeneous Mesh Networks", IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks & Workshops 2009.

[14] Sanjay Kumar Dhurandher, Isaac Woungang, Mohammad S. Obaidat "A Distributed Adaptive Admission Control Scheme for Multimedia Wireless Mesh Networks" IEEE Systems Journal Volume: 9, Issue: 2, June 2015, pp. 595 – 604.

[15] T. G. Basavaraju, S. K. Sarkar, C. Puttamadappa, and M. A. Gautham, "ECARP: an efficient congestion adaptive routing protocol for mobile Ad hoc networks," in Proceedings of the 6th International Conference on ITS Telecommunications Proceedings, pp. 715–718, Chengdu, China, June 2006.

[16] X. Chen, H.M. Jones, and A. D. S. Jayalath, "Congestion-aware routing protocol for mobile ad hoc networks," in Proceedings of IEEE Conference of Vehicular Technology, pp. 21–25,October 2007.

[17] Keerthi D S, T.G. Basavaraju "Load Balancing Routing Mechanisms For Wireless Mesh Networks: A Survey", in Fourth International Conference on "Emerging Research In Computing, Information, Communication And Applications (ERCICA)", 2016. (Published in Springer proceedings)

[18] Sooyeol Yang,Youngmi Baek,Junhyung Kim,Keuchul Cho and Kijun Han "A routing metric for load balance in wireless mesh networks", 11th Intrmational Conference on Advanced Communication Technology (ICACT) Volume: 03,2009,pp.1560-1565.

[19] Ekram Hossain , Kin K. Leung "Wireless Mesh Networks: Architectures and Protocols", Springer, 2008.

.

Journal of
Computational and Theoretical Nanoscience
Vol. 17, 1–5, 2020

Detection and Analysis of Plant Leaf Diseases Using Convolutional Neural Network

K. R. Mamatha[1,*], Seema Singh[2], and S. A. Hariprasad[2]

[1]Department of Electronics an Communication, BMS Institute of Technology & Management, Visvesvaraya Technological University, Bengaluru, India
[2]Department of Electronics and Telecommunication Engineering, BMS Institute of Technology & Management, Visvesvaraya Technological University, Bengaluru, India
[3]Department of Electronics and Communication Engineering, Jain University, Bengaluru, India

India is the second largest food producer in the world. Agriculture plays an important role in Indian economy. Plants that are grown may experience many diseases which if not detected might lead to reduction of crops and further leads to the reduction of quality and quantity of plant products to be delivered to the market. Diseases can occur and the farmer may not be in a condition to identify the disease and he might be unaware of which pesticide to be used and the precautions to be taken in order to over-come it in future. This paper describes as how to detect common diseases a plant might get and the remedies that can be used to overcome them. Convolution neural network is used to classify the plant diseases and suggesting the remedies.

Keywords: Disease Detection, GUI Interface, Convolution Neural Network (CNN), Raspberry Pi (3B), Python, Remedies, Pesticides.

RESEARCH ARTICLE

1. INTRODUCTION

India is a developing country and Agriculture is the backbone of India. According to a survey conducted in 2017–18, agriculture contributes around 18% to India's Gross Domestic Product (GPD). It provides employment to more than 50% of the country's workforce. Indians are directly or indirectly dependent on agriculture which makes it important to maintain the crops for good production. Taking care of crops involves detecting the disorder or diseases related to plants and trying to provide necessary remedies. Monitoring of plants right from the initial stage should be done in order to prevent the attack of diseases. Currently the methods that are used to detect diseases are taken just through naked eye which is not efficient. Plant disease detection manually is more laborious task and at the same time provides less accurate results and can be done in only limited areas. If automatic detection techniques are used, it provides much accurate results in less time.

Some general diseases seen in plants are bacterial, fungal, late and early blight etc. Tomato plants are prone to diseases as they are very sensitive. With the help of Convolution Neural networks, we can apply this problem to classify the disease detected and train a model. CNN is very efficient for image classification. The model is able to recognise different diseases. Here the features of different leaves affected by different diseases are analysed. The main aim of this paper is to provide better accuracy and results of the model developed. All the steps involved in the development of the model has been described in this paper.

2. PROBLEM ANALYSIS

Now a days crop are less immune and are easily affected due to drastic changes in climatic conditions or attack of insects. This problem can be analyzed using CNN techniques by detecting the disease and providing effective results with good accuracy and speed. Detecting disease alone doesn't solve the problem. Remedies can be provided which can be used to eradicate the problem.

3. PROPOSED SYSTEM

To analyze the affected areas, the samples of leaves are taken in the form of dataset. The processing of image is

*Author to whom correspondence should be addressed.

done using various techniques. The stress is given mainly on the leaf of the plant. The preprocessing of the leaf is done in order to classify it as healthy or unhealthy. Once leaf segmentation is done, the preprocessed features will be helpful for providing accurate results. Many classifiers are present in Machine Learning out of which this paper focuses on Convolution Neural Network for automatic leaf disease detection. The inputs are taken with the help of a camera to process the image of the leaf. Now the images are pre-processed and the result is displayed with the help of a user interface.

3.1. Data Set

Data set is nothing but the collection of data. It consists of data pertaining to single element. Data set consists of several images with different attributes with reference to the element which is leaf in our case. Several data related to one element comprises of different parameters which are useful for processing, training and testing. The dataset in this case has been downloaded from Plant village Classification (Fig. 1). This downloaded data set is raw which has to be normalized to same dimensions. Each type of disease has been given a particular name for us to know.

3.2. Preprocessing

1. Shape feature extraction: All the images in the dataset are not of same size due to which the dimensions are converted to universal measurement for example $250*250$. The shape of the leaf is taken into account for processing.
2. Texture feature extraction: The input images are taken and converted into Gray level matrix to compute the pixel value. Contrast, Correlation and homogeneity are taken into account.
3. Color feature extraction: Color is a distinct feature of image representation that is invariant with scaling, translation and rotation. Mean and skewness are used to represent color as a feature.

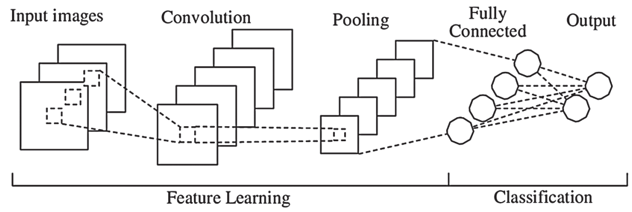


Fig. 2. A basic convolution neural network model.

3.3. Flow of Work

i. Collect the dataset: Dataset is the pre-requisite at all stages of disease recognition. The dataset related to leaf disease detection is downloaded from internet. Here all the duplicate images are removed. Finally the dataset comprises of around 4000 images for training and around 2000 images for testing.
ii. Pre-processing the dataset: Images taken from internet come in different dimensions and resolutions. Hence preprocessing involves cropping of the image and focusing on the leaf by making square boundaries. So that the portion of the leaf which we are focusing on will be highlighted. The images taken in the dataset are scaled to $256*256$ for equal processing.
iii. Splitting the dataset into test and train: One of the common approaches to measure the performance of neural

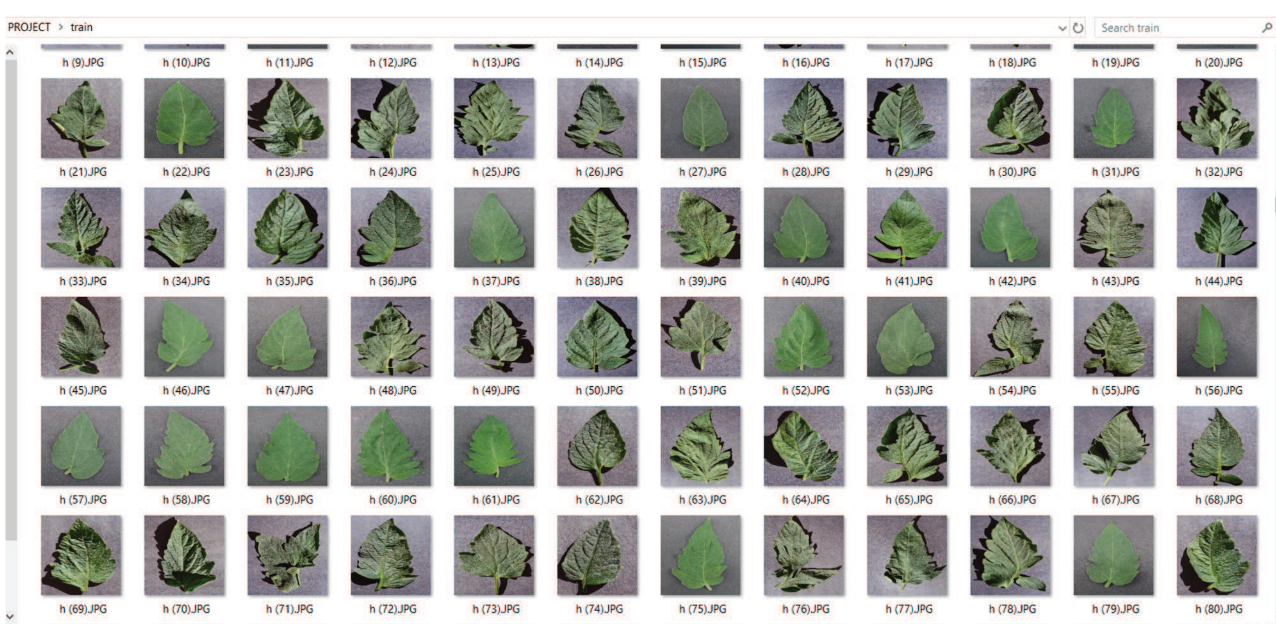


Fig. 1. Dataset of leaves to train the model.

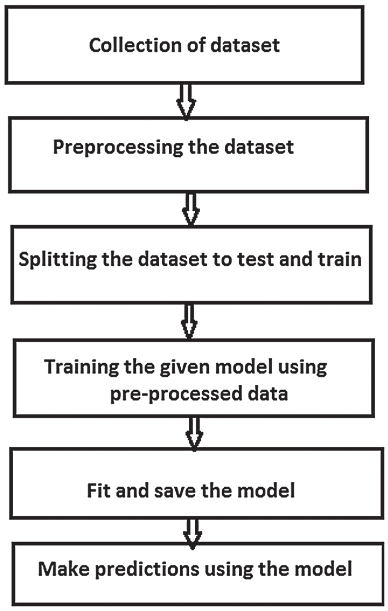


Fig. 3. Block diagram of work flow.

network is to split the dataset into test and train for prediction. The leaves under train are used as inputs to the model for accuracy. The leaves under test dataset are used to compare the outcomes of the model.
iv. Training the given model: Here the major technique used is convolution neural network where there are hidden layers and there is a dense layer to connect all the layers. The block diagram of a basic convolution neural network as in Figure 2.

Convolution Neural Network is a class of deep neural network which plays an important role in image visualization. CNNs have multilayer network in order to minimize the processing. CNN has an input layer, output layer and multiple hidden layers. The hidden layers consist of pooling layers, activation function layers, full connected layers etc. Convolution operation is performed to input layer and then the result is passed on to the next layer. In pooling

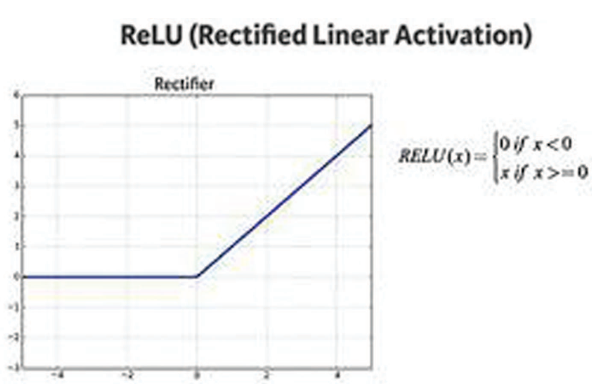


Fig. 4. ReLu (rectified linear activation).

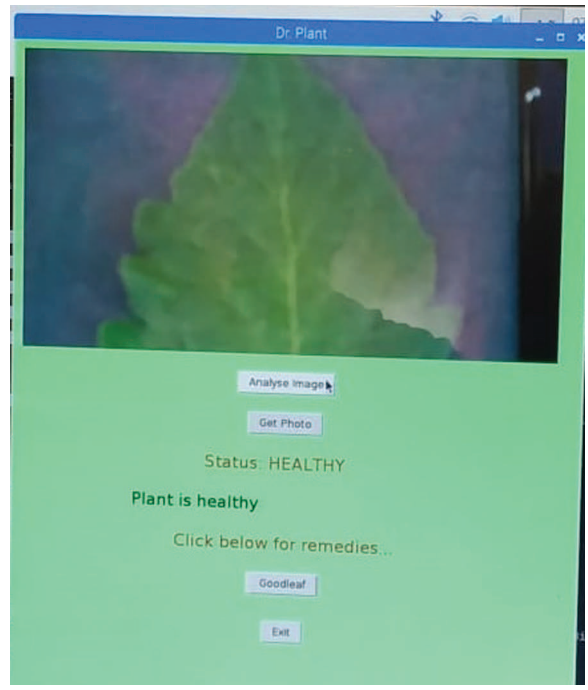


Fig. 5. Status of a leaf suffering bacterial spot.

layer, the output of the convolution layer is taken and the maximum value is taken as pixel value to the next layer. All the values do not come in a certain range. In order to get a normalized value, we use activation function. This paper describes the use of Relu activation function (Fig. 4) which maps the value to zero if the input is negative and maps to the highest value if the input to this layer is positive.

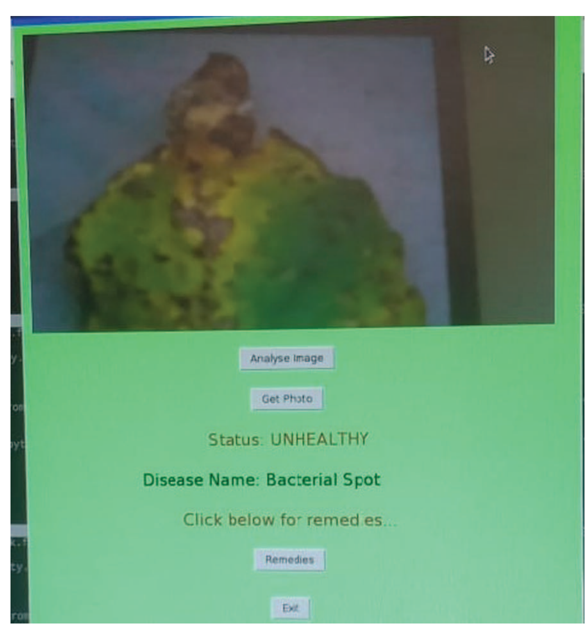


Fig. 6. Status of a healthy leaf.

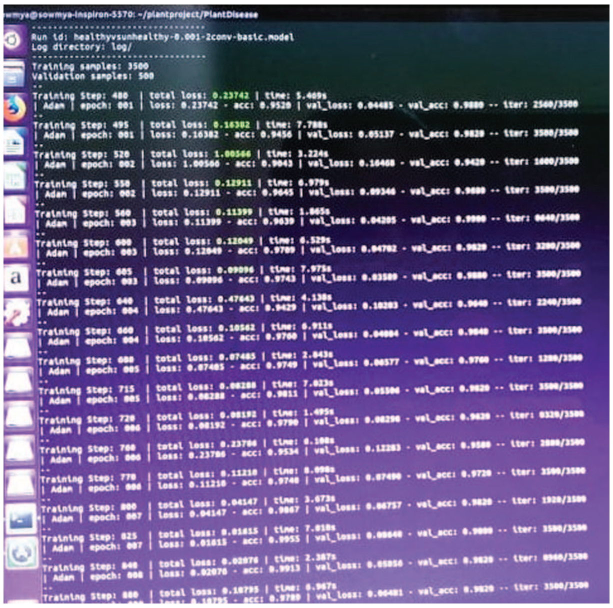


Fig. 7. Accuracy and loss rate during training the model.

A cost matrix function is used to find out the difference in the deviation of the output. Fully connected layers are used to connect all the hidden layers to the input and output layers. In the final layer, since convolution is 2-D, the input to the final layer is converted to 1-D which is an array and is compared to the output to form the error function.

4. RESULTS AND DISCUSSION

This paper depicts how the names of the detected diseases are displayed to the end user. After all the processing of the model, the output is displayed with the help of a monitor and with the use of Graphic User Interface (GUI). The user has two options to see the result.

i. The user has the option to directly select the image from the folder where he/she has stored the clicked picture.

ii. The user also has an option of showing the photo from an external device to a pi camera.

The model is trained to get maximum accuracy and with minimum loss. Testing is done on valid set of images. The model is developed so as to give an accuracy of around 95%. This happens in a fraction of a minute while using Raspberry Pi. In the first go while uploading the code to raspberry pi, it takes two minutes to train the model. Once uploaded, for the next input, it takes not less than a minute to process the information.

Table I. Diseases detected and remedy for the diseases.

sl no.	Disease detected	Combined mixture ratio
1	Bacterial spot	17:17:17
2	Yellow leaf curl	19:19:19
3	Late blight	14:35:14

There will be a checkbox for selecting the option for remedies which could be used to improve the quality of the plant condition. The disease prediction is done by using appropriate classifier and the disease is predicted. Pesticides are also suggested to improve the yield of the plants to maintain its state.

The Table I describes the disease detected and the pesticides to be used that are mixed in the following ratio. The elements used in the mixture are:

- Urea
- Potassium
- phosphorous.

The combination of the above mixture is known as Trichoderma viride. This mixture is used to medicate the plants. Using Dilthane-Z-78 or Dilthane-M-45 spray immunity of the plant can be improved.

5. CONCLUSION AND FUTURE SCOPE

According to various surveys done, there is no proper solution for suggesting the exact remedies for a particular disease in plants. This paper suggests the suitable pesticides that can be used to solve the problem and maintain the condition of the plant. The main advantage of this paper is that it suggests the solution to the problem at a very low cost and not much hardware is required. The outcome can also be shown in an additional screen but the cost may increase. The accuracy of the model is tried to be maintained at a high rate for good possible outputs. Python is used to write the programming part of the work.

The present paper shows the effective features to describe the affected plant. This can also be improved by adding on some of the features. Instead of using pi camera, a better resolution camera with a 360 degree view can be used to provide more accurate results. More datasets can be included to provide details about more diseases. The dataset pertaining to this paper deals with tomato leaf alone as it is very sensitive. Datasets containing to other leaves can also be collected and included.

Acknowledgments: The authors of this paper would like to thank the management of BMS Institute of Technology for providing financial support to execute and publish this paper.

References

1. Dang, K., Sun, H., Chanet, J.P., Garcia-Vidal, J., Barcelo-Ordinas, J.M., Shi, H.L. and Hou, K.M., **2013**. Wireless Multimedia Sensor Network for Plant Disease Detections. *France-China International Workshop*, September 18–20; France, Clermont-Ferrand.

2. Shweta R. Astonkar and Shandilya, V.K., **2018**. Detection and Analysis of Plant Diseases Using Image Processing Technique. *International Research Journal of Engineering and Technology*.

3. Huddar, R., Gowri, S. and Rupanagudi, S.R., **2012**. Novel Algorithm for Segmentation and Automatic Identification of Pests on

Plants Using Image Processing. *Third International Conference Computing Communication & Networking Technologies (ICCCNT)*, July 26–28.

4. Mukherjee, M., Pal, T. and Samanta, D., **2012**. Damaged paddy leaf detection using image processing. *3*(10).
5. Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M. and ALRahamneh, Z., **2011**. Fast and accurate detection and classification of plant diseases. *International Journal of Computer Applications (0975–8887), 17*(1).
6. Sanjay B. Patil, **2011**. Leaf disease severity measurement using image processing. *International Journal of Engineering and Technology, 3*(5), pp.297–301.
7. Martin, V. and Moisan, S., **2009**. Early pest detection in greenhouses. *Computer and Electronics in Agriculture Journal*.
8. Mamatha, K.R., et al., **2017**. Waste segregation system using artificial neural networks. *Helix Journal, 7*(5), pp.2053–2058.

Received: 6 September 2019. Accepted: 2 October 2019.

RESEARCH ARTICLE

Performance Comparison of Conventional Neural Networks and Deep Learning Network for Cervical Cancer Diagnosis

[1] Chandra Prabha R, [2]Seema Singh
[1] Department of Electronics & Communication Engg., [2] Department of Electronics & Telecommunication Engg.
[1]BMS Institute of Technology and Management, [2] BMS Institute of Technology and Management
[1]chandra@bmsit.in, [2] Seemasingh@bmsit.in

Abstract— ***Cervical cancer is the fourth-most common cause for death from cancer in women. Efforts are being made to develop more efficient techniques for the detection of cancer at the initial stage. Conventional methods require expert pathologists to examine the biopsy slide and classify it. In this regard few concerns have risen such as the deficiency of expert pathologists, lack of technical support to doctors and also lack of awareness among women especially in rural areas. Hence there is a requirement for an effective and accurate system that detects cervical cancer which can be used by health worker to detect cancer at initial stage (as a part of basic health check-up). This paper describes and compares two techniques for the cervical cancer diagnosis. The first method involves extraction of key features from complex cytology images using image processing algorithm followed by a neural network classifier with back propagation algorithm using MATLAB tool.The major challenge faced in this method is extracting the key features from complex images with overlapping cells, which is further used by neural network for classification. The other method is based on deep learning that uses inception neural network with tensor flow. A comparative analysis is presented for the same image database which is created with a Bangalore based pathology laboratory. The database is of 460 images of which 197 images are cancerous and 263 are non-cancerous images. The analysis proved that deep learning method was able to provide better classification results.***

Index Terms— Image Processing, Back Propagation neural networks, Convolution Neural Networks, Inception V3, Tensor flow

I. INTRODUCTION

The fourth most common cause of cancer death in women is cervical cancer. Developing countries like India are at a higher risk and a major percentage of cervical cancer patients are from developing and low income group countries.

Cervical cancer is caused due to abnormal cells present at the cervix which multiplies at a faster rate and grows out of control. The cervical cells transform into a precancerous state which is stated as Cervical Intraepithelial Neoplasia (CIN). Depending on the intensity of the cellular degradation it is classified as low-grade CIN and high-grade CIN. Human Papilloma Virus (HPV) is the cause for cervical cancer. Early detection of HPV can be done using either Pap test or HPV test

Liquid Based Cytology method (LBC) [3] is one of the screening methods for cervical cancer. The samples are prepared and diagnosed in a special laboratory. The system is more accurate. As cure rate of cancer is closely related to the stage of the disease at diagnosis time, with a very high probability of fatality if it is left untreated. Therefore, timely identification of the positive cases is very crucial.

Current prevalent method used by pathologists for the diagnosis of cervical cancer is to manually observe the morphological changes in the cells. Due to limited number of skilled and experienced pathologist, the mass screening procedure becomes time consuming and costly. Hence an efficient algorithm is necessary to provide a technical support to the specialists. The Para-medical worker, local health visitor or lab technicians may be trained to take the smear and prepare the slide. As the number of para-medical workers is more, the screening facility may reach masses. Further the deployment of the algorithm can reduce the complexity in handling mass screening. With mass screening program supported with the proposed system, skilled pathologist can focus on critical cases only.

However, the drawback of all the methodology is that despite obtaining the samples for the purpose of testing, there is always an element of human error and hence this will lead to erroneous results. This paper proposes two techniques involved for feature extraction and classification of overlapping cancerous cells. The first method involves extraction of various key features from cervical cytology complex images by applying image processing and then followed by a neural network classifier with MATLAB tool. The other technique uses deep learning network inceptionv3 that produces a high level of accuracy in classification with less processing time (cloud computing).

L Mahanta and K Bora[1] presents a process using pap smear images for the analysis of cervical cancer cells based on Nucleus Cytoplasm ratio which is one of the most important feature of identifying the cancer affected cells. Based on the value of Nucleus Cytoplasm ratio, a normal and an abnormal cell can be identified. This paper concludes Nucleus cytoplasm ratio of a normal cell is less compared to an abnormal cell.

An innovative method for automated system for the diagnosis of cervical cancer by extracting various features from cervical cytology images using MATLAB image processing tool [2] is proposed.in this paper the features like nucleus-to-cytoplasm ratio, shape and color intensity were used to train the neural network using Back-propagation algorithm. The cytology cells were then successfully classified as non-cancerous, low- grade and high- grade cancer cells. Nucleus Cytoplasm ratio is very high for cancerous cells and less for non-cancerous cells. Depending on the shape, the cancer cells are of two types, low-grade which are mostly circular and High-grade which are tapered and possess spindle like structure. Cancerous and non-cancerous cells have distinctively different color intensity distribution.

Z. Lu, G. Carneiro, and A.P. Bradley [3] applied an algorithm for accurate segmentation of the individual cytoplasm and nuclei from a clump of overlapping cervical cells. Existing methods do not provide such a complete segmentation due to severe overlap and poor contrast. In this paper, a scene segmentation is performed to highlight the free-lying cells, cell clump and their nuclei. This segmentation is performed using a joint level set optimization on all detected nuclei and cytoplasm pairs. This technique is constrained by the length and area of each cell, cell shape, the amount of cell shape overlap and the expected grey values within overlapping regions.

Christian Szegedy et. al. [4] applied deep convolutional neural network architecture code called Inception, classification and detection in the Image Net Large-Scale Visual Recognition Challenge. The architectural decisions were based on the Hebbian principle and multi-scale processing for optimization of quality.

Karen Simonyan, Andrew Zisserman [5] investigated the effect of the convolutional network depth for accuracy in a setting that involves large scale image recognition. Their main contribution is a complete evaluation of networks which rises in depth using very small 3 x 3convolution filters.

An enormous amount of work is carried out in the area of cervical cancer diagnosis using neural network with various combination of input features. However, the slides directly from the pathologist were not used. This work proposes two different approach for the classification of same database which is created with a Bangalore based pathology laboratory. The first approach of classification involves extraction of different key features from cervical images using image processing followed by a neural network classifier. The other technique uses deep learning using inception that produces a high level of accuracy in classification and processing time. Based on the comparative analysis of two approaches, usage of deep learning network is appreciated for this application.

II. METHODOLOGY

The two techniques i.e. neural network classifier and deep learning works with the same data set to classify cervical cancer images into cancerous and non-cancerous cells.

The following sections discusses the description of data set used and different method of classifier

DATA SET

Data Set has been obtained from a pathologist:

Total Images collected	460 images
Cancerous images	197 images
Non-cancerous images	263 images
Image size	2040x1528
Horizontal and vertical resolution	96dpi

A. Technique 1: Image processing and neural networks algorithms

Image processing and neural network algorithm are applied to identify cancerous and non-cancerous cells. The following steps are involved. The Fig. 1 describes the flow of steps involved in the process.

Pre-processing

The purpose of pre-processing is to improve the image data by suppressing the unwanted distortions or to enhance some image features. Using the MATLAB tool, pre-processing is done on the images attained.

RGB-to- Grey Conversion

The input images are of high-resolution and in JPEG format. The collected colour image is changed to a grey scale image.

Scene segmentation

There are two stages in Scene segmentation processes: 1) The segmentation of cell clumps. 2) The detection of nuclei and respective cell segmentation.

1) Segmentation of cell clumps

There are four stages in segmentation of cell clumps

Stage 1: Super Pixel Map

The amount of grey value similarities and spatial proximity is determined by finding local maxima using quick shift algorithm [14]. The algorithm segments an RGB image by identifying clusters of pixels in the joint spatial and colour dimensions. Segments are local (super pixels) and it can be applied as a base for further processing. Given an image, the algorithm calculates a forest of pixels whose branches are considered with a distance. This specifies a hierarchical segmentation of the image, with segments corresponding to sub trees. The informative super pixels can be recognised by cutting the branches whose distance label is above a specified threshold (the threshold can be either fixed by hand, or determined by cross validation).

Stage 2: Edge Map

The second stage is applying a canny edge detection on this super-pixel map to detect super-pixel edges and removes most of the background data. Canny edge detection provides good detection, clear response and good localization.

Stage 3: Convex Hull and Clump Boundary

To find candidate cell clumps, the third stage applies an unsupervised binary classifier, where the classes are “background” and “cell clump”. The first assignment is provided by building a convex hull around the connected components of the edge map computed in stage 2.

Stage 4: Clump Boundary Using Gaussian Mixture Models

Gaussian Mixture Models (GMM) starts with an initialization step which will be the convex hull input from stage 3, which assigns the parameters to values based on the

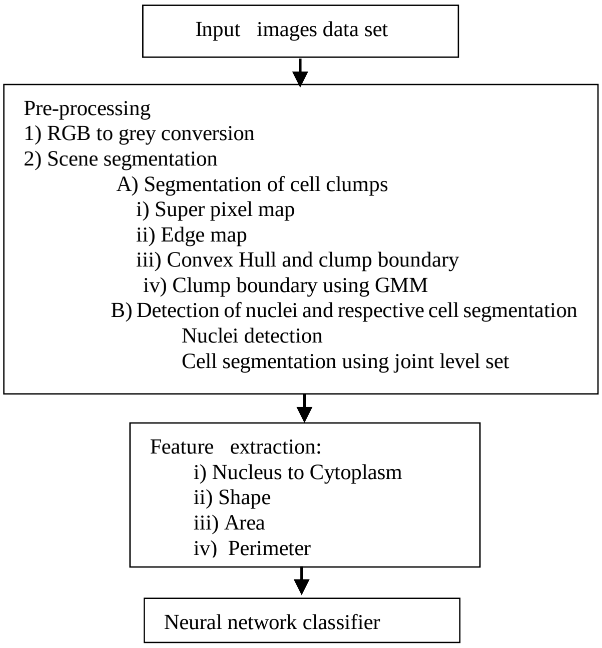


Fig.1. Different stages involved in image processing and neural networks

data. Then, the model iterates over the Expectation (E) and Maximization (M) steps until the parameters estimates converge; where Expectation is given by stage 2.

The result of the convex hull map stage is taken and re-estimated using the GMM. This is iterated until the GMM is stable. In practice, iterating this re-estimation process 10 times produces stable results. At the end of this step the cytoplasm boundaries will be obtained. As the number of iterations for noise removal increases, processing time also increases.

2. Nuclei detection and cell segmentation

Nuclei detection

Nuclei can be described by comparatively low grey values, similar texture, and circular borders. The nuclei detection is primarily an experimental procedure using a thresholding function. After this step all the nuclei of a particular image will be detected. This is then labelled as Nuclei mask and used for later processing.

Cell Segmentation Using Joint Level Set

The segmentation of overlapping cells uses the set of nuclei described in previous section as the initial guess for each level set function. Level Set can be used to efficiently address the problem of curved surfaces/edges propagating in an implicit manner. Thus, by using the nucleus as the initial guess(actual contour), and the obtained cytoplasm mask after GMM as final guess, about 40 or more iterations are performed to get the required accurate cytoplasm mask. For each iteration cytoplasm is considered to be an ellipse.

B. Feature Extraction

Different features need to be extracted from each segmented cell. The extracted features are put in a proper vector form which is used to train the neural network for further classification. Presented in Fig.5.

The pre-processing steps, Nuclei Mask and CytoMask, have to undergo before feature extraction using the steps. Presented in Fig. 2.

Step 1: CytoMask is complimented such that detected Cytoplasm is given by bit 1(white) and background is given by bit 0(black).NucleiMask are such that detected nucleus is represented by bit 0(black) and background is represented by bit 1(white).

Step 2: CytoMask and NucleiMask are given as inputs to the AND function to get a Fused image. Thus, each individual cell is obtained separately.

Step 3: The Fused image and NucleiMask are given as inputs to XOR function to get nucleus of each cell separately. Features of Nucleus can be extracted.

Step 4: From step 2, on applying fill holes function, cytoplasm of each cell is extracted separately. Features of cytoplasm can now be extracted.

The pre-processing is followed by the feature extraction which are used to distinguish between the cell images. Nucleus-to-Cytoplasm ratio: Separation of the nuclei from the cell involves the following steps: a) converting the RGB image to grey- scale, applying a 'Gaussian' filter and setting a threshold to separate nuclei. The threshold value is set by observing the intensity values of the nuclei using 'impixel' function in MATLAB. Area of nuclei and cell is calculated using 'regionprops'. Ratio of these areas are calculated. A threshold value for the ratio is set to distinguish cancerous and non-cancerous cells.

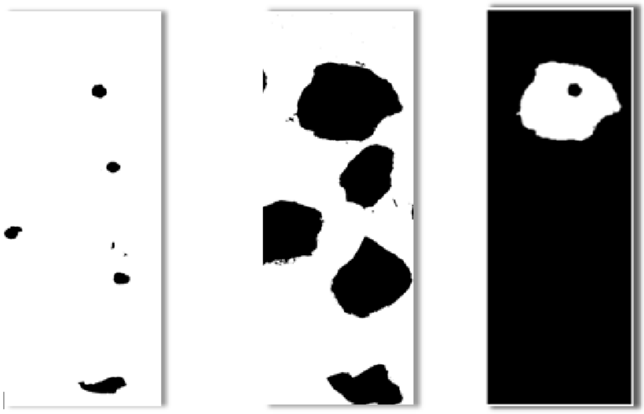


Fig. 2. (a) Nuclei Mask (b) Cyto Mask (c) Fused Image

The key features extracted are shape, area, perimeter, eccentricity, solidity and extent

Shape: The ratios of major-axis length and minor-axis length of the cell is computed. Presented in Fig.3.

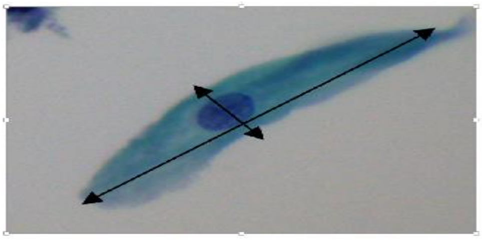


Fig. 3. Major and Minor Axis indicated for a cell

Area: It is the degree of entire number of pixels present in each cell of the image. Presented in Fig.3.

Perimeter: It is distance around the boundary of the region. Its obtained computing the distance between each adjoining pair of pixels around the border of the region.

Eccentricity: It is a scalar whose value is between 0 and 1. It is the ratio of the distance between the foci of the ellipse and its major axis length.

Solidity: It is a scalar specifying the proportion of the pixels in the convex hull that are also in the region. Convex hull is a p-by-2 matrix that specifies the smallest convex polygon that can contain the region. Each row of the matrix contains the x- and y-coordinates of one vertex of the polygon.

Extent: The ratio of pixels in the region to pixels in the total bounding box is specified as extent which is a scalar.

C Neural Network Classifier

Back propagation algorithm is used as a classifier. The method gradient of a loss function is computed with respects to all the weights in the architecture. The computed gradient is fed to the optimization method to update the weights continuously in order to minimize the loss function. Propagation and Weight update are two phases of back propagation. Presented in Fig.8. The other types of cancer can also diagnosed using neural networks [9].

D Technique 2: Deep learning networks

Deep learning is a model which uses neural network architecture to classify the images directly. Conventional neural networks contains only 2 or 3 layers, while deep networks can have more layers.

Convolutional Neural Network (CNN) is a powerful tool that can classify visual inputs into different classes [4]. The CNN is widely successful in the task of classification is because the network can be modelled on animal visual perception to perform classification [5].

CNN contains an input layer, an output layer, and many hidden layers in between. The Feature Detection Layers implements convolution, pooling, or rectified linear unit (ReLU) operation which are executed repeatedly over tens or hundreds of layers, with each layer detecting distinct features. Convolution lays the input images through a set of convolutional filters, each of which initiates certain features

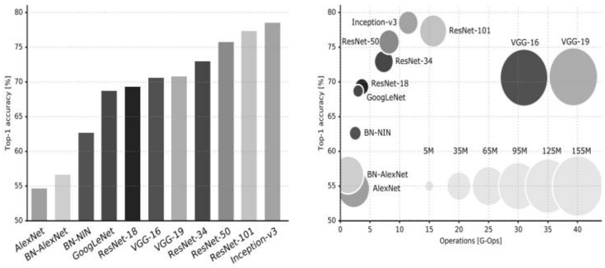


Fig. 4. Comparison of Inception V3 with other CNN models

from the images. Pooling performers nonlinear down sampling to reduce the number of parameters that the network needs to learn about. Rectified linear unit (ReLU) permits for faster and more effective training by mapping negative values to zero and maintaining positive values. After feature detection, the next task is to classify.

The a fully connected layer (FC) which outputs a vector of N dimensions where N is the number of classes that the network will be able to detect. This vector contains the possibilities for each class of any image being classified.

The last layer of the CNN architecture uses a softmax function to provide the classification output.

CNN has less processing time (cloud computing) compared to other image classification methods. Hence the filters are learned instead of being hard-coded in traditional algorithms. In this regard, the Inception V3 which is one of the best image classifier among CNN. The Fig. 4. presents an understanding into the performance of the Inception V3 with respect to other CNN models.

Architecture of inception V3

Inception V3 is one finest image classifiers in CNN [4]. Deeper and wider networks improves the quality of the network.

Computational cost in Inception is lower in comparison to other CNN's [12]. Inception networks are applied in processing large amount of data at practical cost or situations where memory or computational capacity is limited.

It is certainly possible to mitigate parts of these issues by applying specialized solutions to target memory use, or by improving the execution of certain steps through computational tricks. These modification comes with added difficulties. Furthermore, these methods may be applied to enhance the Inception architecture and increase the efficiency gap.

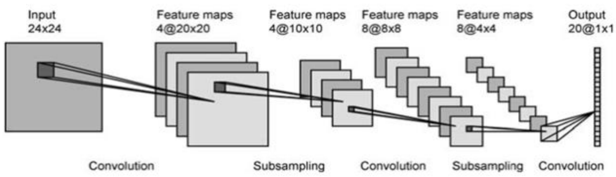


Fig. 5. Convolutional neural network architecture

E Labelling the Images

One of the visual clues to detect a cancerous slide is the nucleus to cytoplasm ratio, which is very high for cancerous cells. Another thing to note in the images is the dense clustering of the cells which is common for cancerous cells. The neural network is been trained to classify two classes cancerous and non-cancerous. The images are trained using Supervised learning algorithm, wherein the images are labelled and the classifier learns the special features from the images fed to it [7].

F Training the Architecture

The network is trained for 1500 steps which results the cross entropy of 0.13201, which is the loss function in machine learning and optimization [8]. Increasing steps results in smaller cross entropy values, leading to more accurate results.

III. RESULTS

A Conventional Neural Network

In this method, processsing time and computational required for the classification of images is high,intermediate image processing results can be verified. Here each cell can be classified separately as cancerous/non-cancerous. Processing of complex images consumes huge amount of time. The Fig.6. describes the outputs of each stages of method 1

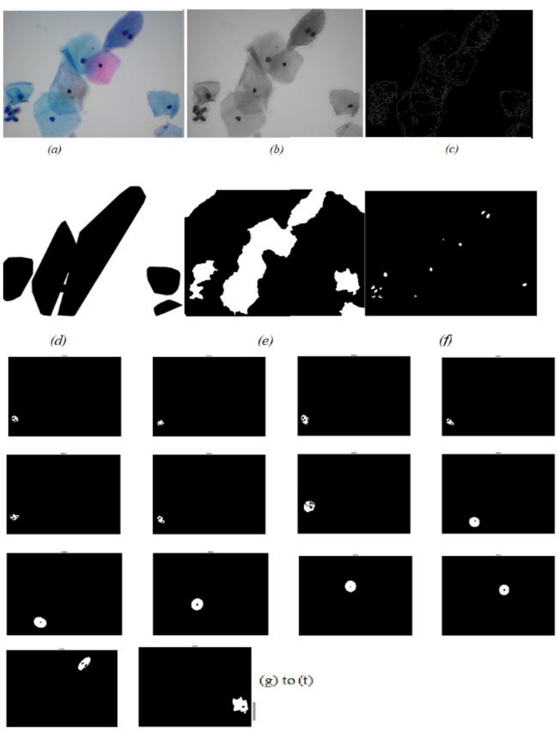


Fig. 6. (a) Input data (b) Super Pixel (c) Edge Detection (d)Convex Hull (e)GMM output (f) Detected Nuclei, (g to t) Segmented cells

	A	B	C	D	E	F	G	H
1	Cyto Area	Nuc Area	Cyto/Nuc Ratio	Major/Minor	Perimeter	Solidity	Eccentricity	Extent
2	9428	1686	5.591933571	1.621739222	386.508	0.950403	0.787259406	0.6951264
3	16024	2193	7.306885545	1.413473696	537.784	0.903524	0.706736461	0.7512424
4	18399	2525	7.286732673	1.555926186	548.456	0.919031	0.766114583	0.6638404
5	9756	1717	5.682003494	1.97766603	421.954	0.916401	0.862740731	0.6088747
6	10726	2035	5.270761671	1.178665975	395.193	0.942365	0.529328777	0.6726452
7	9503	2104	4.516634981	2.048955931	428.851	0.889211	0.872813746	0.5846561
8	23016	1262	18.23771791	1	609.788	0.963577	0.378360393	0.7524273
9	20782	292	71.17123288	1.055517582	516.044	0.985957	0.32004452	0.7821898
10	24848	292	85.09589041	1.244695517	591.959	0.976768	0.595427105	0.7524225
11	26314	757	34.76089828	1	645.318	0.977779	0.403542312	0.7656541
12	24792	118	210.1016949	1.095222674	585.209	0.980734	0.40783346	0.7416904
13	21211	732	28.97677596	1.132529427	521.041	0.987293	0.469411993	0.7769597
14	20149	1817	11.08915795	2.694068096	610.668	0.960345	0.928558589	0.51411
15	43928	1020	43.06666667	1	1037.284	0.810525	0.592948587	0.6364071

Fig.7. Features Extraction

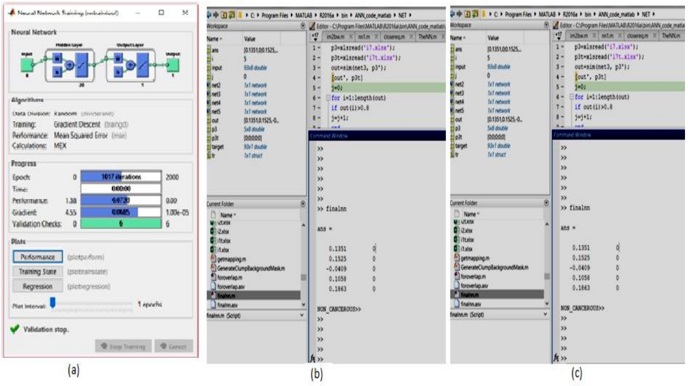


Fig. 8. Neural Network Classifier: (a) Neural Net Training, (b) Cancerous Image,(c) Non-Cancerous Image

B Deep Learning Network

The result of the classifier is very accurate with less processing time and computational cost. By increasing the number of steps smaller cross-entropy can be achieved, which results in better outputs i.e. at step 1499, the cross entropy is 0.013201.

