BMS INSTITUTE OF TECHNOLOGY AND MANAGEMENT

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

RESEARCH COMPENDIUM

2021-22

Compiled By

Prof. Asha .G. Hagargund

SUMMARY OF PUBLICATIONS.

ACADEMIC YEAR 2021-22

SL.NO	PUBLICATION TYPE	COUNT
1	SCOPUS/SCI INDEXED JOURNALS	22
2	SCOPUS INDEXED CONFERENCES	6
3	BOOK CHAPTERS	11
4	REVIEWERS	18

Details of publications ACY-2021-22

- **Publications (Journal) of ACADEMIC YEAR - 2021-22**

SLNO	Faculty	Title of the paper	Journal	Date	Indexing	DOI Number
1	Dr. M. C. Hanumantharaju	Low-Field MRI of Stroke: Challenges and Opportunities	Journal of Magnetic Resonance Imaging	2021	Scopus	https://doi.org/10.1002/jmri.27324
2	Dr. M. C. Hanumantharaju	Gradient, texture driven based dynamic-histogram equalization for medical image enhancement	International Journal Of Biology And Biomedical Engineering	2021	Scopus	DOI: 10.46300/91011.2021.15.36
3	Prof. Saneesh Cleatus T	Epileptic Seizure Detection using Spectral Transformation and Convolutional Neural Networks	Journal of Institution of Engineers - Springer	Feb 2022	Scopus	https://doi.org/10.1007/s40031-021-00693-4
4	Dr.Anitha V R	Parametric Analysis of Compact UWB-MIMO Antenna with Improved Isolation Using Parasitic Reflectors and Protruded Ground Strips	Wireless Personal Communications	October 2021	Scopus	https://link.springer.com/article/10.1007%2Fs11277-021-09235-z
5	Dr.Anitha V R	X-Band Miniaturized Frequency Reconfigurable Travelling Wave Antenna	TETL-International Journal of Electronics Letters	April 2022	Scopus	10.1080/21681724.2022.2068194
6	Dr.Vijayalaxmi V	Bilingual text detection from natural scene images using faster R-CNN and extended histogram of oriented	Pattern Analysis and Applications	2022	Scopus	https://doi.org/10.1007/s10044-022-01066-3

		gradients				
7	Dr.Vijayalaxmi V	An integrated feature frame work for automated segmentation of COVID-19 infection from lung CT images	International Journal of Imaging Systems and Technology	2021	Scopus	https://doi.org/10.1002/ima.22525
8	Dr.Deepareddy	Spectrum sensing in non gaussian noise	Indian Journal of Science and Technology	September 2021	WoS	. https://doi.org/10.17485/IJST/v14i32.1034
9	Mamatha K R	”Placement Management System”	Journal of Chengdu University of Technology,	2021	Scopus	http://cdujournal.com/index.php/vol26num7/
10	Dr.Rashmi N	The peak-to-average power ratio reduction using hybrid scheme with companding and discrete Hartley transform for orthogonal frequency division multiplexing system	International Journal of Intelligent Enterprise,	Jul 2022	Scopus	10.1504/IJIE.2022.123754
11	Prof.Jagganntha KB	A composition-dependent thermal behavior of Si20Te80−xSnx glasses: Observation of Boolchand intermediate phase	Non crytalline solid	Feb 2022	SCI	https://doi.org/10.1016/j.jnoncrysol.2021.121311
12	Prof.Jagganntha KB	Influence of Cu Doping in Si–Te-Based Chalcogenide Glasses and Thin Films: Electrical Switching, Morphological and Raman Studies	IEEE Transactions on Electron Devices	March 2021	SCI	10.1109/TED.2021.3051925

13	Dr. Sabina Rahaman	Nanoarchitectonics earth abundant chalcogenide CTS thin film using spray pyrolysis for visible light driven photocatalysis	Applied physiscs A (springer)	December 2021	SCI	https://doi.org/10.1016/j.optmat.2021.111838
14	Dr. Sabina Rahaman	optimization and fabrication of Cu2SnS3/ZnS thin film heterojuction solar cell	optical material (elsevier)	Jan 2022	SCI	10.1016/j.optmat.2021.111838
15	Dr. Laxmisagar HS	Detection of Breast Cancer with Lightweight Deep Neural Networks for Histology Image Classification	Critical Review in Biomedical Engineering	2022	Scopus	DOI: 10.1615/CritRevBiomedEng.2022043417
16	Dr. Laxmisagar HS	Evaluating the Statistical Stability of POSIT Arithmetic and IEEE 754 Float to Accelerate Data for Detection of Breast Cancer	SSRG International Journal of Electrical and Electronics Engineering	2022	Scopus	DOI:10.14445/23488379/IJEEE-V9I9P106
17	Dr.Prachi Sharma	Synthesis of highly sensitive nanomaterial for ultra-fast photocatalytic activity: A detailed study on photocatalytic capabilities of rod-shaped TiS3 nanostructures	Catalysis Communications	2022	Scopus	https://doi.org/10.1016/j.catcom.2021.106381
18	Asha G Hagargund	Performance analysis of cost effective multi hop time sensitive network for IEEE 802.1Qbv and IEEE 802.1Qbu standards	Journal of physics: Conference series	Jan 2022	Scopus	https://iopscience.iop.org/article/10.1088/1742-6596/2161/1/012002/meta

19	Prof.Shilpahiremath	Assorted noise identification and removal technique to obtain encoded image using CLAHE technique	GIS science journal	March 2022	Scopus	DOI:20.18001.GSJ.2022.V9I3.22.38859
20	Dr.Dankan Gowda V	Smart Agriculture and Smart Farming using IoT Technology	Journal of Physics: Conference Series AMSE 2021	September 2021	Scopus	doi:10.1088/1742-6596/2089/1/012038
21	Dr.Dankan Gowda V	IoT Enabled Smart Lighting System for Smart Cities	Journal of Physics: Conference Series AMSE 2021	September 2021	Scopus	doi:10.1088/1742-6596/2089/1/012037
22	Dr.Dankan Gowda V	Smart villages: IoT technology-based transformation	Journal of Physics: Conference Series ICAPSM 2021	August 2021	Scopus	doi:10.1088/1742-6596/2070/1/012128

Publications (Conference) ACADEMIC YEAR - 2021-22

SLNO	Faculty	Title of the paper	Conference	Date	Indexing	DOI Number
1	Prof. Saneesh Cleatus T	Big EEG Data Images for Convolutional	2021 IEEE Signal Processing in Medicine and Biology	4th December	Scopus	https://doi.org/10.1109/SPMB52430.2021.96

		Neural Networks	Symposium (SPMB)	2021		72272
2	Dr.Anitha V R	Design and Simulation Analysis of a Piezoresistive Cantilever Beam for Low Pressure Detection	2nd International Conference on Micro/Nanoelectronics Devices, Circuits and Systems (MNDCS-2022)	29th to 31st January 2022	Scopus	
3	Dr.Vijayalakshmi G V	A Hierarchical Machine Learning Frame Work to Classify Breast Tissue for Identification of Cancer	11th International Conference on Computer Engineering and Networks	21st -25th October 2021	Scopus	https://link.springer.com/chapter/10.1007/978-981-16-6554-7_56
4	Dr.Vijayalakshmi G V	An ensemble classification based approach for breast cancer prediction	1st International Conference on Frontiers in Engineering Science and Technology	2021	Scopus	DOI 10.1088/1757-899X/1065/1/012049
5	Prof.Jagannttha K B	Switching, Raman and morphological studies on Si20Te74Sn6 chalcogenide glass	1st International Conference on Design and Materials (ICDM) – 2021	28th -29th Jan 2022	Scopus	https://doi.org/10.1016/j.matpr.2022.01.014
6	Jagannttha K B	An Accelerated Approach to Parallel Ensemble Techniques Targeting Healthcare	3rd International Conference on Energy, Power and Environment: Towards Clean	5-7 March 2021	Scopus	10.1109/ICEPE50861.2021.9404519

		and Environmental Applications	Energy			
7.	Dr.Dankan Gowda V; FalguniBakshi; Gahana A; AnushaBabuNaik; Navya HG	Covid-19 Prevention Kit Based on an Infrared Touchless Thermometer and Distance Detector	2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)	2nd -4th December 2021	Scopus	10.1109/ICECA52323.2021.9676014