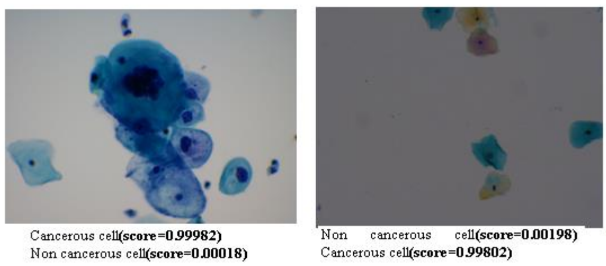


Fig. 9. (a & b) Result scores of Cancerous Image and non-cancerous Image

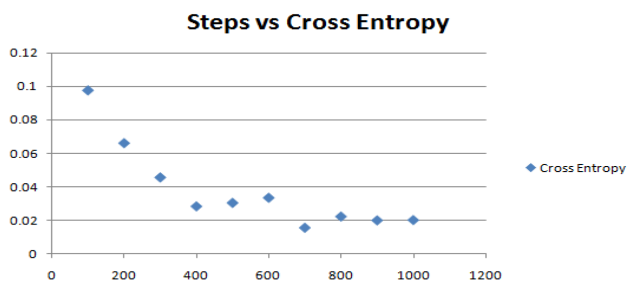


Fig. 10. Performance Curve Of Training Process

C Comparative analysis of simulation results

TABLE I. Comparative analysis of two approaches

Factors	Conventional Neural Networks	Inception neural networks
Platform	Windows	All Platforms as docker platform has been used
Processing Time	Timing differs based on number of clumps and nuclei in image. If there 10 nuclei (overlapped cells)It takes 1 hour.	500 training images, 2 test images timing: *Creating bottle neck-5min 18s *500 training steps -54s *total training time-6min 12s *Display of results -10s
Images processed at a time	One image to be loaded at a time, if more images are loaded computation time increases.(8GB Processor, 2.8GHz)	200 images
Image Processing	Intermediate steps can be recorded to the pathologists for verification	No image processing algorithms used
Status of each cell in the image	Each cell can be classified separately as cancerous /non-cancerous	Not possible: the complete image is classified as cancerous/ non-cancerous

The two methods have been explored for the classification of images as cancerous or non-cancerous. A comparative study of the same for the parameters like processing time, computational cost, platform, cloud support, etc. is presented in the table I.

From the table I, it is observed that Inception Neural Network is the better than conventional neural networks. Due to low processing time and high accuracy, inception neural network is able to process highly complex cervical images with large number of clumps and nuclei

CONCLUSION

The cervical cancer diagnosis is done with the manual observation of morphological changes in the cells, which is very subjective and rises several concerns. Implementation of neural network algorithm and deep learning network in cervical cancer diagnosis for classification of complex images with overlapping cells image into cancerous and non-cancerous cells can reduce critical problems faced by pathologists. This paper presents a comparative analysis of biomedical image classifier for cervical cancer using two different methods for the same dataset of images created with a Bangalore based pathology laboratory. The database has 460 images out of which 197 images are cancerous and 263 are non-cancerous images.

Conventional Neural network method uses image processing algorithm to extract features from complex images with overlapping cell. The key features extracted are area, perimeter, eccentricity, nucleus-to cytoplasm ratio, major axis-to-minor axis ratio of the cells. Further, extracted features are fed to neural network architecture with the structure of 8-20-1. After training, the best performance goal (error) of 0.0738 was achieved. Further the trained network was tested on the unknown cell images to achieve significant accuracy. However, it faces a major issue in extracting the feature from overlapping cells and influences the computational time due to hardware constraints. Deep learning is computationally intensive and complex. However it is capable of giving better accuracy for highly complex images with overlapping cells also. With the support of Cloud computing, deep learning network classifies images as cancerous and non-cancerous with less processing time and computational cost. By increasing the number of steps smaller cross-entropy is been achieved, which results in better output. The deep learning network provides a good accuracy and is robust in comparison with the other image processing techniques used. In order to assist pathologist and to support mass screening with reduced human error, the deep learning network can be used on site. This method can also be used to classify other kinds of cancerous cells after training with those respective cell images.

REFERENCES

[1] L Mahanta, and K Bora. "Analysis of malignant cervical cells based on N/C ratio using pap smear images." International Journal of Advanced Research in Computer Science and Software Engineering 2, no. 11 (2012): 341-346.

[2] S Seema, V. Tejaswini, R P Murthy, and A Mutgi. "Neural network based automated system for diagnosis of cervical cancer." In Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications, pp. 1422-1436. IGI Global, 2020.

[3] Z Lu, G Carneiro, and AP Bradley. "Automated nucleus and cytoplasm segmentation of overlapping cervical cells." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 452-460. Springer, Berlin, Heidelberg, 2013.

[4] C Szegedy, W Liu, Y Jia, P Sermanet, SReed, D Anguelov, DErhan, V Vanhoucke, and A Rabinovich. "Going deeper with convolutions." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9. 2015.

[5] K Simonyan, and A Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[6] K He, X Zhang, S Ren, and J Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition." IEEE transactions on pattern analysis and machine intelligence 37, no. 9 (2015): 1904-1916.

[7] Y LeCun, B Boser, J S Denker, D Henderson, R E. Howard, W Hubbard, and L D. Jackel. "Backpropagation applied to handwritten zip code recognition." Neural computation 1, no. 4 (1989): 541-551.

[8] I Sutskever, J Martens, G Dahl, and G Hinton. "On the importance of initialization and momentum in deep learning." In International conference on machine learning, pp. 1139-1147. 2013

[9] S Singh, J Harini, and B R Surabhi. "A novel neural network based automated system for diagnosis of breast cancer from real time biopsy slides." In International Conference on Circuits, Communication, Control and Computing IEEE, pp. 50-53.2014

[10] Sam Abrahams, DanijarHafner, Erik Erwitt, Ariel Scarpinelli.,"TensorFlow for Machine Intelligence", Bleeding Edge Press, Santa Rosa, CA 95404.

[11] C Szegedy, A Toshev, D Erhan ."Deep neural networks for object detection." In Advances in neural information processing systems, pp. 2553-2561. 2013

[12] D C Ciresan, U Meier, J Masci, L M Gambardella, and J. Schmidhuber "High-performance neural networks for visual object Classification." arXivpreprint arXiv:1102.0183 (2011).

[13] A Karpathy, G Toderici, S Shetty, T Leung, R Sukthankar, L Fei-Fei "Large-scale video classification with convolutional neural networks." In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1725-1732. 2014

[14] Y LeCun, F J Huang, and L Bottou,"Learning methods for generic object recognition with invariance to pose and lighting." In CVPR (2) , pp. 97-104. 2004.

Proceedings of the Second International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020)
IEEE Xplore Part Number: CFP20K58-ART; ISBN: 978-1-7281-4167-1

A Survey on Automated Detection of Breast Cancer based Histopathology Images

Laxmisagar H.S [a]
[a] *BMS Institute of Technology and Mgmt., Yelahanka, Bangalore-560064, India*

Dr. Hanumantharaju M.C [b]
[b] *BMS Institute of Technology and Mgmt., Yelahanka, Bangalore-560064, India*

***Abstract*— The increasing mortality rate in the women population is mainly due to breast cancer. Diagnosing breast cancer in its early stages will always remain crucial. Hence identifying and treating the disease at the earliest will increase the possibilities of survival. Recently, by using ultrasound images a computer-aided diagnosis (CAD) system is being developed to help radiologists to attain higher accuracy for identification. Normally, a CAD system comprises of different phases such as pre-processing, segmentation for regions of interest, feature selection & extraction, and last phase is to do classification. This paper illustrates the various methods used to deploy an automated CAD system development for the early identification of cancer disease. In this paper, various approaches used are abridged and their pros and cons are compared. The performance evaluation of the CAD system is also depicted as well. The dataset of breast cancer histology images (BACH) is made available to participate in a grand challenge aimed at the classification of microscopy and whole slide images, whereas it is made publicly available for the challenge to promote further improvements for developing an intelligent classification system in digital pathology. According to the number of diagnostic classes and image types (Microscopy and whole slide images), an intelligent system is implemented for initial detection for deploying a proper treatment for breast cancer.**

Keywords: *Breast Cancer, Histology images, CAD, BACH, Microscopy, Whole slide images;*

I. Introduction

Breast cancer is considered as the most commonly observed cancer among women, where it reaches approximately 2.1 million women per annum, and also leads to the enormous death rate. Approximately 627,000 women [1] with breast cancer died in the year 2018. Breast cancer disease has increased amongst women in more progressive areas and it is increasing in every place globally. The estimated number of breast cancer rates by 2030 in the U.S. is expected to be 50 percent higher related to what it was in 2011, with calculated 441,000 breast cancer circumstances in 15 years, as per a novel study carried out by the National Cancer Institute (NCI).

To decrease breast cancer rates and survival, early detection remains as a critical part. Early detection approaches focus on providing a suitable method for cancer treatment by reducing difficulties to upkeep and/or refining access to actual detection services. The important technique is to escalate the percentage of breast cancers recognized at an early phase, letting additional effective treatment [2] be used and stopping out from the dangers of fatality from breast cancer. Screening extensively includes testing women to discover cancers before any noteworthy indications that make an appearance. Various techniques have been made use of as tools of breast cancer screening encompassing breast self-exam, clinical breast examination, and mammography.

It is noticed that the mammography technique uses X-rays to identify irregularities within the breast. It is essential to decrease breast cancer deaths by approximately 20% in high-resource samples. clinical breast exam (CBE) is an inspection of both breasts of women made by expert health professionals. Clinical Breast Exam appears to be a very good method for the initial diagnosis of breast diseases. Invasive breast cancer occurs when irregular cells from inside the milk ducts or lobules feast [3] out into neighboring breast tissue. Cancer cells due to abnormality can spread from the breast to other parts of the body through the lymph nodes or the blood canal. They may spread both initial in the process when a tumor is insignificant or far ahead once it is large.

In 2019, the death rate was assessed amongst men and women in the U.S. there were [4]:

Table 1. Depicts Breast Cancer History

Gender	Estimated Invasive Breast cancer	Cancer deaths rate
Women with breast cancer disease	2,68,600	41,760
Men with breast cancer disease	2,670	500

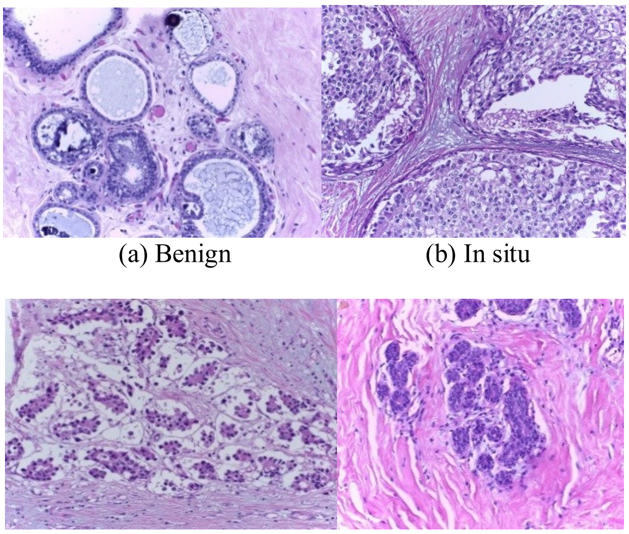
A. Related work

According to the number of diagnostic classes and image types (Microscopy and whole slide images (WSI)), an intelligent system is developed for early detection and treatment of breast cancer which benefits our societies.

B. Histology and whole slide images (WSI)

Developed a system [5] that classifies WSI of breast biopsies into five different diagnostic categories. The initial step was salience finding utilizing a pipeline of four consecutive fully connected convolutional neural networks for multi-level classification. The second approach was a patch-based multi-level convolutional network for analysis. The final step was the saliency and classification plots are attached for pixel-wise labeling and slide level classification. Investigations using 240 whole slide images (WSI) exhibited that both classifier networks and saliency detector achieved improved than contending algorithms. The novelty of five different diagnostic classes is a problem that is discussed which is significant for the clinical applicability of CAD.

This paper [6] explores the challenging of breast tissue arrangement of histology images. The main goal is to categorize histology images into four different classes such as benign, normal, In-Situ and invasive carcinoma. Because of the high resolution (HR) of an individual image in the training dataset, a patch-based technique is implemented which consists of two successive convolutional neural networks. The proposed technique gives 95% validation accuracy on the validation set. The model is trained using [7] BACH dataset. The microscopic image dataset was accessible at https://iciar2018-challenge.grand-challenge.org/dataset/.



(a) Benign (b) In situ

(C) Invasive (d) Normal

Fig.1. Sample histology images of ICIAR 2018 challenge on Breast Cancer Histology [11] (BACH)

In this paper, the computation approach [8] based on deep CNNs is developed for the categorization of H&E stained microscopic images of very small training data. A strong data augmentation technique is used for training with a deep convolutional neural network for feature extraction. The testified results are better for the automated investigation of breast cancer.

To improve the sensitivity [9] of the benign and normal predicted classes of data by designing a Dual Path Network (DPN) which can be used to extract different features. To refine prediction the extracted feature is sent to the second layer of ensemble prediction fusion using GBM, logistic regression and SVM. The result shows that 12.5% development compared to the earlier methods.

In this paper, a transfer learning-based [10] method is used for breast histology image categorization into four tissue subtypes specifically normal, benign, *in situ* and invasive. Using 3-fold cross authentication a classification accuracy was estimated. The inception-V3 network attained an average test accuracy of 97.08% for four different classes of data when it is compared to the ResNet50 network, an average accuracy of 96.66% was achieved. There is an opportunity that malignant cells are existing in only a small part of the image, while the rest of the image represents healthy or benign tissue. Such instances lead to high false-negative rates. Future work is to address this constraint by inventing an appropriate substitute method.

In this paper [11] two different architecture is proposed in that the first method is single task CNN which can be used to predict benign/malignant and second method is multi-task CNN which can be used to predict both benign/malignant along with image magnification level. CNN's are more capable in breast cancer microscopic image classification and the dataset is the main problem to get more robust models. The stain normalization, splitting the network before the last fully-connected layer and deeper architectures for future work could be investigated.

In this paper [12] the performance of four well known deep learning network architecture is evaluated such as GoogLeNet, AlexNet, VGG16, and FaceNet. The two deeper networks such as GoogLeNet and VGG16 achieved the best patch-based classification performance. Deep network structure such as GoogLeNet is generally faster and more stable than VGG16.

With the advent [13] of WSI scanners, which performs high-throughput and cost-effective histopathology image digitization. The most difficult aspect is the segmentation of nuclei in breast cancer microscopic images is the identification of each nuclei when they are clustered very closely with each other and overlap.

C. Challenges

The Grand challenges help in the development of new breast cancer CAD methods to detect breast cancer at an earlier stage which benefits our societies.

II. DIAGNOSIS MODEL FOR BREAST CANCER

During preprocessing a background artifact will be reduced in an image and filtering technique is used to reduce noise in an image for image enhancement. In segmentation for the region of interest is used to detect a mass, microcalcification, and tissue.

The output of the segmentation is given as input to feature extraction for selecting the best features to improve accuracy. Identified ROI using segmentation method is examined independently for special characteristics like analyzation of different gray levels in [14] the region of interest, the texture of the region which is selected, analyzation of location, Size and Boundary in the region of interest and patterns that are found in the selected region of interest. The model is trained using training data and later it can be tested with testing data for validation. The classification algorithm can be designed to distinguish between the different classes of data.

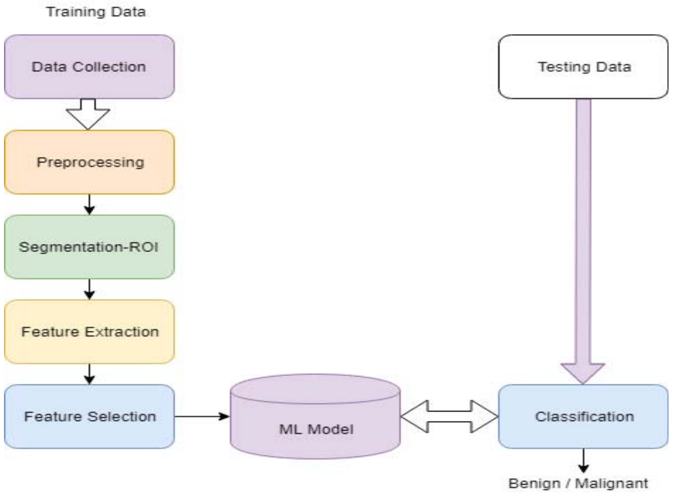


Fig. 2. Model for early detection of Breast Cancer.

III. Techniques used for Breast Cancer Detection

A. Decision Tree (DT)

For diagnosis, a machine learning algorithm called decision tree algorithms is extensively used in the medical field to categorize the clinical data. A feature selection [15] method can be used to increase the classification accuracy.

B. K-Nearest Neighbor (KNN)

The K-nearest neighbor's algorithm is a very simple method theoretically and implementation-wise compared to other methods that can be used to perform this task. The KNN algorithm gives the 1.17% improved [16] classification result compared to the earlier results for this problem. A few drawbacks of the KNN approach are (a) it is required to store all the training data and may take a lot of space (b) it needs more time to do testing.

C. Artificial Neural Network (ANN)

The important refrain used in this paper [17] is that ANN has confirmed to be a motivating and valuable substitute handling approach. Artificial neural network techniques still are not something mystic results with spiritual capabilities that work without any good programmer. In early detection and diagnosis of breast cancer, the programmer must understand their proficiencies and restrictions [18] that can be applied efficiently to various difficulties that may be incurred during cancer detection. Artificial neural networks have various applications in the detection of cancer like liver cancer, lung cancer, and Breast cancer.

D. Support Vector Machine (SVM)

In this method [19] a machine learning technique to do classification tasks such as Support Vector Machines (SVMs) is used for the prediction and identification of breast cancer. The paper offers the execution details along with the consistent results for all the evaluated classifiers. So many relative works of literature were carried out concerning both the forecast and analysis problem representing the superiority of the planned SVM algorithm in terms of sensitivity, specificity, and accuracy. The SVM classification algorithm was instigated for the diagnosis of Wisconsin Diagnostic Breast Cancer datasets. The obtained accuracy with this technique was about 97% along with specificity and sensitivity indices were also instigated which gave satisfactory results.

E. LS-SVM

In this paper [20] the importance of the LS-SVM was inspected using various methods such as k-fold cross-validation, confusion matrix for false negative and positive find outs, classification accuracy, sensitivity, and specificity. The classification accuracy was obtained about 98.53% and gives very good accuracy results compared to other techniques. The new intelligent system can be designed using LS-SVM which can make an effective analysis of data.

F. K-SVM

A hybrid of k-means and SVM (K-SVM) [21] methods can be used to extract suitable information from the data inputs and further this can be helpful for the diagnosis of the breast lump. The K-SVM algorithm can be used to identify the hidden information of the malignant and benign tumor very distinctly. The association between the various class of data needs to be analyzed thoroughly to obtain very good accuracy. Then a very good classification technique such as SVM is used to obtain good accuracy. Using a 10-fold (K-fold) cross-validation this method improves the accuracy of 97.38%. Which is verified on the WDBC dataset. During the training phase, breast tumor features of about six were extracted from the original 32 different features.

G. RS_SVM

Feature selection tool such as RS reduction algorithm is used to eliminate the superfluous features and further by using SVM [22] the diagnostic accuracy can be improved. The proficiency of using this algorithm is noticed on the WBCD dataset using ROC curves, Specificity, Accuracy, Sensitivity and Confusion matrix. The RS_SVM algorithm results suggest that not only it provides very good accuracy but also identifies a grouping of five useful features that can give a small clue to the doctors for breast cancer diagnosis.

H. FS_SVM

In this paper, SVM based classification along with [23] feature selection is proposed for breast cancer identification. The training and testing model was created using WBCD which is used by the researchers who use machine learning techniques for breast cancer detection and diagnosis. The performance matrix was analyzed which is associated with accuracy, specificity, sensitivity, confusion matrix, and ROC curves. The result shows that about 99.5% classification accuracy was obtained using the SVM model that includes five features and this is very important compared with earlier obtained results.

Table 2. Systematic observation of various classification methods

Author (Year)	Method	Classification accuracy (%)
Mehmet Fatih Akay (2009)	SVM-based combined with feature selection (Five features)	99.51%
Polat and Gunes (2007)	LS-SVM (10xCV)	98.53
Hui-Ling Chen et al. (2011)	RS-SVM (Subset of five features)	99.41
Bichen Zheng et al. (2013)	K-SVM (10-fold cross validation)	97.38
Murat karabatak et al. (2009)	AR+NN Model (3-fold cross validation)	95.6
Bennett and Blue et al. (1997)	SVM algorithm (5-fold cross validation)	97.20
Setino (2000)	Neuro-rule 2a (training: 50%, testing: 50%)	98.10

IV. Discussion

Black women having age less than 60 years old are more prone to Breast cancer and it is higher when compared with white women. Breast cancer is more amongst [24] white women having age more than 60 years and less associated with black women. Among white women, death incidences were reduced to 2% per year and among black women, it is reduced to 1% per year. This is a large variance in the percentage rate decrease of deaths among black and white women. Cancer identified at the last stage for black women due to various reasons compared to white women. Black women were diagnosed with about 45.7% in the more curable stage when it is compared with white women.

The collected data suggest that both white and black breast cancer patients in regions with lesser income values /average education are recognized at a later stage of disease when compared to patients in regions with higher income values /middle education. Compared to white women, black women had very little chance of detecting breast cancer disease. These collected data confirm that women of very poor socioeconomic grade benefit from mammography screening.

As the density of Breast growths, it becomes very problematic to notice minor cancers with mammography as cancers have a similar X-ray attenuation as fibro glandular breast tissue. As anticipated, breast density increases result in a decrease in the sensitivity of mammography. The mammography sensitivity for a woman with nearly completely fatty breasts denoted as ‘Dense-1’ is 88% as associated with [25] almost 82% for women having dispersed fibro glandular densities which can be denoted as ‘Dense-2’, 69% for women with heterogeneously dense breasts denoted as ‘Dense-3’, and 62% for women having extremely dense breasts denoted as ‘Dense-4’. It is noticed that women having a density of breast 75% or more were likely to have breast cancer about 17.8 times extra compared with women having breast density which is less than 10% detected within one year of the last screen examination. This is suggestively improved danger of breast cancer within a year of a mammography screening.

Table 3. Comparison of different classification algorithm [26]

Classification algorithm	Accuracy (%)
Multilayer Perceptron	96.70
KNN	96.27
Classification regression trees	93.62
Gaussian Nave Bayes	96.42
SVM	98.24

Mathematically, we can represent the accuracy of the classification algorithm to evaluate the performance of each classification algorithm.

$$\text{Accuracy} = (x/n) \times 100\%$$

In some recent research, the paper author mentioned that women having a density of breast which is greater than 75% had a 3.25-fold possibility of occurrence of breast cancer [27] compared with women having breast density less than 5%. A biopsy is one of the best methods which can be used to detect breast cancer. When a biopsy is done an expert radiologist uses a specialized needle device to extract a small amount of tissue from the ROI or doubted area directed by X-ray or Ultrasound imaging. A difference can be visualized concerning the dense and non-dense breast as shown in fig. 2. It is investigated through a thorough study of the number of incidences of cancer among black women is more compared to white women. The reason behind this is due to illiteracy, less awareness about breast cancer and various socio-economic problems. In the rural side, women are more prone to breast cancer due to less awareness about the disease.

Table 4. Summary of different techniques with limitations and solution

Method	% Accuracy	Limitations	Solution
Classification Improvement through feature selection using KNN, [28] Linear discriminant analysis and probabilistic neural network	99.17	The standard deviation of the average retained features decreased as the training proportion increased	Recommended to use the KNN algorithm with as much data to train the model as possible
A hybrid [29] approach using linear discriminant analysis (LDA) and then applying the new reduced feature dataset to SVM	98.82	Less disease option and A less practical method for assisting doctors with a quick second opinion	Integrating significant dimensionality reduction methods with ML classification
Sigmoid based Naïve Bayes [30] algorithm	99.20	SVM with sigmoid shows worse performance	Proper feature selection methods must be used for the realization of the algorithm
Majority based voting [31] mechanism	99.42	Performing accurate classification	To evaluate different feature selection algorithms to determine the smallest subset of features in accurate classification
ANN with [32] logistic algorithm for diagnosis & detection of breast cancer	98.50	Large features are effect on the cost of model implementation only 16 features are used	In the future, we can try on all UCI and achieve the best accuracy

V. Conclusion and Future scope

The literature review is carried out to understand the development procedure of CAD systems for the detection and analysis of single and multilevel breast cancer disease. Various techniques to classify with the WBCD dataset using a machine learning algorithm is been explored in this paper. It is noticed that the support vector machine learning algorithm gives the best result to classify the early identification of breast cancer. Some limitation of using ultrasound, mammography image classification is also discussed. It is possible to improve the accuracy to 99.8%, which was achieved using SVM single or hybrid technique to reach 100% accuracy. It is noticed that by using ANN, the MRI delivers 100% accuracy [33] in the early identification of Breast cancer. This paper has also surveyed some limitations on using mammography, ultrasound and MRI scanning images for the early detection of cancer in the case of the dense breast. Besides, using SVM with feature extraction is a new place [34] for future research work. We have noticed that from so many research papers the histology based image classification is observed as the best method in early detection and proper treatment for breast cancer. BACH is arranged in a systematic way to promote research for automatic breast cancer microscopic and whole slide image analysis. By increasing [35] dataset with more trained labels using data augmentation technique has the promise to increase the training accuracy. The proper CNN design is required to attain high performance in the case of H&E stained image analysis. It is observed that color normalization is not crucial to achieving high accuracy. Proper training with the data augmentation technique is required to infer the correct class of data and to reduce the misclassification of data. At the outset, the present deep neural network methods have problems dealing with large, HR images and further improvement on development approaches for whole slide image analysis is required. In the future, it is required to design and implementation of a deep neural network such as MobileNet using FPGA with large data sets with additional comprehensive training labels that have the potential to result in systems that are beneficial to the pathologist in clinical applications. Even though MobileNet architecture gives little less accuracy compared to all other methods discussed above but for fast processing and mobile application it is required to classify different classes of data using 7 series FPGA kit such as Kintex-7. The Grand challenges help in the progress of new breast cancer CAD approaches to detect breast cancer at an earlier stage which benefits our societies.

References

1. National Cancer Institute: The Burden of Cancer in the United States. https://www.cancer.gov/about-cancer/understanding/statistics
2. Monticciolo, Debra L., Mary S. Newell, Linda Moy, Bethany Neill, Barbara Monsees, and Edward A. Sickles. "Breast cancer screening in women at higher-than-average risk: recommendations from the ACR." Journal of the American College of Radiology 15, no. 3 (2018): 408-414.
3. Didkowska, Joanna, and Urszula Wojciechowska. "WHO position paper on mammography screening." Oncology in Clinical Practice 11, no. 1 (2015): 16-19.
4. Breast Cancer Facts and Statistics: https://ww5.komen.org/BreastCancer/WhatisBreastCancer.html
5. Gecer, Baris, Selim Aksoy, Ezgi Mercan, Linda G. Shapiro, Donald L. Weaver, and Joann G. Elmore. "Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks." Pattern recognition 84 (2018): 345-356.
6. Nazeri, Kamyar, Azad Aminpour, and Mehran Ebrahimi. "Two-stage convolutional neural network for breast cancer histology image classification." In International Conference Image Analysis and Recognition, pp. 717-726. Springer, Cham, 2018.

7. ICIAR 2018 grand challenge: In 15th International Conference on Image Analysis, Recognition. https://iciar2018-challenge.grand-challenge.org/
8. Rakhlin, Alexander, Alexey Shvets, Vladimir Iglovikov, and Alexandr A. Kalinin. "Deep convolutional neural networks for breast cancer histology image analysis." In International Conference Image Analysis and Recognition, pp. 737-744. Springer, Cham, 2018.
9. Vang, Yeeleng S., Zhen Chen, and Xiaohui Xie. "Deep learning framework for multi-class breast cancer histology image classification." In International Conference Image Analysis and Recognition, pp. 914-922. Springer, Cham, 2018.
10. Vesal, Sulaiman, Nishant Ravikumar, AmirAbbas Davari, Stephan Ellmann, and Andreas Maier. "Classification of breast cancer histology images using transfer learning." In International Conference Image Analysis and Recognition, pp. 812-819. Springer, Cham, 2018.
11. Bayramoglu, Neslihan, Juho Kannala, and Janne Heikkilä. "Deep learning for magnification independent breast cancer histopathology image classification." In 2016 23rd International conference on pattern recognition (ICPR), pp. 2440-2445. IEEE, 2016.
12. Wang, Dayong, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H. Beck. "Deep learning for identifying metastatic breast cancer." arXiv preprint arXiv:1606.05718 (2016).
13. Veta, Mitko, Josien PW Pluim, Paul J. Van Diest, and Max A. Viergever. "Breast cancer histopathology image analysis: A review." IEEE Transactions on Biomedical Engineering 61, no. 5 (2014): 1400-1411.
14. Houssami, Nehmat, Georgia Kirkpatrick-Jones, Naomi Noguchi, and Christoph I. Lee. "Artificial Intelligence (AI) for the early detection of breast cancer: a scoping review to assess AI's potential in breast screening practice." Expert review of medical devices 16, no. 5 (2019): 351-362.
15. Bindhu, V. "Biomedical Image Analysis using Semantic Segmentation." Journal of Innovative Image Processing (JIIP) 1, no. 02 (2019): 91-101.
16. Sarkar, Manish, and Tze-Yun Leong. "Application of K-nearest neighbor algorithm on breast cancer diagnosis problem." In Proceedings of the AMIA Symposium, p. 759. American Medical Informatics Association, 2000.
17. M. Karabatak and M. C. Ince, "Expert Systems with Applications An expert system for detection of breast cancer based on association rules and neural network," Expert Syst. Appl., vol. 36, no. 2, pp. 3465–3469, 2009.
18. M. Sarkar, T. Leong, L. Kent, and R. Road, "Application of K-Nearest Neighbors Algorithm on Breast Wisconsin of That priority," pp. 759–763, 2000.
19. I. Maglogiannis and E. Zafiropoulos, "An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers," pp. 24–36, 2009.
20. K. Polat and S. Güne, "Breast cancer diagnosis using least square support vector machine," vol. 17, pp. 694–701, 2007.
21. Zheng, Bichen, Sang Won Yoon, and Sarah S. Lam. "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms." Expert Systems with Applications 41, no. 4 (2014): 1476-1482.
22. H. Chen, B. Yang, J. Liu, and D. Liu, "Expert Systems with Applications A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis," Expert Syst. Appl., vol. 38, no. 7, pp. 9014–9022, 2011.
23. Akay, Mehmet Fatih. "Support vector machines combined with feature selection for breast cancer diagnosis." Expert systems with applications 36, no. 2 (2009): 3240-3247.
24. Hall, Susan A., and Beverly Rockhill. "Race, poverty, affluence, and breast cancer." American journal of public health 92, no. 10 (2002): 1559-1559.
25. Leconte, Isabelle, Chantal Feger, Christine Galant, Martine Berlière, Bruno Vande Berg, William D'Hoore, and Baudouin Maldague. "Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density." American Journal of Roentgenology 180, no. 6 (2003): 1675-1679.
26. Bataineh, A. A. "A comparative analysis of nonlinear machine learning algorithms for breast cancer detection." International Journal of Machine Learning and Computing 9, no. 3 (2019): 248-254.
27. Vierkant, Robert A., Amy C. Degnim, Derek C. Radisky, Daniel W. Visscher, Ethan P. Heinzen, Ryan D. Frank, Stacey J. Winham et al. "Mammographic breast density and risk of breast cancer in women with atypical hyperplasia: an observational cohort study from the Mayo Clinic Benign Breast Disease (BBD) cohort." BMC Cancer 17, no. 1 (2017): 84.
28. Khuriwal, Naresh, and Nidhi Mishra. "Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm." In 2018 IEEMA Engineer Infinite Conference (eTechNxT), pp. 1-5. IEEE, 2018.F
29. Fogliatto, Flavio S., Michel J. Anzanello, Felipe Soares, and Priscila G. Brust-Renck. "Decision Support for Breast Cancer Detection: Classification Improvement Through Feature Selection." Cancer Control 26, no. 1 (2019): 1073274819876598.
30. Omondiagbe, David A., Shanmugam Veeramani, and Amandeep S. Sidhu. "Machine Learning Classification Techniques for Breast Cancer Diagnosis." In IOP Conference Series: Materials Science and Engineering, vol. 495, no. 1, p. 012033. IOP Publishing, 2019.
31. Mushtaq, Zohaib, Akbari Yaqub, Ali Hassan, and Shun Feng Su. "Performance Analysis of Supervised Classifiers Using PCA Based Techniques on Breast Cancer." In 2019 International Conference on Engineering and Emerging Technologies (ICEET), pp. 1-6. IEEE, 2019.
32. Assiri, Adel Saad, Sergio Velastin, and Saima Nazir. "A Hybrid Ensemble Method for Accurate Breast Cancer Tumor Classification using State-of-the-Art Classification Learning Algorithms." (2019).
33. Ghaddar, Bissan, and Joe Naoum-Sawaya. "High dimensional data classification and feature selection using support vector machines." European Journal of Operational Research 265, no. 3 (2018): 993-1004.
34. V. Bevilacqua et al., "An Optimized Feed-forward Artificial Neural Network Topology to Support Radiologists in Breast Lesions Classification," pp. 1385–1392, 2016.
35. A. G. Howard and W. Wang, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2012.

International Journal of Advanced Science and Technology
Vol. 29, No. 3, (2020), pp. 11717 - 11723

Design and Implementation of Efficient Routing Algorithm for Wireless Sensor Networks

[1]Raghunandan.G.H, [2]Dr.A.Shobha Rani,

[1]*Department of Electronics and Telecommunication Engineering,*

[2]*Department of Electronics and Communication Engineering,*

[1,2]*BMS Institute of Technology and Management, Bengaluru, India*

[1]*Email: raghunandangh@bmsit.in*

Abstract

A minor node with wireless communications, computation capabilities, sensing Applications in the field of Wireless sensor network has broadened in various fields. These applications necessitate precise information gathering and also unremitting, expanded active service. Overall energy consumption is one the noteworthy impact of sensor networks in routing protocol. Appropriate Energy. Efficient routing algorithms require characteristics of networks to be inherited. Because of limited resources in WSN, increasing the lifetime of the network will always have a great interest. A large amount of the sensor node's energy is used for data transmission to base station. Consequently, the energy drains more rapidly. In this project, Agglomerative cluster based approach is used between sensor nodes, base station and gateway nodes to reduce the energy used by the cluster heads. It was proved that the efficiency of the network lifespan, residual energy of network has been improved in the in the simulation results. The performance of WSN of the proposed scheme is compared to other routing scheme and it shows improvement in the WSN.

Keywords: *WSN, cluster head, sensor node, LEACH.*

1. Introduction

Major developments in digital signal processing (DSP) which has led to growth of micro-sensors. Previously few industries use wired sensors; implementation provides deployment of sensor nodes more viable than before. Previously, there has been study regarding applications of WSN such as agricultural field, vehicle monitoring, machine monitoring, military surveillance and home automation. Many researches are undergoing considering the power constraints in WSNs by large deployment of sensors. To ensure real-time and reliable data transmission. Recently there has been exposure in the field of WSNs and their applications because they are easy to deploy and are of low cost, have flexibility. A WSN have distinctive set of resource curtailment like limited battery power, processing ability and limited communication bandwidth. Since sensors are battery-powered, energy efficiency is of vital importance in WSNs. Algorithms are used to solve the problem of power constraint without altering the standard. Local collaboration among sensors, suppression, data compression, redundant data, avoidance of direct transmission to far distant sensors are of the major factors that influence algorithm designers to device unique distributed, scalable and energy competent solution for WSNs.

In common, the sensor nodes measure environmental conditions. The sensor node extracts some useful information by processing the raw sensor signals. The output of this processed signal is transmitted the through direct communication or multi-hop

ISSN: 2005-4238 IJAST

communication with access point across other sensor nodes. In some situations, repeaters (RPs) are used for multi-hops, to support sensors installed outside the radio range. One of the components of WSN is the base stations which have more energy, computational, communication resources. Forwarding of data from wireless sensor network on to a server is done by the BS which acts as a gateway. Energy is a limited resource of WSN, and it determines the lifetime of WSNs. The computation subsystem has less energy consumption when compared to the communication subsystem. The energy required for transmitting one bit may consume as same as executing a few thousands instructions. Hence, communication must be traded for computation. In different environment, together with remote and hostile regions, in which ad-hoc communications are the key element of WSNs that is to be deployed in large numbers. Due to this reason, algorithms and protocols should concentrate on the subsequent issues:

- Lifetime maximization: Sensor nodes must be energy efficient and consumption of energy of the device should be less due to the limited energy resources. The radio power supply when not in use should shut off to conserve power of the node
- Fault tolerance and Robustness.
- Self-configuration.

2. Existing Systems

In this system [2] Network is divided into areas as per geographical locations based on clustering for Cluster Head (CH) selection and formation in WSNs. CH. The cluster head energy level reduces after several trials because of more energy consumption. The data is forwarded to base station after aggregating at cluster level. Hence, it is no more able to be a cluster head. Thus, formation of cluster takes place when CH remaining energy is below threshold goes; hence cluster reformation is not that efficient comparatively. In [3] it aims at that in the network level and sensor node level in a WSN they should minimize energy consumption. The distance between the receiver and the transmitter is estimated before available transmission, and then, the very minimum transmission power needed to transmit, to minimize the energy consumption of the sensor node, but here the neighbor status awareness is less in [4] extending the network life time of the network was done considering mobile base station. Even though it shows better performance but mobile base station is not feasible for all the situations. In [9] for transmission of data to the Base Station majority of the energy of sensor nodes is used. Thus, there is fast depletion of energy. Here agglomerative to reduce energy consumption of cluster heads a portable base station is utilized along with cluster approach. But movement of Base Station is not always feasible.

In [5] this it discusses up several technical challenges and many application possibilities which occurs when the sensor networks interconnect several nodes when wide networks are established. This wireless sensor networks communicates using multiple -hop wireless communications systems. In [6] In order to guarantee reliable multi-hop communication and to maintain the routes in the network routing protocols for wireless sensor networks are used. It provides us with an idea on routing protocols for Wireless Sensor Networks and compares their relative strengths and limitations provide better energy efficiency or increase wireless. The proposed model in [10] discusses about WSN using a gateway nodes which will further increase the number of nodes. If gateway nodes dies then there will chances of losing the data which is already aggregated for transmission to base station. Routing in WSNs [5] is generally, classified into many basis. The classification considered in this paper is based on network structure.

ISSN: 2005-4238 IJAST

3. Proposed System

The following assumptions are considered for the proposed system:

- The base station and sensor is considered to be static.
- The nodes have equivalent initial energy and are left as unattained after deployment, i.e. battery recharge is not possible. But a limitation on energy, memory and computation is not considered for Base Station. The nodes have the capability to vary the power transmission with reference to the distance.
- Using wireless radio signal strength distance can be calculated.
- Usually, energy reduction is one of the reasons for node failure.

In the recent past wireless sensor network has attracted substantial research attention as WSN's are fast growing in hierarchical-based routing algorithms, Cluster Heads are in-charge of compressing, forwarding and, gathering, data to the BS. Cluster heads plays an important role in reducing the congestion of the network. CHs are also elected based on certain criteria. Thus efficiency, network lifetime and network stability of WSN are increased. In proposed algorithm, the sensor member senses the data and sends the data to the individual Cluster Heads.

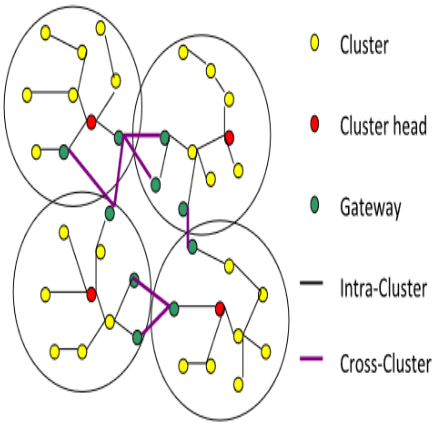


Figure 1. Clustering in WSN

Cluster Head processes this data and sends it to gateway node which in turn forwards it to the Base Station. Hence the lifespan of the CH increases which increases the network lifetime. Clustering is a process of connecting nodes using a specific topology to perform certain tasks as per the requirements. The algorithm used for wireless sensor network discovers a set of distinguished nodes to construct the appropriate topology of the network.

Next step after the deployment of the sensor nodes is grouping the sensors into cluster. In proposed algorithm, cluster formation is same as that employed in the LEACH algorithm. Once Cluster is formed considering nodes of the network, the Cluster Head (CH) is elected at the beginning of every round. The proposed technique is fixed with threshold value for the cluster head selection. Data transmission is continuously monitored with energy updating after each round. We have agglomerative based clustering approach, the distance between each single node to each of other nodes is determined and the node which is at almost equal distance with neighboring nodes is considered to be CH. Cluster Head is elected based on following set of rules:

- CH's residual energy: The CH is selected based on the maximum amount of energy it has.

ISSN: 2005-4238 IJAST

- Cluster head to the base station distance is considered. The more distance from each other the more energy required for data transmission.
- Depends on the number of nodes in the cluster that is in a cluster if there are more number of members data processing required is also more.

4. Simulation and Evaluation

The proposed system is simulated using MATLAB software. We consider 100 sq.m area and 100 nodes are distributed randomly. MATLAB Simulation is done for 1200 rounds. The proposed algorithm is compared with LEACH. Figure 2 shows sensor deployment, Identification of cluster head in each cluster and presence of base station. Figure 3 indicates transfer of data from the sensors to the cluster head, from cluster head to base station. Figure 4 shows the presence of dead sensors after around 950 iterations. Figure 5 shows the number of dead sensors present after few more iteration.