Publications (Book chapter) ACADEMIC YEAR - 2021-22

Sl. No.	Author Name	Book title	Chapter title	Online publication link	Publisher

1	Dr.Anitha V R	Evolution in Signal Processing and Telecommunication Networks	Circle-to-circle Fractal Antenna Fed by Microstrip for Tri-band Applications	10.1007/978-981-16-8554-5_37	Springer
2	Vijayalakshmi GV Mahesh, Alex Noel Joseph Raj, Ruban Nersisson	Aiding Forensic Investigation Through Deep Learning and Machine Learning Frameworks	Implementation of Machine Learning-Aided Speech Analysis for Speaker Accent Identification Applied to Audio Forensics	DOI: 10.4018/978-1-6684-4558-7.ch008	IGI Global
3	R Uma, Vijayalakshmi GV Mahesh, P Ramkumar	Aiding Forensic Investigation Through Deep Learning and Machine Learning Frameworks	Investigation of Deep Fake Images Using Pre-Trained CNN Frameworks	DOI: 10.4018/978-1-6684-4558-7.ch007	IGI Global
4	Santrupti Bagali, Sanjana, Sara Sadiya, Priyanka and Deepa Reddy	Disruptive developments in biomedical applications	Prediction and classification of Alzheimer's disease using machine learning models.	DOI: 10.1201/9781003272694	CRC Press

5	Dr.Rashmi N	Air traffic management and control	Deep Learning Network for Classifying Target of Same Shape using RCS Time Series	10.5772/intechopen.99619	Intechopen
6	Dr.Rashmi N	smart Data Intelligence	Doppler Radar Technique for Geriatric	10.1007/978-981-19-3311-0_29	Springer
7	Rashmi N, Hamsavahini R, Mamatha K R	New approaches in Engineering Research,	Studies on RF Energy harvesting for low power devices.	https://www.bookpi.org/bookstore/product/new-approaches-in-engineering-research-vol-7/	B P international
8	Sabina Rahaman, Jaganatha K B, Thyagaraj T, Lakshmisagar	Current Approaches in Science and Technology Research	Investigating the Effect of Annealing on the properties of Cu2SnS3 thin film using spin coating	https://doi.org/10.9734/bpi/castr/v14/2562F	B P international
9	Shilpa Hiremath, Chandra Prabha R, and Sushil Kumar I	Advancing the Investigation and Treatment of Sleep Disorders Using AI	An Investigation Screen for Sleep Disorders Using AI/ML	https://www.igi-global.com/chapter/an-investigation-screen-for-sleep-disorders-using-aiml/285274	IGI Publisher
10	Dr.Dankan Gowda V	Book Title: Artificial Intelligence for Smart Cities and	Chapter Title: Intelligent IoT Enabled Lighting	Book Chapter https://benthambooks.com/future-series-by-	Bentham Science

		Villages:Advanced Technologies, Development, and Challenges, 202	System 1	subject/computer-and-information-science/sub-category/artificial-intelligence-and-machine-learning/	Publishers
11	Dr.Dankan Gowda	Book Title: Artificial Intelligence for Smart Cities and Villages:Advanced Technologies, Development, and Challenges, 2021	Chapter Title: IoT Enabled Smart Village for Sustainable Development	Book Chapter https://benthambooks.com/future-series-by-subject/computer-and-information-science/sub-category/artificial-intelligence-and-machine-learning/	Bentham Science Publishers

Reviewers of various Journals.