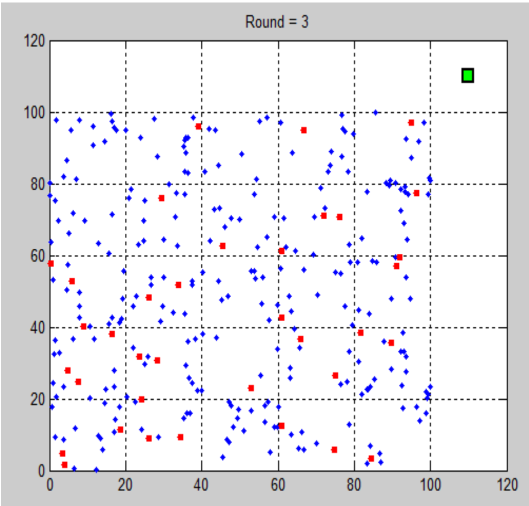


Figure 2. The Sensor Deployment and the Presence of Cluster Heads

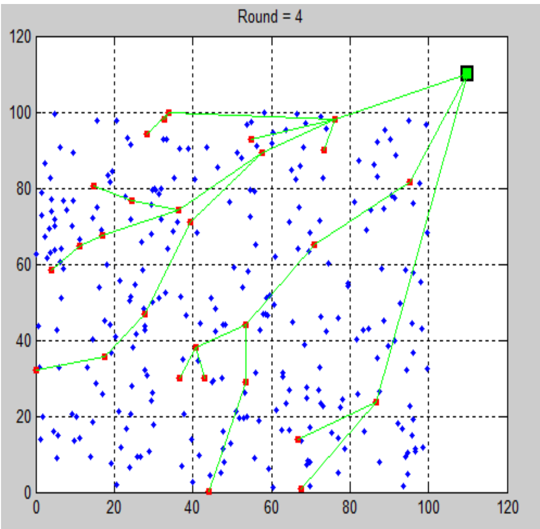


Figure 3. Communication of the Cluster of Sensors with Cluster Head and BS

ISSN: 2005-4238 IJAST

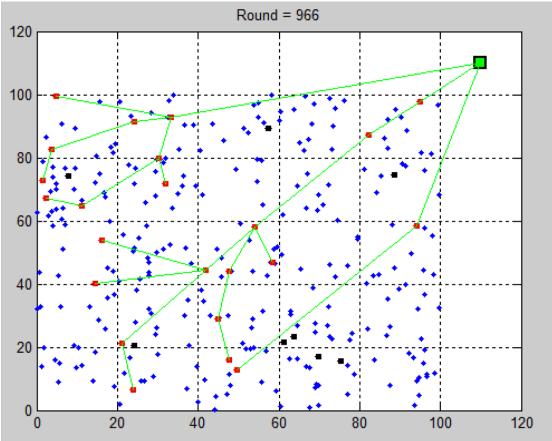


Figure 4. Presence of few dead sensors after little iteration

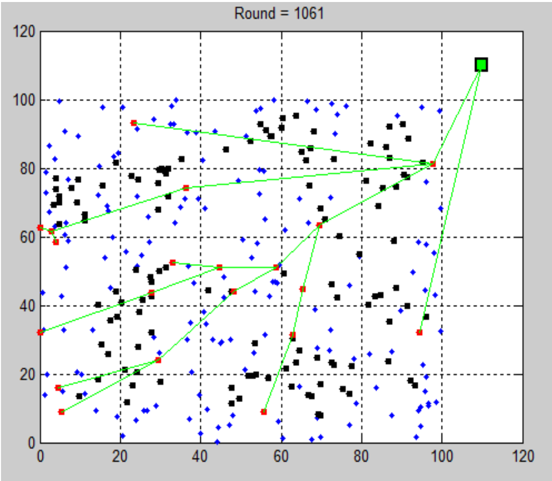


Figure 5. More Number of Dead Sensors after Further More Iteration

Here we are comparing our proposed system with leach protocol. Figure 6 indicates the presence of alive nodes after corresponding iterations. The number of dead sensors appears nearly around 1000 iteration as shown in figure 7. Hence the network lifetime is improved and the residual energy after each round provides us the information about the amount of energy consumed as in figure 8. Hence there is an improvement in the lifetime of the network of our proposed system.

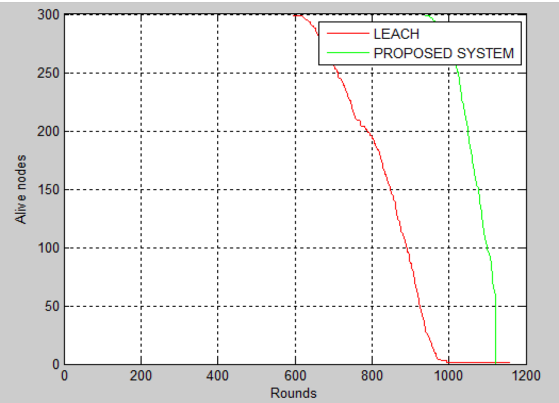


Figure 6. Number of Alive Sensor after Corresponding Rounds

ISSN: 2005-4238 IJAST

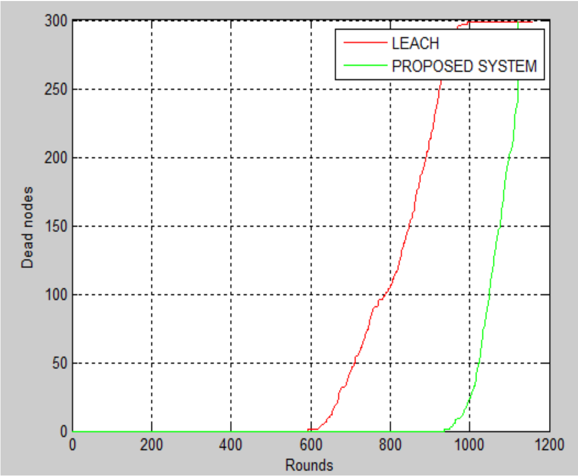


Figure 7. Graph of the Number of Dead Nodes Present After the Respective Rounds

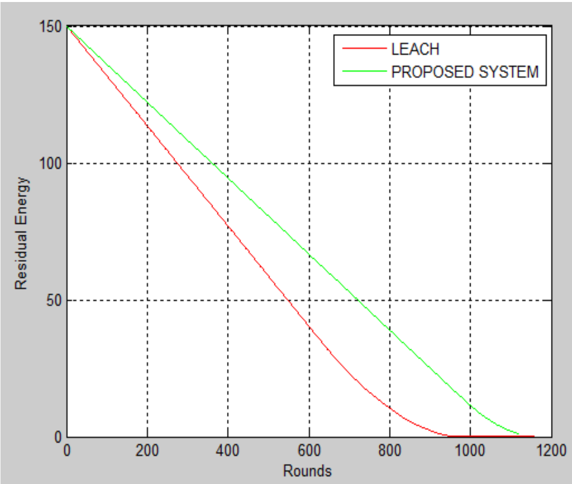


Figure 8. Residual Energy after Each Round

5. Conclusion and Future scope

In agglomerative clustering, the CH selection and power consumption is a major challenge. The energy utilization can be decreased by properly designing the cluster head selection mechanism. The selection of cluster head proposed in this paper is a new technique discussed with respect to distance from base station. The proposed system shows its performance in increasing lifetime of the network and overall performance of the network, which stay alive for the maximum amount of time. For the purpose of increase in lifetime of wireless sensors there is a need of significant research done in this field. An efficient manner is required for reducing the energy utilization by the nodes in wireless sensor networks.

References

[1] Moumita ,Paramita C, Sarmistha, Chowdhury "Analysis of Energy Efficient WSN Routing Schemes", International Conference of EAIT ,March 2014.

[2] Abhishek C, Sumedha, "Minimization of Avg Energy Consumption to Prolong Lifetime of WSN", 2014 IEEE Global Conference on WCN, July 2014.

ISSN: 2005-4238 IJAST

[3] Ruqiang, Sun,Yuning, "Energy-Aware Sensor Node Design With Its Application in WSN", IEEE Transactions on instrumentation and measurement, Vol. 62, No. 5, May 2013.

[4] Raghunandan., Sagar M, "A Novel Approach to increase overall Efficiency in WSN ",2012 IEEE ICCEET proceedings, pp 699-703, March 2012.

[5] Raghunandan.G.H, "A Comparative Analysis of Routing Techniques for WSN", IEEE NCOIET-2011, IEEE Conference Proceedings, February 2011.

[6] Akkaya and Younis, "A Survey of Routing Protocols in WSN" Elsevier Ad Hoc Network Journal, vol. 3, pp. 325-349, 2005.

[7] Ying H , Chin Y, Ting C, " AER Protocol for mobile sink in WSN," IEEE Conference on Media Computing, conference proceeding, pp. 44-49, August 2008.

[8] Raghunandan. G H, Lakshmi "Secure Routing protocols for WSN", International Conference on FCS proceedings, August 2011.

[9] Raghunandan G H, Shobha Rani, S Y Nanditha, " Hierarchical clustering based routing Technique for overall efficiency of WSN," In Proceeding of the IEEE 2017 International Conference on RTECIT, May 2017.

[10] Raghunandan G H, Shobha Rani, S Y Nanditha, G Swathi "Hierarchical agglomerative clustering based routing algorithm for overall efficiency of WSN" In Proceeding of IEEE ICICT -2017, July 2017.

Authors

Prof.Raghunandan.G.H is presently working as Assistant Professor in Department of Electronics and Telecommunication Engineering of BMSITM, Bengaluru. Wireless Ad-Hoc and Sensor Networks, Embedded system are some of his research interest. He has authored six Textbooks and published more than 20 papers at various International conferences and journals.

Dr.A.Shobha Rani is presently working as Associate Professor in Department of Electronics and Communication of BMSITM, Bengaluru. Her areas of interest are Wireless Ad-Hoc and Sensor Networks. She has published many papers at various International conferences and journals.

ISSN: 2005-4238 IJAST

IMPLEMENTATION OF AUTOMATIC AIR TRAFFIC CONTROL SYSTEM: AN OVERVIEW

Dankan Gowda V[1] , Sridhar S B[2] , Ramesh Naidu P[3] Naveen Pai G[4], , Vinay Kumar B C[5]

[1,2,4] Department of Electronics and Communication Engineering

[3] Department of Computer Science and Engineering

[5]Department of Information Science and Engineering

[1] B.M.S Institute of Technology and Management Bengaluru,

[2] Vijaya Vittala Institute of Technology, Bangalore

[3]Nitte Meenakshi Institute of Technology Banaglore

[4]Yenepoya Institute of Technology, Moodbidri

[5]Sri Krishna Institute of Technology, Bangalore

Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India

Abstract

The airspace system's capacity and safety are highly dependent on the skilled coordination of air traffic control System (ATCS) and flight desk personnel. which means that the rate of air traffic will increase exponentially leading to significant congestion, flight delays, and pollution. To keep these numbers of aircrafts at safe distances from each other, to direct them during takeoff and landing from airports, to guide them around bad weather and ensure that traffic flows smoothly with minimal delays. The main objective of this research is to provide an automatic Communication between the Airport and the Aircraft since the present situation in India is completely manual. The Problem of manually checking ***c****limatic conditions, runway parameters, air traffic, and various other information can be solved by using GSM technology. To reduce the manual efforts and human errors. Before landing, the arrival time of the aircraft is announced automatically.*

Keywords: Aircraft, Air Traffic control System, pollution, Traffic Flow

1. INTRODUCTION

The present situation in Indian airport is all using manual communication between the airport and aircraft. The airport personnel will manually check out for the climatic conditions, runway parameters, air traffic, and various other information and report it to the pilot. Having all the runway parameters, then the pilot decides whether to land immediately or wait for some time. In this process of manual operations, there may be some human errors occurring leading to disasters.

Therefore, to reduce the manual efforts and human errors, we need to have some kind of automated system monitoring all the parameters and functioning of the communications between the pilot and the airport personnel. The automation system should have the ability to control the air traffic, having the information about the airport. It should acquire the flight information, compare with the pre-stored information about the flight and take appropriate actions for its landing.

The communication between the air traffic controller and the pilot is setup using Global System for Mobile Communication Technology (GSM).

2. RELATED WORK

We analyzed so many papers by doing Literature Survey. From the study of these papers we came to know about the non-linear Coordinated Turn model exhibits better performance than linear Uniform Motion model for all estimation techniques. We also analyzed about the effect of different window sizes on the performance of the adaptive filters. We studied about the ATC (Air Traffic Control) and ATFM (Air Traffic Flow Management).

Aiman Javed proposed RMSE [1] results, it has found that the non-linear Coordinated Turn model exhibits better performance than linear Uniform Motion model for all estimation techniques. Scaling factor based g-Adaptive technique has been employed for adaptation purposes. The effect of different window sizes on the performance of the adaptive filters has also been studied.

Aakash Desai proposed a novel methodology for routing and management of aircraft within the airspace managed by the ATC (Air Traffic Control) of an airport. The system examines various parameters such as current wind conditions, air time, taxi time, type of aircraft, etc. to determine the most efficient combination of runways in terms of overall fuel consumption. An automated fuel-optimal air traffic management system[2] for airport ATC (Air Traffic Control) is proposed.

Kai-QuanCai proposed air traffic flow management (ATFM) which aims to facilitate the utilization of airspace and airport resources and is critical in air transportation systems. This paper addresses the problem of alleviating the airspace congestion and reducing the flight delays in ATFM simultaneously [3].

3. METHODOLOGY

The pilot will sends a message via GSM transmitter requesting for the runway for its landing. This message is received by the aerodrome controller GSM receiver. The aerodrome controller analyses the flight information, and compares it with the pre-defined values. The controller shall check for the authenticity of the arriving flights with the stored database of all the flights and pilots. The message sent by the pilot would consist of pilot identification, flight identification, flight class, number of onboard passengers, and flight technical details. To this message, the ground based air traffic controller would reply with the parameters with which the flight has to approach the runway, and the exact geographical location of the runway, and the time at which it has to descend. The controller could add miscellaneous information about the climatic conditions, and the condition of the runway. During this process, the controller simultaneously has to announce the arriving flight name, and its time. Using the GSM MODEM, the air traffic controller communicates with the pilot. Every pilot in his aircraft is provided with another GSM MODEM for retrieving and sending the messages. The GSM MODEM is interfaced with the controller using RS-232 interface, using LM 232.

4. DESIGN CONSIDERATIONS

Our motivation is to build an embedded system where airport and the aircraft are monitored and controlled continuously and automatically, the LCD is used to display the messages. The micro controller will sense the runway and send the availability (condition) of the runway and platform numbers etc. The main feature used in our Project is GSM Technology. We have also designed using Microcontroller which will sense the runway and the availability of the runway Parameters. For the checking of Climatic Conditions we are using LDR circuit, Relay circuit and Sensors. For Displaying of the messages we are using LCD display.

5. BLOCK DIAGRAM

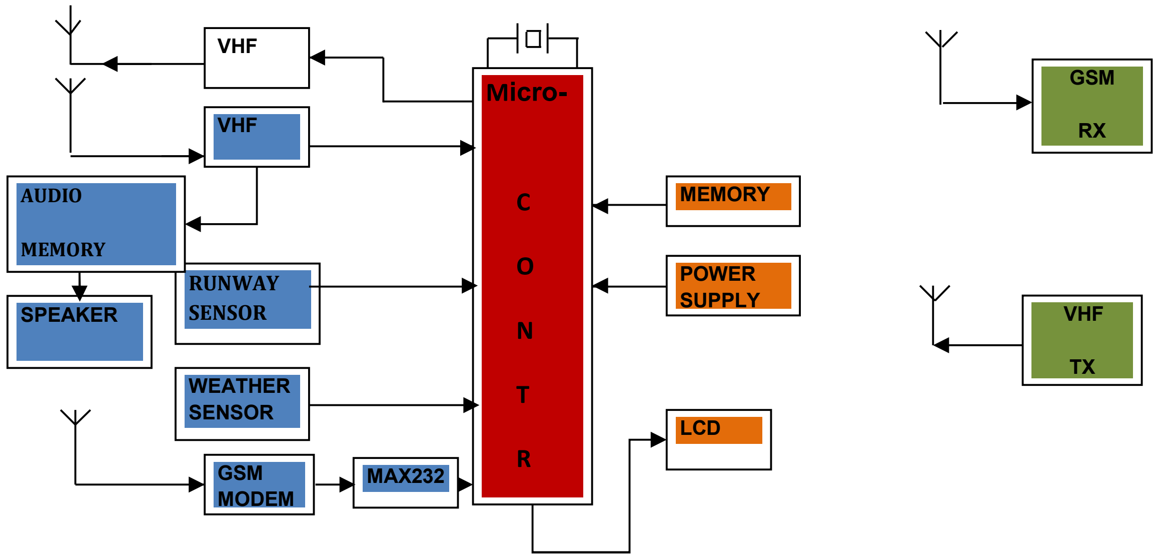


Figure. 1. Block Diagram Representation of Aircraft Traffic Controller

Microcontroller

In this project we use an 89S52 micro controller, which belongs to the Intel 8051 family Architecture, it plays an important role and can also be called as 'heart of our project' and this function are controlled by standard AT commands of GSM technology. It contains a ROM burner to burn program into the microcontroller. ROM burner can erase the flash ROM in addition to burning a program into it.

VHF Transmitter and Receiver

We are mainly using very high frequency modulation called, FREQUENCY MODULATION technique in our approach. It is working under 27 MHZ of frequency. It contains RF amplifiers for amplification purposes. Its main purpose is to transmit the message through an antenna from the transmitter part and receives the message through an antenna by the receiver part.

GSM Modem and Max232

These two units help in sending and receiving of messages. GSM stands for GLOBAL SYSTEM FOR MOBILE COMMUNICATION. The main feature of MAX232 is it converts RS232 logic to TTL logic while transmission, TTL logic to RS232 logic while reception.

LCD (LIQUID CRYSTAL DISPLAY)

Here we are using 2X16 LCD display to display the message of 2lines and 16 characters. The contrast of the LCD can be varied through the variable resistor.

Runway Sensor

For runway sensing purpose we are using IR rays, which helps in checking the runway availability**.** If the runway is free it sends particular signal to the microcontroller.

Weather Sensor

In this we are using LDR (light dependent resistor), which forecasts the weather conditions. If the conditions are satisfied for landing it sends positive signal to the If the runway is free it sends particular signal to the microcontroller.

Audio Memory Amplifier and Speaker

As soon as the aircraft lands the arrival time of the aircraft will be announced automatically through the speakers. The audio memory amplifier contains many recorded messages in its memory. Before announcing it amplifies the message.

Memory and Power Supply

Here we use memory to store different pilot's information messages, which are required for announcements etc. Here we use AC 220V power supply from which we can derive +5V and +12V of power supplies respectively. +5V is required for LCD and the microcontroller and +12V is required for other devices.

Solid State Voice Recording and Playback Circuit

When the aircraft reaches the vicinity of the airport, there should be provision for automatic announcement for the benefit of passengers. So this circuit provides that mechanism of automatic announcement. In this circuit we have partitioned the memory depending on the number of messages required. There is the presence of switches connected to the APR 9600 for the 8 messages to be stored. The AGC circuit automatically controls the gain .As the distance from the mouth and mike varies, there is signal variation, the AGC produces constant output irrespective of the variation, which is required for the microcontroller. Voltage is regulated as required by the components connected by voltage regulator. Opamp acts as amplifier. The amplified signal is carried through speaker. The block diagram of the solid state voice recorder and playback circuit is shown in Figure.2.

SOLID STATE VOICE RECORDER AND PLAY BACK

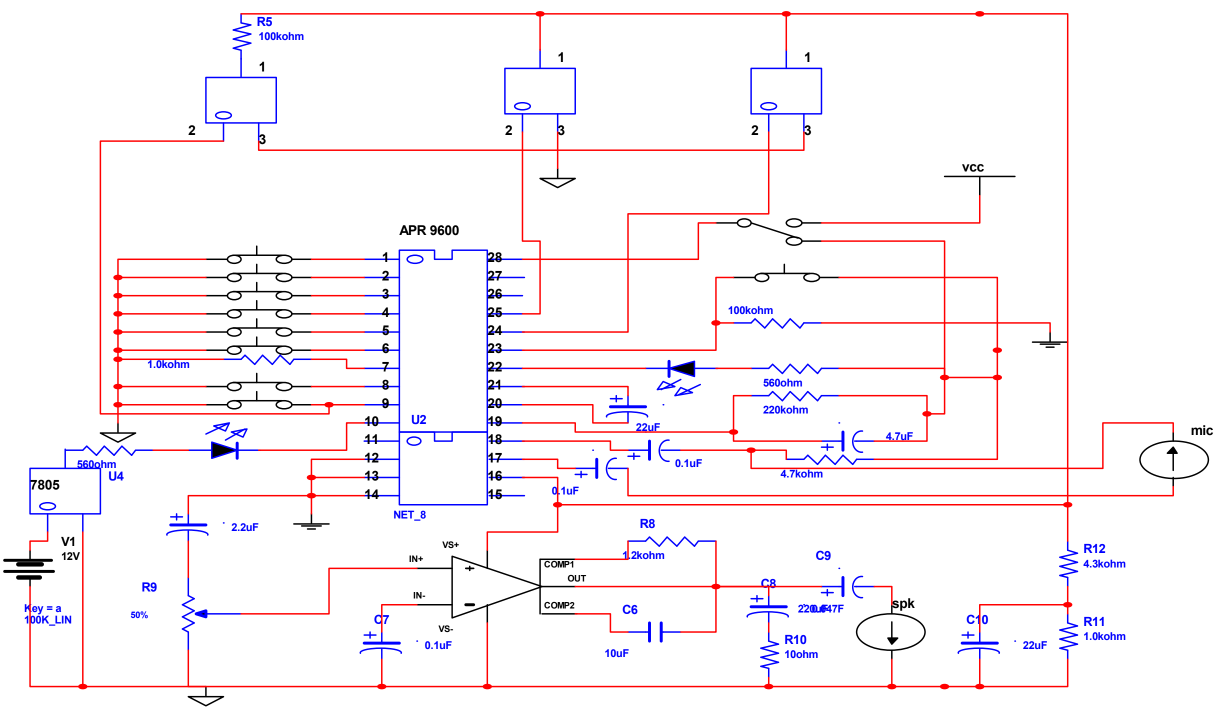


Figure.2. Solid State Voice Recorder And Play Back

IR Reciever

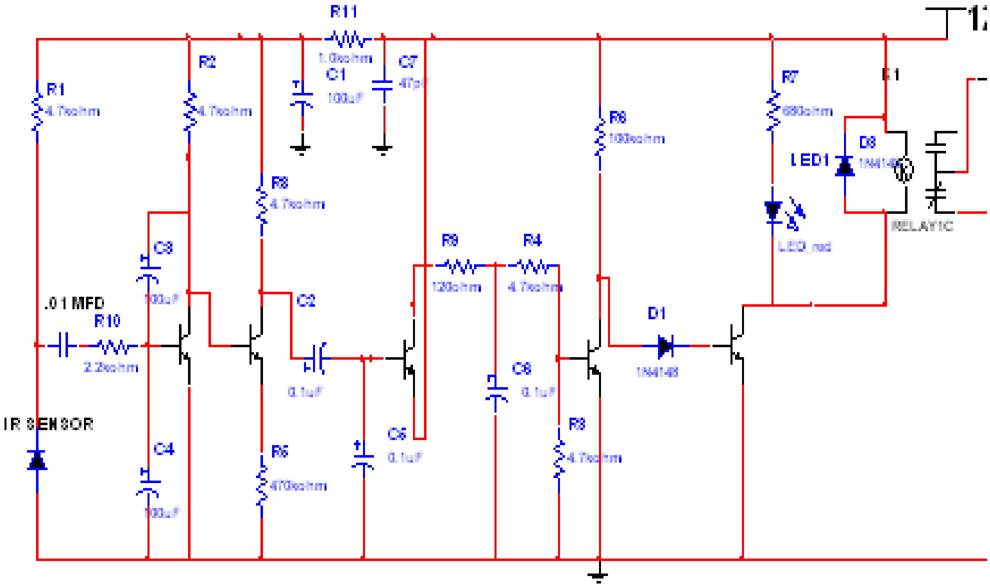


Figure: 3. Circuit Diagram Of IR Receiver

The above figure shows the block diagram of the IR receiver. In receiver there are 5 transistors in this circuit. Transistors 1-4 are used as amplifiers and transistor 5 is used as a driver circuit to drive the relay and it provides negative voltage to the relay through emitter to collector.

Working

The receiver IR LED (photo transistor) is placed in the base of transistor Q1. The received IR signals is amplified at transistor Q1 and output is taken from collector and given to the base of transistor Q2. The transistor Q2 is amplified input signal and fed to the base of transistor Q3 through coupling capacitor Q4 and output is taken from collector and fed to the base transistor Q4 through resister R8 and R9 and it is working as a coupler, the output from collector is given transistor Q5 through the diode D1.

The diode D1 is providing positive half cycle to the transistors. The diode D2 used as a polarity diode. The negative voltage is extended from emitter to collector to the relay. When there is no input signals the transistor 1-4 are not conducting so transistor Q5 also will not operate so negative voltage extended to the relay. One end of the relay coil is directly connected to the positive terminal and the other end is connected to the collector of transistor Q5 that is providing negative voltage to the relay for activation. When there is no input signal the transistor Q1 to Q5 are in forward bias and the transistor Q5 will produce negative voltage from emitter to collector the and the relay gets activated .

When the IR sensor receives the input signal from the transmitter, the received signal is amplified by the transistor Q1 to Q4 is amplified then fed to the base of transistor Q5 through a diode which de- operating the transistor Q5 so there is no negative voltage from collector. So the relay gets no negative signal and it is in ideal condition (de activated).

Transmitting of Dual tone Multi-frequency

The DTMF IC NO. 91214 generate various types of 16 tones which are modulated in modulator section after amplification by the RF section is radiated by the antenna. Transistor BC 494 are used as RF amplifier and modulator.

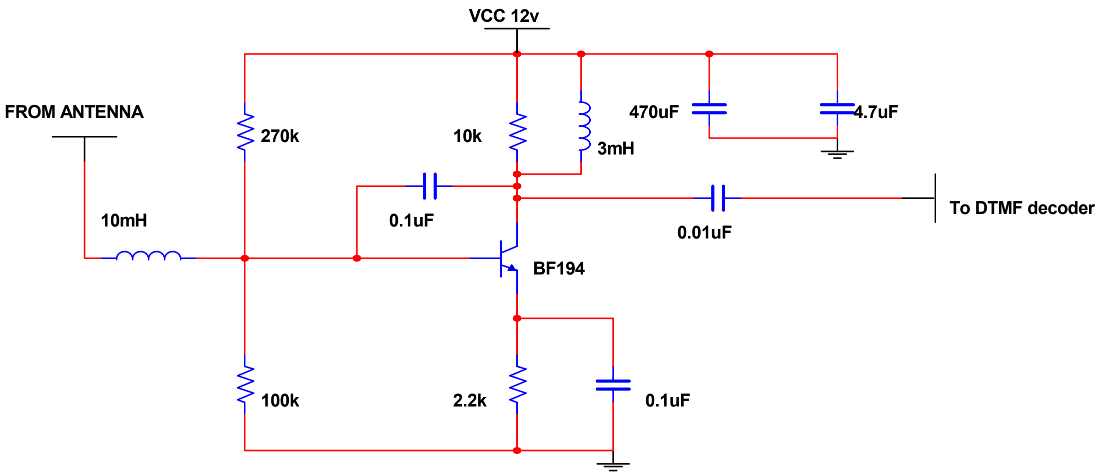


Figure: 4. Circuit Diagram of FM Receiver

FM receiver receives the FM modulated signal from the antenna and it is modulated by the transistor BF194. Then it is transmitted by the antenna to the DTMF decoder for further operations.

The received FM signal (27 MHz) is de coded by the DTMF de coder is send to the micro controller then appropriate driver section for activate the motor. IC 8870 is used as DTMF de coder and all twelve transistors are working as driver to drive the DC motor. Here we use pre amplifier and Darlington amplifiers to give more amplified version of signal at the relay input.

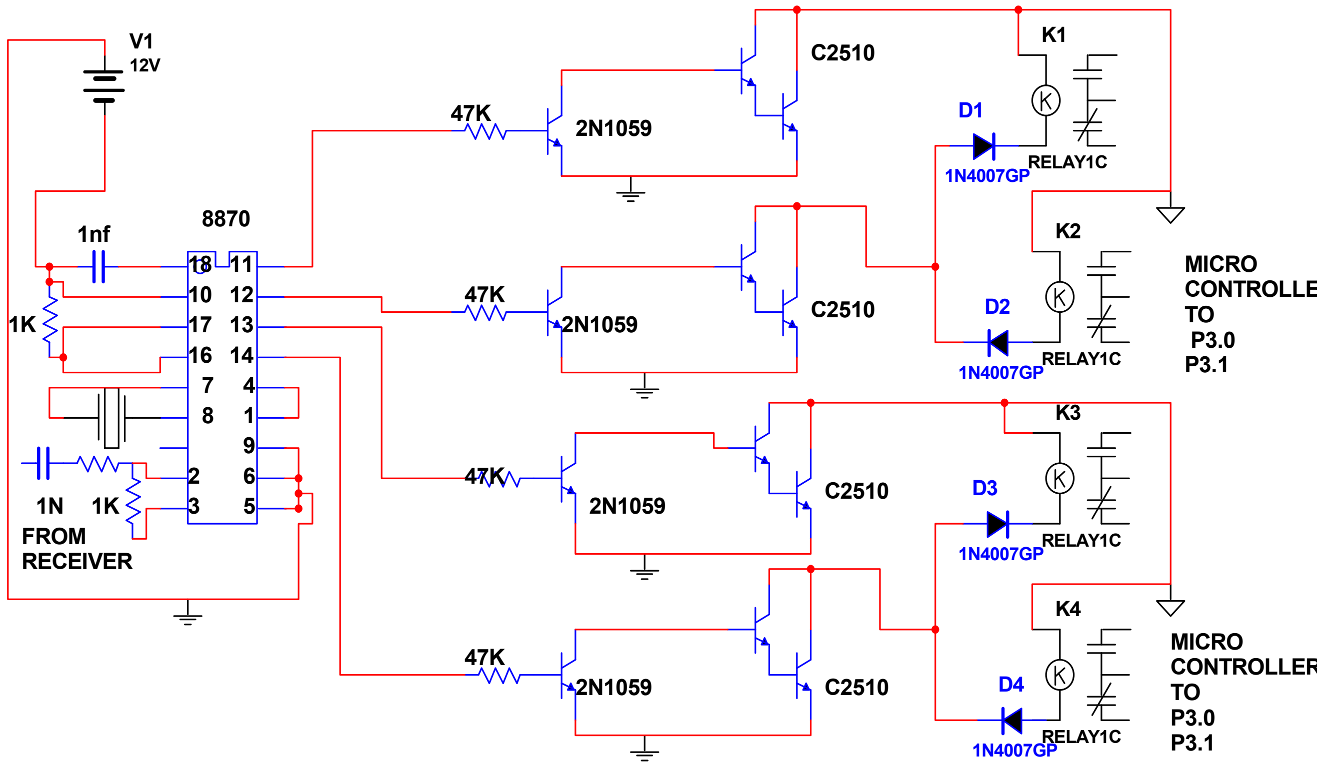


Figure.5: Circuit Diagram of DTMF DECODER

It is not easy detect and recognize DTMF with satisfactory precision Often, dedicated Integrated circuit is used, although a functional solution for DTMF transmission and receiving by a microprocessor (a PIC in most cases) exists. It is rather complicated, so it is used only marginally. Most often, a MT 8870 or compatible circuit would be used. Most decoders detect only the rising edges of the sine waves .So, DTMF generated by rectangular pulses and RC filter works reliably. The mentioned MT 8870 uses two 6^{th} order-band pass filters with switched capacitors. These produce perfect sine waves even from distorted input, with any harmonics suppressed.

6. RESULTS AND DISCUSSION

As a result, a well-organized system will exist so that there will not have any kind of malfunctions during landing or takeoff. The air traffic controller could authorize the incoming flights for its landing at the initial stage, and could guide the pilot for the safe landing of the aircraft. And, also would guide the passengers who are arriving with their luggage and also with the departing passengers. Future work will continue on many levels. The automation in our prototype will be improved to address the conflict detection/resolution deficiencies uncovered in this study, and to provide the additional functionality requested by the controllers. Research, technology, and procedure development will continue to improve the function allocation between air and ground and automation and controllers. Mixed equipage operations and off-nominal situations will also be studied.

7. CONCLUSION

In this paper, we are presenting a system that is known as Air traffic control System (ATCS) for our client from different parts of the world. To reduce the manual efforts and human errors, we need to have some kind of automated system monitoring all the parameters and functioning of the connections between the pilot and the airport personnel. In this research work we implemented a prototype model of an airport automation system within the limited available source and economy. The system can be subjected to further development using advanced techniques.

REFERENCES

[1] *Aiman Javed, Javadpur and Nilanjan Patra, Kolkatta Electrical Engineering Department,Published in India Conference (INDICON) 2016 ,IEEE Annual.*

[2] *Aakash Desai Department of Computer Engineering, Vidyalankar Institute of Technology, Mumbai, India published in Recent Trends in Electronics Information and Communication Technology (RTIECT), IEEE International Conference on May 2016.*

[3] *Kai-Quan Cai School of Electronics and Information Engineering, Beihang University, published in IEEE Transactions on intelligent Transportation Systems on March 2017.*

[4] *R. Jain, F. Templin, and K.-S. Yin, "Analysis of L-band digital aeronautical communication systems: L-DACS1 and L-DACS2" in Proc. IEEE Aerospace. Conf., Mar. 2011.*

[5] *Allendoerfer, K. R., Mogford, R. H., & Galushka, J. J. (1999). Comparison of the Plan View Display and Display System Replacement system baselines.*

[6] *Cardosi, K. M., & Murphy, E. D. (1995). Human factors in the design and evaluation of air traffic control systems (DOT/FAA/RD-95/3). Washington, DC: Federal Aviation Administration Office of Aviation Research.*

[7] *Federal Aviation Administration. (1996). Project configuration management guidelines. Unpublished manuscript. Atlantic City International Airport, NJ: DOT/FAA William J. Hughes Technical Center.*

[8] *Federal Aviation Administration. (1997). Human factors job aid. Washington, DC: Office of Chief Scientific and Technical Advisory for Human Factors.*

[9] *Galushka, J., Frederick, J., Mogford, R., & Krois, P. (1995). Plan View Display baseline research report (DOT/FAA/CT-TN95/45). Atlantic City International Airport, NJ: DOT/FAA Technical Center*

[10] *Vossen T, Michael B. Optimization and mediated bartering models for ground delay programs. Naval Research Logistics, 2006.*

[11] *Dankan V Gowda, D. V. Kishore, Shivashankar, A. C. Ramachandra, and C. Pandurangappa, "Optimization of motorcycle pitch with non linear control," in Proceedings of the 1st IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT '16), pp. 1656–1660, Bangalore, India, May 2016.*

[12] *Gowda, Dankan V., and Sadashiva Chakrasali., " Comparative Analysis of Passive and Semi-Active Suspension System for Quarter Car Model using PID Controller," Int. Conf. on Recent Trends in Signal Processing, Image Processing and VLSI (ICrtSIV), Banre, India, 2014*

[13] *Dankan Gowda V., Ramachandra A C., "Slip Ratio Control of Anti-Lock Braking System with Bang-Bang Controller" International Journal of Computer Techniques ,Volume 4 Issue 1, Jan – Feb 2017.*

[14] *Kishore, D.V., Gowda, D.V., Shivashankar, Mehta, S.: MANET topology for disaster management using wireless sensor network. In: International Conference on Communication and Signal Processing, April 6–8, 2016, India.*

[15] *Mahaveer Penna, Dankan V Gowda, J J Jijesh and Shivashankar, "Design and Implementation of Automatic Medicine Dispensing Machine", 2nd IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT), pp. 1962-1966, 2017.*

[16] *Gowda, D.V. and Ramachandra, A.C. (2018). Importance of non-linear controller in implementing anti-lock braking system-a technical review. International Journal of Advanced Research in Computer Science, 9(2), 193-199.*

[17] *D. Gowda, M. Penna, S. Palle, KA, S.P and K.N. Chaithra, "Digital Intercommunication System in Advanced Light Helicopter (ALH)", Advances in Computational Sciences and Technology, vol. 10, no. 5, pp. 699-708, 2017.*

[18] *Dankan V Gowda,Shivashankar,Varun Reddy S,Varun C A, Sahana M and T Rajesh,"Implementation of swarm intelligence in obstacle avoidance," Recent Trends in Electronics, Information & Communication Technology (RTEICT) in 2017.*

[19] *Gowda, D.V., Pandurangappa, C., Kishore, D.V.; and Ramachandra, A.C. (2016). Two wheeler vehicle model development for driving simulator application. GIT Journal of Engineering and Technology, 9(1), 47-51.*

[20] *Dankan V Gowda, Ramachandra A C, Thippeswamy M N, Pandurangappa C, Ramesh Naidu P 'Modeling and Performance Evaluation of Anti-lock Braking System', Journal of Engineering Science and Technology, Taylor's University, Vol 14(5), Pages: 3028-3045. Indexed:Scopus. 2019. ISSN: 1823-4690.*

[21] *Dankan V Gowda, Ramachandra A C, Thippeswamy M N, Pandurangappa C, Ramesh Naidu P 'Synthesis and Modelling of Antilock Braking System using Sliding Mode Controller', Journal of Advanced Research in Dynamical and Control Systems', Vol 10(12),Pages:208-221. Indexed:Scopus. 2018.*

[22] Dankan V Gowda, Ramachandra A C, Thippeswamy M N, Pandurangappa C, Ramesh Naidu P 'Automotive braking system simulations V diagram approach', International Journal of Engineering & Technology (Scopus Indexed)', Vol 7(3),Pages:17-40.Indexed:Scopus.2018.ISSN: 2227-524X.

[23] S. B. Sridhara and B. Ramesh, "An adaptive multicast routing for varied services in wireless cellular ad hoc network", 2015 IEEE International Conference on Computational Intelligence and computing Research, 2015,pp.1-5

[24] S. B. Sridhara and B. Ramesh, "Energy efficient device to device multicast routing for wireless cellular ad hoc network", 2015 IEEE International Conference on Applied and Theoretical Computing and Communication Technology , 2015, pp.708-712 IEEE Part Number: CFP15D66-USB IEEE ISBN: 978-1-4673-9222-8. DOI: 10.1109/ICATCCT. 2015.7456975, Publisher: IEEE.

[25] S. B. Sridhara, B. Ramesh, " An Access Fairness Resource Provisioning of Services for Energy Efficiency in Wireless Cellular Ad hoc Network", International Journal of Computer Science and Information Security, Vol.14 No.6,2015,pp.735-747, IJCSIS ISSN 1947-5500, Pittsburgh, PA, USA.

[26] Sridhara S.B and Ramesh B, "Energy efficient QoS multicast routing for varied services in wireless cellular ad hoc network", International Journal of Engineering Technology and Computer Research, Vol. 3, Issue 6, 2015, pp.14-24.

[27] Sridhara S.B and Ramesh B, "QoS Provisioning for Energy Efficiency in Mobile Ad hoc Network", Global Journal of Computer Science and Technology, Vol.15, No 5-E,2015, pp.41-52. ISSN Numbers: Online: 0975-4172 Print: 0975-4350 DOI: 10.17406/GJCST. Publisher: Global Journals Inc. (USA)

[28] Sridhara S.B and B. Ramesh, "Enhancement Throughput of Ad hoc Network using DSDMAC Protocol", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Issue 1, 2014, pp.9521-9528.

[29] Sridhara S.B, "A Novel Hybrid Routing Based Energy Efficient Routing Protocols for Mobile Ad-Hoc Network to improve the performance and throughput". National Conference on Wireless Control & Communication Technologies, [NCWCT-2014], 24th – 25th April 2014, Dr. AIT, Bangalore

[30] Sridhara S.B., Ramesh M., Veeresh Patil, Adaptive Scheduling Design for Time Slotted Channel Hopping Enabled Mobile Adhoc Network, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No 3, 2020 Pages272-277 The Science and Information (SAI) Organization United Kingdom: 19 Bolling Road, Bradford, BD4 7BG, West Yorkshire, United Kingdom

ADV MATH SCI JOURNAL

Advances in Mathematics: Scientific Journal **9** (2020), no.7, 4485–4493
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.7.17 Spec. Issue on RDESTM-2020

DESIGN AND IMPLEMENTATION OF CRYPTCLOUD SYSTEM FOR SECURING FILES IN CLOUD

P. RAMESH NAIDU[1], N. GURUPRASAD, AND DANKAN GOWDA V

ABSTRACT. This research paper deals with an efficient method of securing files and folders in cloud using the method of cryptography. Some features such as Advanced Encryption System (AES) algorithm is used to convert plain text to cipher text and cipher text to plain text of the confidential data using one of the most promising and secured technique known as Cipher text-Policy Attribute-Based Encryption (CP-ABE). In this research paper, we will be analyzing the different ways of misusing credentials to access: one among is the cloud user and the other is on the semi-trusted network authorization side. To reduce risk, we recommend the CryptCloud which is secure and more liable trusted authority and able to cancel CPABE based cloud storage system with trace and audit the white-box. We demonstrate how to analyze security issues and further show the usage of our system through experiments, an enhanced AES cryptosystem to handle the failures decryption of the original AES. Then we will be developing a reliable and accurate scheme based on the enhanced AES and confidential distribution of key for each file. As per the data owner new access policy is specified, the stored cipher text will be updated directly in the cloud server without decryption and the update at the cloud will be able to verify by data owner. The proposed scheme can verify the distributed confidential information to avoid the users from Hacking and can reject various attacks such as the collusion attack.

[1]*corresponding author*

2010 *Mathematics Subject Classification.* 68P25, 94A60.

Key words and phrases. Advanced Encryption System, CryptCloud, CPABE, Public key Encryption.

1. Introduction

This paper is about a stepwise walk through to solve a problem of securing files and folders in cloud using cryptography.Some time data which is stored in the cloud server is sensitive that can be hacked because we cannot trust cloud server so prominently [1]. So that, the cipher text data will be stored in the cloud. When we are Updating the stored cipher text which is in the cloud when a data owner chosen a new policy to access data and user need authenticity to access data[2] and to verify data of a user to access cipher text data. The data owner and cloud user first login to the cloud using their credentials. Owner uploads a file in cloud [3] and user requests to that file. File gets uploaded in cloud assigning a unique key to each file when its uploaded. Key is generated using AES algorithm which is of 128 bits. The time of file uploading gets recorded in the database[4]. The user needs the key to access the file and it can be accessed for a certain time limit. User requests the access of file to the admin, admin either accepts the requests or it can be denied [5]. User can download the file and view it using a private key of the file. This paper comprises of 5 major sections [6], which will guide you step by step to the research and findings. These are: 1) Introduction 2) Research Elaborations 3) Results or Finding 4) Conclusions

2. Research Elaborations

This section of the paper involves with different strategies and approaches for encrypting and decrypting the generalized model of the cloud [7] architecture using symmetric key encryption system. The steps involved in the development [8] are as mentioned below which will be discussed in detailed. The list are mentioned below:

(1) Uploading and encryption of file
(2) Key Generation
(3) Decryption
(4) Request and Response

2.1. **Uploading and encryption of file.** As the owner uploads the file in cloud after login in, a unique private key is assigned to a file. Key generation and file gets encrypted [9] using Advanced Encryption System (AES) algorithm at the time of uploading a file. The generic approach for file uploading [10] and validating is described below using flow diagram. Encryption Using AES:

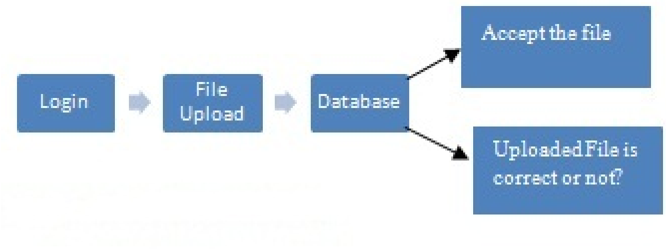


FIGURE 1. Uploading and validating

Cipherfun(byte in[4*Nbs], byte put[4*Nbs], word
wd[Nbs1*(Nrs+1)])
begin
byte stateone[4,Nb]
stateone = get
AddRKey(stateone, wd[0, Nbs-1])
for round = 1 step 1 to Nrs–1
SbBytes(stateone)
Rowshift(stateone)
Columnsmix(stateone)
AddRKey(stateone, w[round*Nbs, (round+1)*Nbs-1])
end for
SbBytes(stateone)
Rowshift(stateone)
AddRKey(stateone, wd[Nrs*Nbs, (Nrs+1)*Nbs-1])
out = state
end

Initially, decrypted data is taken as an input and send to state array element. After AddRKey(add round key), stateone array is converted to implement a function round of 10,12 or 14 times, with final little different [11] from first Nrs H1 rounds. The final Stateone output is then copied and displayed. The function round key is used in terms of parameterized key that consists of single array of

4 bytes of words will be generated by using Key generation technique. The individual transformations - SbBytes(), Rowshift(), Columnsmix(), and AddRKey() process the state array and are described as follow:

a. SbBytes()-SubBytes uses S-box which is applied [12] independently on each byte of 4x4 matrix. First Four bit represents the row number and next four bit represents column number .Size of S-box is 16x16.Get the corresponding value from the S-box and then covert it into 8 bits to substitute it into the block. Consider first four bit 0000 as row number which 0 and next four bit 0101 as column number which is 5.0th row and 5th column in s-box has value of 52.it is again converted in 8 bits. This will be stored in state array.
b. Rowshift()-ShiftRows function performs circular right shift operation of rows of state array. There will be no change in 0th row. Row 1 ,2,3 will be shifted 1bits,2bits,3bits respectively.]
c. Columnsmix()- Output of Rowshift() will be considered as input to MixColumns().Consider each word as one column and apply multiplication operation with predefined 4x4 matrix. The result will be one word which is again stored in state array.
d. AddRKey()-AddRoundKey function perform xor operation. Output from Columnsmix is xored With key .First column of output of columnmix xored with first column of the key and the resultant is stored in state array.

2.2. **Key Generation.** Key generation routine is applied on cipher key K in AES algorithm. Generate the words of Nbs (Nrss + 1) using key expansion. Initially there will be Nbs words requiring Nr rounds. The resultant will be array of 4 bytes words key schedule which is denoted as [wi],i covers range from 0 to Nb (Nrs + 1).Key Expansion algorithm consists of two functions Rotword() and Sbwords. Based on input of Rotwords function it performs cyclic permutation and produces an output. Subwords function performs S box on each 4 byte of input to produce an output word.

Key Expansion :

ExpansionKey(byte keys[4*Nks], word wd[Nbs*(Nrs+1)], Nks)

begin

word tempval

```
i = 0
while (i < Nks)
wd[i] = word(key[4*j], key[4*j+1], key[4*j+2], key[4*j+3])
j = j+1
end while
j= Nks
while (j < Nbs * (Nrs+1)]
tempval = wd[j-1]
if (j mod Nks = 0)
tempval = SubWord(RotWord(tempval)) xor Rcon[i/Nks]
else if (Nks > 6 and i mod Nks = 4)
tempval = SubWord(tempval)
end if
wd[j] = wd[j-Nks] xor tempval
j = j + 1
end while
end
```

2.3. **Decryption or Inverse cipher.** The transformation of the encryption can be reversed and then compiled in the back order to create the inverted Cipher of the AES algorithm. The Inverse Cipher functions used in are - InverseShiftRows (), InverseSubBytes (), InverseMixColumns () and AddRKey ().

Inverse Cipher:

```
InverseCipher(byte in[4*Nbs], byte put[4*Nbs],
word wd[Nbs*(Nrs+1)])
begin
byte stateone[4,Nbs]
stateone= get
AddRKey(stateone, wd[Nrs*Nbs, (Nrs+1)*Nbs-1])
for round = Nrs-1 step -1 goto 1
InverseShiftRows(stateone)
InverseSubBytes(stateone)
AddRKey(stateone, wd[round*Nbs, (round+1)*Nbs-1])
InverseMixColumns(stateone)
end for
```

InverseShiftRows(stateone)
InverseSubBytes(stateone)
AddRKey(stateone, wd[0, Nbs-1])
put = stateone
end

The individual transformations - InverseSubBytes(), InverseShiftRows(), InverseMix-Columns(), and Inverse of AddRKey() process the state array and are described as follow:

a. InverseShiftRows()-function performs circular inverse right shift operation of rows of state array. There will be no change in 0th row. Row 1 ,2,3 will be shifted 1bits,2bits,3bits respectively.
b. InverseSubBytes()-InverseSubBytes uses InverseS-box which is added independently on each byte of 4x4 matrix of state array.
c. InverseMixColumns()-It is inverse of MixColumns() .InverseMixColumns () Follow column column by column, treat four-term polynomial as a each column.
d. Inverse of AddRoundKey()-Inverse of AddRoundKey function perform Inverse xor operation. Output from InverseMixColumns is xored With key .First column of output of InverseMixcolumn xored with first column of the key and the resultant is stored in state array .

2.4. **Request and Response.** The general architectural design or the workflow of request and response of the user and admin. User requests for accessing the file and admin sends the response towards it.

3. Results And Findings

Each file is getting encrypted and a key is assigned to every file through which user can access the file. We have used the time scheduling algorithm, so the user can download the file only during the specified time period, and if the user fails to download the file within the time period then the pass-code expires and the user again needs to send the request to the admin and again the admin needs to accept the request.

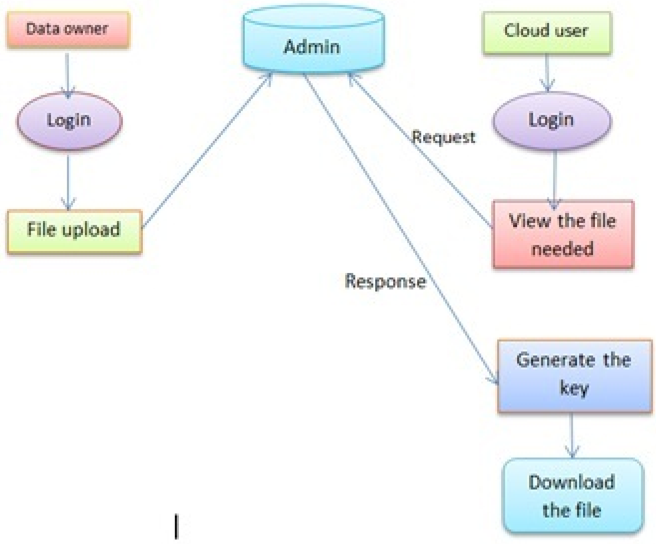


FIGURE 2. The request and response

FIGURE 3. Users request accepted by the admin

4. CONCLUSION

When the data is being stored by the admin in the cloud there is always an issue if the data is accessed securely by the user or not. There are number of exterior threats that can lead to data leaks, including harmful deceptions from cloud service providers or cloud user accounts. The best way is to encode files which can be achieved with use of hybrid cryptography. Hybrid cryptography with the help of AES algorithm provides security to the files stored in the cloud and make it less vulnerable to threats. It can be seen that AES is the most secure

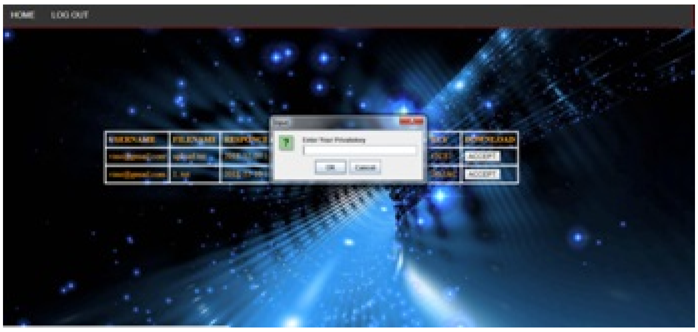


FIGURE 4. Passkey entered by the user to access the file

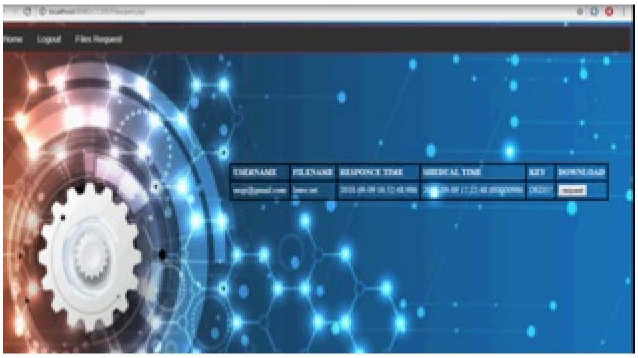


FIGURE 5. File downloaded after entering the passkey

symmetric encryption technology. User can access files from cloud only if they have the key which is sent by the admin after secure authorization. It guarantees a method for safe cloud environment.

Future work:

The main aim for future is to implement code through which user can get to know if there request has been accepted or not. We can extend our scope by including backup and recovery actions to help prevent data loss in the event of an attack.

REFERENCES

[1] M. ALI, R. DHAMOTHARAN: *Sedasc: Secure data sharing in clouds.* , IEEE Systems Journal, **1**(2) (2017), 395–404.

[2] M. ARMBRUST, A. FOX: *Security in cloud computing: Opportunities and challenges*, Information Science and Technology Journal, **2** (3) (2015), 357–383.

[3] M. ALI , S. U. KHAN: *A view of cloud computing*, Communications of the ACM Journal, **53** (4) (2010), 50–58.

[4] N. ATTRAPADUNG, H. IMAI: *Attribute-based encryption supporting direct/indirect revocation modes*, Journal of Cryptography and Coding, **3** (14) (2009), 278–300.

[5] D. V. KISHORE, D. V. GOWDA : *MANET topology for disaster management using wireless sensor network*, 2016 International Conference on Communication and Signal Processing (ICCSP), (2016), 0736–0740.

[6] D. V. GOWDA, D. V. KISHORE : *Optimization of motorcycle pitch with non linear control*, 2016 International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT), (2016), 1656–1660.

[7] M. PENNA, D. V. GOWDA : *Design and implementation of automatic medicine dispensing machine*, 2017 International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT), (2017), 1962–1966.

[8] D. V. GOWDA, C. A. VARUN : *Implementation of swarm intelligence in obstacle avoidance*, 2017 International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT), (2017), 525–528.

[9] D. V. GOWDA, A. C. RAMACHANRDA: *Synthesis and Modeling of Antilock Braking System Using Sliding Mode Controller*, Journal of Advanced Research in Dynamical and Control Systems, **11**(12) (2018), 208–221.

[10] D. V. GOWDA, A. C. RAMACHANRDA: *Modelling and Performance Evaluation Of Anti-Lock Braking System*, Journal of Engineering Science and Technology, **14**(5) (2019), 3028–3045.

[11] D. BONEH, X. BOYEN: *Short signatures without random oracles* , Journal of EUROCRYPT, **2**(1) (2004), 56–73.

[12] H. CAI, B. XU: *Iot-based big data storage systems in cloud computing*, Perspectives and challenges. IEEE Internet of Things Journal, **4**(1) (2017), 75–87.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY, VISVESHWARAYA TECHNOLOGICAL UNIVERSITY, BANGALORE, KARNATAKA STATE, INDIA

Email address: `ramesh.naidu@nmit.ac.in`

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, GLOBAL ACADEMY OF TECHNOLOGY, VISVESHWARAYA TECHNOLOGICAL UNIVERSITY, BANGALORE, KARNATAKA STATE, INDIA

Email address: `nguruprasad18@gmail.com`

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING, B.M.S INSTITUTE OF TECHNOLOGY AND MANAGEMENT, VISVESHWARAYA TECHNOLOGICAL UNIVERSITY, ADDRESS: BANGALORE, KARNATAKA STATE, INDIA

Email address: `dankan.v@bmsit.in`

Smart Moto Helmet

Shilpa Hiremath[1], Thyagaraj Tanjavur[1],

Akkiraju Subbaram Harshitha[2], Bindu H P[2], Girish D.S[2], Dhanya V Nair[2]

[1]Assistant Professors, Department of ECE, BMS Institute of Technology and Management, Bangalore-560064, India.
shilpasharankh@bmsit.in, thyagaraj_tanjavur@bmsit.in
[2]UG Students, Department of ECE, BMS Institute of Technology and Management, Bangalore-560064, India.
harshiakkiraju@gmail.com, binduhp4@gmail.com, girishds1997@gmail.com, dhvnair@gmail.com

ABSTRACT

Safety on a two-wheeler is the main reason why we came up with this smart motohelmet idea. The smart motohelmet would help any citizen follow the rules of the road and thus lead to safe and easy driving for everyone. In our research findings, two-wheelers account for 25% of total road crash deaths in India- the top five reasons to this includes drunken driving and avoiding safety gears like helmets or seat belts. As engineers, it is our duty to implement a design that would solve a societal concern, is cost-effective without compromising on safety. The need for this project in our country is high, as annually there is 10% increase on the population that buys a new two wheeler, which caters the need of a helmet and due to its safety features people opt for this smart motohelmet. The benefitted audience would be all riders of two-wheelers: motorbikes, scooter.

KEYWORDS: Smart motohelmet, bike, force sensing resistor, GSM, GPS, accelerometer, Zigbee, MQ3, ignition control

I. INTRODUCTION

Motorcycles are one of the most widely used vehicles in India. Hence it has a major contribution to the accidents and road crash deaths. This was the birth of the idea to make a Smart Motohelmet which will help in practicing safe riding, not just to keep the rider safe but also other fellow riders.

Hence, our motive is to make a helmet that is easy to use by all age groups with a cost-effective design by making the set-up on a general PCB to benefit the users with continuous checks on alcohol consumption & helmet being worn to control the ignition of bike, Fall/Accident detection with immediate alert system to the rider's emergency contacts and with night driving visibility for the rider. The arrangement of the circuitry would be embedded inside the helmet which include the general PCB and few other sensors placed in various locations of the helmet. This solution is very much required to prevent and reduce such accidents in a country wherein the purchase of a new two-wheeler increases with a large percentage every year.

II.EXISTING METHODS AND LIMITATIONS

The research performed by our team concludes that the existing system ensures safety of the rider but has comprised on features offered, as cost was an issue and the implementation of the project is not embedded into a helmet through a less interconnected circuit such as using wire wrap board.

Features on the helmet provided are limited to use of only alcohol detection, only accident detection. The compactness of the system is not investigated as there is no use of wire wrap board to prevent interconnections. The existing technology focuses on four-wheeler vehicle's safety and is limited in perspective on terms of reach and cost. The cost of a smart helmet available goes as high as $2000.The size of the system is not considered as key.

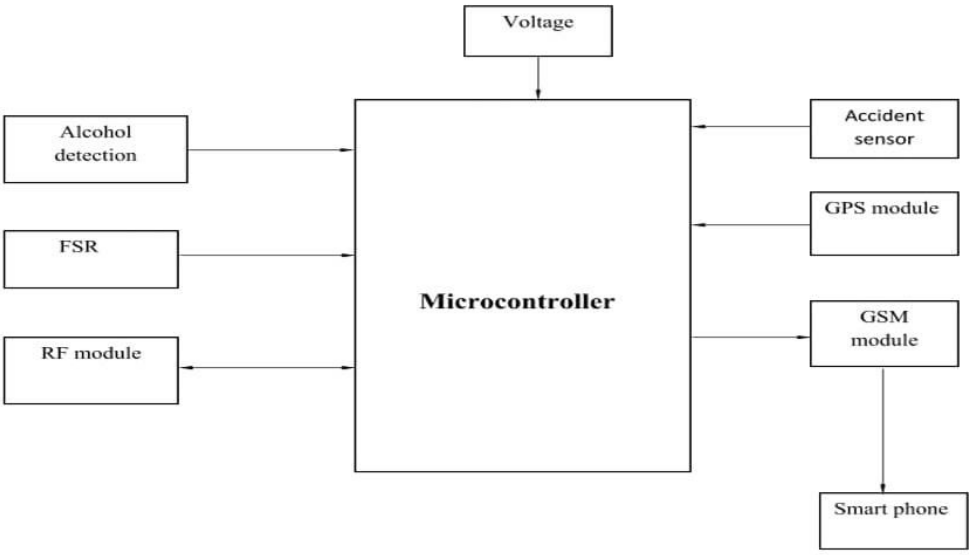
III. PROPOSED SYSTEM

A. Overview of the features of helmet

The features included in the project are:

Night driving visibilities is used to make the rider's helmet thoroughly visible to rear fellow travellers as LED panels are created to indicate the left and right indicators.
Helmet is worn checks are performed before switching ON the bike's ignition by means of a FSR embedded in the core of helmet, with this data we determine whether to ignite the bike or not.
Accident sensor A GPS module sends the location of the rider via text message sent from a GSM module when an accident has occurred which is sensed using this sensor.
Alcohol consumption detector MQ-3 sensors present in the helmet are used to detect alcohol levels of the rider - if alcohol levels are found to be high, the ignition of the bike is not switched ON.
The sensors, modules are placed inside the helmet and padding is done such that the rider will not feel uncomfortable.

The block diagram of the proposed system is as follows:

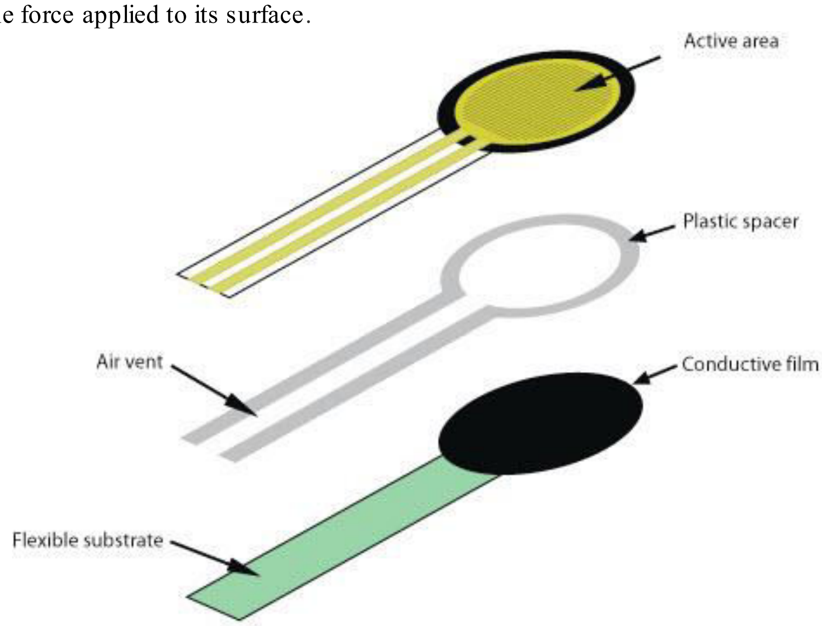


B. The features in detail

We are using Arduino Nano board in both the helmet and the bike sections. The reason for this is its small size, it's compatibility with all the sensors, modules used and arduino boards are easy to work with.

Two wheeler riders avoid the use of helmet even though it is made compulsory in almost all the cities in India. We came up with the idea of using the basic principles of force/pressure and resistance in order to check if the rider has his helmet on and thus switching on the ignition [3], [8], [9]. The alcohol detector and the helmet worn detector work in conjunction with the Zigbee module to control the ignition.

Resistors are passive components that are widely used in electrical and electronics circuits. A force sensing resistor (FSR) is a material whose resistance changes due to application of force, pressure or mechanical stress. It is sometimes known as force-sensitive-resistor. The FSR sensors are made of conductive polymer which has a property of changing its resistance based on the force applied to its surface.



If force is applied to a surface of sensing film, then the particles touches the conducting electrodes and thus resistance of the film changes. There are several resistive based sensors but force sensing resistors operate satisfactorily in difficult environments and also require a simple interface compared to other resistive based sensors.

Even though there various types of force sensors, the force sensing resistors are having several advantages such as thin size (less than 0.5mm), very low cost and also good shock resistance. The only disadvantage of FSR sensors is low precision; there will be approximately 10% or more difference in measurement results.

The problem of drunk and driving cases can be eliminated to a great extent with the use of the alcohol detector present in the helmet with the help of which the alcohol levels of the rider will be constantly monitored and ignition will be controlled accordingly [1]. We are using a simple MQ3 sensor for this detection which has a threshold value set. It is a low cost semiconductor sensor which can detect the presence of alcohol gases at concentrations from 0.05mg/L to 10mg/L. the sensitive material used for this sensor is SnO_2, whose conductivity increases as the concentration of alcohol gases increases. It provides both analog and digital outputs. It is suitable for detecting alcohol concentration on breath just like the common breathalyzer and has a high sensitivity and a fast response time. It can operate between 10°C-70°C. It is fixed near the buckling portion of the helmet so that the levels of alcohol can be determined clearly.

If the amount of alcohol is lower than the threshold value then the bike starts else the transmitter Zigbee module is used to turn off the ignition by communicating the message to the receiver Zigbee module present in the bike. Zigbee is an IEEE 802.15.4 based specification for a suite of high level communication protocols used to create personal area networks. It is a low-power, low data rate, and close proximity wireless ad hoc network. The technology defined by the Zigbee specification is intended to be simpler and less expensive than other wireless personal area networks (WPANs), such as Bluetooth or Wi-Fi.

Accidents during night times can be avoided with the help of the LED lights present on the top of the helmet which help displaying the turns of the bike. An accelerometer present near the handle of the bike is used to detect the amount of tilt and then indicate the left and right directions. ADXL337 is a low-power 3-axis accelerometer which operates on a supply in the range of 1.8V to 3.6V and can be used in tilt-sensing applications. It is capable of measuring acceleration with a minimum full-scale range of +-3g. It is capable of measuring static acceleration due to gravity. We have checked the amount of tilt by placing the accelerometer in various directions and set threshold values for both the directions.

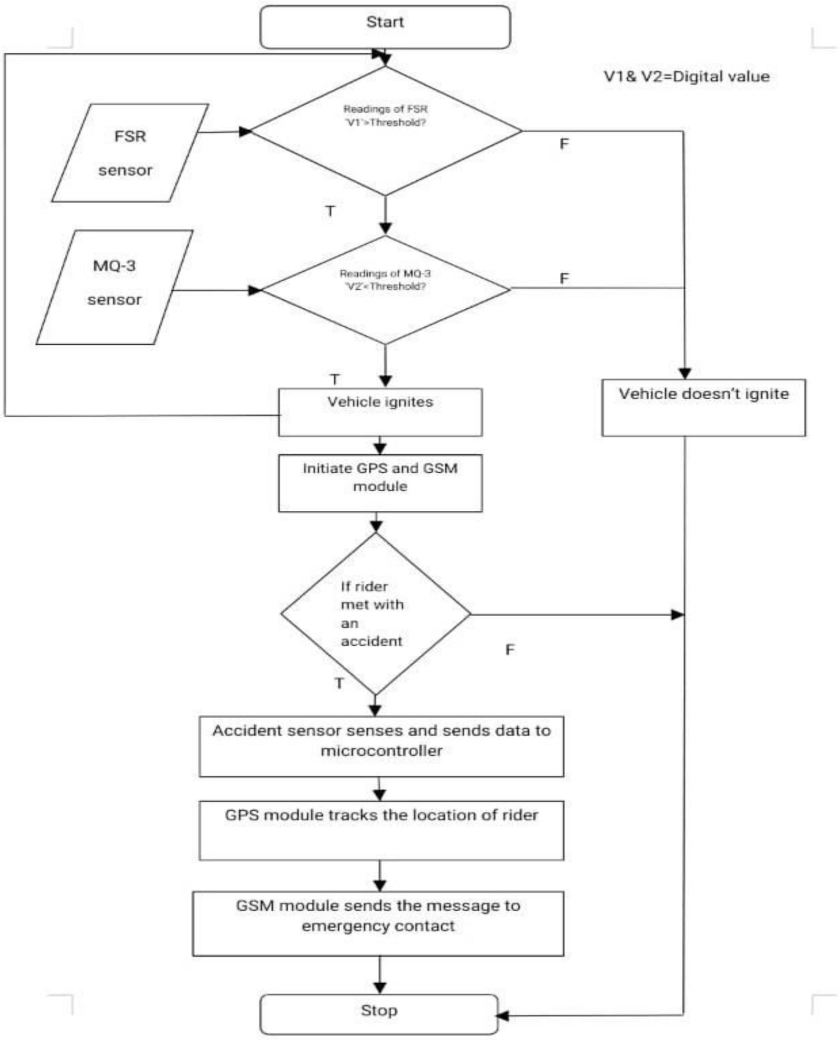
The reason for not reporting the accidents in time is one of reasons of deaths [10]. Another accelerometer is used inside the helmet and an accident condition is recorded if there's a sudden change in the threshold value of the accelerometer during the ride [5]. Using the GPS and GSM modules in the helmet the rider can be rescued by sending the location of accident to his emergency contacts which is already fed in the module. [2], [4], [6], [7]

Due to the wide availability of GSM it is chosen as the medium for transfer of location information. The simple and inexpensive short message service (SMS) allows users to send up to 160 characters. SIM800L is a miniature cellular module which allows for GPRS transmission, sending and receiving SMS and making and receiving voice calls. After connecting power module boots up, searches for cellular network and login automatically. On board LED displays connection state (no network coverage - fast blinking, logged in - slow blinking). It has a 3.7-4.2V voltage requirement.

GPS receivers use a constellation of satellites and ground stations to compute position and time almost anywhere on earth. At any given time, there are at least 24 active satellites orbiting over 12,000 miles above earth. The positions of the satellites are constructed in a way that the sky above your location will always contain at most 12 satellites. The primary purpose of the 12 visible satellites is to transmit information back to earth over radio frequency (ranging from 1.1 to 1.5 GHz). With this information and some math, a ground based receiver or GPS module can calculate its position and time**.** NEO-6M GPS module is a well-performing complete GPS receiver with a built-in 25x25x4mm ceramic antenna, which provides a strong satellite search capability. With the power and signal indicators, we can monitor the status of the module. The module can save the data when the main power is shut down accidentally due to the presence of data backup battery. It requires 3-5V of supply and the default baud rate is 9600bps. We are using 3.7V 1000mAH Li-Po battery to power up this module. We are using a boost regulator to use the same battery but supply power to all other modules in the helmet as they require 5V. The batteries are rechargeable and this comes as an added advantage.

Wire wrap board is used for housing components to make a circuit for compactness, simplicity of servicing and case of interconnection which is later embedded inside the helmet.

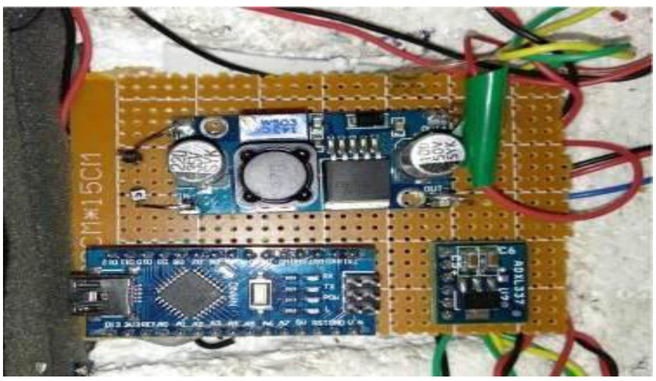
Flow chart for the proposed system:



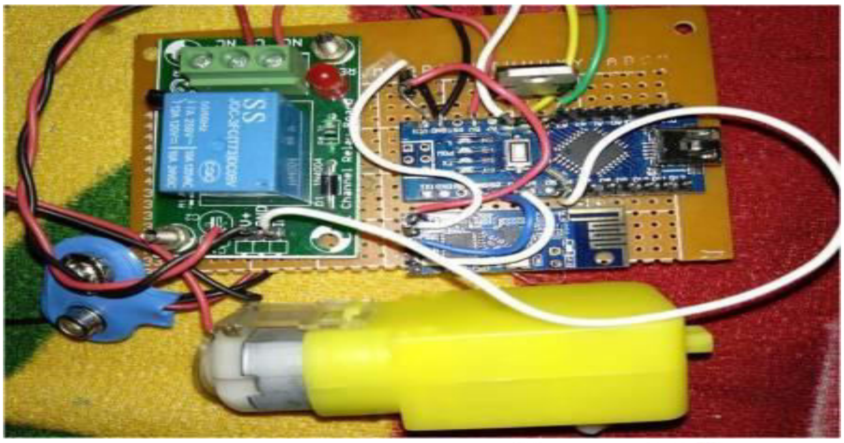
IV.RESULTS AND DISCUSSION

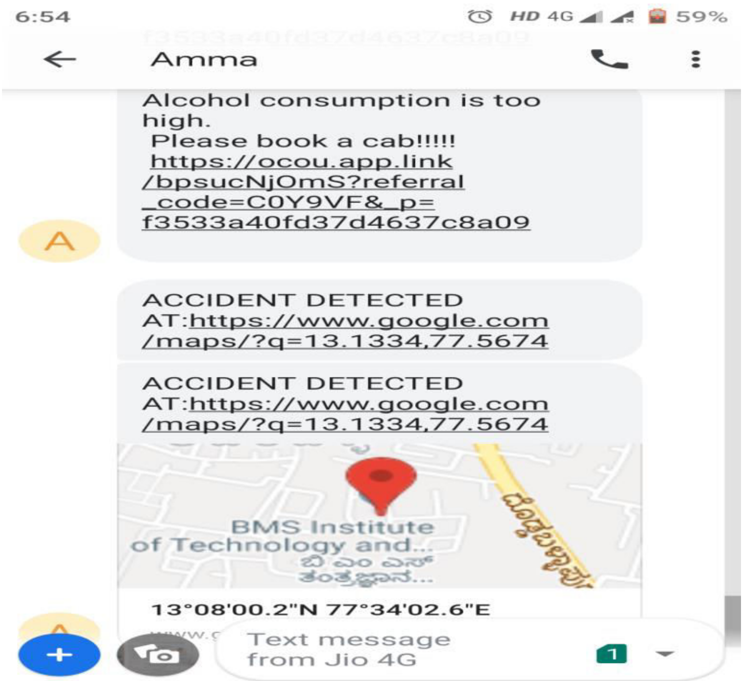
Space optimization is achieved as we selected the smallest available modules and made sure they are fit at various places inside the helmet and are soldered to the wire wrap board. We were able to check the ignition and the accident conditions.

General PCB:



Zigbee receiver side (controls ignition):





V.CONCLUSION AND FUTURE WORK

In the paper we talked about the smart motohelmet prototype. A system for the helmet has thus been developed which, through communication between a module in the helmet and one on the vehicle ensures safety aspects. The accident feature can be extended to sending the location to the nearby hospitals and police station or calling the ambulance. We can keep sending the location in particular time intervals until the emergency contact acknowledges until the acknowledgment is received. We can make use of the solar panels for powering up the modules in the helmet. Another alternative could be the use of wind energy. We can also add another helmet worn check by using a switch near the buckle of helmet that warns the rider if the helmet is not buckled up. We can implement various bioelectric sensors on the helmet to measure various activities. We can use small camera for the recording the driver’s activity. We can add a heat sink or a fan to reduce the temperature. An air bag system can also be incorporated for additional safety purpose.

ACKNOWLEDMENT

We students would like to express our heartfelt gratitude towards our guides Prof. Shilpa Hiremath and Prof. Thyagaraj Tanjavur of ECE department, BMS Institute of Technology and Management, Bangalore. We would love to thank the other professors from our department who encouraged us to continue our work. On a personal note, we would love to thank our parents, family and friends whose support has been so valuable.

REFERENCES

[1] Sayan Tapadar , Shinjini Ray , Himadri Nath Saha , Arnab Kumar Saha , Robin Karlose, "Accident and alcohol detection in bluetooth enabled smart helmets for motorbikes", IEEE, 2018.

[2] Neha Mangle, G Shivananda, Aishwarya Kashyap, "A GPS-GSM predicated vehicle tracking systemmonitored in a mobile app based on Google maps", IEEE, 2017.

[3] Durga K Prasad Gudavalli, Bh. Sudha Rani, C. Vidya Sagar "Helmet operated smart E-bike", IEEE, 2017.

[4] P karthik, B Muthu Kumar, K Suresh, "Design and implementation of helmet to track the accident zone and recovery using GPS and GSM", IEEE, 2016.

[5] Sreenithy Chandran, Sneha Chandrashekar, N Edna Elizabeth, Konnect: "An Internet of things based Smart helmet for accident detection and notification", IEEE, 2016.

[6] Sagar Patil, Medhini Ganesh Hegde, Saikt Bhattacharjee, "Smart Motorcycle security system", IEEE, 2016.

[7] Pradip V Mistar, R H Chile, Real time Vehicle tracking systembased on ARM7 GPS and GSM technology, IEEE, 2015.

[8] Amitava Das, Soumitra Goswami, Priti Das, Design and implementation of intelligent helmet to prevent bike accident in India, IEEE, 2015.

[9] Mohd Khairul Afiq Mohd Rasli, Nina Korlina Madzhi, Juliana Johari, "Smart helmet with sensors for accident prevention", IEEE, 2013.

[10] IEEE Maker Project Smart helmet Maker: Sandeep- https://transmitter.ieee.org/makerproject/view/4b317

Automation of Object Sorting Based on Colour

[1]Shashikala.J ,
[1]Asst. Professor, BMS Institute of Technology & Management, Bengaluru, India
Email: [1]shashikala_j@bmsit.in

Abstract: *Automation or automatic control, is the use of various control systems for operating equipment such as machinery, processes in factories, boilers and heat treating ovens, etcThe paper presents sorting of object using microcontroller based on color.*

Keywords: *Automation, sorting, colour sensors, microcontrollers*

1. INTRODUCTION

Automation has been achieved by various means including mechanical, hydraulic, pneumatic, electrical, electronic devices and computers, usually in combination. The benefits of automation [1] include labor savings, savings in electricity costs, savings in material costs, and improvements to quality, accuracy and precision

In the existing system the objects are sorted manually mostly by human beings. This creates a tendency for human errors to come into account and thus result in the work going wrong. If objects or parts in industries are not sorted correctly, then there is a high chance of huge chaos and the final product being defective.

2. REVIEW OF LITERATURE

For many packaging industries, color object counting and sorting is the main job that needs to be completed. Traditionally, the item sorting process was performed by the manually. However, this method has some disadvantages such as increase in the cost of the product, slow, and inaccuracy due to the human mistake.

Nowadays, the competition is so intense that the efficiency of the product is regarded as the key to success. The efficiency of the product includes the speed of the production, lowering material and labour cost, improving quality and decreasing the rejection. Taking into account all the issues this project is being built which is very useful for the industries. This project is aimed at obtaining fully automatic material handling system. This is achieved using microcontroller machine. This device synchronizes the robotic arm 's movement to select the moveable objects on the conveyor belt.

3. PROPOSED SYSTEM

An automated sorting system has primary task of sorting components by number. It also consists of conveyor belt which reduces material handling efforts. These procedures also take place concurrently, viz. handling and testing of materials [2]. The figure below shows the block diagram for the proposed program.

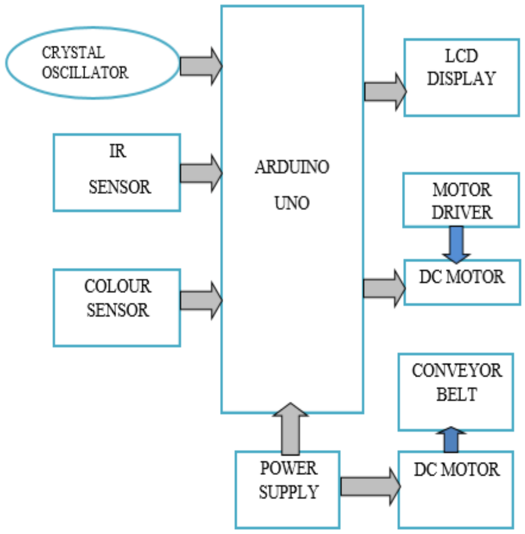


Figure 1: **Block diagram for object sorting**

As the DC motor (12V, 3.5rpm) receives a 3.4V supply it starts spinning. It will monitor the conveyor belt movement on which the product is placed. When an object exists, the sensor generates an output frequency that is proportional to the object's color Therefore, when red filter is chosen, the sensor gives maximum frequency for red objects, and similarly other colored objects[3] are also sensed by corresponding filters.

As the light falls on the object the color sensor is reflected back. As stated earlier, the TCS2300 color sensor has 4 color filters for green, red, blue and black (no colour), which is chosen by its selected pins. The software stored inside the microcontroller selects filters Frequency output from the color sensor depends on the object's color, as well as the microcontroller's select pin configuration information. Select pin can choose one of the four photo diode[4] filters that can provide output according to the object's color.

If an object exists, the sensor generates an output frequency proportional to the color of the object and the configuration of the selected photo diode in such a way that the respective photo diode is given maximum frequency for the respective color. Therefore, when red filter is chosen, the sensor gives maximum frequency for red objects, and similarly other colored objects are also sensed by corresponding filters. Frequency obtained during each filter selection is counted[5] and stored in separate registers and these values are analyzed to use the larger one to describe the object's color.

The second and third DC motors are used to control the gateway of the particular color object. If placed object is red, the color sensor sense the color of the object and it sends the signal to the microcontroller and display the name RED ITEM with the help of the LCD Display. In our project, red object go straight ,blue drop the left side and green object drop the right side with the help of the gateways. The products will finally fall to the corresponding sections in the container.

Then the IR sensor sense how many objects moved on the conveyer belt and it sends the signal to the microcontroller. At the last displays the total number of object moved on the conveyer belt with the help of the LCD display unit. Once the power up dc motor starts, moving the conveyer belt so that the objects on the conveyer belts also starts moving. Color sensor starts sensing the objects which are moving on conveyer belt. Then the sensed signal sends to the display unit via Arduino for displaying the color of the object.

Figure 2: Conveyer belt

For this project the Arduino UN0 microcontroller is used to monitor motion and detect objects. The microcontroller operates on collection of preprogrammed and stored instructions in the memory. This then takes the instructions and one by one from its programme.

4. RESULT & DISCUSION

The paper presents design, development of the sorting object. Using microcontroller Arduino UN0 the objects are sorted as per the color. The color detection is identified by the color sensor. The sensed signal is sent to a microcontroller unit and sent to 16x2 LCD display for displaying the color of the sensed object. Here IR sensor is used for counting the number of objects which are moved on the conveyor belt that count value is also displayed on the LCD.

In this, objects of 3 colors red, blue and green are chosen for demonstration purpose. The system output is displayed on the LCD display i.e name of the project - automatic color sorting, color of the object- red or blue or green and final count of the objects – item number which are sorted.

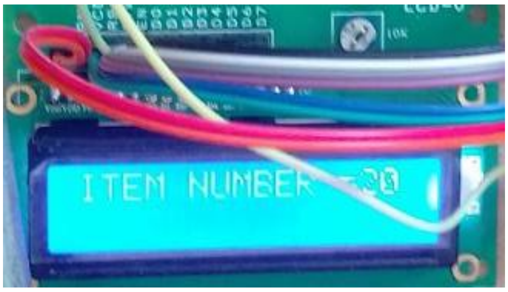


Figure 3: LCD Display

Above figure shows final result displayed on the LCD display of count value of the objects which are passed on the conveyor belt.

5. CONCLUSION

It is very useful with the help of color sensor and conveyor belt in large varieties of industries, particularly in the packaging sector. Automatic sorting machine[6] increases operator performance, practicality and health. This guarantees exceptional processing power and peerless efficiency including color detection. We will of course add high speed DC motors and sensors.

object sorting system using embedded system are an embedded system approach for object color detection and object sorting can be successfully implemented. Due to use of embedded system for color determination, manual efforts are reduced which produces result in improving accuracy as well as saves money and time.

REFERENCES

[1] Amitesha Sachdeva, Mahesh Gupta, Manish Pandey, Prabham Khandelwal. Development Of Industrial Automatic Multi Colour Sorting and Counting Machine Using Arduino Nano Microcontroller and TCS3200 Colour Sensor.

[2] Dhanoj M, Reshma K V, Sheeba V, Marymol P (2015), Colour sensor based object sorting robot using embedded system..

[3] Kunhimohammed C. K, Muhammed Saifudeen K. K, Sahna S, Gokul M. Sand Shaeez Usman Abdulla.(2015), Automatic Color Sorting Machine Using TCS230 Color Sensor And PIC Microcontrolle.

[4] Ali joy (2014) , Object Sorting Robotic Arm Based on Colour Sensing, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 3, pp-7741-7746.

[5]Vishnu r. Kale, V. A. Kulkarni,(2014), Automation of object sorting system using pick & Place Robotic arm & image processing , Proceedings of 3rd IRAJ International Conference, Mumbai, India. ISBN: 978-93-82702-51-1.

[6] Vindhya Devalla, Dr . Rajesh Singh, Amit Kumar Mondal, VivekKaundal ,(2012), Design and Development of Object Recognition and Sorting Robot for Material Handling in Packaging and Logistic Industries, International Journal of Science and Advanced Technology (ISSN 2221-8386) Volume 2 No 9 , pp 30-35.