August 2021 to July 2022

Reviewer				
Sl.No	Name of the faculty	Journal	Date	Indexing
1	Dr. Sabina Rahamn	Journal of Molecular Structure	13th November, 2021	SCI
2	Dr. Vijayalakshmi G V	Virtual Conference on Futuristic Communication and Network Technologies	December 10-11, 2021	Scopus
3	Dr.Anitha V R	International Journal of Electronics	01 July 2021	Scopus
4	Dr.Anitha V R	International Journal of Informatics and Communication Technology (IJ-ICT)	01 August 2021	Scopus
5	Dr.Anitha V R	IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)	13th – 16th December 2021	Scopus
6	Dr.Anitha V R	Indian Conference on Antennas and Propagation (InCAP) 2021	13th – 16th December 2021	Scopus
7	Dr.Anitha V R	2nd International Conference on Symposium, Information and Computing (ISAIC 2021)	December 3-6, 2021	Scopus
8	Dr.Anitha V R	4th International Conference on Inventive Research in Computing Applications (ICIRCA 2022)	21st – 23rd September 2022	
9	Dr.Anitha V R	IEEE WIECON-ECE, IEEE SPICSCON, IEEE BECITHCON, IEEE PEEIACON, IEEE RAAICON by IEEE Bangladesh Section.	3rd – 5th December 2021	Scopus

10	Dr. Deepa N Reddy	IEEE-CONECCT	Julu-14-16th 2023	
11	Dr.Rashmi N	International Conference on Applied Intelligence and Sustainable Computing (ICAISC-2023)	16th-17th June, 2023	scopus
12	Dr.Rashmi N	IEEE BIT ICAECIS-2023	19th-21st April 2023	scopus
13	Dr.Rashmi N	3rd IEEE International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)	19-20 July 2023.	scopus
14	Dr.Rashmi N	Intelligent Decision Technologies	5 November 2022	Scopus
15	Dr. Vijayalakshmi G V	IEEE DISCOVER-2022	14-15, October 2022	Scopus
16	Dr. Vijayalakshmi G V	VICFCNT-2021	December 10th and 11th 2021	Scopus
17	Dr. Vijayalakshmi G V	Computers in Biology and Medicine	Nov 2021	WoS
18	Dr. Vijayalakshmi G V	IEEE Access	Nov 2021	Scopus

SCOPUS/SCI INDEXED JOURNALS

2021-22

Review J Magn Reson Imaging. 2021 Aug;54(2):372-390. doi: 10.1002/jmri.27324.
Epub 2020 Aug 22.

Low-Field MRI of Stroke: Challenges and Opportunities

Seema S Bhat [1], Tiago T Fernandes [2], Pavan Poojar [1 3], Marta da Silva Ferreira [2], Padma Chennagiri Rao [1], Madigondanahalli Chikkamaraiah Hanumantharaju [4], Godwin Ogbole [5], Rita G Nunes [2], Sairam Geethanath [1 3]

Affiliations
PMID: 32827173 DOI: 10.1002/jmri.27324

Abstract

Stroke is a leading cause of death and disability worldwide. The reasons for increased stroke burden in developing countries are inadequately controlled risk factors resulting from poor public awareness and inadequate infrastructure. Computed tomography and MRI are common neuroimaging modalities used to assess stroke with diffusion-weighted MRI, in particular, being the recommended choice for acute stroke imaging. However, access to these imaging modalities is primarily restricted to major cities and high-income groups. In the case of stroke, the time-window of treatment to limit the damage is of a few hours and needs a point-of-care diagnosis. A low-cost MR system typically achieved at the ultra-low- and very-low-field would meet the need for a geographically accessible and portable solution. We review studies focused on accessible stroke imaging and recent developments in MR methodologies, including hardware, to image at low fields. We hypothesize that in the absence of a formal, rapid stroke triaging system, the value of timely on-site delivery of the scanner to the stroke patient can be significant. To this end, we discuss multiple recent hardware and methods developments in the low-field regime. Our review suggests a compelling need to explore further the trade-offs between high signal, contrast, and accessibility at low fields in low-income communities.
LEVEL OF EVIDENCE: 4 TECHNICAL EFFICACY STAGE: 6.

Keywords: MR accessibility; accessible stroke MRI; diffusion imaging at low-fields; machine learning; mobile stroke unit; open-source imaging.

LinkOut - more resources

INTERNATIONAL JOURNAL OF BIOLOGY AND BIOMEDICAL ENGINEERING
DOI: 10.46300/91011.2021.15.36 Volume 15, 2021

Gradient, Texture Driven Based Dynamic-Histogram Equalization For Medical Image Enhancement

Vidyasaraswathi H N [1*], Hanumantharaju M C [2]

[1]Dept. of ECE, Bangalore Institute of Technology, Karnataka, India
[2]Dept. of ECE, BMS Institute of Technology & Management, Karnataka, India
Email: vidya.gowdaec@gmail.com

Received: May 28. Revised: July 19, 2021. Accepted: July 21, 2021. Published: July 22, 2021.

Abstract: In many clinical diagnostic measurements, medical images play some significant role but often suffer from various types of noise and low-luminance, which causes some notable changes in overall system accuracy with misdiagnosis rate. To improve the visual appearance of object regions in medical images, image enhancement techniques are used as potential pre-processing techniques. Due to its simplicity and easiness of implementation, histogram equalization is widely preferred in many applications. But due to its mapping function based image transformation during enhancement process affect the biomedical patterns which are essential for diagnosis. To mitigate these issues in medical images, a new method based on gradient computations and Texture Driven based Dynamic histogram equalization (GTDDHE) is accomplished to increase the visual perception. The spatial texture pattern is also included to ensure the texture retention and associated control over its variations during histogram modifications. Experimental results on MRI, CT images, eyes images from medical image datasets and quantitative analysis by PSNR, structural similarity index measurement (SSIM), information entropy (IE) and validated that the proposed method offers improved quality with maximum retention of biomedical patterns across all types of medical images.

Keywords: Gradient, Local binary pattern, Image enhancement etc.

I. INTRODUCTION

Image enhancement is one of the essential methodology used widely in many digital image processing system, has been an active area of research in recent years due to the invention of vision based applications. Most of the existing image enhancement studies developed some spatial transformations to improve the quality of the image and reducing the noise presence for better visualization since the visibility level is poor in any low-contrast image features.

In recent, imaging processing is steadily emerging in medical fields for several diagnostic measurements and plays significant role in examination of the patient health condition. Different type's medical images are used for various clinical diagnoses. In general medical images are comprise of several complicated regions, which required some unique pre processing models to simplify the complications that arises in various stages of CAD system. Moreover medial images are not carrying fine details for an accurate classification due to poor lighting conditions, external noise interferences and other restrictions related to imaging devices which makes the clinical diagnosis a difficult task to accomplish. Medical images itself comes in different format namely magnetic resonance imaging (MRI), chest X-ray (CXR), computerized tomography (CT) and mammographic images etc. During classification only the pixel intensity differences among various region and object boundary information are playing vital role.

Therefore, for improved classification rate in medical images both quality and contrast need to improve collectively in such way that each object can be easily observed during analyses.

Though the core objective of any image enhancement technique is to improve quality while suppressing noise, it is essential to restore the basic details without causing any significant changes. Even in some cases image restoration is also considered as image enhancement; but still it is only deals with appropriate estimation of the actual image from observed one. Most of the existing image-enhancement techniques not consider degradation process, in general, it is not modelled; however, for medical images degradation process may causes some notable performance loss with the inclusion of enhancement techniques.

Among various methods Histogram based techniques for image enhancement is widely investigated in many works which is based on equalization of the histogram of the values. Histogram Equalization (HE) has some inherent characteristics of changing the intensity values and spatial patterns which cause some notable problems in medical imaging. To overcome this issue, Contrast limited adaptive HE (CLAHE) is proposed as potential model for medical image analysis. Shome et al. (2011) [1] proves that contrast limited image enhancement using histogram equalization explores the abnormalities with improved quality.