ISSN 2321 3361 © 2020 IJESC

Research Article Volume 10 Issue No.7

IoT Based Smart Security Home Automation System using BLYNK Framework

Dr. Surekha R. Gondkar[1], Nayana[2], Ashwini P. S.[3], Kavyashree[4], Niveditha M.[5]
Associate Professor[1]
Department of Electronics and Communication Engineering
BMS Institute of Technology and Management, Doddaballapura, Bangalore, India

Abstract:
The Internet of Things moved beyond the trends and became an autonomous technical and social relevance. As per the Gartner Research, the number of connected things will have reach 26.66 billion by 2019 and by 2025 this number will have increased up to 75 billion. Right at the moment, people can feel the effect of the iot in their lives. Smart phones, smart bulbs, smart appliances and even smart security systems — all that smart stuff which can operate without human interaction. This project discuss the design circuit to control devices which we are connected to circuit and also monitor the status of electrical devices using local server and BLYNK framework app. The method creating a prototype consisting of sensors and actuators it may be humidity sensor, motion sensor, fan, bulb etc. Nodemcu is an open source IOT platform. It includes esp8266 express system. It based on esp-12 module it is used to controlfan, lights, temperature sensor and can use many more controls.

Keywords: IoT, ardiuno, nodemcu, wi-fi, smart phone.

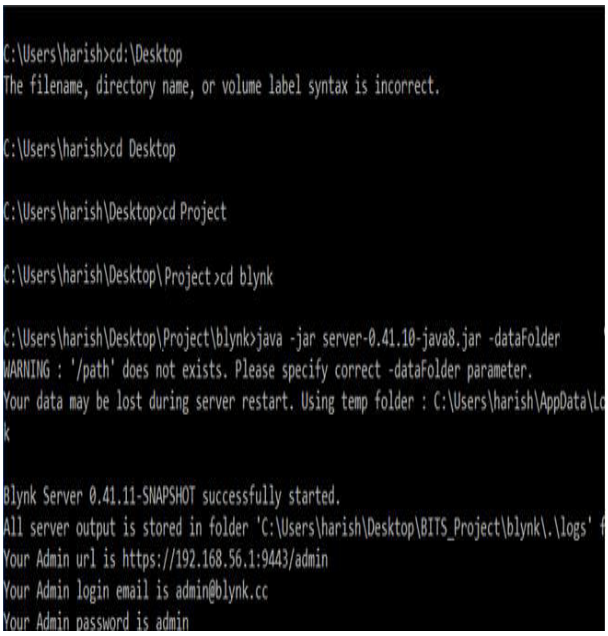
I. INTRODUCTION

Home automation refers to handling the electronic devices and controlling some home appliances' using computer technology. Now days automation is very popular it provides ease security and efficiency. Sensors sense the status of appliances and update the local server. If person away from the home he can change the status using blynk app which is installed in user app. this project will approach the controlling the appliances using local sever. This IOT based project smart home automation system trying to achieve comfort combined with simplicity. the currently built prototype of the system sends alerts to the email using internet. in other hand if owner identifies the that person entering his house. and user can make arrangements such as opening the door, switching on various appliances inside the home. IOT is one of the new technology that will have the biggest influence on our global and urban design. Large part of the home/apartment automation system including- lighting, temperature controller, power management, security management can be controlled by using local IOT server which provides scalability, low-cost maintainance and less down time.

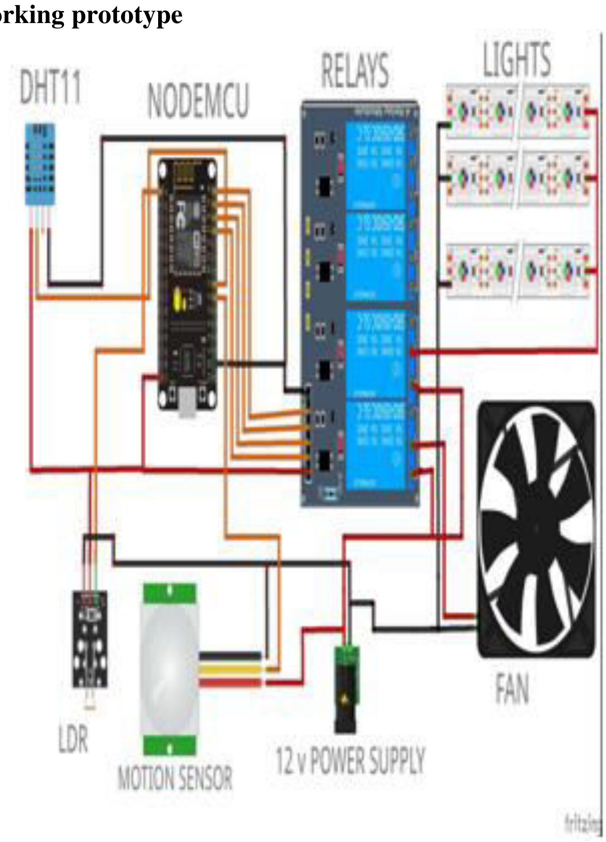
II. PROPOSED SYSTEM

Proposed system enables secure local IOT server with an IOT App which will control the IOT devices locally in home/ apartment complex. The user can control the lights, fans, AC, security alarms and can read Temperature, humidity based on the need using the BLYNK IOT app installed on their Android Phones. The local BLYNK IOT server can be deployed on Raspberry Pie or on a laptop which acts as local cloud for all control / monitoring of the devices. All these are controlled by using a Local server without Internet. The control speed is good as everything happening locally. Proposed system is implemented using BLYNK sever.

Creating Local IOT Server using BLYNK: BLYNK is a platform with IOS and Android apps to control Arduino, ESP8266, Raspberry Pi and the likes over the Internet. We can easily build graphic interfaces for all your projects by simply dragging and dropping widgets. BLYNK Server is an Open-Source **Natty** based Java server, responsible for forwarding messages between BLYNK mobile applications and various microcontroller boards and SBCs (i.e. Arduino, Raspberry Pi. etc.). Installation of the BLYNK server on our laptop using the following command as per the screenshot below



Working prototype

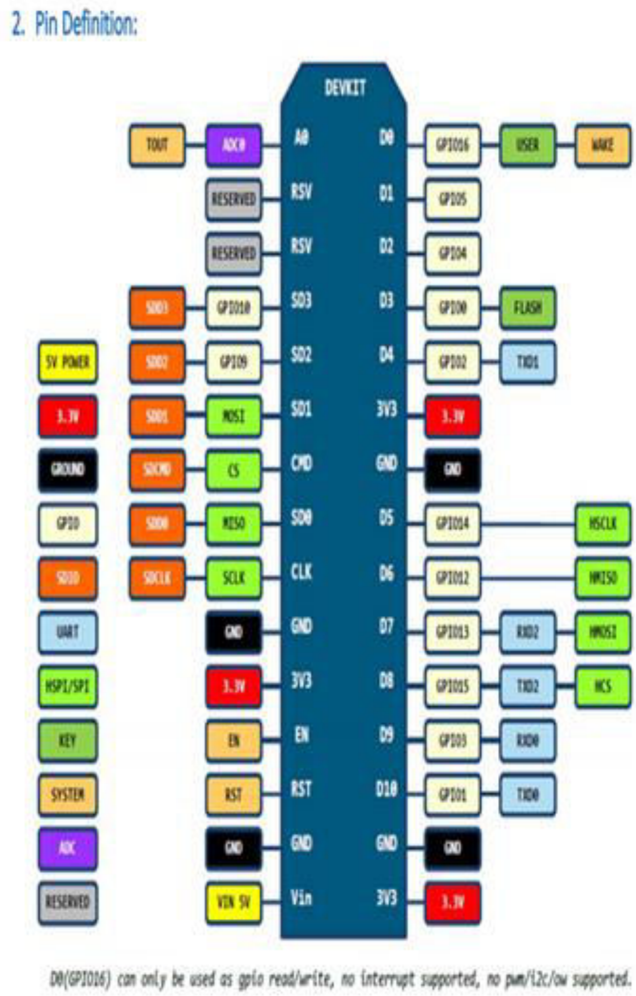


Components required

Arduino Uno, Nodemcu
Dht11 sensor
Fan
Light
Relay 2 channel
Motion sensor
LDR sensor
Power board

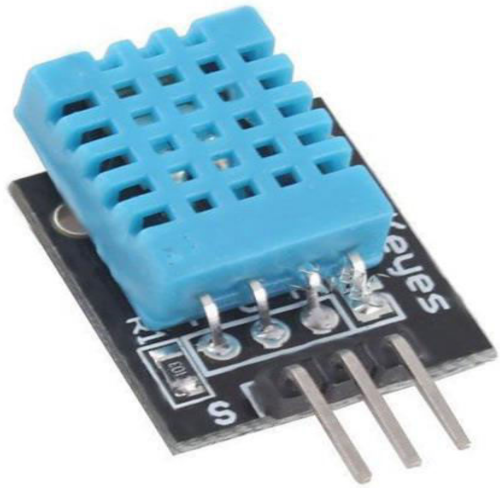
Arduino: this is first and foremost is freely access open-source chip. This Arduino Community refers to the project and user community that chip utilizes microcontroller-based microchip boards. These boards are known as Arduino Modules, those are open-source prototyping platform. The simplified microcontroller board comes in a variety of development packages

NODEMCU



use of ESP8266 chip will used as serial commands. This chip will basically a Wi-Fi/Serial transceiver. However, that is not convenient. that is prescribed to using the very cool Arduino ESP8266 project. Esp8266 is a modified version of esp8266. the Arduino IDE that will need to install your computer. This helps very convenient way of using the ESP8266. now we are ready to using the well-known Arduino IDE. Following the below step to install ESP8266 library to work in Arduino IDE environment.

DHT11 SENSOR



The DHT11 is a basic ultra low-cost digital temperature and humidity sensor. This sensor can be easily interfaced with any microcontroller. Such as arduino and raspberry pi to measure temperature and humidity simultaneously.Dht11 humidity sensor and a temperature is available as a sensor and a module. The difference between this and a module is pull up resistor and a power on led. Dht11 sensor temperature and humidity sensor measure in surrounding air.

Windows PC/Laptop: Minimum of 2GB RAM and 80GB of HDD

Implementation

Arduino Uno IDE is used for developing the programs on embedded systems like Arduino Uno boards, Node MCU boards. The Arduino Uno is an open-source microcontroller board based on the Microchip ATmega328P microcontroller and developed by Arduino.cc. The board is equipped with sets of digital and analog input/output pins that may be interfaced to various expansion boards and other circuits.

ADVANTAGE:

1)Security concern:

It is not enough to know how to protect your online privacy, manufacturers will also have to take steps to ensure users' data security also

2)Fast growth of IOT:

Smart manufacturing is gaining traction and IOT technology uses it for its benefit. The top drivers of the rapid growth are improved operational efficiency, productivity, optimization of asset utilization and reduction in downtime.

3)Developing 5g networks to impact use:

The mobile networks that outclass the cable ones in speed, connection stability and availability open a great prospect to the growth of the IoT projects. Ideas such as "smart cities /smart apartments" where all civic amenities are united into one chain and collecting, analyzing and transmitting data to each other in order to create a more efficient system

4)Edge computing:

Edge computing is the type of architecture where the data stores locally, in micro-centers for processing. In other words, instead of mindlessly sending data to the cloud storage, its primary processing will be carried out in the processors of the devices themselves. A lot of data collected by IoT things and sent to the cloud turns out to be useless. Take an outdoor home security camera, it normally transfers terabytes of data to the server, while perhaps a few megabytes of these records that demonstrate suspicious activity may prove useful in the end.

III. RESULT

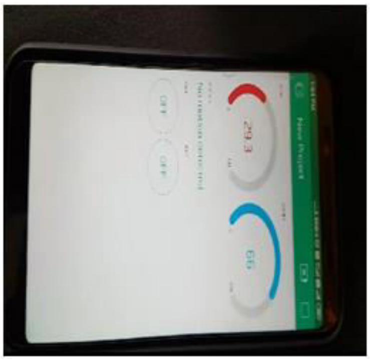


Figure.1. Shows the blynk app output. In the we seen controls electrical devices using on or off using blynk.

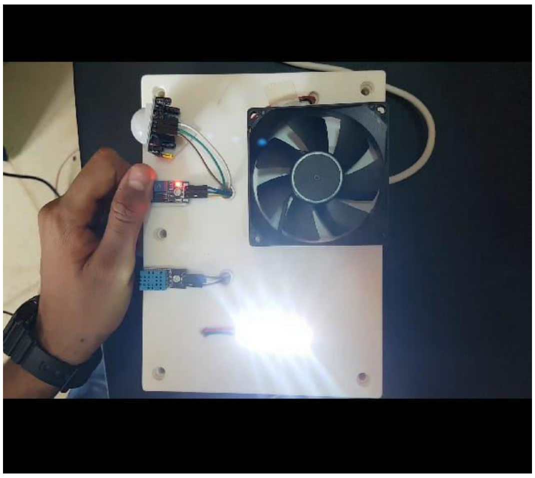


Figure.2. We are controlling output led by using led sensor.

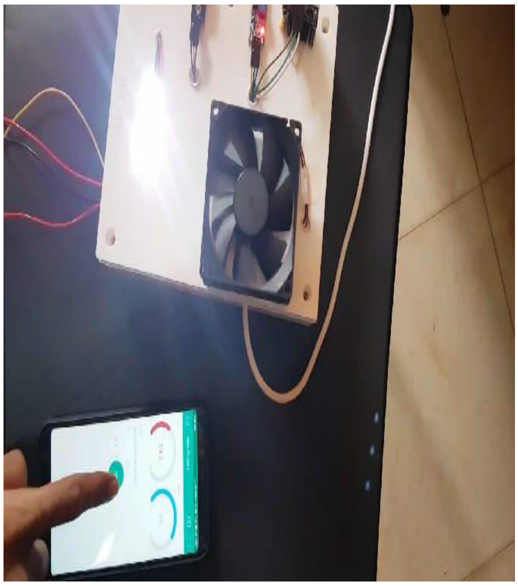


Figure.3. We are controlling electrical bulb using blynk app by making on and off.

IV. CONCLUSION:

In this paper, we are creating prototype to control the home appliances such as electrical bulb, fan, relay in the fig1, fig2, fig3. The design of a control system for home automation has been presented. This paper mainly focuses on the use of WIFI to minimize installation, and the designed prototype can be applied to real-time control of home appliances to monitor &remotely control them.

V. REFFRENCE:

[1]. Satpathy, L. Smart Housing: Technology to Aid Aging in Place. New Opportunities and Challenges. Master's Thesis, Mississippi State University, Starkville, MS, USA, 2006.

[2]. M. Chan, D. Esteve, C. Escriba and E. Campo, ―A review ―A review of smart homes―Present state and future challenges,Computer Methods and Programs inBiomedicine, vol. 91, no. 1, pp. 55-81, Jul 2008.

[3]. M. G. Golzar, H.Tajozzakerin, ―A N A New Intelligent Remote Control System for Home Automation and Reduce Energy Consumption, 2010 Fourth Asia International Conference on Mathematical /Analytical Modeling and Computer Simulation, pp.174- 180, 2010.

[4].A.E.Nikolaidis, S.S.Papastefanos, G.I.Stassinopoulos,M.-P.D rakos, and G.A.Doumenis, —Automating remote configuration mechanisms for home devices, IEEE Trans. on Consumer Electronics,vol.52,no. 2, pp. 407-413, May. 2006.

[5].A. Kamilaris, V. Trifa, and A. Pitsillides, —Homeweb: An application framework for—Homeweb: An application framework for web- based smart homes, 2011 18th International Conference in Telecommunications (ICT), pp. 134 –139, May 2011.

[6.] K. Bing, L. Fu, Y. Zhuo, and L. Yanlei, —Design of an internet of things-based smart home system,, 2011 2nd International Conference in Intelligent Control and Informa tion Processing (ICICIP), vol. 2, pp. 921 –924, July 2011.

[7].Internet of Things (IoT): BLYNK Framework for Smart Home Ermi Media's, Syufrijal, and Muhammad Rif'an 2018.

ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **9** (2020), no.7, 4405–4414
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.7.11 Spec. Issue on RDESTM-2020

INTERNET OF THINGS: INTERNET REVOLUTION, IMPACT, TECHNOLOGY ROAD MAP AND FEATURES.

V. DANKAN GOWDA [1], S. B. SRIDHARA, K. B. NAVEEN, M. RAMESHA, AND G. NAVEENA PAI

ABSTRACT. Internet of Things (IoT) also called as internet of everything, that everything includes human being as well so that becomes an Internet of everything. IoT is one of the greatest phenomena of this century. It has gained a lot of popularity due its emerging growth. In this Paper we have made an attempt to describe an overview, Impact of Internet of Things and also present an IoT importance in the Current scenario and also presented the Impact of Internet of Things and then extended to Internet Revolution, Typical Views of IoT, Need of IoT , Various Benefits of IoT and Technology road map of an IoT, Features of IoT.

1. INTRODUCTION

Upcoming and latest Technology IoT, Internet of Things, almost all every area, every device, every sensor, every software are connected to each other and those devices we can able to access remotely through a smart phone or through computer is called Internet of Things. It is a rocking feature, Internet of Things So, the Figure 1. Here which is given below, it is showing some kind of sense. Almost all it is connected to every device, so these devices can be accessed remotely.

The some of the very popular words, in software industry which are the smart phone, it is an electronic device acting like a computer which is connected to

[1] *corresponding author*

2010 *Mathematics Subject Classification.* 68T09.
Key words and phrases. Internet revolution, Impact,Big data, Artificial Intelligence.

internet and another popular term is smart homes that is the home which is connected to all devices connected to each other and these devices can accessed through a smart phone [1]. similarly some more popular terms are smart city, smart village. In industrial automation using Artificial intelligence, may be hundred members work may going to do with Robot, So that is a big challenging area called Robotics technology.

Now a days Every where we are getting this term is called IOT and in fact why IOT so important, first imagine there was a time in a way back when we had a phone and we had a machine and we cannot connect our phone with any of our machines right, because technology was not grown to the level of inventing connectivity software s like as Bluetooth or Wi-Fi.

In the Later years, time has arrived, here we can connect our desktop with your mobile phone or maybe we can connect your laptop with a mobile phone so that we can control our mobile phones operation [2] from laptop or can control laptop operations from mobile phone right because they are connected each other.

Imagine your home if everything is connected start from your tube light, AC,refrigerator, maybe desktop,electronic shoes so everything will be connected and if we want to connect all these devices we require a connection. To establish the connection a network between them is used. now we have a technology where we can have multiple devices connected to each other example we can control AC temperature with a mobile phone.

with the help of Infrared sensors now, which is not just about sensors it is about let's say we want to control our AC. say we are not at home we are in at our workplace and want to control our home AC temperature so we can even do that with mobile phone, if it is connected with AC [3]. connection between them requires Internet that needs unique IP address that means phone knows about the IP address of AC, AC knows about the IP address of refrigerator.

For example, if we are going inside house and suddenly if we open the door ultimately that the light should be ON and once we go out the light should be OFF how to do that again we can use your IOT. Now all these various sensors are connected together to generate huge amount of data and all this data is stored on this server and again that is the reason to handle this huge amount of generated data, we are stepping into the new technology called Big Data. Technology specifies that every device needs an a unique IP address and that is why we require ipv6. when we have ipv4 now we are working with ipv6 so everything will be connected with Internet now. We know IOT is a booming technology in today's world. so first of all Internet of Things from the name itself will come to know that it is an internet or the interconnection of different things now this thing's essentially are devices or like nonliving things that can be categorize it. The Internet of Things is the categorization or classification where we interconnect this nonliving things say just devices which we clubbed together and communicate to each other by establishing the protocols between them.

FIGURE 1. IoT Definition

2. INTERNET REVOLUTION

An internet Revolution, the father of Internet Vint Cerf and the meaning of Internet is basically an international network or collection of TCP/IP communications Protocols. In geeks it was very basic level information after that internet of masses that is heavy things, everything online entry completely available here onwards, social media slowly entered, video communications, perfect digital media entered, and then technology introduced a smart phones or mobile internet, Figure 2. gives an brief idea about Revolution of IoT. In the Mobile communication, the first generation (1G), it was completely analog communication and supported for analog data transmission. In second generation (2G) supported for digital communication and digital data transmission,third generation (3G) multimedia services came into picture and fourth generation (4G) explored wireless that is we can say broad band connectivity, wireless broad band connectivity and this what named after long term evolution. In fifth generation (5G) is a international mobile telecommunication and the upcoming technology that is sixth generation (6G) is completely on air fibered technologies. The fourth generation (4G) onwards there is new term we are familiarizing that is called internet of things.

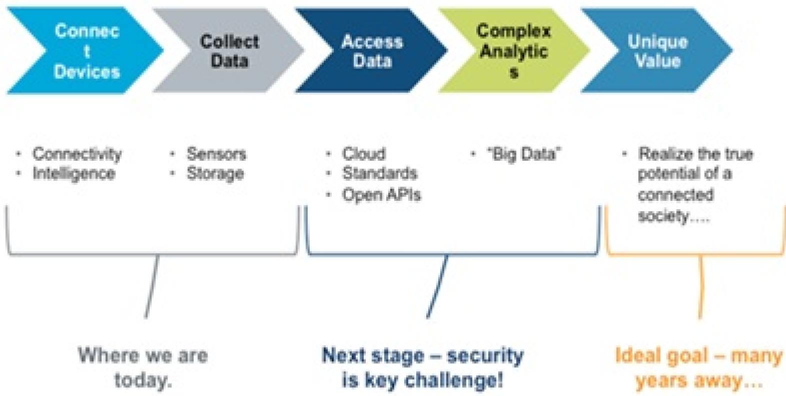


FIGURE 2. Revolution of IoT

3. IMPACT OF IOT

The year 2012 and onwards, the impact of the internet, education, business, communication, entertainment, health or medical becoming most popular. Now a days observe education in online become very popular. The e-learning is very demand wherever the trainer and trainee are going to meet so we can share our resources each other. The Business in almost all the developed countries [4], say 70 to 90 percent prefer to do only through online mode and is so popularly called e-commerce. Figure 3 shows the impact of IoT technology. Communication now a days our communication are clear, comfort, compatible. These communications is possible because of internet. Perfect video conference, audio chatting became most common to use in our daily day to day life. Entertainment, So whenever we require games or some kind of video applications, a many gaming applications available in the smart phones. The Medical or Healthcare area drastically developed including developing in countries also. The definition is already provided, smart health is it is an application platform for electronic health record system.

The data related to health care system are virtually maintained, every precautionary actions also listed in smart record or e record.For example doctor is in other area say in US or in deli or in Chennai or in UK and patient in particular place but still patient can take the suggestions.

Another Important application area of IoT Technology is the smart city, it is an vision develop urban to integrate information and communication technology.

FIGURE 3. Impact of IoT

4. TECHNOLOGY ROAD MAP OF IOT

In the business or the industry of an organization uses some of the tools for robotics, process automation and many other things. The human intervention involved in this now moreover we essentially call it as the intelligence of things where the device has that intelligence in itself that when it has to communicate with the other suitable technologies around itself so that is basically the intelligence or the artificial intelligence comes into picture in IOT. The term IOT was very first point in the year of 1999 and the progenitor of this term was Kevin Ashton so he is the person who was first coined the term IOT. The Technology road map that is for the IoT [5], we have at the time axis on the X plane and technology which on the y axis that is shown in Figure.5. it is an increasing graph now at the base level you have the RFID tags that the radio frequency identification tags for facilitating the routing inventory and loss prevention now this as part of application, we will apply it to the dairy and product industry there for example say we have a cattle in that cattle farm we have different kinds of breeds of like animals like cows, buffaloes which are milk Brady now for example same if we made an RFID chip and put on top of this animals whether on the horns [6] or at the body of these animals then we will have the entire control about the health of that animal the entire dashboard of that cattle form will be available so we will be able to know at which point of time at which breed is being suffering from what kind of disease and what is the health status. The complete information for that particular animal so that is where the RFID tags from the example basis it considered and depicted in the Figure 4.

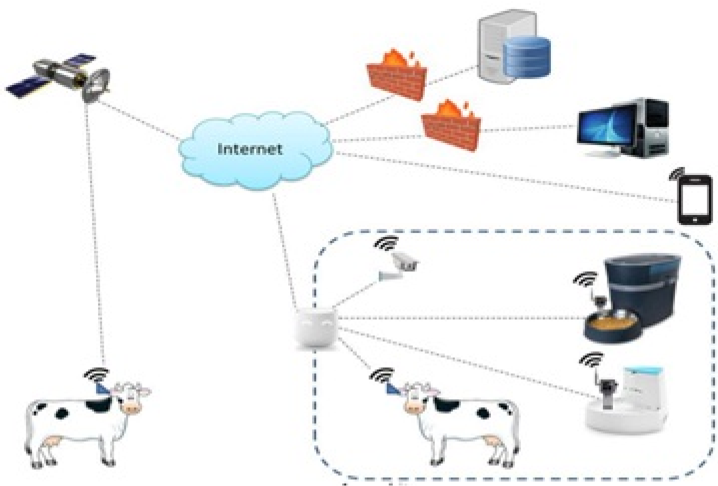


FIGURE 4. Animal tracking and Caring using IoT

The surveillance and health care becomes a vertical marketing applications like for the we have CCTV cameras and all steps included installed in our societies or any organizations or even the healthcare whenever we go so these will provide real-time data of all the persons in and out movement of this places so this can provide a vital role where the internet-of-things can come into picture and can communicate with the necessary persons or the authorities which they want to communicate and the cost reduction [7] in this level is very much reduced leading to diffusion into second wave of application that will lead to locating people and everyday objects suppose if some person or a suspicious person has been called in surveillance camera and then you have to locate it now in the current day it's not very easily possible to identify which person or which like or a theft or robbery if something have happened in a bank then it is very difficult to identify that person or such gang so with IOT we can locate and monitor real-time activity of person. It is just a road map like where we have reached or where we have taken the IOT from 2015 to 2020 so initially it was started in the year 1999 but then to its presence and existence came in the year 2015.

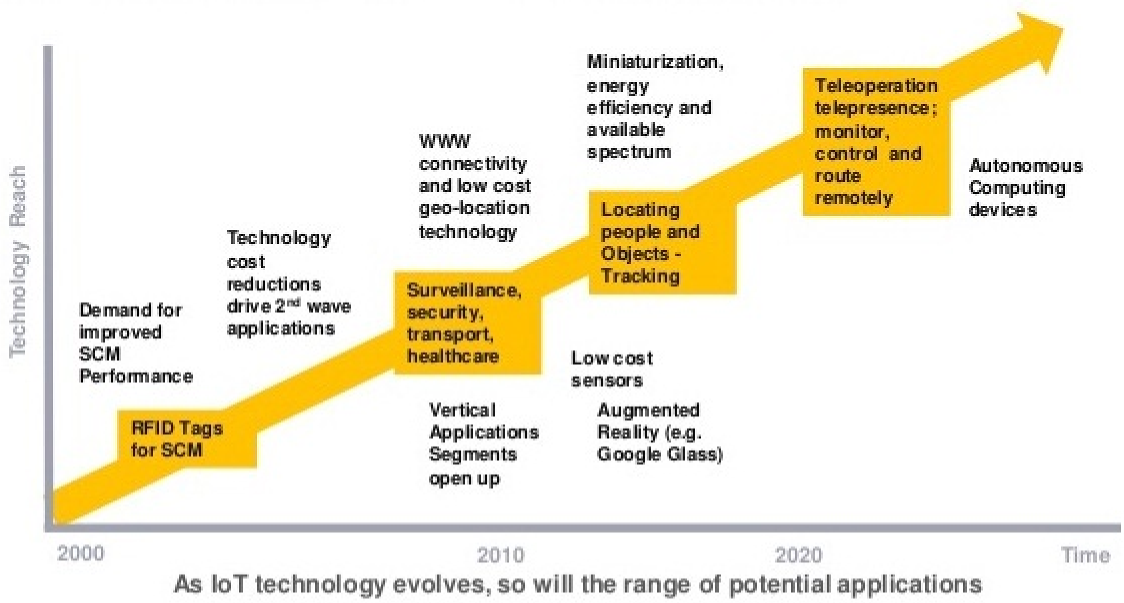


FIGURE 5. Technology Road Map of IoT

5. FEATURES OF IOT

Internet of Things has many features, in this section, we describing a basic predominant features which are very essential. The artificial intelligence that is the AI plays an important role in IoT. What gives an intelligence is the artificial intelligence now in IOT we have physical devices, however make them create virtual devices and incorporate some kind of intelligence so that these devices will act as smart it means we can virtually create anything as smart and we can enhance any aspect of this intelligence that is data collection processing, logging everything and can incorporate into these devices which are not naturally [8] there but by creating virtual instances of those devices and by incorporating this AI we can create well-established network between these devices. The feature of IoT is depicted in Figure.6.

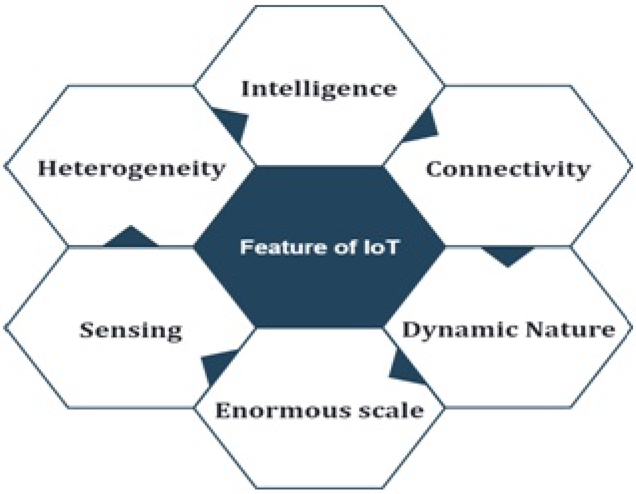


FIGURE 6. Features of IoT

Artificial intelligence in connectivity we have a similar example like we have the refrigerator we have one smart phone now we have for example we have a remote mi phone as a feature which has inbuilt infrared sensor or IR beam inside that so with the help of that beam what we can do is actually we can control number of devices we can control our AC air-conditioner we can control projector we can control a microwave oven and many things so that is what we call the collection of small devices so that is where the connectivity plays into picture. now we have N number of commodity hardware and commodity items in our home so all these we can clip together and we can put in a centralized repository and can make it work in a well coordinated [9] manner so that is where we talk about the connectivity of IOT devices. To summarize all about these IoT features we have the artificial intelligence, have connectivity, sensors and active engagement that all controlled with the help of small devices.

6. Conclusion

IoT became a booming Technology in the IT industry, introduction of big data and artificial intelligence came into existence. Connecting machines has been something that has been around for many years and it took decades to build an overnight sensation so it's not not anything new what is new now is the connectivity so we have the sensors we have the network we have the cloud and eventually all this connects back to the process and back to the people we think that is where this whole aspect and this whole ecosystem coming together that is what makes it extremely interesting.

References

[1] D. V. Kishore, D. V. Gowda: *MANET topology for disaster management using wireless sensor network*, 2016 International Conference on Communication and Signal Processing (ICCSP), (2016), 0736–0740.

[2] D. V. Gowda, D. V. Kishore: *Optimization of motorcycle pitch with non linear control*, 2016 International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT), (2016), 1656–1660.

[3] R. Dhall, V. K. Solanki: *An IoT based predictive connected car maintenance approach*, International Journal of Interactive Multimedia and Artificial Intelligence, **4**(3) (2017), 16–22.

[4] M.Penna, D. V. Gowda: *Design and implementation of automatic medicine dispensing machine*, 2017 International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT), (2017), 1962–1966.

[5] D. V. Gowda, C. A. Varun: *Implementation of swarm intelligence in obstacle avoidance*, 2017 International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT), (2017), 525–528.

[6] D. V. Gowda, A. C. Ramachanrda: *Synthesis and Modeling of Antilock Braking System Using Sliding Mode Controller*, Journal of Advanced Research in Dynamical and Control Systems, **11**(12) (2018), 208–221.

[7] Vatan, A. Sharma, S. Goyal: *IoT Standards and Applicability to Human Life* , International Journal of Innovative Technology and Exploring Engineering (IJITEE), **8**(10) (2019), 3147–3153.

[8] A. Sakshi, A. Sharma, S. Goyal: *Issues and Concerns In Security Of Cloud Environment In Internet Of Things*, Journal of Computational and Theoretical Nanoscience, **16**(10) (2019), 4374–4378.

[9] D. V. GOWDA, A. C. RAMACHANRDA: *Modelling and Performance Evaluation Of Anti-Lock Braking System* , Journal of Engineering Science and Technology, **14**(5) (2019), 3028–3045.

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
B.M.S INSTITUTE OF TECHNOLOGY
VISVESHWARAYA TECHNOLOGICAL UNIVERSITY
BANGALORE, KARNATAKA STATE, INDIA
Email address: dankan.v@bmsit.in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
VIJAYA VITTALA INSTITUTE OF TECHNOLOGY
VISVESHWARAYA TECHNOLOGICAL UNIVERSITY
BANGALORE, KARNATAKA STATE, INDIA
Email address: sridharasb1947@gmail.com

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
B.G.S INSTITUTE OF TECHNOLOGY
ADICHUNCHANAGIRI UNIVERSITY
KARNATAKA STATE, INDIA
Email address: naveenkb@bgsit.ac.in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
GITAM SCHOOL OF TECHNOLOGY
GITAM (DEEMED TO BE UNIVERSITY)
BANGALORE, KARNATAKA STATE, INDIA
Email address: rameshmalur037@gmail.com

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
YENEPOYA INSTITUTE OF TECHNOLOGY
VISVESHWARAYA TECHNOLOGICAL UNIVERSITY
MOODBIDRI, KARNATAKA STATE, INDIA
Email address: naveenpai@yit.edu.in

ISSN 2347 - 3983
Volume 8. No. 7, July 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter158872020.pdf
https://doi.org/10.30534/ijeter/2020/158872020

FPGA Implementation of Low Power High Speed BTED Algorithm for 8 Bit Error Correction in Cryptography System

Ramesha M[1], Dankan Gowda V[2], Sridhara S.B[3], Naveena Pai G[4], Bharathi Gururaj[5]
[1] Assistant Professor, GITAM School of Technology, GITAM University, Bengaluru, India, rameshmalur037@gmail.com
[2]Assistant Professor, B.M.S. Institute of Technology and Management, Bengaluru, India, dankan.v@bmsit.in
[3] Professor, Vijaya Vittala Institute of Technology, Bengaluru, India, sridharasb1947@gmail.com
[4]Assistant Professor, Yenepoya Institute of Technology, Mangalore, India, naveenpai@yit.edu.in
[5]Assistant Professor, ACS College of Engineering, Bengaluru, India, bharathigururaj@gmail.com

ABSTRACT

There are so many Error Correction Codes (ECC) have been using since decades to rectify one bit or multiple bits errors in the memory designs. To overcome the MCUs issue, Bose–Chaudhuri–Hocquenghem (BCH), Reed-Solomon codes and Punctured Difference Set (PDS) codes have been currently employed. In these conventional codes encoding and decoding is more complicated and need extra power, additional area, and huge delay and also in case of Content Addressable Memory (CAM) inserting cannot be employed due to tight coupling in between the cells. To substantiate the MCUs, issue a single-error rectification and two-fold location of integrated current Sensors are employed. To practically correct the MCUs error the new method called Bit Transition Encoder and Decoder (BTED) scheme is employed which are two dimensional matrix codes of data size 32 bits, which divides the information into the numerous sub information's like symbols of each 4-bits. The proposed BTED algorithm is implemented on Artix-7 FPGA development board and which is comparatively less delay and power in comparison with various existing methodologies. The simulation results shows that there is 18% improvement in delay, 15% in power reduction and 67% improvement in hardware resources utilizations compared to conventional algorithms.

Key words: BTED, Error Correction Codes, Cryptography, Security Issues.

1. INTRODUCTION

In memories, errors occur and those errors are identified and rectified by different approaches like Hamming codes and self-checking methods. However these approaches are not capable of the present production requirement of precision and extreme speed memories [1]. In these approaches it's easy to identify errors but it is hard to rectify them because rectifying errors again a fault error will generate a false positive error that cannot be recognized and because of this reason, it becomes a huge issue. The system developed for figuring out the faults must assure that output codes given are not the fault codes [2]. The choice of the output data code is an extremely critical task [3]. The selected code that has high error recognition capability that can reduce fault by realizing the fault steady property; however, it includes a wide number of outputs and also more system cost. To select a code of fewer error detection capacities would encompass fewer supplementary outputs at the same time, to achieve shortcoming secureness, it may be necessary to alter the circuit structure. To obtain accurate output result the decision of the generated code can be examined by the particular circuit. Figure.1 depicts a well-known structured design of a concurrent error detection plan [4].

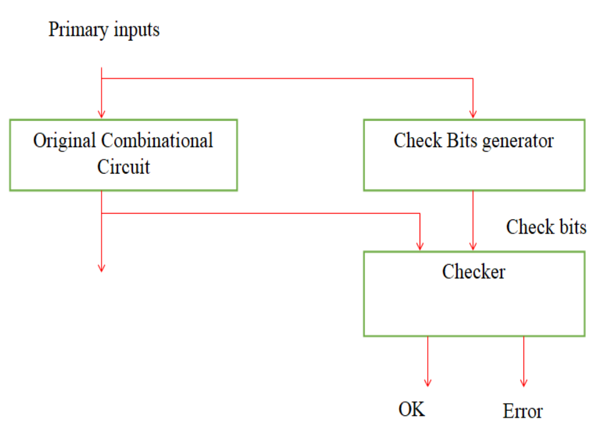


Figure 1: Architecture for concurrent error detection.

Table 1: Error Correction capability of different existing schemes

Techniques used for correction of errors	Data bit size	Redundant bits	error correction capability
DMC(Decimal Matrix Codes)	32	6	4
Matrix Code	32	8	2
Hamming Code	32	7	1

The results show that error modification capacity is increased by employing decimal matrix codes (DMC) to 4 bits. Error Correction capability of different existing schemes described in Table.1.The Hamming code occupies more area as long as cut registers and four input LUT's are compared with Matrix and Decimal Matrix Codes. However the Matrix Codes have less delay, control and occupies less area than cut registers and four inputs LUT's than the other two codes. However, it has a significant load that it rectifies only 2 bit errors [5].The DMC has a fundamental advantage that it consumes less power and has intermediate delay and area between the other two codes [6]. In the DMC, most importantly, the gap and sort out lattice are accomplished, that is the information N-bit is separated into K picture of length m-bits. The articulation can be composed as (N=K × m) where (K= K1 × K2, lines and sections). It's not important to change the physical structure of the memory to actualize the DMC [7], [8]. It can accept 32-piece of information for instance and clarify the DMC plot.

Absolute 32 bits are separated into 8 pieces of size 4 bits. From D0 to D31 are 32 data bits, check bits are 10 Horizontal (Ho-H9) and 16 vertical bits. The recognition and rectification capacity relies upon the how we pick the estimations of k and m. Exchange off happens in picking these qualities for greatest execution [9], [10]. Consequently, m and k should be carefully changed as per the enlarge review capacity and furthermore lessen the amount of overabundance bits. Right now, this circumstance, when m = 8 and k = 2×2, at that point just 1-piece issue is cured, right now tedious bits is 40[11] [12]. At the point when m = 2 and k = 4×4, at that point 2-piece flaw is helped, right now monotonous bits is 32. At the point when m = 4 and k = 2×4, at that point 5-piece deficiency is cured, right now monotonous bits is 36. The particular ultimate objective to overhaul the steadfastness of memory, the slip-up cure limit is at first considered, so k = 2 × 4 and m = 4 are utilized to create [13][14].

2. BTED ENCRYPTION AND DECRYPTION ALGORITHM

BTED algorithm is implemented in the projected BTED encryption and decryption to enhance the safety level, capability of errors location identification and their corrections. In this algorithm, the power consumed is not more than other detections approaches. It includes bitwise integer subtraction and integer addition. In the bitwise algorithm, all bits are divided into symbols of each 4bits and it has been arranged in the form of matrix for final execution. The N-bit word is defined in separate forms of m bit symbols (N = n × m). The created symbols are placed in a two-dimensional matrix of n= n1 × n2 (n1- a range of columns, n2- a range of rows). Employing decimal number addition on symbols per row the horizontal redundant bits 'H' is achieved. A binary operation on the bits per column provides vertical redundant bits 'V'. The divided symbol and arrange-matrix are represented in a logical format. Figure.2 depicts encoder module having 32-bit input and producing an output of 20-bit horizontal redundant number and 16-bit vertical redundant number.

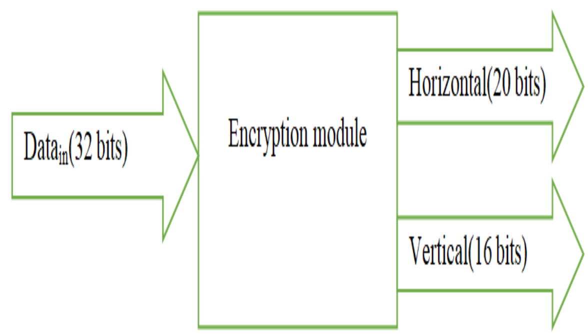


Figure 2: Proposed encryption module with 32 bits input and 20bits of horizontal and 16 bits of vertical

3. WORKING OF BTED ALGORITHM

The horizontal and vertical check bits are produced from input message bit 12345678 and this is 32-bit random number which can be the tag id. The output 2198e is a 20-bit horizontal number and 444c is the 16-bit vertical number of an encoder module as shown in Figure.2. The 32-bit word is given as input to the encoder module; it produces an output frame of 68-bit length. The frame consists of a 16-bit vertical redundant number, a 20-bit horizontal redundant number, and 32-bit input. In the BTED technique, a 32-bit word is taken as input, where data bits are characterized in cells from D0 to D31. This 32-bit word is taken as eight symbols of 4-bits each, n1= 2 and n2 = 4 are selected at the same time. Bits from H0 to H19are horizontal check bits and V0 to V15 are vertical check bits. The data is 32 bit in word (D0 to D31) and it is divided into 8 symbols and every symbol is 4 bits as exhibited below Table 2 and 3.