In general the cumulative effect of image enhancement process in medical image spatial patterns and its influence in clinical diagnostic accuracy is investigated in many existing works. This work presents the gradient and texture driven dynamic sub-histogram equalization technique based on statistical measurements of various image attributes such as intensity bound, texture driven spatial correlations and gradient driven object orientations etc.

Several existing works [2-3] well proved that the texture retention and boundary details are essential measures which play significant role in clinical diagnostics measurements [4]. In this paper, intensity, gradient and texture measures are used for contrast limited adaptive non linear histogram equalization which has metrics as follows:

- The spatial orientations of object pixels are remain unaltered while visual appearance is improved.
- The statistical significance of texture information is considered and its retention is ensured in enhanced image.

II. Related Works

Bhateja et al. (2013) [5] introduced non-linear framework which employs a robust polynomial filter for enhancing the mammographic abnormalities. Through filtering approach mammographic masses are improved irrespective to the background tissues. The experimental results prove that contrast is increased with significant suppression of noise level and preserved edges without causing any notable blurring artifacts. Singh et al. (2014) [6] proposed exposure based Sub-Image Histogram Equalization (ESIHE) to improve the enhancement of low exposure image. Here by dividing the input image into sub-images using pre defined thresholds and intensity levels. To control enhancement rate the histogram is clipped using a threshold value and gray level occurrences.

colour image enhancement technique proposed by Wong et al. (2016) [7] increase the colour information and suppressing the artefacts. Here equalization is performed recursively to extend the intensity ranges optimally using the golden section search algorithm. The histogram values are equalized independently over sub images. Color image enhancement developed by Lee et al. (2017) [8] carried out intensity HE based on hue and saturation values in HSI color space. As compared to conventional HE the visual colourfulness is well preserved in color space conversion driven enhancement process. The performance metrics are validated using both back-lighting and night-time image sets are used for experimentation. Wan et al. (2018) [9] introduced histogram partition and brightness correction for infrared image enhancement technique. Here based on of human visual system visual correction factor are introduced and grayscale re-mapping process is carried out to suppress the noise level. In addition to this, brightness correction is also accomplished using particle swarm optimization algorithm.

Histogram shifting based contrast enhancement is developed in Ying et al. (2019) [10] for reversible data hiding with minimal distortions over embedded image. Here transfer matrix is constructed by maximizing the entropy of the histogram and embedding is performed accordingly to increase the embedding capacity. Srinivas et al. (2020) [11]

combines the contrast and Color enhancement using spatial contextual similarity histogram model. Based on the deviation of the pixel values in luminance channel cumulative distribution function is formulated and contrast-enhancement is achieved through intensity transformation. And chromatic information is used for adaptive color restoration to restore the consistent color information's. The hierarchical equalization model developed by Bai et al. (2020) [12] comprises of four stages namely centre regionalization, local equalization, global equalization and multi-scale fusion. Here equalization strategy is used to perform both for both contrast improvements and correct the color characteristics of the each channel. Kandhway et al. (2020) [13] proposed highly optimized contrast and edge enhancement methodologies for bio medical images to explore the different anatomical characteristic of the diseases. Here plateau limit is used to select the histogram based on min, max, mean, and median of it with some adjustable parameter. Based on relative vacancy of histogram bins the residue blocks are generated and meta-heuristic algorithm is incorporated based on a fitness function which includes two functions with edge, entropy, gray level co-occurrence matrix (GLCM) for contrast, and energy for visual impairments.

Gamma correction based dynamic clipped HE proposed by Subramani et al. (2020) [14] control the color enhancement rate using mean values. Here both clipping and gamma correction is performed over decomposed image sub blocks to preserves basic image attributes like entropy, colors, saturation etc. the same histogram clipping operations proposed in Bhandari et al. (2020) [15] used salp swarm algorithm (SSA) optimization to formulate the optimal plateau limit (PLs) to overcome the performance degradations caused by illuminations and backlighting effect and suppress the associated artifacts.

In all above discussed models for contrast enhancement cause some undesirable effects which will reduce the diagnostic measurements of the biomedical images. Histogram based image enhancements have some limitations like performance trade off over texture retention, boundary diffusion and artifacts related problems etc. Moreover enhancement from conventional methods required some statistical characteristics of the input images which are not available for most biomedical image modalities. Therefore it is essential to narrow down the information loss and parametric requirements of existing HE techniques without compromising the reconstruction quality. In this work image attributes driven histogram equalization is proposed which includes various image characteristics namely luminations, gradient and texture information. Here the trade off in the SSIM texture measure is narrow down and the penalty gap due to over enhancement is also suppressed during contrast enhancement. Here equalization is processed over luminance channel and texture and object boundary details are preserved using threshold bound computed in identical channels to accomplish contrast limited image enhancement: the proposed GTDDHE approach can improves the contrast and explore precise biomedical abnormalities in each medical images.

III. Proposed Method

The image enhancement through GTDDHE method modifies both the dynamic range and contrast levels to formulate the histogram in some desired shape using some mapping function as cumulative distribution function. This stretching of histogram peak values directly relates the intensity level changes and contrast is adjusted accordingly as shown in Figure 1.

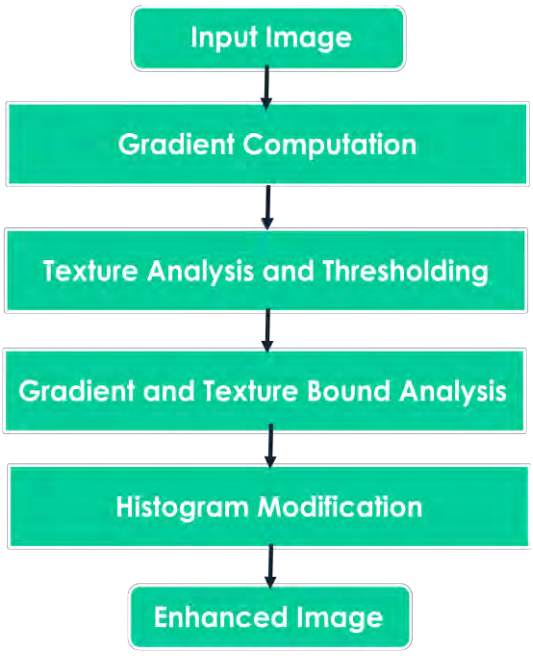


Figure.1 Proposed GTDDHE method flow graph.

A. Gradient Approach

Gradients of pixels are used to express the co-occurrence variations of pixels inside every region. The computation of local pixel co-occurrence in the similar region discloses the correlative nature of the

neighbouring pixels precisely. The gradient measure in each object model contributes in identifying the magnitude shift between the complexes and smooth region. Here, a significant change in a gradient indicates object discriminations. This simple linear variation is most effective for modelling the object of interest. The gradient with the variance measure is shown in Figure 2(a).

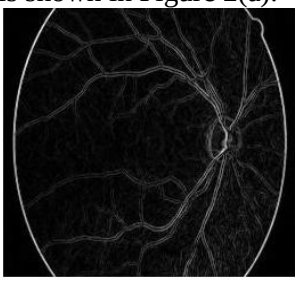


Figure.2 (a) Gradient

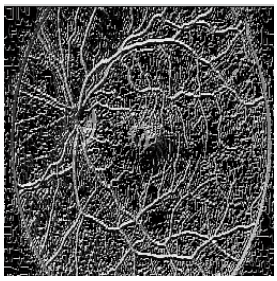


Figure.2 (b) Texture Output

B. Spatial invariant Texture Model

Texture models are used to decide which region contains the region of interest in medical objects. For texture analysis in vision applications, the local binary pattern (LBP) operator has been used as a standard feature-based approach. Even small changes in pixel intensity values (due to illumination or dynamic changes) give a changed LBP code. Also invariant and uniform textures are efficient in opposition to any dynamic changes. The vital mathematical explanation of the LBP operation is described in the accompanying equations below.