Table 2: Direction of horizontal, vertical check and data bits

D15	D14	D13	D12	D11	D10	D9	D8	D7	D6	D5	D4	D3	D2	D1	D0
D31	D30	D29	D28	D27	D26	D25	D24	D23	D22	D21	D20	D19	D18	D17	D16
D47	D46	D45	D44	D43	D42	D41	D40	D39	D38	D37	D36	D35	D34	D33	D32
D63	D62	D61	D60	D59	D58	D57	D56	D55	D54	D53	D52	D51	D50	D49	D48
V_{15}	V_{14}	V_{13}	V_{12}	V_{11}	V_{10}	V_9	V_8	V_7	V_6	V_5	V_4	V_3	V_2	V_1	V_0
V_{31}	V_{30}	V_{29}	V_{28}	V_{27}	V_{26}	V_{25}	V_{24}	V_{23}	V_{22}	V_{21}	V_{20}	V_{19}	V_{18}	V_{17}	V16
H15	H14	H13	H12	H11	H10	H9	H8	H7	H6	H5	H4	H3	H2	H1	H0
H31	H30	H29	H28	H27	H26	H25	H24	H23	H22	H21	H20	H19	H18	H17	H16
H39	H38	H37	H36	H35	H34	H33	H32								

Table 3: Characterization of symbols along with Data Bits

symbol 0= D_0 to D_3	symbol 1= D_7 to D_4	symbol 2= D_{11} to D_8
symbol 3= D_{15} to D_{12}	symbol 4= D_{19} to D_{16}	symbol 5= D_{23} to D_{20}
symbol 6= D_{27} to D_{24}	symbol 7= D_{31} to D_{28}	symbol 8= D_{35} to D_{32}
symbol 9= D_{39} to D_{36}	symbol 10= D_{43} to D_{40}	symbol 11= D_{47} to D_{44}
symbol 12= D_{51} to D_{48}	symbol 13= D_{55} to D_{52}	symbol 14= D_{59} to D_{56}
symbol 15= D_{63} to D_{60}	------------	-----------

Bits from H_0 to H_{39} are called check bits in direction of horizontal and V_0 to V_{31} are called check bits in vertical direction. As shown in Equation (1) to (8) by decimal addition in horizontal redundant bits ‘H’ are obtained. Figure.3 shows proposed 32 bits BTED structure and its internal modules along with their widths.

$H_4\ H_3\ H_2\ H_1\ H_0 = D_3\ D_2\ D_1\ D_0 + D_{11}\ D_{10}\ D_9\ D_8$ (1)
$H_9\ H_8\ H_7\ H_6\ H_5 = D_7\ D_6\ D_5\ D_4 + D_{15}\ D_{14}\ D_{13}\ D_{12}$ (2)
$H_{14}\ H_{13}\ H_{12} H_{11}\ H_{10} = D_{19}\ D_{18}\ D_{17}\ D_{16} + D_{27}\ D_{26}\ D_{25}\ D_{24}$ (3)
$H_{19} H_{18}\ H_{17} H_{16}\ H_{15} = D_{23}\ D_{22}\ D_{21}\ D_{20} + D_{31}\ D_{30}\ D_{29}\ D_{28}$ (4)
$H_{24} H_{23}\ H_{22} H_{21}\ H_{20} = D_{43}\ D_{42}\ D_{41}\ D_{40} + D_{35}\ D_{34}\ D_{33}\ D_{32}$ (5)
$H_{29} H_{28}\ H_{27} H_{26}\ H_{25} = D_{47}\ D_{46}\ D_{45}\ D_{44} + D_{39}\ D_{38}\ D_{37}\ D_{36}$ (6)
$H_{34} H_{33}\ H_{32} H_{31}\ H_{30} = D_{59}\ D_{58}\ D_{57}\ D_{56} + D_{51}\ D_{50}\ D_{49}\ D_{48}$ (7)
$H_{39} H_{38}\ H_{37} H_{36}\ H_{35} = D_{63}\ D_{62}\ D_{61}\ D_{60} + D_{55}\ D_{54}\ D_{53}\ D_{52}$ (8)

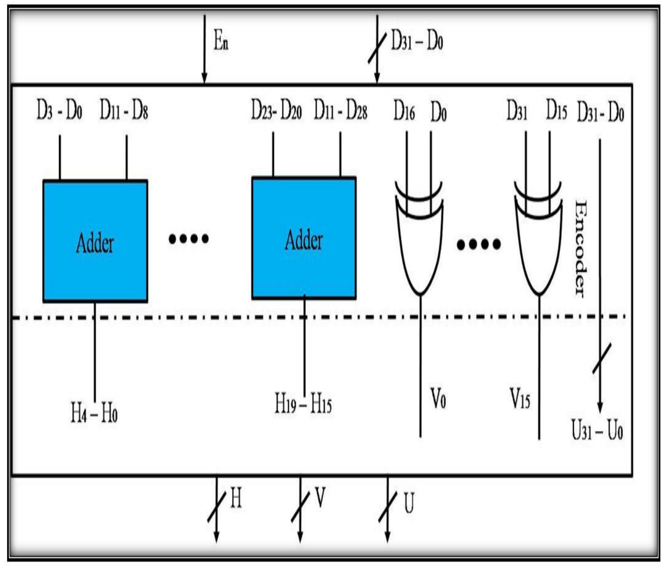


Figure 3: Proposed 32 bits BTED structure and its internal modules along with their widths

Similarly, remaining horizontal redundant bits are obtained. Here ‘+’represents bitwise integer addition. The vertical redundant bits ‘V’ may be obtained from XOR operation as given below in the table 4:

Table 4: Vertical redundant bits ‘V’ obtained from XOR Operation

$V_0=D_0 \oplus D_{32}$	$V_1=D_1 \oplus D_{33}$	$V_2=D_2 \oplus D_{34}$	$V_3=D_3 \oplus D_{35}$
$V_4=D_4 \oplus D_{36}$	$V_5=D_5 \oplus D_{37}$	$V_6=D_6 \oplus D_{38}$	$V_7=D_7 \oplus D_{39}$
$V_8=D_8 \oplus D_{40}$	$V_9=D_9 \oplus D_{41}$	$V_{10}=D_{10} \oplus D_{42}$	$V_{11}=D_{11} \oplus D_{43}$
$V_{12}=D_{12} \oplus D_{44}$	$V_{13}=D_{13} \oplus D_{45}$	$V_{14}=D_{14} \oplus D_{46}$	$V_{15}=D_{15} \oplus D_{47}$
$V_{16}=D_{16} \oplus D_{48}$	$V_{17}=D_{17} \oplus D_{49}$	$V_{18}=D_{18} \oplus D_{50}$	$V_{19}=D_{19} \oplus D_{51}$
$V_{20}=D_{20} \oplus D_{52}$	$V_{21}=D_{21} \oplus D_{53}$	$V_{22}=D_{22} \oplus D_{54}$	$V_{23}=D_{23} \oplus D_{55}$
$V_{24}=D_{24} \oplus D_{56}$	$V_{25}=D_{25} \oplus D_{57}$	$V_{26}=D_{26} \oplus D_{58}$	$V_{27}=D_{27} \oplus D_{59}$
$V_{28}=D_{28} \oplus D_{60}$	$V_{29}=D_{29} \oplus D_{61}$	$V_{30}=D_{30} \oplus D_{62}$	$V_{31}=D_{31} \oplus D_{63}$

The encoding is performed via decimal and binary addition operations. The encryption calculates the discharged bits employing XOR gates and multi-bit adders. The horizontal redundant bits are H19-H0 and vertical bits are represented as V31-V0. Data bits are directly taken from D63 to D0.

4. RESULTS AND DISCUSSION

The proposed work is intended and established for hardware prototype module having low power, low area and reduced delay employing ECC, BTED encryption and decryption by point addition and point doubling to produce 256 points to form S-Box. The advanced prototype module comprises 256 points key generation, read-only storage memory creation; BTED encryption and decryption possesses lessor amount of hardware resources. The point on the ECC curve is the 32bits data "d" which is applied for BTED encryption, which employs integer addition between various symbols to obtain encryption output. BTED encryption specifies 20 bits horizontal bits and 16 bits vertical bits and these 2 bits information is for error bit detection in decryption sections.

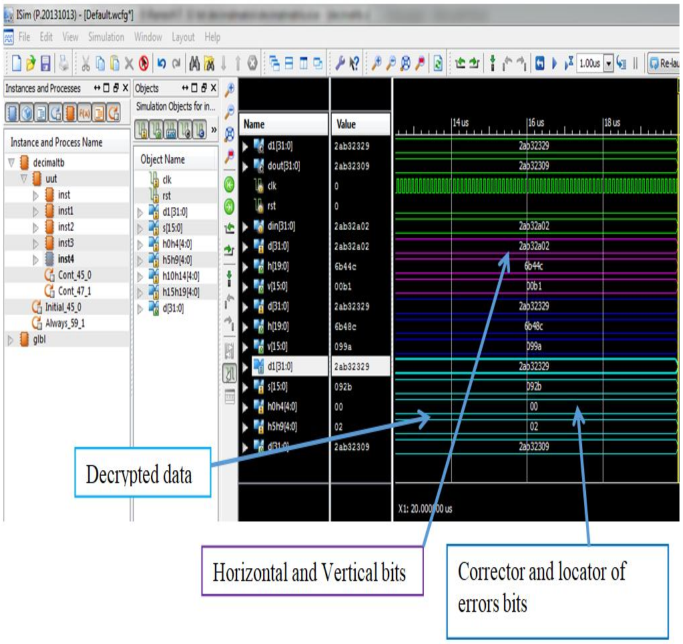


Figure 4: BTED Decryption output in Vivado Xilinx Ism simulator

The simulation result of BTED Decryption output on Vivado Xilinx Ism simulator as shown in Figure.4. Here the vertical bits of the encryption module is given as input to decryption module which contains the same integer operation to obtain 16 bits conditional bits for identification of error bit at the error locator. The'sʼ symbolized in the Figure.4 is the syndrome bits which is obtained by performing XOR operation between 'v' and 'vd' as depicted in Figure.4. The horizontal bits of encryption unit are given as input to decryption module which performs the same integer operation to obtain 16 bits conditional bits for identification of error bit at error locator. In the Figure.4 the condition bits are indicated by 'h' which is obtained by performing XOR operation between 'h' and 'hd' and is indicated in Figure.4. The 'v' and 'h' are inputs to the error locator and error detector modules and it illustrates the outputs of encryption and decryption. There are several other parameters are taken to considerations for comparison purpose namely total power, delay and device utilization summary are indicated in the Table 5 for different ECC's. This proposed research work has satisfactory performance in terms of power and speed compared to other ECC's.

Table 5: Power, slices registers, slice FF's, LUT's and delay Comparison summary of the proposed techniques

Type of ECC used	Slice Registers	Slice flip flops	LUTs	Bounded I/O	Delay(ns)	Power Utilization
Proposed work	1628	637	934	44	27.190	0.086W
Punctured distinc tion set (PDS) codes [1]	2391	926	982	NA	76.23	221.1mW
Decimal Matrix Codes [2]	2093	784	1932	NA	45.89	10.8mW
Matrix Codes [3]	1782	1027	1027	NA	70.1	24.7mW
Matrix Code [4]	3291	2081	4581	96	140.548	0.121W
Hamming code	7621	3201	2682	84	170.133	0.163W

5. CONCLUSION

In the proposed research work, the projected processor for a generation of S-Box and its encryption and decryption has been discussed. Various techniques have been adopted in the design of the processor to minimize the power consumption, area and also to enhance the speed. The bit transition encryption and decryption has also been discussed. The Encoder re-use approach reduces the area overhead of additional circuits. Simulation and synthesis outcomes exhibit that Bitwise Matrix Code requires 0.1mW of power and has a delay of about 3.109ns and the area utilization is reduced by 45%. The design is tested and demonstrated on Artix-7 FPGA prototype boards.

Simulation and synthesis shows that the projected base band-processor can complete it's fruitfully with a power consumption of about 5mW on 1.2V supply. In this proposed work a method for mapping any alphanumerical characters and any type of data for error correction is done by employing a non- singular matrix is showed. The mapping focuses are disorganized and decoded by employing the ECC approach and displays in Read Only Memory.

REFERENCES

1. Arunkumar, S.S. Tyagi, Manisha Rana, Neha Aggarwal, Pawan Bhadana, Manav Rachna, **A Comparative Study of Public Key Cryptosystem based on ECC and RSA.** *International Journal on Computer Science and Engineering (IJCSE)*, ISSN: 0975-3397, Vol. 3 No. 5 May 2011, pp-1904-1909.
2. O. Srinivasa Rao, **Efficient Mapping Methods for Elliptic Curve Cryptosystems**, *International Journal of Engineering Science and Technology*, Vol. 2(8), 2010, ISSN: 0975-5462, 3651-3656.
3. C. Wang, M. Daneshmand, M. Dohler, X. Mao, R. Q. Hu and H. Wang, **Guest Editorial - Special**

Issue on Internet of Things (IoT): Architecture, Protocols and Services. *IEEE Sensors Journal*, vol. 13, issue. 10, pp. 3505–3510. https://doi.org/10.1109/JSEN.2013.2274906

4. S.Raza, **Secure communication for the Internet of Things - a comparison of link-layer security and IP sec for 6LoWPAN**, *Security and Communication Networks*, vol. 7, no. 12, pp. 2654–2668. https://doi.org/10.1002/sec.406
5. Gonzalez G. Organero M, Kloos C. **Early in infrastructure of all Internet of Things in space for learning**, *8th IEEE International Conference on Advance Learning Technologies*, pp-381-383.
6. Dankan V Gowda, Ramachandra A C, Thippeswamy M N, Pandurangappa C, Ramesh Naidu P ,**Synthesis and Modelling of Antilock Braking System using Sliding Mode Controller**, Journal of Advanced Research in Dynamical and Control Systems', Vol 10(12),Pages:208-221.
7. Dankan V Gowda, Ramachandra A C, Thippeswamy M N, Pandurangappa C, Ramesh Naidu P, **Modeling and Performance Evaluation of Anti-lock Braking System,** Journal of Engineering Science and Technology, Taylor's University, Vol 14(5), Pages: 3028-3045.
8. Kishore, D.V., Gowda, D.V., Shivashankar, Mehta, S.: **MANET topology for disaster management using wireless sensor network**. *International Conference on Communication and Signal Processing*, April 6–8, 2016, India. https://doi.org/10.1109/ICCSP.2016.7754242
9. Dankan V Gowda, D. V. Kishore, Shivashankar, A. C.Ramachandra, and C. Pandurangappa, **Optimization of motorcycle pitch with non linear control**, in Proceedings of the *1st IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology* (RTEICT '16), pp. 1656–1660, Bangalore, India, May 2016.
10. Ramesha, M. & Ramana, T. **Performance analysis of fbmc transceiver architecture over OFDM system.** *Far East Journal of Electronics and Communications*.17.1353-1372.10.17654/EC017061353.
11. Ramesha, M. & Ramana, T, **A Novel Architecture of FBMC Transmitter using Polyphase Filtering and its FPGA Implementation**. *Indian Journal of Science and Technology*. 9.10.17485/ijst/2016/v9i48/94148.
12. Trinidad, Emmanuel. (2019). **Juxtaposition of Extant TV White Space Technologies for Long-Range Opportunistic Wireless Communications.** *International Journal of Emerging Trends in Engineering Research*.7. 209-215. 10.30534/ijeter/2019/17782019.
13. B, Sridhara & M, Ramesha & Patil, Veeresh. **Adaptive Scheduling Design for Time Slotted Channel Hopping Enabled Mobile Adhoc Network.** *International Journal of Advanced Computer Science and Applications*.11. 10.14569/IJACSA.2020.0110333.
14. Lakshmi, Boggula. (2019). **Energy Efficient Routing Mechanism for Harsh Environment in Wireless Sensor Networks**. *International Journal of Emerging Trends in Engineering Research*. 234-238. 10.30534/ijeter/2019/04792019.

ANALYSIS OF QUANTUM KEY DISTRIBUTION IN CRYPTOGRAPHY AND ITS APPLICATIONS

Dr. Ambika R[1*], Dr. Anilkumar D[2], Shashikala J[2]

1,2 BMS Institute of Technology and Management, Bengaluru, India

(Email id:ambikr@bmsit.in, anilkumard81@bmsit.in, shashikala_j@bmsit.in)

(Corresponding author: Dr. Ambika R: Email id: ambikar@bmsit.in)

Abstract: Safe communication is a growing concern in almost every sector. The data was encrypted using cryptographic techniques. Compared to traditional cryptography, quantum cryptography is a strong and optimistic step. Our data would be safer than in the past if quantum encryption is used in the right way. To protect the confidential information to be transmitted, cryptographers are trying to develop more sophisticated techniques. But hackers, code breakers and eavesdroppers are working furiously to break the systems. As security is achieved by either cryptographers or security is breached. The success is, however, provisional. The method of protecting the message using a deciphering device and of breaking the system. In this article, various quantum encryption and distribution methods are discussed and analyzed.

Keywords: **Classical Cryptography, Quantum Cryptography, Quantum Key Distribution, Avalanche Effect**

1. INTRODUCTION

Protecting significant and important data is of utmost concern to the organization or multiple input, multiple output transmitter-receiver-based communication system. Cryptography is one of the key techniques of protecting the data. Novel concept in cryptographic security is the Quantum Encryption. Quantum cryptography uses quantum fluctuations of laser light at the physical layer which is introduced into the existing network. This enables extreme-secure communication and perfect security. Security is an important characteristic of any network with special reference to mobile - adhoc networks. The wireless networks are potential for hacking using mobile devices. There is no clear line of defence for protecting any mobile network. Development of mobile application security system, which uses a layered security approach and strong cryptographic techniques is the probable and low-cost solution to protect such application-based wireless networks.

Quantum key distribution is a cryptographic protocol and it allows two communicating parties to distribute a secret key in the presence of an eavesdropper [1]. Quantum cryptography uses quantum mechanical concepts such as the Heisenberg uncertainty principle and the no-cloning theorem in order to ensure that Eves dropper cannot gain non-negligible amount of information without **being detected, even if his/her computational power is** unlimited. This paper discusses about the review of the work carried out using quantum cryptography and defines various terms related to it.

2. LITERATURE REVIEW

In this section, related articles on key generation, distribution and quantum key distribution have been reviewed and analyzed.

Ambika R. et. al., introduced a data protection multi-transceiver system [2]. They considered three users and implemented RSA commutative and proved RSA algorithm's commutative nature. In this paper, authentication is accomplished by key exchange approach and the limitation of the same is discussed. In most of the existing data authentication or security systems, overhead of key exchange has been increased. To achieve the goal of data security with individual encryption/decryption without affecting the data security and its integrity, a modified RSA has been developed and this mechanism is known as Commutative RSA. The device communicated is secureand can not be colluded . In general, the use of cryptographic techniques is favored, hence the proposed Multi FPGA core protocol in this paper adopts the commutative RSA algorithm.

In this paper, a security or authentication was implemented for multiple MIMO or transceiver terminals using the public key cryptography technique called Commutative RSA. The goal of data security is implemented in multiuser communication environment. The commutative RSA approach has been implemented with multiple FPGA cores that function as an individual transceiver terminal and perform its individual encryption without affecti ng the original data. Serial Montgomery multiplication was planned and simulated with multiplier with multiplier Radix-2. The results achieved have been compared, and the new Serial Montgomery architecture has been found to work better than previous architectures.

In this paper, the commutative RSA core was implemented for simulation and illustration of data authenticity among multiple user terminals in a communication environment on multi ple FPGA devices.

Normand J. Beaudry et al., demonstrated a general method for proving the security of quantum key distribution protocols in two ways [3]. First is based on super dense coding, and second is based on the LM05 dual-way protocol. They proved that an eavesdropper secures these two protocols against other, more common forms of attacks.

Highly robust and optimized system architecture has been proposed for implementing a commutative RSA algorithm for data authentication among multiple MIMO terminals, simulated on FPGA devices in [4]. A noble commutative RSA approach has been implemented to facilitate secure data communication among multiple input multiple output channels or transceiver systems. Commutative RSA states the order in which encryption is performed does not affect the encryption output.

Authors implemented and simulated the complete code on several FPGA devices. The robust Montgomery modular multiplication mechanism has been adopted with radix-2 multiplication architecture to maximize device output with limited space and higher speed. Authors have proposed the implementation of the CRSA cryptography core based on serial Montgomery and parallel Montgomery with the aim of improving device efficiency for its lower memory occupancy, rapid pace, higher throughput and lower power consumption.

Conventional cryptosystems such as ENIGMA, Data Encryption Standard or even RSA are based on a mixture of mathematics and guesswork [5]. In formation theory has shown that traditional secret-key cryptosystems can not be completely secure unless the key, which is only used once is at least as long as the cleartext.

On the other hand, the principle of computational complexity is not well established to prove the computational security of public-key cryptosystems. The writers in this paper use a radically different cryptographic base. They used a quantitative physics uncertainty theory.
Through modern information theory and cryptography, digital messages can often be passively tracked or copied through practice, even by someone ignorant of what they say. However, when in non-orthogonal quantum states information is encoded, for instance single photons with polarization directions 0, 45, 90 and 135 degrees, a communication channel is obtained whose transmission cannot, as a rule, be copied, read or copied reliably by an eavesdropper who is ignorant of such essential information used to render transmission. Without altering a random, uncontrolled manner that could be detected by the legitimate users of the channel, the cryptanalyst or attacker can not obtain partial information about such a transmission either.
Information security is important in today's world and in future for data confidentiality since we expect our digital opponents to be fingerprinted with more powerful computers and new algorithms. The advent of quantum computers that are capable of launching effective attacks on conventional techniques, such as the commonly used types of public key cryptography, is of particular concern [6].

Since the protection of quantium cryptography relies solely on the local Equipment of the legitimate users, the key challenge in the implementation of quantum cryptography is to determine how much information these equipment leaks to possible opponents. If this information leakage can be restricted under a certain standard, a security technique called the data protection extension can be used to recover it. This compresses a partly secret bit-sequence to a highly secure key and depends on the estimated information leakage, the compression. Thus, the security promise of the theoretical protocol against available technologies can be restored by properly characterizing a true framework.

Data protection enhancement is not the only resources available to ensure the safety of quantum cryptography implemented. Hardware and protocol modifications can dramatically reduce the lack of information and the potential for lateral channels and active attacks. In addition, thehardware of a real system can be tested with quantum correlations. These tests may be demanding to perform but have the advantage of immunity from a wide range of problems with implementation [7].

Makhamisa Senekane et. al, suggested the six-stage quantum key distribution protocol to implement its optical implementation [8]. This protocol uses more detector range and improves its consistency, increases the likelihood of discrimination, and improves detector.

Mohd Asad Siddiqui, Tabish Qureshi have proposed Quantum Key Distribution method. In this, user A sends user B pairs of qubits and each is in one of four states. User B generates a secure key with one qubit and a subsidiary key with the other. He randomly determines for what pair which key to use. In order to match the safe key of user A, an auxiliary key must be added to the safe key of user B. This scheme gives the BB84 standard protocol an additional security layer [9]. Between two remote parts, a secret key is generated, and a quantum channel is used. The hidden key may be used Vernam Cipher or single pads to send encrypted messages.

Cipher Vernam is very secure given that a key is exchanged. Secure messaging applications are based Android and Web-based systems that are used to securely send messages among registered users to any company using cryptographic algorithms [10]. U ntil sending the message the application can be protected by user authentication. The safe messaging system uses minimal overhead processing while maintaining safety. Authentication of each user is strengthened using Salt in the database to store sensitive credentials for each user. Message encryption and decryption was done using the Advanced Encryption Standard monoalphabetic algorithm.
This algorithm is actually less secure than the key- public encryption method.
This is the principal restriction of this work.
An eavesdropper breaking in into the message returns an insignificant message. encryptio n and decryption are clearly one of the easiest ways to conceal the meaning of a message from interferers in a network environment.

3. IMPERFECT ENCODING

That Quantum Key Distribution Protocol needs to determine how the quantum states to be sent over the unstable public channel are to be prepared. This is analogous to the initial selection of good prime numbers in the RSA algorithm, where a poor selection can jeopardize overall protection.

The prepared countries may differ from those specified in the protocol for the purposes of Quantum Key Distribution due to physical equipment imperfections. This suggests that the impact on security should be small.

However, the outcome does not standardize the usual losses of a communications channel when standard security tests are implemented, which results in a much higher rate of lowering than expected.
This condition has been overcome by recent safety proof, which is available in the finite-size case as well as asymptotically endless key blocks.
These proofs can be used to calculate the difference between the ideal model and to restore the protection of a Quantum Key Distribution system and achieve almost the same key generation rate as for faultless encoding.

4. PHASE CORRELATION BETWEEN SIGNAL PULSES

In multi-photon emissions, a replacement of the ideal sngle-photon light sources with reduced laser is common in the Quantum key distribution. One of the consequences of this substitution is that the electromagnetic phase of each light pulse emitted may partly be linked to the other pulse phase because of its las er consistency. Therefore, it is very necessary to have phase correlation in quantum key cr yptography between signal or quantum codes.

A. Bright-light attack

The single photon detectors are key components of most Quantum Key Distribution syste ms. The weak optical signal received by the transmitter via a communications channel is detected in the Quantum Key Distributor receiver. Lawrence photodiodes that work in Geiger mode

are the most common detectors for single photons.
The Avalanche Photo Diode is biased over the decompose voltage to make a single photon easily detectable through sudden growth in the output current by triggering a self-sustaining avalanche.
However, any detector needs to spend some of its time under a collapsing voltage, becaus e at some point the detection avalanche needs to be shut down to reset the detector. The detector enters the linear mode when it is no longer sensitive to single photon radiation.

It has been demonstrated that Eve can use the linear regime to monitor the detector output deciding bits of the final key unnoticed. Even can also force a single photon detector into t he linear mode in some situations, send a bright light into the receiver module and take ad vantage of its control abilities and steal key bits.

It is important to analyze this group of attacks carefully and make efforts to distinguish between the incorrect running of a detector and real loopholes. For example, when handling gated Avalanche Photo diodes in a specific mode, bright illumination from a continuous-wave laser of 1FW – 10mW was easily detected. The APD parameters such as photo current, bias tension, temperature, after-pulsing rate or quantum efficiency can be also monitored on an on-line Quantum Key Distribution system.

These countermeasures dramatically reduce information Eve is able
to obtain from the key, and current research continues to calculate its effect precisely on the safe key rate.

B. Efficiency mismatch and time-shift attacks

The Quantum Key distribution configuration consists of two detectors, which are correlat ed with either 0 or 1. If an adversary is able to figure out which of the detectors answered the input light, he will learn the key bit. It is therefore necessary to distinguish two detecto rs from the point of view of the opponent. This is challenging because it is impossible that 2complex objects such as photo detectors would be the same. general the opponent has o ptions to attack the Quantum Key Distribution configuration if the response curves of the devices are different.
In addition, if the detector efficiencies vary in wavelength, an opponent can attempt to attack "efficiency-match," while an opponent may launch the "time-shift" attack if the time responses are different.
Another attack was recently described to deliberately induce a detector malfunction.

The combination of PA and the differences in the parameters of the two detectors can be used to prevent this kind of attack. The safety evidence considers the difference between detector response curves and removes the additional information leaked through PA to an opponent.
The symmetrisation of the detectors with regard to a single photon signal condition can al so deny this attack. The two detectors will be virtually identical by randomly switching th e bit allocation between the two detectors, thereby preventing one opponent from assignin g bit values to the information it gets.

C. Back-flash attack

The backflash attack is a passive way to let an opponent learn the bit values associated wit h detection events. During the avalanche of charging carrers due to a detection event, the s econdary photons emitted byan Avalanche PhotoDiode that travel back to an opponent fro m the detectors through the transmitter communication canal.This effect was demonstrate d in gated detectors InGaAs / InP routinely used in QK systems with a wavelength of tele communications.

The backflash attack is a passive way to let an opponent learn the bit values associated wit h detection events. During the avalanche of charging carrers due to a detection event, the s econdary photons emitted by an Avalanche PhotoDiode that travel back to an opponent fr om the detectors through the transmitter communication canal.This effect was demonstrat ed in gated detectors InGaAs / InP routinely used in QK systems with a wavelength of tel ecommunications.

In order to prevent a back-flash attack, proper design and analysis of Quantum Key distribution systems can be carried out. Some solutions may involve low-loss passive optical devices such as insulators, circulators, or special filters, which link the likelihood of a photon leaking back from the quantum key distribution system to a low level in order to recover security using passive attack. The use of a short gate will also reduce the light emission intensity in fast-moving detectors significantly.

5. CONCLUSION

Intense and prolific research over the past twenty years has been triggered by quantum cry potgraphy and Quantum Key Distribution. Quantum Key Distribution allows secret key cr eation by means of a combination of a conventional channel and a quantum channel, s uch as an optical fibre connection or an optical open- space linking.
Quantum Key Distribution's primary value is that all eavesdropping can be observed in th e rows, connected intrinsically to the "quantum" of the transmitted signals in the quantum channel.

This property gives rise to a cryptographic property that can not be obtained by classical techniques, which enables the Key Establishment to operate under an extremely high security standard, called unconditional security or theoretical information security. The natural candidates for Quantum Key Distribution based security solutions are therefore highly security applications.
However, many important issues remain to be solved. A further prospect is to examine and demonstrate the integration of quantum key distribution into real security infrastructures alongside the existing challenges related to theory and experimentation of Quantum Key distribution.

In order to take full advantage of the opportunities provided by point-to - point distance-restricted quantum key distribution links, it is important to develop a network architectures. However, the Quantum Key Communication Networks are not all-round networks they are locked, hidden and distant, and their characteristics are closely related to their physical layer's quantum characteristics.
Such networks therefore differentiate essentially from the conventional Key Distribution infrastructure. We nevertheless find the use of Quantum Key Delivery networks in high-security environments which until now relies solely on trustworthy couriers from Key Establishme nt to be promising.In the context of safe networks based in symmetric-key encryption systems, Quantum Key Distribution-based systems can also be regarded as an alternative to public-Key session key exchange.
In conclusion, the Quantum Key Distribution Network is a security network that enables the theoretical distribution of key information over a worldwide network.
We hope their evolution can be combined successfully with "classical cryptography"

ideas that
open promising paths towards progress in cryptography and safety in the networks.

FUTURE SCOPE

In this paper we have analysed various quantum key distribution techniques along with classical key distribution. Our work will continue in enhancing some of the techniques using GPGA for better efficiency and throughput.

REFERENCES

[1] *N. Sasirekha, M.Hemalatha (2014), Quantum Cryptography using Quantum Key Distribution and its Applications, International Journal of Engineering and Advanced Technology, ISSN: 2249 – 8958, Volume-3, Issue-4.*

[2] *Ambika R et.al., "Data Security Using Serial Commutative RSA Core For Multiple FPGA System"(2014), Second International Conference on Devices, Circuits and Systems, DOI: 10.1109/ICDCSyst.2014.6926198*

[3] *Normand J. Beaudry et. Al., "Security of two-way quantum key distribution" (2013), American Physical Society, A 88, 062302.*

[4] *Ambika R et.al., "Securing Distributed FPGA System Using Commutative RSA Core" (2013), Global Journal of Researches in Engineering, Vol 13, Issue 15.*

[5] *Charles H. Bennett and Gilles Brassard, "Quantum cryptography: Public key distribution and coin tossing", Theoretical Computer Science 560 (2014) 7–11.*

[6] *Marco Lucamarini et.al., "Implementation Security of Quantum Cryptography Introduction, challenges, solutions" (2018), ETSI White Paper No. 27,*

[7] *SECOQC, White Paper on Quantum Key Distribution and Cryptography*

[8] *Makhamisa Senekane et. al., "Six-State Symmetric Quantum Key Distribution Protocol" (2015), Journal of Quantum Information Science, 5, 33-40*

[9] *Mohd Asad Siddiqui, Tabish Qureshi, "Quantum Key Distribution with Qubit Pairs" (2014), Journal of Quantum Information Science, 4, 129-132, http://dx.doi.org/10.4236/jqis.2014.43014*

[10] *Rahman MM et. Al., "Development of Cryptography-Based Secure Messaging System" (2016), Journal of Telecommunications System & Management, Vol 5, Issue 3, DOI: 10.4172/2167-0919.1000142.*

[11] *Ambika R and Hamsavahini R, "A Survey on Hardware Architectures for Montgomery Modular Multiplication Algorithm" (2013), International Journal of Emerging Technologies in Computationaland Applied Sciences(IJETCAS), Vol 5, Issue 3, pp. 217-221.*

International Journal of Electrical Engineering and Technology (IJEET)
Volume 11, Issue 5, July 2020, pp. 83-90, Article ID: IJEET_11_05_009
Available online at http://www.iaeme.com/ijeet/issues.asp?JType=IJEET&VType=11&IType=5
ISSN Print: 0976-6545 and ISSN Online: 0976-6553
DOI: 10.34218/IJEET.11.5.2020.009

DRONE INTEGRATED WEATHER SENSORS FOR AGRICULTURE PURPOSE

Dr. Anil Kumar D
Assistant Professor, Department of Electronics and Communication Engineering,
BMS Institute of Technology & Management, Bengaluru, India

Dr. Ambika R
Professor, Department of Electronics and Communication Engineering,
BMS Institute of Technology & Management, Bengaluru, India

ABSTRACT

The conditions within an ecosystem as well as weather of a field affect crop productivity greatly. Different weather conditions produce different effects and impact on the quality of the crop yield or on the ecosystem. Weather elements form a chain reaction, as the atmosphere is not the only one being affected, atmospheric air temperature, vapor pressure and relative humidity or moisture content can act together and form diverse effects on crops. These diverse effects in turn reduces radiation which is necessary for plants, or increases rainfall patterns. Consistent high temperatures can increase the heat transfer to local water bodies in addition to heating the air. Monitoring the climate and the weather conditions are important not only as an environmental baseline, but to maintain quality working conditions, marine studies and recreational safety. The parameters of climate are measurable, for example, atmospheric vapor pressure, temperature, precipitation and solar radiation can be captured and recorded daily on the Arduino Sensor Integrated Drone. Means and extreme datasets, maximum and minimum weather trends with deviations of lengthy time series would be calculated for each of these climate parameters which are considered in this study. These results are a simple form of climate indices, as they already describe changes in climate. All the readings and datasets are recorded on a cloud platform, as well as, in an installed microchip on the drone. Data synchronization is done with Arduino Programming Rule.

Key words: Arduino, sensor, weather data, environment.

Cite this Article: Dr. Anil Kumar D and Dr. Ambika R, Drone Integrated Weather Sensors for Agriculture Purpose. *International Journal of Electrical Engineering and Technology*, 11(5), 2020, pp. 83-90.
http://www.iaeme.com/IJEET/issues.asp?JType=IJEET&VType=11&IType=5

1. INTRODUCTION

Climate and weather data monitoring systems provide vivid and exact information for farming purposes which goes beyond weather stations. Interesting enough, every farmer needs to know and monitor the weather if high productivity has to be met.

Agricultural productivity relies directly and indirectly on the weather and climatic conditions. Therefore, failure to adhere to the weather conditions before, during and after farming renders crop production susceptible to unsuspected disease attacks that are caused or influenced by certain weather conditions. This study is intended to meet the demand for easy-to-use, easy-to-access, and accurate weather data for farmers using a drone (UAV) with mounted Arduino Sensors to give accurate climate information to farmers in crop production especially rainfall, solar (duration and intensity) and temperature.

This Arduino drone is being built and developed with state of-the-art multi-parameter weather sensors which focus on four main weather parameters—radiation, temperature, rainfall, and humidity which is an all in one compact lightweight instrument called Arduino Drone. These readings of weather parameters focus on the above four data sets as they have directly or indirectly effect on crop production.

2. LITRATURE SURVEY

Literature Survey is an important phase in the system development life cycle as we collect and acquire the necessary information to handle or develop a project during this phase. A literature review is a description of literature relevant to a particular field or topic. It gives an overview of what has been said, who the key writers are, what are the prevailing theories and hypothesis and what methods and what methodologies are appropriate and useful.

In this chapter research is done prior to taking up the project and understanding the various methods that were used previously. A detailed analysis of the existing system was performed. This study helped to identify the benefits and drawbacks of the existing systems. In agriculture usage of drones is very limited. Presently the drones are used in following areas

- **Soil and Field Analysis** - Drones obtain 3D maps of existing soil which will help farmers to determine the most effective patterns for planting, managing crops, soil and more.

Seed Planting - Drone planting is new technology which helps to minimize the need for on-the-ground planting, which can be costly, time intensive and strenuous work.

- **Crop Spraying and Spot Spraying** - Drones can be equipped with large reservoirs, which can be filled with fertilizers, herbicides, or pesticides. Using drones for crop spraying is much safer and cost-effective.
- **Crop Mapping and Surveying** - One of the biggest advantages of using drone technology is the ease and effectiveness of large-scale crop and acreage monitoring. With drone mapping and surveying, technology decisions can now be made based on real-time data, not outdated imagery, or best-practice guesswork.
- **Irrigation Monitoring and Management** - Drones that are equipped with thermal cameras can help to spot irrigation issues, or areas that are receiving too little or excessive moisture. With this information, crops can be better laid out to maximize drainage, adhere to natural land runoff, and avoid water pooling, which can damage sensitive crops.

This project will give first-hand information about the weather to farmers directly on the farm so their mobile and crop performances can be monitored.

3. BLOCK DIAGRAM OF PROPOSED SYSTEM

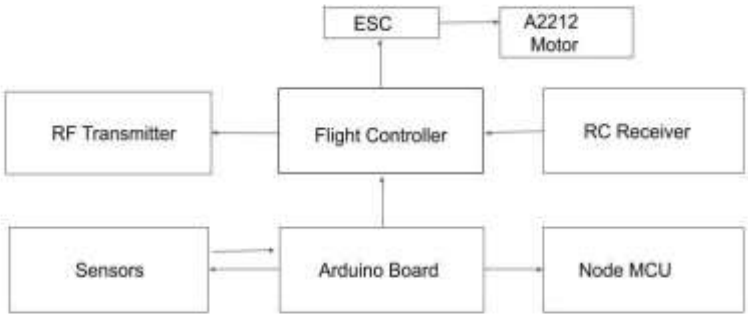


Figure 1 Proposed System

4. COMPONENTS AND EXPERIMENTAL PROCEDURES

1. For our experiment, we designed a quadrotor using the **propeller frame F450 frame**.
2. **Flight Controller**- CC3D autopilot running Ardupilot firmware as our flight controller.

Figure 2 Flight Controller

3. **RC Receiver and RF Transmitter**- FS-iA6B This receiver system uses low power electronic components and a sensitive receiver chip for the drone. The RF Transmitter modulation uses intermittent signals thus reducing even more power consumption.

Figure 3 RC Receiver and RF Transmitter

4.**Electronic Speed Controller (ESC)**- An Electronic Speed Controller (**ESC**) is an electronic circuit that acts as the interface between the pilot's commands and the individual **drone** motors.

5. **DHT11**- is a low-cost humidity and temperature sensor with a single wire digital interface. The sensor is calibrated and doesn't require extra components so you can get right to measuring relative humidity and temperature.

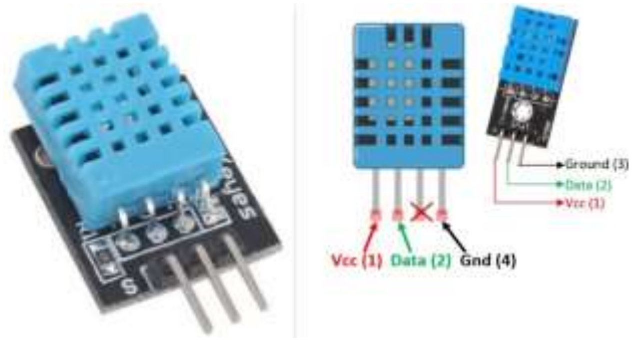


Figure 4 DHT11

6. **BMP180** - The BMP180 barometric pressure sensor is a great sensor that can be used to predict the weather, detect altitude, and measure vertical velocity. It is perfect for weather stations, remote controlled vehicles, weather balloons, and lots of other projects. It is an extremely sensitive sensor too. As you will see in a minute, it can detect changes in altitude of just a few inches.

Figure 5 BMP180

7. **Raindrop Sensor**- is a tool used for sensing rain. It consists of two modules, a rain board that detects the rain and a control module, which compares the analog value, and converts it to a digital value. The raindrop sensors can be used in the automobile sector to control the windshield wipers automatically, in the agriculture sector to sense rain and it is also used in home automation systems.

Figure 6 Raindrop Sensor

8. **ESP8266(Node MCU)** -NodeMCU is an open source IoT platform. It includes firmware which runs on the ESP8266 Wi-Fi SoC from Espressif Systems, and hardware which is based on the ESP-12 module. The term "NodeMCU" by default refers to the firmware rather than the development kits. The firmware uses the Lua scripting language. It is based on the eLua project, and built on the Espressif Non-OS SDK for ESP8266. It uses many open source projects, such as lua-cjson and SPIFFS.

Figure 7 ESP8266(Node MCU)

5. METHODOLOGY

Data Captured by the Arduino Board- When the drone flies, the weather data is recorded by the sensors which is mounted on the Arduino board. The temperature and humidity sensor (DHT11) consist of a capacitive humidity sensing element and a thermistor for sensing temperature. The humidity sensing capacitor has two electrodes with a moisture holding substrate as a dielectric between them. Change in the capacitance value occurs with the change in humidity levels. The IC measure, process this changed resistance values and change them into digital form.For measuring temperature this sensor uses a Negative Temperature coefficient thermistor, which causes a decrease in its resistance value with increase in temperature. To get larger resistance value even for the smallest change in temperature, this sensor is usually made up of semiconductor ceramics or polymers. The rain drop sensor has two cases: -

Case1: When the input of the inverting terminal is higher than the input of the non-inverting terminal.

Case2: If the input of the inverting terminal is lower than the input of the non-inverting terminal.

When the rain board module's surface is exposed to rainwater, the surface of the rain board module will be wet, and it offers minimum resistance to the supply voltage. Due to this, the minimum voltage will be appearing at the non-inverting terminal of LM393 Op-Amp. The comparator compares both inverting and non-inverting terminal voltages. If the condition falls under case (1), the output of the Op-Amp will be digital LOW. If the condition falls under case (2), the output of the Op-Amp will be digital HIGH. The below diagram shows the equivalent circuit of both the conditions.

The pressure sensor senses the pressure using (BMP180). The BMP180 is a piezoresistive sensor that detects pressure. Piezoresistive sensors are made up of a semiconducting material (usually silicon) that changes resistance when a mechanical force like atmospheric pressure is applied. The BMP180 measures both pressure and temperature, because temperature changes the density of gasses like air. At higher temperatures, air is not as dense and heavy, so it applies less pressure on the sensor. At lower temperatures, air is denser and weighs more, so it exerts more pressure on the sensor. The sensor uses real-time temperature measurements to compensate for the pressure readings for changes in air density. The BMP180 outputs an uncompensated temperature (UT) value and an uncompensated pressure

(UP) value. The temperature measurement is taken first, followed by a pressure measurement. This flow chart outlines the steps the sensor takes when performing a measurement:

All the sensors within the Smart Arduino Drone would be calibrated using the Arduino Software and data analysis done by Arduino. After all the sensors are mounted on the Arduino board, the Arduino board needs to be aligned, connected to a data logger (computer) and power source (Battery). Flying height of the drone depends on the individual but it is estimated that flying higher allows you to cover more ground in a shorter amount of time with less battery usage.

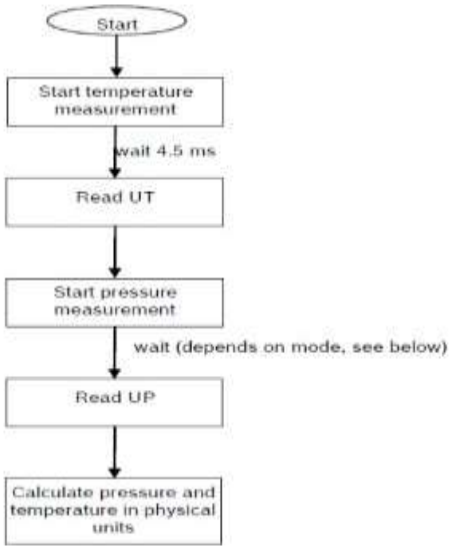


Figure 8 Flowchart for the steps taken by sensor when performing measurement

Data Analysis using Software- The software or any program used for the operation of any system can be written in any language considering various factors. The choice is made on the basis of following decisive factors: -

- Required execution speed
- Accurate control of peripheral devices
- Programming experience of team members
- Time available

Assembly language is considered to be the best for projects that need minimum memory, the highest execution speed, and precise control of peripheral devices but since writing in this language is a tedious task with more knowledge in C programming, we choose to write our source code in the C language. Not only that C programming is used knowing us the more about it but it has also several advantages over assembly language.

5.1. Choosing C++ language

C++ is a powerful, flexible language that provides fast program execution and imposes few constraints on the programmer. It allows low level access to information and commands while still retaining the portability and syntax of a high-level language. These qualities make it a useful language for both system programming and general-purpose programs. Its flexibility comes from the many ways the programmer has to accomplish the same tasks. C++ includes bitwise operators along with powerful pointer manipulation capabilities. C++ imposes few constraints on the programmer. The main area this shows up is in C's lack of type checking. This can be a powerful advantage to an experienced programmer but a dangerous disadvantage to a novice. Another strong point of C++ is its use of modularity. Section of code can be stored in libraries for re-use in future programs. This concept of modularity also helps with ‘c++’ portability and execution speed. The core C++ language leaves out many features included in the core of other languages. These functions are instead stored in the C++ standard Library where they can be called on when needed. An example of this concept would be C’s lack of built in I/O capabilities. I/O functions tend to slow down program execution and also be machine independent when 23 running optimally. For these reasons, they are stored in a library separately from the C language and only included when necessary.

- Arduino IDE
- Blynk android application in android mobile phone.

Blynk was designed for the Internet of Things. It can control hardware remotely, it can display sensor data, it can store data, visualize it and do many other cool things.

There are three major components in the platform:

Blynk App - allows you to create amazing interfaces for your projects using various widgets we provide.

Blynk Server - responsible for all the communications between the smartphone and hardware. You can use our Blynk Cloud or run your private Blynk server locally. It’s open-source, could easily handle thousands of devices and can even be launched on a Raspberry Pi.

Blynk Libraries - for all the popular hardware platforms - enable communication with the server and process all the incoming and outcoming commands.

6. RESULTS AND DISCUSSION

6.1. Picture of Module

It is based on the autopilot data generated, the Arduino Sensor Integrated Drone will predict the average means of temperature and humidity, rainfall, wind pressure. This Drone in agriculture will be a big boost to farmers since it carries the potential of completely transforming and revolutionizing the farming and agriculture industry.

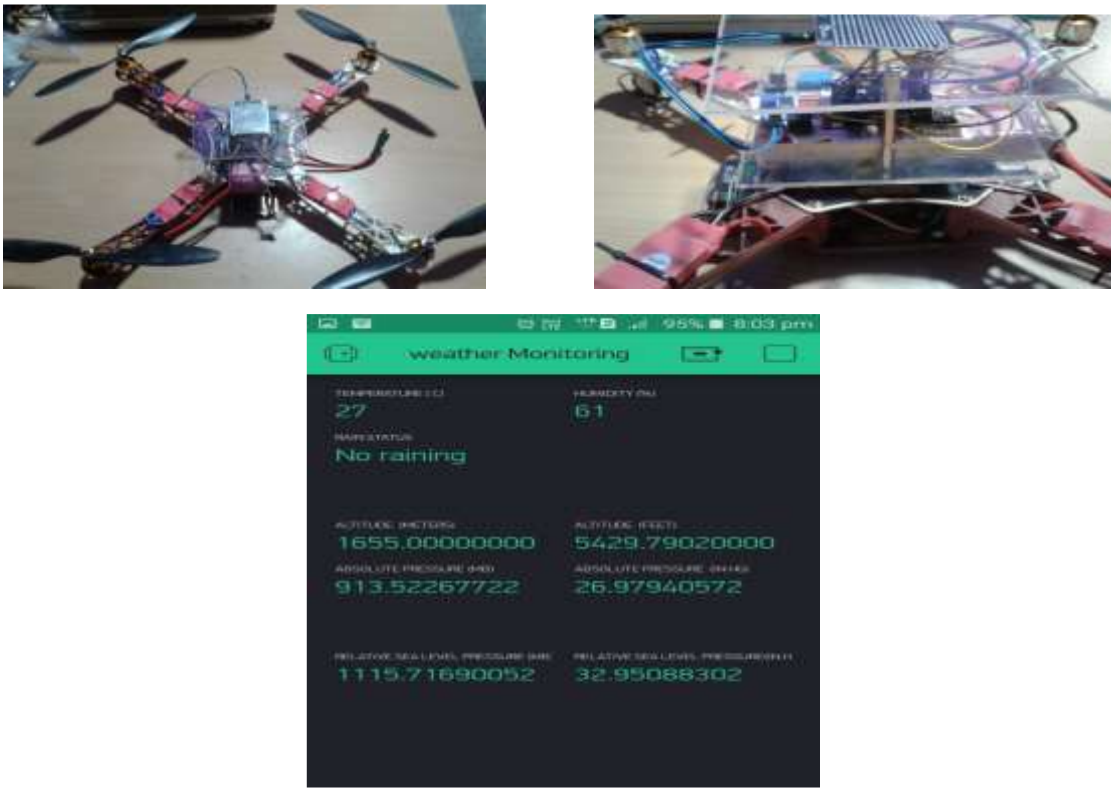


Figure 9

Weather parameter reading, tracking weather and estimating yields, and then collecting the data and analyzing it for prompt action are some elements considered in this project

7. CONCLUSIONS

This Arduino Integrated Drone can mechanize every step of farming and eliminate the costs of human errors and enable farmers to react quickly to threats (such as drought conditions and flood), helping with maximization of income and returns on investment crop productivity.

This Drone in agriculture is a big boost to farmers because it carries the potential of completely transforming and revolutionizing the farming and agriculture industry.

FUTURE ENHANCEMENT

Weather parameter reading, soil health scans, monitoring crop health, applying fertilizers and watering the fields, even tracking weather and estimating yields, and then collecting the data and analyzing it for prompt action are some elements considered in this ongoing project.

REFERENCES

[1] Vining KC. (1990). Effects of weather on agricultural crops and livestock: An overview. International Journal of Environmental Studies, 36 (1-2):27-39

[2] Darack E. (2012). UAVs: The new frontier for weather research and prediction. Weatherwise, 65(2): 20-27

[3] Acheampong AA, Fosu C, Amekudzi LK, Kaas E. (2017). Perceptible water comparisons over Ghana using PPP techniques and reanalysis data. South African Journal of Geomatics, 6(3):449

[4] Jensen J. (2019). Agricultural Drones: How Drones Are Revolutionizing Agriculture and How to Break into this Booming Market. https://uavcoach.com/agricultural-drones/

[5] Herrmann JW, Lin ET, Al E. (2001). Affordable space systems manufacturing: Intelligent synthesis. Paper presented at the CD-ROM ASME Design Engineering Technical Conferences.

[6] Arakliotis S, Nikolos DG, Lawris KE. (2016). A Rule Based Arduino Programming System for Young Students. Paper presented at the International Conference on Modern Circuits and Systems Technologies

[7] Curry JM, Ohnson ER, Tarnes JH. (2015). Effect of dropped plies on the strength of graphite-epoxy laminates. AIAA Journal, 30(2):449-456 Cereveny R.Hamilton and the hurricane. Weatherwise. 2016;69(5):42-48

[8] S.Khan, L.Arago and J. Iriarte,"A uavlidar system to map amazonian rainforest and its landscape transformations," International Journal of Remote Sensing , vol. 38, no. 8-10,pp. 2313-2330, 2017.

[9] X. Guo, S. Denman, C. Fookes, L. Meijias, and S. Sridharan,"Automatic uav forced landing site detection using machine learning," in 2014 *International conference on digital image computing: Techniques and Applications (*DICTA), Nov 2014,pp. 1-7.

[10] Polly, J.; Hegarty-Craver, M.; Rineer, J.; O'Neil, M.; Lapidus, D.; Beach, R.; Temple, D.S. The use of Sentinel1 and -2 data for monitoring maize production in Rwanda. In Proceedings of the Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI. Int. Soc. Opt. Photonics 2019, 11149, 111491.

[11] Zhang, C.; Kovacs, J.M. The application of small unmanned aerial systems for precision agriculture: A review. Precis. Agric 2012, 13, 693–712.

[12] Hung, C.; Xu, Z.; Sukkarieh, S. Feature Learning Based Approach for Weed Classification Using High Resolution Aerial Images from a Digital Camera Mounted on a UAV. Remote Sens. 2014, 6, 12037–12054.

International Journal of Advanced Science and Technology
Vol. 29, No. 03, (2020), pp. 12283 - 12291

Interpretation of One Dimensional Hilbert Transform in terms of Fourier Transform and its Applications

Dankan Gowda V[1], Sridhara S B[2] , P.K Prakasha[3], Naveena Pai G[4], Naveen K B[5]
[1]Assistant Professor, Department of Electronics and Communication Engineering, B.M.S. Institute of Technology, Bangalore, Karnataka (State), INDIA.
[2]Professor & HoD, Department of Electronics and Communication, VijayaVittala Institute of Technology, Bangalore, Karnataka (State), INDIA
[3]Associate Professor, Department of Electronics and Communication, Donbosco Institute of Technology, Bangalore, Karnataka (State), INDIA.
[4]Assistant Professor, Department of Electronics and Communication, Yenepoya Institute of Technology, Moodbidri, Karnataka(State), INDIA
[5]Associate Professor, Department of Electronics and Communication, BGS Institute of Technology, Adichunchanagiri University, Karnataka(State), INDIA
E-mail: dankan.v@bmsit.in

Abstract

This Paper presents an mathematical and numerical study of the one Dimensional Hilbert transform method and signal analysis. In the one-dimensional case the basic relations of Hilbert transforms are reviewed and derived. The interpretation of Hilbert transform domain in conditions of Fourier renovate is analyzed and derived. Data processing techniques using Hilbert transforms have been discussed. The theory of the one-dimensional Hilbert transform is reviewed and the application of 1D Hilbert transform in quantitative interpretation of magnetic field as well as their derivatives are also investigated.

Keywords: *Hilbert transforms, one-dimensional, two-dimensional, magnetic field, Signal Processing.*

1. Introduction

This paper cares with the application of Hilbert integral transform and associated problems on two dimensional potential field data and two dimensional signal analyses. After the event of 1 dimensional Hilbert transform was made by Gold and skim Cizek, this renovate has create more and more applications in hypothesis and carry out of digital signal dispensation [1]. The fundamental relations of 1D Hilbert transform are reviewed and derived. These relations have become very significant especially in the Signal Processing Applications. In magnetic data application is elaborated as two specific methods; the first one uses point of intersections between the vertical magnetic effect of anomalous bodies and its Hilbert transform, several models are tested with different polarization angle and the second method deals with the magnetic anomaly and its Hilbert transform envelope and phase of such an anomaly[2].

In gravity processing , the physical parameters of a dipping fault are often obtained by Fourier transform, second straight derivative of the enormity outcome of the dipping liability its Hilbert transform [3]. This method is extended from the theoretical models to the field data using 1D discrete Hilbert transform and cubicspline method[4]. The gravity effect of a vertical fault will also be analyzed as a special case of the dipping fault.

In Signal application, the 1D signal is constructed by using the complex envelope (amplitude and

ISSN: 2005-4238 IJAST

phase) of same signal, which can be obtained by using a discrete Hilbert transform algorithm. The minimum phase characteristics of a wavelet, which can be calculated by using Hilbert transform, has many uses[5]. It may be used for feedback filtering such as inverse and recursive filtering. The discussion will also be extended to Lowpass and bandpass filtering which can be directly articulated in stipulations of Hilbert renovate.

2. Review of Literature

The development of the one-dimensional Hilbert transform is closely patterned after Gold and Radar (1969); Cizek (1970)[6]. The 1-D Hilbert transform is often defined as a Quadrature filter which introduces a 90 degree phase shift. In a complex function the real and imaginary parts are related together through the Hilbert transform.

"In almost every field where Fourier transform techniques are used to represent and analyze physical processes, one finds that there are situations where exist associations flanked by the real and imaginary parts or the magnetic and the phase of the fourier transform".[7] These relationships are frequently represented by one or other of the Hilber transform relations.

"In this paper we shall derive and review a number of such relationships that are important in both the theory and application of digital data processing"[8, 10]. We shall see, for instance , that if a sequence is causal, then the important and imaginary part of its Fourier transform or the even and odd components of the function are related by a Hilbert transform. Such a transform can be used in constructing the envelope and phase of a time function [11, 15].

The task of designing a 1-D Hilbert transform with different length operators and several windows such as Hamming, Triangular, and Blackman windows is always to obtain the best approximation to the exact transform.

3. One Dimensional Hilbert Transforms Equations

In this section we will review some relations before attempting to derive Hilbert transform equation in integral form which is considered as a starting point of discrete Hilbert transform algorithm. "Recent years, the Hilbert transform has found more and more applications in hypothesis and carry out of digital signal dispensation".

As we know the Fourier change of a function f(t) is F(w), then

$$F(w) = \int_{-\infty}^{\infty} f(t)e^{-jwt}\ dt \qquad \text{for all } w \quad (3.1)$$

And

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(w)e^{jwt}\ dw \qquad \text{for all } t \quad (3.2)$$

But equation (3.1) may also be written as

$$F(w) = \int_{-\infty}^{\infty} f(t)\,[\cos(wt) - j\sin(wt)]\,dt \qquad (3.3)$$

where;

$$e^{-jwt} = \cos(wt) - j\sin(wt)$$

Then

$$F(w) = \int_{-\infty}^{\infty} f(t)\cos(wt)\,dt - j\int_{-\infty}^{\infty} f(t)\sin(wt)\,dt \qquad (3.4a)$$

So equation (4a) takes the following form;

$$F(w) = F_r(w) - jF_i(w),$$

Where $F_r(w)$ and $F_i(w)$ are the real and imaginary parts of the equation (3.4a), such that

$$F_r(w) = \int_{-\infty}^{\infty} f(t)\cos(wt)\,dt$$

And

ISSN: 2005-4238 IJAST

$F_i(w) = \int_{-\infty}^{\infty} f(t)\sin(wt)\,dt$(3.4b)
It also follows that
$F(-w) = F_r(w) + jF_i(w)$(3.5)
Because

$$F_r(-w) = F_r(w)\text{[for even function]}$$
$$F_i(-w) = -F_i(w) \qquad \text{[for odd function]}$$

Therefore we should be able to write
$F_r(w) = \frac{[F(w)+F(-w)]}{2}$ (3.6a)
$F_i(w) = \frac{j[F(w)-F(-w)]}{2}$ (3.6b)
From equation (3.2), the inverse Fourier transform may be divided into regions of positive and negative frequency

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{0} F(w)e^{jwt}\,dw + \frac{1}{2\pi}\int_{0}^{\infty} F(w)e^{jwt}\,dw \qquad \text{for all } t$$

$= \frac{1}{2\pi}\int_0^{\infty} F(-w)e^{-jwt}\,dw + \frac{1}{2\pi}\int_0^{\infty} F(w)e^{jwt}\,dw$ (3.7)
Upon substituting equation (3.6), we have equations with positive frequency only
$f(t) = \frac{1}{\pi}\int_0^{\infty} F_r(w)\cos(wt)\,dt + \frac{1}{\pi}\int_0^{\infty} F_i(w)\sin(wt)\,dw$ (3.8)
This may be printed in stipulations of a complex time $t_c = t + j\sigma$
$\lim_{\sigma\to 0} f(t) = \lim_{\sigma\to 0}\frac{1}{\pi}\int_0^{\infty}[F_r(w)\cos(wt) + F_i(w)\sin(wt)]\,e^{-\sigma w}dw$ (3.9)
Next let us define a complex function;
$f_c = f(t,\sigma) - j\,f_H(t,\sigma)$(3.10)
Such that the complex function is made of the original real part f(t) and a part called the quadrature function, $f_H(t)$, which will be shown to be the Hilbert transform of f(t). The quadrature is defined so that it introduces a 90^0 phase shift because we will sometimes want to use it to obtain the envelope of the real time function. Therefore, sines are converted to cosines and cosines are converted to negative sines in equation (3.8);
$f_H(t,0) = \frac{1}{\pi}\int_0^{\infty}[F_i(w)\cos(wt) - F_r(w)\sin(wt)]dw$ (3.11)
And
$f_c(t,\sigma) = \frac{1}{\pi}\int_0^{\infty}[F_r(w) - jF_i(w)]\,e^{jwt}\,e^{-\sigma w}\,dw$(3.12)
Substitution of F_r and F_i from equation (3.4b) into the equation (3.11) will relate f_H and f more clearly.

$$f_H(t) = \frac{1}{\pi}\int_0^{\infty}\int_{-\infty}^{\infty}[f(T)\sin(wT)\cos(wt) - f(T)\cos(wT)\sin(wt)]dT\,dw$$

$= \lim_{w'\to\infty}\frac{1}{\pi}\int_0^{w'}\int_{-\infty}^{\infty} f(T)\sin w(T-t)\,dT\,dw$(3.13)
Integrating first with respect to dw we obtain;
$f_H(t) = \lim_{w'\to\infty}\frac{1}{\pi}\;P\int_{-\infty}^{\infty} f(T)((1-\cos w'(T-t))/(T-t)\,dT$(3.14)
Where, P indicates the Cauchy principal value. It can be shown that;
$\lim_{w'\to\infty}\frac{1}{\pi}\;P\int_{-\infty}^{\infty} f(T)(\cos w'(T-t))/(T-t)\,dT = 0$ (3.15)
Using the Riemann-Lebesque Lemma we have;
$f_H(t) = -\frac{1}{\pi}\;P\int_{-\infty}^{\infty}\frac{f(T)}{t-T}dT$ (3.16)
Equation (3.16) is known as the HT in integral form. Substitution of F_r and F_i from equation into equation (3.11) gives

$$f_H(t) = \frac{j}{2\pi}\int_0^{\infty}[F(w) - F(-w)]\cos(wt)\,dw - \frac{1}{2\pi}\int_0^{\infty}[F(w) + F(-w)]\sin(wt)\,dw$$

ISSN: 2005-4238 IJAST

$= \frac{j}{2\pi}\int_0^{\infty}\left[F(w)e^{jwt} - F(-w)e^{-jwt}\right]dw$ (3.17)

$= \frac{1}{2\pi}\int_0^{\infty} j\, F(w)e^{jwt}\; dw - \frac{1}{2\pi}\int_{-\infty}^{\infty} jF(w)e^{jwt}\; dw$(3.18)

If we let

$F_H(w) = j\, F(w)\, sgn(w)$ (3.19)

Where Sgn is a signum function

$Sgn = \begin{cases} 1; & w > 0 \\ 0; & w = 0 \\ -1; & w < 0 \end{cases}$(3.20)

Then equation (3.18) shows that f_H and F_H are inverse Fourier transforms

$f_H(t) = \int F_H(w)e^{jwt}\, dw$ (3.21)

From equation (3.19) and (3.11) it is also possible to obtain the inverse Hilbert transform.

$f(t) = \frac{1}{\pi}\; P\; \int_{-\infty}^{\infty}[f_H(T)/(T-t)\; dT$(3.22)

In the equation (3.19) we can define a Fourier transform, $Q(w)$ as

$Q(w) = j\, Sgn(w)$(3.23)

The Fourier series of $Q(w)$ is q_n

$$q_n(n\Delta t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} Q(w)e^{-jwt}\quad dw$$
$$= j\,\frac{1}{2\pi}\int_{-\pi}^{0} e^{-jwn\,\Delta t} - j\,\frac{1}{2\pi}\int_{0}^{\pi} e^{-jwn\,\Delta t}\quad dw$$
$$= \frac{1}{2\pi}\left(-1 + e^{(jn\pi\Delta t)} + e^{(-jn\pi\Delta t)} - 1\right)$$

$q_n(n\Delta t) = \begin{cases} 0; & n \text{ even} \\ -\frac{1}{n\pi\Delta t}; & n \text{ odd} \end{cases}$ (3.24)

As $\Delta t \to 0$ the Fourier series becomes a Fourier integral and $n\Delta t$ becomes the continuous time variable t, i.e.,

$q(t) = \frac{-1}{\pi t}$ (3.25)

The function $q(t)$ and its Fourier transform, $Q(w)$, are plotted in the Figure. 1. By equation (3.19) the Hilbert transform in the FD is given by the product $F(w)$ and $Q(w)$.

$F_H(w) = F(w)Q(w)$ (3.26)

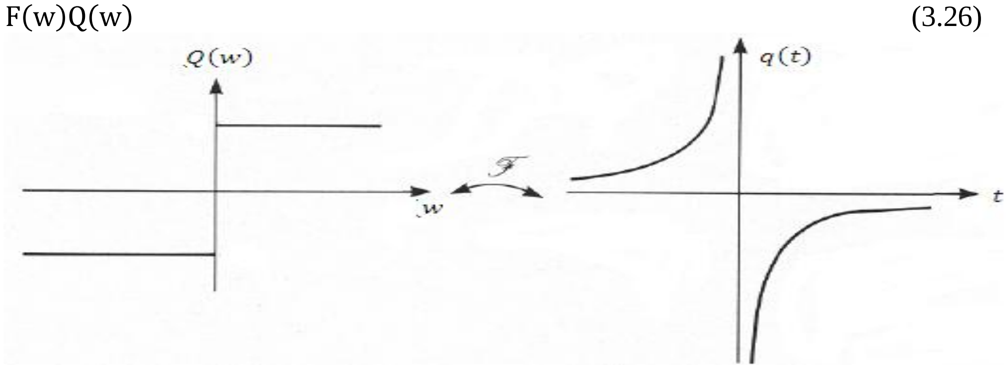


Figure.1. Represents the function $Q(w)$ and its inverse transform, $q(t)$.

Table 1. Some Simple functions and their Hilbert Transforms

Function	Hilbert Transform
$f(t)$	$F_H(t)$

ISSN: 2005-4238 IJAST

$\delta(t)$	$\frac{-1}{\pi t}$
$\sin t$	$\cos t$
$\cos t$	$-\sin t$
$(\sin t)/t$	$(\cos t - 1)/t$

By the convolution theorem it is possible to obtain the Hilbert transform in the TD by convolving $q(t)$ with $f(t)$.
$f_H(t) = q(t) * f(t) = (-\pi t)^{-1} * f(t)$ (3.27)
Writing out the convolution indicated by the star in equation (3.27) we can obtain the equation (3.16) again. It can be also shown that if Hilbert transform is applied twice successively we should reverse all the phases in the Fourier harmonics and obtain the original function back.
$f(t) = -(-\pi t)^{-1} * f_H(t) = -(-\pi t)^{-1} * (-\pi t)^{-1} * f(t)$(3.28)
Hilbert transforms of few simple functions are listed in Table 1.

4. One Dimensional Hilbert Transform Relations in Wave Envelope and Phase Study

The envelope, E, of any real function, $f(t)$, can be obtained by taking the modulus of equation (3.10)
$E(t) = [f^2(t) + f_H{}^2]^{1/2} = [f_c f_c{}^*]^{1/2}$ (4.1)
The instantaneous phase, $\emptyset$, can be also given as a function of time
$\emptyset(t) = \arctan\left[\frac{f_H(t)}{f(t)}\right]$, (4.2)
The instantaneous frequency, $\emptyset(t)$, is obtained from the rate of change of phase. The complex function f_c may then be written as follows:
$f_c = E(t)\exp[-j\,\emptyset(t)]$ (4.3)
Solving for the phase we have
$\emptyset = j\log\frac{f_c}{E(t)}$. (4.4)
Letting
$F_c = \frac{f_c}{E}$ (4.5)
And differentiating with respect to t, yields the instantaneous frequency
$\emptyset'(t) = \frac{d\emptyset}{dt} = \frac{j\left(\frac{dF_c}{dt}\right)}{F_c}$ (4.6)
Following Claerbout, this may be approximated by the imaginary part, I_m , of the difference equation;
$\emptyset(t_i) \cong I_m\left[\frac{2\{Fc(t_i)-Fc(t_{i-1})\}}{\{\Delta t\,(Fc(t_i)+Fc(t_{i-1})\}}\right]$(4.7)
Where, i=0,1,2,3,4,……
The instantaneous frequency is useful for functions that show dispersion. It is often used in reflection Seismology to detect zones of interferences indicating rapid changes in the elastic impedance in the earth.

5. The Minimum Phase Concept

"A linear system is said to be physically realizable and its impulse response is causal if it vanishes if the time is negative". [16]
$f(t) = 0\ ;\ t < 0$ (5.1)
As we know, any function can be expressed as the sum of an even and odd function.
$f(t) = f_e(t) + f_o(t)$(5.2)
$= \frac{[f(t)+f(-t)]}{2} + \frac{[f(t)-f(-t)]}{2}$(5.3)

ISSN: 2005-4238 IJAST

But from equation (5.1) , we have
$f_e(t) = -f_o(t)$ for $t < 0$ *and* $f_e(t) = f_o(t)$ for $t > 0$ (5.4)
For a physically realizable linear system. Therefore, for causal signals, we can introduce the signum function such that
$f_e(t) = f_o(t)\ \mathrm{sgn}(t)$ and $f_o(t) = f_e(t)\mathrm{sgn}\ (t)$(5.5)
Similarly, the real part of the Fourier transform of $F(w)$ is even whereas the odd is the imaginary part.

$$F_e(w) = \frac{-j}{\pi}\ P \int_{-\infty}^{\infty}\left[\frac{F_0(w)}{\omega - W}\right] dw = F_r\ (\omega)$$

$$F_o(w) = \frac{-j}{\pi}\ P \int_{-\infty}^{\infty}\left[\frac{F_e(w)}{\omega - W}\right] dw = jF_j\ (\omega) \qquad (5.6)$$

Thus as a result of (5.5) and because of the similarity between $F(w)$ in equation (3.19) and the even and odd properties of $F(w)$, the causal function has the real and imaginary parts are indeed Hilbert transforms;
The transfer function may be written as
$Y(w) = |Y(w)| \exp[j\emptyset(w)]$. (5.7)
Taking log on both sides, we have
$\mathrm{Log}\ Y(w) = \mathrm{Log}\ |Y(w)| + j\emptyset(w)$. (5.8)
If the impulse response is causal function it follows that we can compare $\mathrm{Log}\ Y(w)$ to $F(w)$ in equation (3.4b), thus $\mathrm{Log}\ |Y(w)|$ and $\emptyset(w)$ are Hilbert transforms just as $F_r(\omega)$ and $F_i(\omega)$ in equation (5.7) and (5.8)

$$\mathrm{Log}\ |Y(w)| = \frac{-1}{\pi}\ P \int \left[\frac{\emptyset(T)}{T - w}\right] dT \qquad (5.9)$$

$$\emptyset(w) = \frac{1}{\pi}\ P \int \left[\frac{\mathrm{Log}\ |Y(w)|}{T - w}\right]\ dT \qquad (5.10)$$

The Log of the transfer function, $Y(w)$, must be analytical and bounded in the minor half of the z-plane must no poles or zeros so that it will be a minimum phase system.

6. Application of 1-D Hilbert TransformMagnetic and Gravity Data Interpretation

In recent years, Hilbert transform techniques have found more and more applications in magnetic and gravity data interpretation particularly in fast quantitative elucidation of seriousness and magnetic data. The transformation of vertical gradients (vertical magnetic field) into horizontal gradients (horizontal magnetic field) and vice-versa with help of Hilbert transform.
Derived simple relations between the vertical magnetic anomalies of sheets (finite and infinite depth extent) and their Hilbert transforms. Similar relations were also extended to simple dikes and horizontal circular cylindrical bodies.
A simple automatic direct method of interpretation which is intended for determining uniquely the physical parameters of two dimensional structures (eg. Semi-infinite dikes and the) horizontal cylinders. The method is flexible enough to encompass total vertical and horizontal field measurements at all latitudes. Linear regional gradients do not hamper the application of the method. Remnant magnetization does not affect the determination of the main parameters, i.e. depth and location. Unlikely most other methods no knowledge of origin or datum level is required in this method.

6.1 Seismic signal Analysis

“In recent years, the discrete Hilbert transform has found more and more applications in hypothesis and carry out of digital signal dispensation”. The representation in hypothesis and carry out of digital signal dispensation. “The representation of a sign because the real a part of a posh function in time may be usefuldevice in many areas of signal analysis”[17].

ISSN: 2005-4238 IJAST

“In seismic application, the traditional seismic signal are often viewed because the real component of a posh signal which may be uniquely calculated under usual conditions”. Application of the Hilbert transform (1-D) in seismic signal analysis has been very useful, especially once we represent the signal as “The real a part of a posh function in time”. “In general terms, the benefits of the complex envelope stem from the natural separation of amplitude information from angle information”. “In a true signal, these are blended in such how which may be confusing to visual analysis”[18-19]. In. this section we'll see how the complex envelope of a seismic wavelet are often calculated.

6.2 Inverse Filters

“If we assume Y is that the known output of a filter B and X is an unknown input, then we've a drag that one-often-haswith a transducer/recorder system”.[20]For example, The output of a seismometer is widely line from which the seismologist might need to work out the displacement velocity or acceleration of the ground”. “To undo the filtering operation of the filter B(Z), we'll attempt to find another filter A(Z) where B(Z)A(Z)=1, meaning A(Z) is inverse of B(Z)”.
Let’s take an example, if B(Z)=1-Z/2, then, by Taylor’s series formula, we have

$$A(Z) = \frac{1}{1-\frac{Z}{2}} = 1 + \frac{Z}{2} + \frac{Z^2}{4} + \frac{Z^3}{8} + \cdots \tag{6.1}$$

We can represent this polynomial in a computer since the filter coefficients will drop off rapidly in magnitude. But if $B(Z) = 1 - 2Z$, then

$$A(Z) = \frac{1}{1-2Z} = 1 + 2Z + 4Z^2 + 8Z^3 + 16Z^4 + \cdots \tag{6.2}$$

Here the coefficients of the series increases without bound which in fact will produce a serious problem. The minimum-phase function has further applications in geophysics such as the feedback filtering or inverse filters, Homopophic filtering or Homomorphic deconvolution, zero-lag inverse filter, Minimum entropy decnvolution, etc.
It determines the minimum amount of dispersion in viscous wave propagation which is implied by causality. “The Homomorphic system may be a system which obeys a generalized principle of superposition, that's wont to divide the wavelet and therefore the impulse rejoinder of the transmission path from a seismic record”. “If the wavelet or the impulse rejoinder of the channel is minimum-phase, a seismic evidence is usually represent because the convolution of a wavelet with the impulse rejoinder of the communication path”. “Since the wavelet or the impulse retort of the broadcast path obeys the minimum-phase assumption, the seismic record are going to be a minimum phase, during which case, the Homomorphic filter works nicely but a drag will arise when the seismic record may be a mixed-phase i.e both of the wavelet and therefore the impulse reaction of the transmission aren't minimum phase, during which the Homomorphic system fails”. to unravel this problem we've to vary the seismic record which is mixed phase to minimum phase and this might be achieved by application of Hilbert transform.

6.3 Design of 1-D Hilbert Transform Operator

The ideal Hilbert Transform can numerically be obtained in the case of finite duration approximations, the standard techniques of windowing, frequency sampling and equi-ripple approximation can be applied in approximating the characteristics of the ideal Hilbert transformer.
In the case of N=10 the impulse response appears to contains more than usual ripple in both sides of the main lobe and these ripples which appear in ht(n) have affected on the magnitude of H(ω). The benefit of using the windows is to reduce or attenuate the ripples, which inturn leads one to obtain a smooth ht(n) and H(ω). Therefore, it is necessary to know which window is more suitable in obtaining the best approximation to the ideal Hilbert Transform.

ISSN: 2005-4238 IJAST

When we truncated ht(n) by Triangular window with N=10, it gives better resolution than impulse response itself. The $|H(\omega)|$ obtained with a triangular window is better because no ripples are noticeable when compared with $|H(\omega)|$ which is initially truncate by a boxcar.
In the case of Hamming window with N=10, the ripples are decreased in both sides of main lobe of ht(n) but the $H(\omega)$ is not smooth as we wish, and the $|H(\omega)| \neq 1$. Therefore, it is found that if we want to use the operator length of N=10, the window which will give the best resolution for ht(n) and $H(\omega)$ is the Blackman window.
The smaller the N is, the wider the ripples, and as N increases the ripples or the lobes become narrower. But the amplitude remains almost constant. In other words, as N increases the main lobe of ht(n) approaches to a spike and $|H(\omega)|$ approaches to true values.
In the conclusion we can therefore say that for any value of N the best approximation can be obtained when the Blackman window is incorporated in designing the Hilbert transformer.

6.4 One Dimensional Minimum-Phase

"The minimum-phase function has further applications in geophysics such as the feedback filtering or inverse filters, Homomorphic filtering or Homomorphic deconvolution, zero-lag inverse filter, minimum entropy deconvolution, etc. it determines the minimum amount of dispersion in viscous wave propagation which is implied by causality. In this section will discuss one of these uses which is related to seismic exploration".
The Homomorphic system is employed to disconnect the wavelet (seismic source) and therefore the impulse reaction of the transmission path from a seismic record". "If the wavelet or the impulse reaction of the channel is minimum-phase, a seismic evidence is usually represent because the convolution of a wavelet with the impulse reaction of the transmission path".[21,22] "Since the wavelet or the impulse reaction of the broadcast path obey the minimum-phase assumption, the seismic record are going to be a minimum phase, during which case, the Homomorphic filetr works nicely, but a drag will arise when the seismic record may be a missed phase that's both of the wavelet and therefore the impulse response of the transmission aren't minimum phase, during which case the Homomorphic system fails". [23,24] to unravel this problem we've to vary the seismic record which is mixed phase to minimum phase and this might be achieved by application of Hilbert transform or the opposite appropriate methods mentioned.
"The Hilbert transform has become important adjunct to the more familiar integral transforms, such s Laplace and Fourier transforms, in analyzing signals and systems". "It are often shown that a lot of ideal filtering processes are often expressed in terms of Hilbert transform". "By studying artificially constructed signals and proceeding to seismic records, it are often seen that the complex envelope which is obtained by 1-D Hilbert transform is more amenable to visual interpretation than the important signal itself". this is often attributed to the natural separation of amplitude information from phase information.

7. Conclusion

The Hilbert transform can be defined in different ways, and through different. Mathematical relations, but essentia1ly it is a quadrature filter which introduces a 90 degree phase shift. The real and imaginary components of a complex function, or the even and odd parts of a real function, are related through the Hilbert transform relationships. The integral equations of the transform are reviewed in this research work as a first step towards, the application discrete Hilbert transform relations. The one Dimensional Hilbert transform has many applications in Geophysics, especially in seismology, gravity and magnetic which are discussed.

ISSN: 2005-4238 IJAST

References

[1]. Fryze, S., Wirk–und Blind–und Scheinleistung in Elektrischen tromkreisen mit nichtsinusoidalenformigen Verlauf von Strom und Spannung. Elektrotech. Z., 53 (25), **(2004),** pp. 696–599.

[2]. Jackson, L. B. On the relationship between digital Hilbert transformers and certain low-pass filters. IEEE Trans. ASSP, vol. 23, no. 8, **(2005)**, pp. 381–383.

[3]. Krylov, W. W. and Ponomariev, D. M. Definition of the signal delay in linear two-ports using the Hilbert transform (in Russian). Radiotechnika i Elektronika, no. 5, **(1980),** pp. 204–206.

[4]. Rabiner, L. R. et al. FIR digital filter design techniques using weighted Chebyshev approximation. Proc. IEEE, vol. 63, no. 4, Apr. (2004), pp. 585–610.

[5]. Carson, J. R. and Fry, C. F., Variable frequency electric circuit theory with application to the theory of frequency modulation, Bell Syst. Techn. J., vol. 6, no. 4, **(1994),** pp. 513–530.

[6]. Dankan V Gowda, D. V. Kishore, Shivashankar, A. C. Ramachandra, and C. Pandurangappa, "Optimization of motorcycle pitch with non linear control," in Proceedings of the 1st IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT '16),**(2016),** pp. 1656–1660.

[7]. Gowda, Dankan V., and Sadashiva Chakrasali., " Comparative Analysis of Passive and Semi-Active Suspension System for Quarter Car Model using PID Controller," Int. Conf. on Recent Trends in Signal Processing, Image Processing and VLSI (ICrtSIV), **(2014),** pp. 156–162.

[8]. Dankan Gowda V., Ramachandra A C., "Slip Ratio Control of Anti-Lock Braking System with Bang-Bang Controller" International Journal of Computer Techniques ,Volume 4 Issue 1, Jan – Feb **(2017),** pp. 146–151.

[9]. Kishore, D.V., Gowda, D.V., Shivashankar, Mehta, S.: MANET topology for disaster management using wireless sensor network. In: International Conference on Communication and Signal Processing, April 6–8, **(2016)**, pp. 134–160.

[10]. Mahaveer Penna, Dankan V Gowda, J J Jijesh and Shivashankar, "Design and Implementation of Automatic Medicine Dispensing Machine", 2nd IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT), **(2017),**pp. 1962-1966,

[11]. Gowda, D.V. and Ramachandra, A.C., Importance of non-linear controller in implementing anti-lock braking system-a technical review. International Journal of Advanced Research in Computer Science, 9(2), (2018), pp.193-199.

[12]. D. Gowda, M. Penna, S. Palle, KA, S.P and K.N. Chaithra, "Digital Intercommunication System in Advanced Light Helicopter (ALH)", Advances in Computational Sciences and Technology, vol. 10, no. 5,**(2017)**, pp. 699-708.

[13]. Dankan V Gowda,Shivashankar,Varun Reddy S,Varun C A, Sahana M and T Rajesh,"Implementation of swarm intelligence in obstacle avoidance," Recent Trends in Electronics, Information & Communication Technology (RTEICT) in **(2017),** pp. 712-718.

[14]. Gowda, D.V., Pandurangappa, C., Kishore, D.V.; and Ramachandra, A.C., Two wheeler vehicle model development for driving simulator application. GIT Journal of Engineering and Technology, 9(1),**(2016),** pp. 47-51.

[15].Dankan V Gowda, Ramachandra A C, Thippeswamy M N, Pandurangappa C, Ramesh Naidu P 'Modeling and Performance Evaluation of Anti-lock Braking System', Journal of Engineering Science and Technology, Taylor's University, Vol. 14(5), **(2019),** pp. 3028-3045.

[16]. Dankan V Gowda, Ramachandra A C, Thippeswamy M N, Pandurangappa C, Ramesh Naidu P 'Synthesis and Modelling of Antilock Braking System using Sliding Mode Controller', Journal of Advanced Research in Dynamical and Control Systems', Vol 10(12),**(2018)**, pp. 208-221.

[17]. Ramesha M, T .Venkata Ramana "Design and Implementation of fully pipelined 64-point FFT Processor in a FPGA". International Journal of Applied Engineering Research [IJAER], ISSN 0973-4562 Volume 11, Number 6 **(2016)** pp 3940-3943.

[18]. Dankan V Gowda, Ramachandra A C, Thippeswamy M N, Pandurangappa C, Ramesh Naidu P 'Automotive braking system simulations V diagram approach', International Journal of Engineering & Technology, Vol 7(3),**(2018),** pp.17-40.

[19]. Ramesha M, Dr. T .Venkata Ramana "A Novel Architecture of FBMC Transmitter Using Polyphase Filtering and its FPGA Implementation". Indian Journal of Science and Technology [IJST], Vol9 (48), **(2016),** pp.117-126.

[20]. Ramesha M, Dr. T .Venkata Ramana "Performance Analysis of FBMC Transceiver Architecture over OFDM System" Far East Journal of Electronics and Communication [FJEC], Volume 17(6), **(2017),** pp 1353-1372.

[21]Sridhara S.B, Ramesha M, Veeresh Patil, "Adaptive Scheduling Design for Time Slotted Channel Hopping Enabled Mobile Adhoc Network", International Journal of Advanced Computer Science and Applications [IJACSA], Vol. 11, No. 3, **(2020)**,pp 272-277

[22]. S. L. Hahn, "Comments on `A Tabulation of Hilbert Transforms for Electrical Engineers'," IEEE Trans. on Commun., vol. 44, July **(1996),** p. 768-774.

[23]. S. L. Hahn, \Hilbert transforms," in The Transforms and Applications Handbook (A. Poularakis, Ed.), Boca Raton FL: CRC Press, vol. 5,**(2004)**, pp.163-275.

[24]. J.S. Bendat: "The Hilbert Transform and Applications to Correlation Measurements", Brüel&Kjær, **(2005).**

ISSN: 2005-4238 IJAST

Automated Computer Vision based Weed Removal Bot

Gaurav Sethia
Dept of Computer Science and Engineering
BMS Institute of Technology and Management
Bangalore, India
gaurav08@ieee.org

Harish Kumar S Guragol
Dept of Computer Science and Engineering
BMS Institute of Technology and Management
Bangalore, India
harish.guragol@ieee.org

Swati Sandhya
Dept of Computer Science and Engineering
BMS Institute of Technology and Management
Bangalore, India
sadikas2000@ieee.org

Shruthi J
Assistant Professor, Dept of Computer Science and Engineering
BMS Institute of Technology and Management
Bangalore, India
shruthij@bmsit.in

Rashmi N
Assistant Professor, Dept of Electronics and Communication Engineering
BMS Institute of Technology and Management
Bangalore, India
rashmiswamy@bmsit.in

Abstract— **Weeds are a dangerous factor for a good yield of crops. The conventional method of removing weeds was either plucking manually or spraying herbicides uniformly all over the field. Spraying herbicides not only contaminates crops but also gives rise to many health-related issues. The purpose of the paper is to develop a mobile model that can detect weeds in real- time with their position coordinates and scrape them off. The model first scans the specific area for leaf detection and classifies it as weed or crop with a prediction accuracy of 99.5%. If the classified leaf is a weed, the coordinates are found and the robotic arm removes them with the help of a high-speed rotating blade, without harming the crops and environment. The left outs can further be utilized as fertilizer and no harmful chemicals have been used.**

***Keywords*—Weed Removal, Image Processing, Smart weeding, Mask RCNN, VGG 16, Smart Agriculture, Delta Robot**

I. Introduction

India's economy is dependent on Agriculture for more than 70%. The source of income for more than half of the population is directly related to the yield obtained in the fields. Farmers working in the field face many challenges in the whole period of crop generation, among which weeds are one of the major threats to the natural environment. They are destroying native habitats and the desired crop which leads to poor production of crops and raised economic challenges. It's a major need for farmers to remove weeds from the crop field to have a better crop production but in a more efficient manner.

As mentioned, one of the major issues in agriculture is the control of weeds growing among the plantation crops. Weed tends to snatch space, nutrients, sunlight, water, and hardly get affected by changing the natural environment. Moreover, the extreme conditions of the environment help them, even more, to sustain as it affects native crops easily. At present, these kinds of plants are being removed manually, wherever possible, or weedicides are being sprayed uniformly all over the field to keep them under check. In conventional weed control systems, herbicides are sprayed uniformly all over the field. This technique is very inefficient as only about 20% of the spray reaches the plant and less than 1% of the chemical contributes to weed control, leading to wastage, contamination of the environment, and health problems in people [1]. To avoid these consequences, an effective weed control system should be employed. These systems must be capable of locating weeds in the field and remove them. The left-outs are eventually decomposed and can serve as fertilizer.

As the major goal of Sustainable Farming is to increase the yield-reducing its reliance on pesticides, herbicides, and to control the growth of the weed. Precision farming techniques are required to address this challenge with many researchers working on this with their vision to create an effective solution. With the development of AgriBot to perform the various agriculture activities are communicated using Wi-Fi technology [2]. Researchers have developed a weed detection and classification method for weed control robots in cornfields with a computer vision algorithm to classify plants like weeds or crops with their properties [3]. An Autonomous Robotic System for Mapping Weeds in Fields [4], is used to identify by aerial image analysis of areas with high bio-mass density, thus indicating areas with weed infestations. It also sparks an idea to identify weeds based on spectral analysis. Visual odometry System is used for weed detection by capturing images with the help of mobile robots [5]. An effective classification system of plants and weeds [6], aims to focus on a vision-based perception system to identify the value and distinguish it from the weed plants. Traditional methods such as Plant identification using leaf images [7] extract features and classify based on geometrical parameters extracted by digital image processing. A self-supervised training method in the context of RGB imaging provides an apt framework for hyperspectral crop/ weed discrimination with prior knowledge of seeding patterns [8]. Besides, much research on autonomous vision systems using features like color, texture, and shape analyzed with different algorithms and techniques such as crop row detection by principal component analysis [9]. Currently, manual weeding is a tedious task to perform mechanical weed control is considered for automation achieved by field robot BoniRob [10] with a very high-speed vision-based weed control and development of a low-cost delta robot [11] for the weed control in organic farms. Advanced robots like Oz weeding robot, which helps during weeding and hoeing chores to increase farm profitability while respecting the environment

978-1-7281-6828-9/20/$31.00 ©2020 IEEE

[12], are contributing enormously for better productivity. A lot of models have already been generated in the market as a product, like ecoRobotix which is an autonomous weeding robot for row crops, meadows, and intercropping cultures which detect and selectively spray the weeds with a microdose of herbicide [13]. Consequently, a vast variety of autonomous weed control systems are available and are being developed to address the challenges.

The first step involves image segmentation which separates the plant area from background pixels and classifies the leaf as weed or crop. Spraying herbicide uniformly leads to contamination of crops as well as wastage of costly chemicals. Patch spraying can help in saving chemicals but it's again a disadvantage for the environment. Excessive use of chemicals leads to health problems, affected lungs, skin problems, etc. Instead, removal or uprooting of weed can be a better solution. A mobile system that could make use of the real-time data to find the coordinates of leaf position with computer vision and help in the removal of weeds using the robotic arm would be much more efficient and minimize environmental damage.

II. METHODOLOGY

The proposed system removes the weed in three major steps. Fig. 1 shows the process flow of the system. First, an image is captured employing a Raspberry Pi camera, that is mounted facing downwards, i.e., towards the ground. The image is passed to Raspberry Pi with the assistance of Camera Serial Interface (CSI). With the help of the RGB image obtained from the camera, the leaf is assessed as weed or crop using a set of geometrical parameters. If the identified leaf is a weed, then a delta robotic arm with a high-speed rotating blade as the end-effector reaches the leaf coordinates and cuts the leaf. If the identified leaf is a crop and not a weed, then the robot moves forward and captures another image, such that the antecedent captured space isn't captured once more.

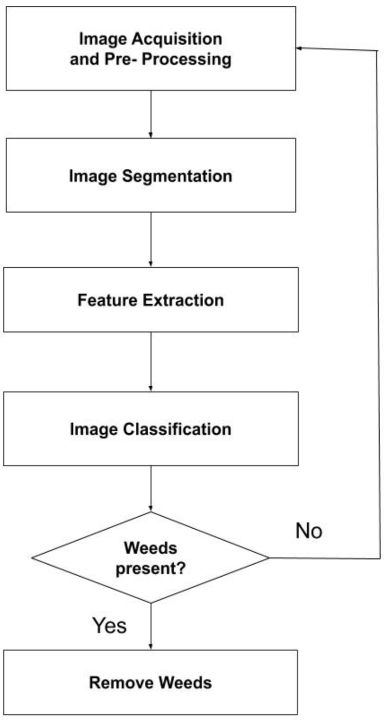


Fig 1 Process Flow

III. WEED IDENTIFICATION

A. Image Acquisition and Pre-processing:

The image captured from the raspberry pi camera is rescaled to the image size on which the algorithm is trained. The image is then filtered using low and medium pass filters to remove noise. Gaussian blur, with a kernel size of 5*5, is applied to the filtered image. All these steps help us get a clear image of the leaf, which results in a more accurate prediction.

B. Image Segmentation:

Image Segmentation is the process of partitioning a digital image into multiple segments. The pre-processed image contains a lot of background details that are not required by the algorithm and has to be removed. To provide better input to the next automated image processing technique we need to extract the leaf pixels from the image and discard others. We extract our region of interest (ROI) using semantic segmentation. Mask Region Convolutional Neural Network (Mask RCNN) [14] has been proven to outperform other semantic segmentation algorithms, and hence Mask RCNN neural network is used to extract the leaf region from the preprocessed image. Fig. 2 explains the architecture of the Mask RCNN neural network. ResNet 101 architecture is used as the backbone model for Mask RCNN and features extracted are passed to a Region Proposal Network. The ROI obtained the network computes Intersection over Union (IoU) to filter weak predictions. And finally, we add a mask branch which returns the region mask for all the detected objects. The segmented image mask is fed for image classification, hence processing one leaf at a time will overcome the problem of multiple leaves in a single frame.

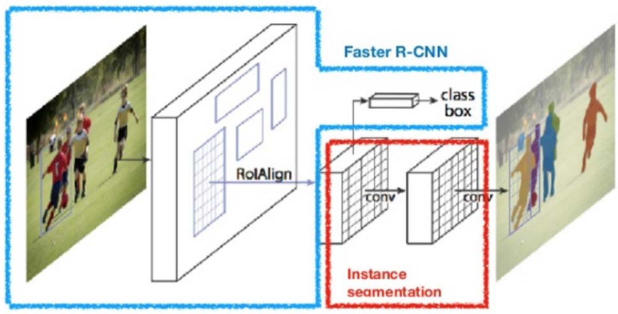


Fig 2 Mask RCNN Architecture

C. Feature Extraction:

Feature extraction is employed to extract relevant features for the recognition of plant leaves. The redundancy is far away from the image and therefore the leaf images are represented by a set of numerical features. They are calculated after the image pre-processing. The subsequent features are derived from the geometric parameters. The subsequent ratios are extracted and computed from the leaf image [15].

1. Rectangularity - It is the ratio of the product of major and minor axis of the ROI to the area of the ROI.

$$\text{Rectangularity} = \frac{Length*Width}{Area}$$

2. Narrow Factor - It is the ratio of the product of twice the radius of the leaf to the major axis.

$$\text{Narrow Factor} = \frac{Diameter}{Length}$$

3. Perimeter to Diameter Ratio - It is the ratio of the perimeter of the ROI to the diameter.

$$\text{Perimeter to Diameter Ratio} = \frac{Perimeter}{Diameter}$$

4. Form Factor - It is the ratio of 4 times the product of PI and area of the ROI to the square of perimeter of the ROI.

$$\text{Form Factor} = \frac{4 * \pi * Area}{Perimeter^2}$$

5. Eccentricity - It is the ratio of the foci of the ellipse to the major axis length of the ROI.

$$\text{Eccentricity} = \frac{Foci}{Length\ of\ Major\ Axis}$$

6. Aspect Ratio - It is the ratio of the major axis to the minor axis of the ROI.

$$\text{Aspect Ratio} = \frac{Length}{Width}$$

7. Perimeter to length and Breadth Ratio - It is the ratio of the perimeter of the ROI to the sum of length and breadth of the ROI.

$$\text{Perimeter to Length and Breadth} = \frac{Perimeter}{Length+Breadth}$$

D. Image Classification:

The feature values were calculated and passed to a classification algorithm, to make the prediction. Many algorithms were used to precisely predict the class of the leaf. If the algorithm found a similarity between leaf and the features of images in defined classes of weed, the model classified leaf as weed or else trained model for the detected leaf. The algorithm used includes k Nearest Neigbhors (kNN) with K value as 1, Support Vector Machine (SVM) with radial basis function kernel, Random Forest with 150 trees in the forest, and Decision Tree with a maximum depth of 7. The test set is 20% of the dataset. In the second approach, we use a Convolutional Neural Network (CNN). We use VGG16 architecture, which has 16 layers that have weights and approximately 138 million parameters [16] shown in Fig. 4.

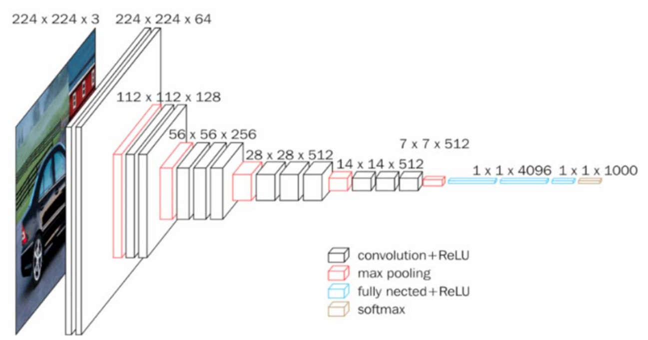


Fig 3 VGG 16 Architecture

We train the model with 100 epochs. initial learning of 0.001 and batch size of 32 images. We directly pass the segmented image to the network for classification. Both approaches were tried on a dataset with 31 different species of leaves, having approximately 60 images per class [17].

IV. HARDWARE

The robot hardware includes a raspberry pi board as the brain, a delta arm that cuts the weed and land moving robot which helps in traversing the field. Both the delta arm and the land robot are controlled by the raspberry pi. Fig. 4 shows the CAED model of the robot.

Fig 4 Top: CAED Model of the robot;
Bottom: Fabricated Model

A. Delta Arm:

Delta arm is used to move the blade in the 3D world. The end-effector of the arm has a high-speed brushless motor that powers the blade. The blade rotates at high speed and cuts the weed plants. Fig. 5 shows the delta structure.

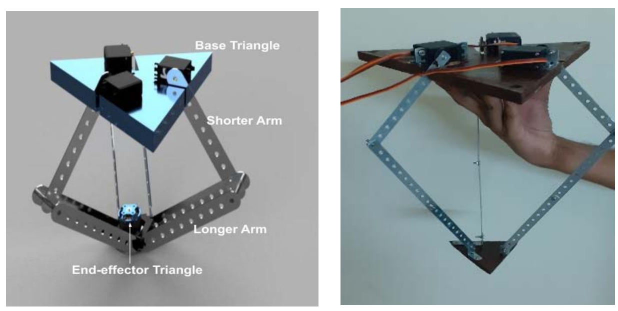


Fig 5 Structure of Delta Arm

Table 1 Dimensions of Delta Arm

Component	Dimension (cm)
Base Equilateral Triangle (s_B)	29
End Effort Equilateral Triangle (s_P)	10
Bicep / Shorter Arm (L)	14
Forearm/ Longer Arm (l)	28

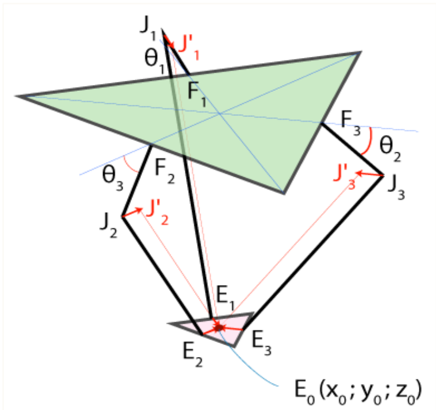


Fig 6 Kinematic Labeling

The angle of joint i to place the end effector at the desired location [18], is given by

$$\theta_i = 2\tan^{-1}(t_i)$$

where, i = 1, 2 & 3

and $t_i = \frac{-F_i \pm \sqrt{{E_i}^2+{F_i}^2-{G_i}^2}}{G_i - E_i}$

where,

$E_1 = 2L(y+a)$

$F_1 = 2zL$

$G_1 = x^2 + y^2 + z^2 + a^2 + L^2 + 2ya - l^2$

$E_2 = -L(\sqrt{3}(x+b) + y + c)$

$F_2 = 2zL$

$G_2 = x^2 + y^2 + z^2 + b^2 + c^2 + L^2 + 2(xb + yc) - l^2$

$E_3 = -L(\sqrt{3}(x-b) - y - c)$

$F_3 = 2zL$

$G_3 = x^2 + y^2 + z^2 + b^2 + c^2 + L^2 + 2(-xb + yc) - l^2$

where,

x, y, z are the coordinates of the destination point

and,

$a = w_B - u_P$

$b = \frac{s_P}{2} - \frac{\sqrt{3}}{2} w_B$

$c = w_P - \frac{1}{2} w_B$

where, $w_B = \frac{\sqrt{3}}{6} s_B$

$u_B = \frac{\sqrt{3}}{3} s_B$

$w_P = \frac{\sqrt{3}}{6} s_P$

$u_P = \frac{\sqrt{3}}{3} s_P$

B. *Land Moving Robot:*

The land moving robot has 3 legs. The middle leg moves between the two target crop rows, and the other 2 legs move on the sides of target rows, hence the robot covers two crop lanes at a time. The robot is powered by a hybrid mechanism in which a 12Ah Sealed Lead Acid Battery, with an output of 12V supplies the required power, and a solar panel of power 50W with 12V output continuously charges the battery. A camera is mounted facing downwards, under the chassis so that the lighting conditions are almost constant.

V. RESULTS

The image was segmented from the background, and the leaf species was classified. If the classified leaf belonged to the weed family it was removed by the robot using a highspeed rotating blade. Fig. 6 shows the output of the pre-processing and segmentation step.

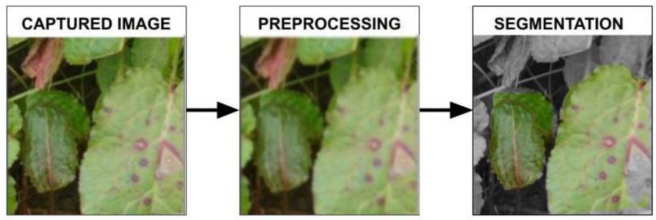


Fig 6 Stages of Weed Identification

In this model, we are using accuracy to measure how close is the data to the true value, precision i.e., how close two or more measurements are to each other and Root Mean Square Error (RMSE)[19] which is a standard way to measure the error of a model in predicting quantitative data. Formally it is defined as follows

$$\text{RMSE} = \sqrt{\sum \frac{(y`_i - y_i)^2}{n}}$$

Where, $y`_1, y`_2 \ldots y`_b$ are predicted values
$y_1, y_2 \ldots y_n$ are observed values
n is the number of observations

Table. 2 shows the classification accuracy using different algorithms.

Table 2 Classification Accuracy

Algorithm	Accuracy	Precision	RMSE
kNN (k = 1)	0.844	0.852	1.390
SVM	0.777	0.710	1.906
Decision Tree	0.788	0.795	1.453
Random Forest	0.900	0.916	1.337
CNN (VGG 16)	0.995	-	-

Clearly, Convolutional Neural Network (VGG16) outperforms other algorithms and predicts with 99.5% accuracy, and hence we use VGG16 architecture to classify leaf species. Fig. 7 shows training and validation loss vs epochs plot for Mask RCNN and training and validation accuracyvs epochs plot for the VGG16 CNN.

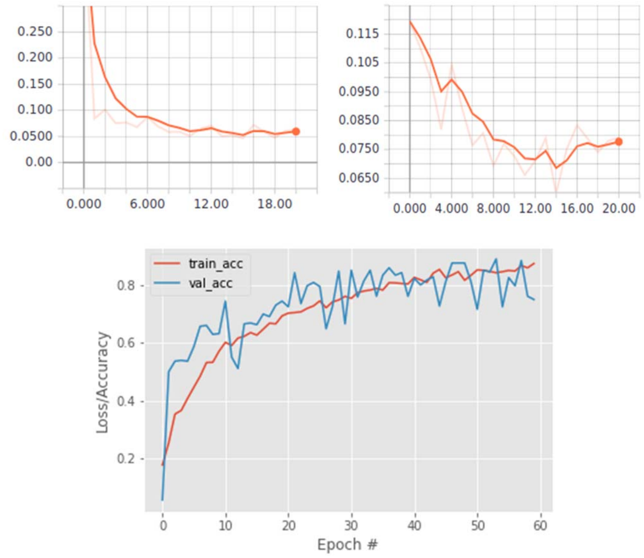


Fig 7 Upper Left: Train Loss Mask RCNN Upper Right: Validation Loss Mask RCNN Bottom: Training and Validation Accuracy VGG16

VI. CONCLUSION

An algorithm for automation of weed removal task was done successfully and the prototype of the model for automatic leaf detection as weed or crop was successfully designed. The method used to identify the correct leaf species resulted in a correct identification rate above 99.5%, which is better than most of the previous works' results. Mask RCNN neural network has significantly improved the segmentation results and directly contributed to better classification accuracy. Our approach uses a high-speed rotating blade to remove the weeds, which eliminates the usage of weedicides and other chemicals, which wasn't ensured in any of the earlier systems. Hence, with low cost and avoiding harmful chemical substances, it is efficient to use a mobile and compact system with a robotic arm cutter with high precision and easy operation.

The system gives good results but has a lot of scope for improvement. A more robust algorithm can be developed for plant identification which can recognize more species of leaves irrespective of their color and shape. The design can be further optimized to suit the needs of farmers and provide maximum area coverage at the same time. The control mechanism for the delta arm can be made more precise and hence improving its precision.

The other aspects of Autonomous robot are:

1. It can be used in the detection of the Drug (Narcotic) leaves among the crops and removal of the same.
2. It can be used in the detection of leaf diseases caused by the plant.
3. It can be used to detect and study rare plants and their variations.
4. It can be used in the soil analysis with required modification in the software and hardware design.
5. The robot can also be used in the plant seeding and plowing with necessary modifications.

ACKNOWLEDGMENT

We express our gratitude to Dr. S.S. Patil, Head of Department, Department of Agricultural Statistics, Applied Mathematics & Computer Science, GKVK, and Mr. A.N.S Shrengesh, Dept of Botany, GKVK for their help and support to provide a deep insight about leaf features. We thank Mr. Kartikay Awasthi, Dept. of Mechanical Engineering, BMSIT and Mr. Shrinidhi N Hegde, Dept of Electronics and Communication Engineering for their technical guidance during the mechanical design process and electronic circuits respectively. The authors of this paper would like to thank BMS Institute of Technology & Management for the facilities, encouragement, support offered during this project.

REFERENCES

[1] R Aravind, M Daman, and B S Kariyappa, "Design and development of automatic weed detection and smart herbicide sprayer robot", published in 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS).

[2] Gulam Amer, S.M.M.Mudassir and M.A Malik, "Design and Operation of Wifi Agribot Integrated System" published in 2015 International Conference on Industrial Instrumentation and Control (ICIC).

[3] Amir H. Kargar B. and Ali M. Shirzadifar, "Automatic Weed Detection System and Smart Herbicide Sprayer Robot for corn fields" published in 2013 RSI/ISMInternational Conference on Robotics and Mechatronics.

[4] Karl D. Hansen, Francisco Garcia-Ruiz, Wajahat Kazmi, Morten Bisgaard, Anders la Cour-Harbo, Jesper Rasmussen, and Hans Jorgen Andersen, "An Autonomous Robotic System for Mapping Weeds in Fields" in 2013 IFAC Intelligent Autonomous Vehicles Symposium.

[5] Preeti Mehta, "Automation in Agriculture: Argibot in the Next Generation Weed Detection and Herbicide Sprayer-A Review" in Journal of Basic and Applied Engineering Research.

[6] P. Lottes, M. Hoeferlin, S. Sander, M. Muler, P. Schulze Lammers, and C. Stachmiss, "An Effective Classification System for Separating Sugar Beets and Weeds for Precision Farming Applications" in 2016 IEEE International Conference on Robotics and Automation (ICRA).

[7] Sachin D. Chothe and V.R. Ratnaparkhe, "Plant Identification Using Leaf Images" in International Journal of Innovative Research in Science, Engineering, and Technology.

[8] Alexander Wendel and James Underwood, "Self-Supervised weed Detection in Vegetable Crops Using Ground-Based Hyperspectral Imaging" in 2016 IEEE International Conference on Robotics and Automation (ICRA).

[9] S. I. Cho, D. D, Lee, and J. Y. Jeong, "Weed-plant Discrimination by Machine Vision and Artificial Neural Network"

[10] Andreas Michaels, Sebastian Haug, and Amos Albert, "Vision-Based High-Speed Manipulation For Robotic Ultra-Precise Weed Control" in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

[11] Stephan Hussmann, Florian J. Knoll, Vitali Czymmek, Andre Meissner and Tim Holtorf, "Development and evaluation of a low-cost delta robot system for weed control applications in organic farming" in 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)

[12] Naio Technology Website – https://www.naio-technologies.com/en/agricultural equipment/weeding-robot-oz/

[13] ecoRobotixs Website –
https://www.ecorobotix.com/en/avo-autonomous-robot-weeder/

[14] Kaiming He, Georgia Gkioxari, Piotr Dollar and Ross Girshick, “Mask R-CNN”, in 2017 IEEE Transactions on Pattern Analysis and Machine Intelligence.

[15] “Plant Identification Using Leaf Images”, International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) Vol. 4, Special Issue 6, May 2015.

[16] Kiran Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, International Conference on Learning Representations, 4 Sep 2014.

[17] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, and Qiao-Liang Xiang, “A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network” in 2007 IEEE International Symposium on Signal Processing and Information Technology.

[18] Robert L. Williams II, “The Delta Parallel Robot: Kinematics Solutions”.

[19] Statistics How To Website -
https://www.statisticshowto.datasciencecentral.com/

Third International Conference On Emerging Trends In Science And Technologies In Engineering Systems, ICETSE-2020, Organized By SJCIT, Bangalore

July 19 2020

Efforts to overcome the existing impairments in implementing Helicopter Satellite Communication System (HSCS)

Rashmi N[1*], Hamsavahini R[1], Abhinandini H R[2], Sanjana B A[2]

[1]Assistant Professor, Department of Electronics and Communication Engineering

BMS Institute of Technology and Management, Bangalore, Karnataka, India

Abstract

Communication can be broadly classified as analog and digital. These two types only deal with terrestrial networks. But, when we want communication to be achieved with very high efficiency it is very much essential to deal with satellites. Satellite Communication is becoming more advantageous day by day. Almost in all the areas, the importance of satellite communication is being recognized. The role of satellite communication in the disaster management has played an important role these days. This has gained very high popularity because of the failure or delay of the organization to detect the disaster at the right time. Satellite communication is generally useful for information gathering during the disaster [1]. Helicopters are used in the satellite communication technology and this technique is called as Helicopter Satellite Communication System (HSCS). There are few technical issues existing in this system too. Thus to overcome these issues and make it more efficient many schemes have been implemented. One such scheme is hybrid modulation with Automatic Frequency Control (AFC). [2]

I. Introduction

Communication is the most important part of our daily life. Not only owing to one's life, communication also plays a major role when it comes to avionics. Avionics is the science which deals with aircrafts and rotorcrafts. Helicopter is the most important rotorcraft which is used in Indian Army, Navy and Air force. Hence communication is very much important between the crew members of the helicopter as well as the pilots of other helicopters. Thus to facilitate such communication, Analog Communication System was introduced.

Analog Intercommunication System provides the direct mode of communication. The problems associated with the analog system are, deformation and the loss appearing in the signal and the proper control over the audio, that is, its volume and frequencies were not taken care of. Thus to overcome these problems, Digital Intercommunication System was introduced. Digital intercommunication system provides reliability and resistance against tightly packed voice signals. It also provides the communication of data from air to air, air to ground and ground to air using high frequency technologies such as Very High Frequency and Ultra High Frequency. The foremost aim of the Digital Intercommunication System is providing two way communication amidst the crew members (pilot and co-pilot) in the flight or helicopter (any aircraft or rotorcraft). In addition to this, the system has the provision to communicate the signals to the crew members at the ground station using a jack externally during the sustentation operations at any time necessary.

This system is a chief control for all the communication equipment established on the Advanced Light Helicopter. The Digital Intercom System connects itself with the various radio channels, warnings and other types of audio outputs. This system also provides the provision to communicate with the pilot and co-pilot of the same or different helicopters, and also with the Air Traffic Controller (ATC). [3]

There were some of the problems associated with Digital Communication Systems too. Thus replacement of this old technology communication with Helicopter Satellite Communication is done in the recent years. Helicopter Satellite Communication plays a vital role in today's world because; helicopters are not just limited to one specific application. The Helicopters are used over a large area covering a wide range. They are used to acquire the information and broadcast the same in the cases of incidents such as disasters which take place over a large scale and also in the case of other emergencies. According to the present day scenario, the repeater stations are widely used to transfer the information such as video clippings and the photographs captured by the helicopter. Capturing of the videos using the helicopter is one of the best methods to obtain on scene information for relief and the management of disaster. [4]

There are few technical issues in this system. Many solutions are being provided to overcome these issues and declare the Helicopter Satellite Communication System (HSCS) as the best technique to provide communication using Helicopters.

II. Predicaments associated with the Digital Communication System

The high requirement of bandwidth is the major problem in these systems. The proportion or the amount of information which can be transmitted per unit time depends on the bandwidth of the media usually used in the respective transmission. The analog systems have less requirement of bandwidth when compared to the digital systems. This is because, the digital signals, for example, the square pulse, delta function and many other digital signals, have many rising and falling pulses. These pulses have a lot of variation in their voltages within a very small amount of time. Thus we can say that, almost within a very few milliseconds, there will be a huge variation in its voltage. Because of this reason, these signals have very high frequency during their rise and fall. As the frequency increases, bandwidth also increases. Thus they occupy very large bandwidth.

High power consumption is also a major issue in digital systems. The digital circuits consume a lot of power. This can be proved by the power dissipation in integrated circuits that is, in digital VLSI (Very Large Scale Integration). With regard to the integrated circuits, three major causes for the large power utilization are given by- utilization of the power dynamically, utilization of the power statically and utilization of the power in the short circuit.

Improvement or rise in the power density and fall in the hardware measurements of the transistor are the stamp of authentication of the present day computer chips. Both of these technologies are increasing the thermal conductivity management challenge inside the chip and the package surrounding it. Both of these technologies are increasing the thermal conductivity management challenge inside the chip and the package surrounding it, as well as escalating research related progress on the materials which possess very high conductivity. The management of the thermal conductivity is largely acknowledged as an important characteristic of the computer design, due to which, the temperature significantly affects the performance of the device. Adding on to this, the lifetime of the device can be reduced extremely because of the excess thermal stresses which take place at the interfaces. [5]

Hardware complexity in the digital systems is another hurdle for the communication. The logic gates serve as an important factor in the designing of a digital circuit. These circuits put together manifest a digital system. The major problem while designing any kind of digital circuit is the hardware complexity of that circuit. When the concept of conventional gate logic is adopted in the design of the digital circuit, then the utilization of more number of components is the major challenge which it has to face. Thus these systems have more complex circuitry and require a sophisticated device making. [6]

The digital systems can be easily broken or damaged, that is, even if a small amount of digital data is unable to be recovered or if it is not interpreted exactly, the meaning of the large amount of data can change completely.

III. Role of Satellite Communication for disaster management

Whenever the earthbound communication resources are inadequate it means that there is a huge damage caused to the infrastructure or it can also be due to the enormous traffic. Hence we cannot depend entirely on the earthbound or the terrestrial communications for the disaster management operations. And it is also essential that for the disaster management operations the communication should be very flexible and reliable as it is the part of emergency communications. Thus everyone who desires to have such type of communications is able to experience perfect working of the communication system every time and everywhere under any circumstances. Furthermore, the broadband communications have been playing an important role in increasing the communications during the emergencies. By considering the radio links as a basis, the systems which are based on the satellites have a lot of benefits which include reliability, availability, robustness, quickness, deploy ability and accessibility from any location.

The Satellite Communication System is a core network which is used for disaster management and its recovery which includes the topology of the satellite network, configuration of service and also for the technical applications. The communication provides the favorable path for managing the critical situations such as occurrence of the disaster and also helps to provide relief during these emergency situations. [7]

IV. The Satellite Communication System employing Helicopters

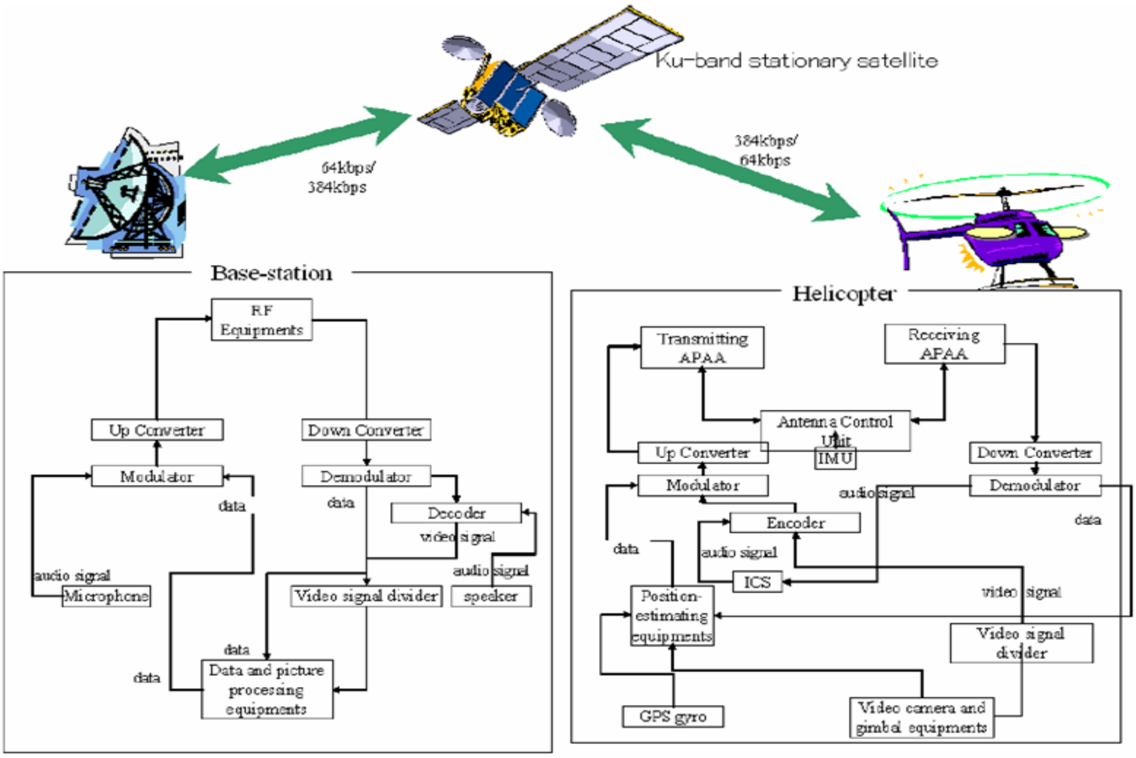


Figure 1 Diagram showcasing how exactly the communication takes place by adopting the principles of satellite communication in the Helicopter

V. How exactly the System looks like?

The continuous prevention of the radio waves being reached to the destination in the Helicopter and Satellite Communication is mainly due to the blades of the Helicopter. These helicopter blades intervene or interrupt the helicopter directly by communicating with the satellite in order to check whether the antenna is situated right under the blade of the helicopter. The time break is about 6.93 milliseconds and the amount of interception taking place is for 39.1 milliseconds. By picturing the case of the helicopter transmission, the system which usually transmits at the exact time of the blade has to be designed and this is considered to be an alternative to solve this complicated issue. [4]

VI. The designing requirements for Helicopter Satellite Communication System

While designing the system based on Satellite Communication and Helicopter, the most important conditions to be followed were; 1. The system which has to be designed must be very small in size; 2. The system to be designed should not be bulky and heavy, instead light weight is preferred; 3. A very effective link must be established between a helicopter and the satellite for the communication to take place, and it should be irrespective of the helicopter direction; 4. The power due to transmission should not harm the pilot, co-pilot and other crew members; 5. The transmitting power should be maintained such that it should not prevent the radio waves reaching the destination due to the blockages in the rotor blade; 6. The video capturing objectives must be based on the positioning function. [4]

VII. Technical Issues ceasing to implement the system

1. Avoidance of the shadowing due to rotor blades; 2. Satellite tracking; 3. Polarization tracking; 4. Avoidance of interference with other satellites [1]

VIII. Helicopter Videoing

The technique of capturing the video using helicopter is the most popular method to obtain the information for emergency communications such as disaster management. A new system has been developed which captures the information related to the disaster and immediately accounts for the transmission of this information. Such a system is called as Ku band utilization in the satellite communication with helicopter. As we have already discussed about the major issue in such type of communications, that is, interruption due to rotor blades of the helicopter, there are techniques proposed to overcome these drawbacks. These techniques are called as time diversity transmission technique which is applicable in the forward link, and the other technique namely blade synchronized transmission technique which can be applied at the return link. The estimation of the position is also implemented for on scene videoing pictures. [8]

IX. Techniques to improve the efficiency of HSCS

In this system, there exists two major issues and they have to addressed properly; 1. Rotor blades of the helicopter is the major part which is leading to the blockage of the received signal; 2. The Doppler Shift in the system which accounts for the fluctuation of the carrier frequency. New techniques have been proposed to overcome these two drawbacks. There were many other techniques which were implemented to overcome these issues; but they were inefficient. The continuous blockage of the signal received can be cancelled by using a hybrid modulation scheme with Automatic Frequency Control (AFC). This technique is better when compared to the use of former diversity technique based on time, as this increases the amount of data transmitted per unit time; whereas AFC transmits the original amount of data per unit time. [2]

Most of the communications in the helicopter are handled by a radio which operates in very high frequency range and it has the limits such as its coverage area and the range at which it can operate. As the amount of area occupied by the relay station which is established on the ground is only around 40km (radius), there is a huge requirement for many numbers of such stations to be established. If we consider an idea of establishing relay stations, it doesn't work out in the areas covered completely with mountains or sea. Thus to overcome all these geographical limitations, the satellite itself is made as a relay station in this system. Whenever the helicopter is outside the terrestrial frequency range, the pilot operations must be undertaken with a very high efficiency because even it includes the satellite communication with the Air Traffic Controller. [2]

This system which is based on Satellite and Helicopter comprises of the two important stations; one at the ground and one at the helicopter. Communication between the two stations is considered to be bidirectional and it can be achieved using a transponder. The transmission link is established between these two stations, that is forward and reverse link. The location of the Antenna should be such that it should not interfere with the rotor blades of the helicopter. It would be considered the best way if it is situated at the tail of the helicopter. Some of the experiments have proven that, as the antenna is located at a large distance from the main rotor blades, then the amount of interruption of signal with the blades would be avoided. [2]

Establishing a system which communicates using the satellite technology with a helicopter is a tedious task, because the rotor blades of the helicopter is the major obstruction in such systems, as this leads to the Doppler shift. Thus because of these major impairments, many applications such as Global Positioning System, degradation in the performance of bit error rate and the discontinuity in the received signal over the recovered carrier. [2]

X. Results and Discussions

The transmission from the helicopter can automatically stop when 1. The tracking error angle becomes large 2. The receiving level becomes below the limited value 3. The posture of the helicopter inclines beyond the anticipation 4. The onboard equipment fails. The final flight test included the following ventures 1. Verification of the satellite tracking performance 2. Measurement of the communication quality 3. Operation check of transmission protection 4. Picture transmission and position fixing. [1]

The parameters of simulation in order to calculate BER is given below:

Helicopter	
Rotor revolutions	ω = 6.45 rev/s (387 rev/min)
Blades	$M = 4$
Maximum speed	V = 68.4 m/s (246$km/hour$)
Maximum acceleration	A = 3.09 $m/s2$ (11.1$km/hour/s$)
Modem	
Modulation	BPSK with time diversity
Time diversity factor	$L = 2$
Information data rate	R_b = 3600 bit/s
Carrier frequency	f_C= 12.5 GHz
Demodulation	Coherent detection
Forward error correction	
Code	Convolutional code with
Viterbi decoding	
Rate	$R = 1/2$
Constraint length	$K = 7$

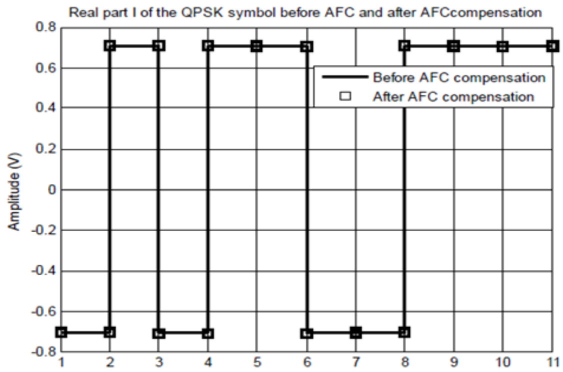


Figure 2 Phase symbol before and after AFC

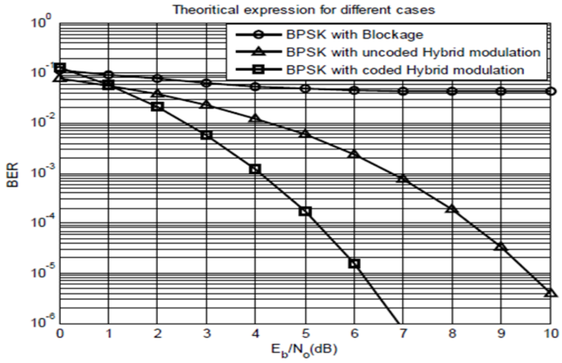


Figure 3 BER behavior for different schemes

XI. Conclusion

The difficulties in analog as well as digital communication systems led to the development of new technology in communication era which is called as Helicopter Satellite Communication System (HSCS). Thus the helicopter satellite system was developed successfully and it was also demonstrated for the first time. The key technologies of the HSCS for disaster control operations were experimentally demonstrated. The developed HSCS system provides a real time information collection and transmission directly via a satellite. Research and Development aiming at small-size, light-weight and wide-band HSCS is proceeding. Continuous research and development collaborated with disaster management bodies will be carried out. [1]

This paper concludes that the Helicopter Satellite Communication system is one of the best ways to realize communication. A novel modulation scheme based on hybrid techniques for the accurate cancellation of the periodic or continuous blockage in Helicopter Satellite Communication System is implemented instead of conventional time diversity scheme. The proposed AFC scheme incorporates adaptive modulation in an efficient way. The traditional time diversity scheme increases the transmission bit rate to reduce errors, but in the novel AFC scheme, similar type of signal is made to transmit just by changing the number of time slots. [2]

XII. References

[1] Wataru Chujo, "Helicopter Satellite Communication System for Disaster Control Operations", Kashima Space Research Centre, National Institute of Information and Communications Technology, 893-1, Hirai, Kashima, Ibaraki 314-8501, Japan.

[2] Emad. A. Elaty, Sami A El-Doli, Mohammed Abd-Elnaby, Fathi E Abd El-Samie, "An Adaptive Modulation Scheme in Helicoper-Satellite Communication", The Mediterranean Journal of Electronics and Communications, Vol.9, No.3, September 2013.

[3] Dankan Gowda, Mahaveer Penna, Satyasrikanth Palle, Sai Priya K A and Chaithra K N, "Digital Intercommunication System in Advanced Light Helicopter", IEEE, Advances in Computational Sciences and Technology, ISSN 0973-6107 Volume 10, Number 5, pp. 699-708, Research India Publications, 2017

[4] Masaki Satoh et al, Huan-Bang Li, Yoshiyuki Fujino, Hiromtsu Wakana, Amane Miura, Yutaka Ozaki, Hiroyuki Satou, Eiji Watanabe, Manabu Sawa, "Helicopter-Satellite Communication System Developed for Transmission of Disaster and Emergency Information", Communication Research Laboratory, American Institute of Aeronautics and Astronautics, Inc., AIAA-2003-2319, 2003.

[5] Patrick K. Schelling, Li Shi and Kenneth E. Goodson, "Managing heat for Electronics", Materialstoday, June 2005.

[6] Muhammad Khalid, Shubhankar Majumdar and Mohammad Jawaid Siddiqui, “Reduction of Hardware Complexity of Digital Circuits by Threshold Logic Gates Using RTDs”, Information and Communication Technology for Intelligent Systems, Smart Innovation, Systems and Technologies 106.

[7] Yang-Min Lee, Bon-lun Ku and Do-Seab Ahn, “A Satellite Core Network System for Emergency Management and Disaster Recovery”, ICT Standardization program of MKE (The Ministry of Knowledge Economy) [20l0-Pl-02, Standardization on PPDR communication for next generation], IEEE, 2010.

[8] Huan-Bang Li, Masaki Sato, Amane Miura, Shinichi Taira, and Hiromitsu Wakana, “Ku-band Helicopter Satellite Communications for on scene disaster information transmission”, National Lnstitute of Information and Communications Technology 893- 1 Hirai, Kashima-shi, Ibarah 3 14-8501 Japan, 0-7803-8523-3/04/$20.0, IEEE, 2004.