$$C = \sum_{d=0}^{7} 2^d * B_d \qquad (1)$$

$$B_d \ for\ 0 = 0: \sum_{d=0}^{7} td > c \ \ , d \in [0,7] \qquad (2)$$

$$B_d \ \ for\ 1 = 1: \sum_{d=0}^{7} td < c \ \ , d \in [0,7] \qquad (3)$$

Generally, the binary pattern is termed uniform in case the LBP binary pattern comprises of two bitwise transitions from 1 to 0 or 0 to 1 atleast and have circular binary pattern. For static textures, rotation-invariant features of this kind are not dependent of input image angle as shown in Figure 2(b). Dynamic textures (DT) and uniform textures, combined with rotational-invariant feature description, is the key element where local features obtain the ability to describe motion and provide the most discriminative patterns for object modelling.

C. Detailed Steps of Proposed Method

Here input histogram values are decomposed into several groups based on threshold bounds and analyzes its variations with gradient and texture bounds for equalization process as given below:

//Input – Low contrast Image

//Output – Enhanced image

Step 1: Set sigma (σ) //variance value

Step 2: Extract gradient channel using Gaussian derivative - CG channel

$$Gauss = \frac{1}{\sqrt{2\pi\sigma}} e^{x^2/(-2\sigma^2)} \qquad (4)$$

Step 3: Extract Texture channel using LBP - T channel

$LBP=2^7p0+2^6p1+2^5p2+2^4p3+2^3p4+2^2p5+2^1p6+2^0p7$

Where, p0-p7 true value comparison of centre pixels with 8 surrounding pixels in 3x3 macro block

Step 4: Extract 3x3 overlapping window

Let P_{ij}- current pixel as well as center pixel in 3x3 macro block

CG_{ij}- current pixel gradient value

T_{ij}- current pixel texture bound value

if ($P_{ij} > T_{lower}$) && ($P_{ij} > T_{upper}$) && ($CG_{ij} < P_{ij}$)

OR(1,T_{gr})+1 // Spatial Orientation measure for boundary preservations

end

Step 5: Apply histogram equalization over each sub groups of P_{out}

if ($T_{ij} > T_{lower}$) && ($T_{ij} < T_{upper}$)

P_{out} =(2^L-1) x OR(1,T_{gr})/2^Mx2^N //Texture Preservation

end

Where,

T_{lower},T_{upper} –Histogram threshold bounds for groups.

T_{gr} – number of histogram groups.

L- histogram levels

M,N- size of input images

IV. Experimental Results

The performance metric evaluation of the GTDDHE method is experimented over the different biomedical dataset which includes eye images, MRI and CT images which are collected from various sources. From each dataset 30 images of different level of textures and complication in patterns. Performance metric evaluation of the proposed GTDDHE is experimented using different quantitative parameters.

The peak signal-to-noise ratio (PSNR) measure is used to validate the reconstruction quality of proposed image restoration through SR as given below:

$$PSNR(dB) = 10\log_{10}\left[\frac{(2^n-1)^2}{MSE}\right] \qquad (5)$$

Where MSE denotes the mean squared error.

The structural similarity index (SSIM) is used to measure the structural similarity between input LR and HR images based on independent parametric comparisons such as luminance (μ), contrast (σ) as follows:

$$\mathrm{SSIM(e,f)} = \frac{(2\mu_e\mu_f+C_1)(2\sigma_{ef}+C_2)}{(\mu_e^2+\mu_f^2+C_1)(\sigma_e^2+\sigma_f^2+C_2)} \qquad (6)$$

And the measure the structural similarity is evaluated based on its correlations.

Entropy measures the highness of information in enhanced image as shown in Eqn. (7) and its improvement is measured using Measure of enhancement by entropy (EMEE) value as shown in Eqn. (8)

$$Entropy\ \ (E) = \sum_{x,y}^{255} Pr(b,c)\big(-\ln Pr(b,c)\big) \qquad (7)$$

where Pr(b,c) denotes the probability of the difference between two successive pixels b and c.

$$EMEE = \frac{1}{N}\sum\frac{I_{\max}}{I_{\min}}20\log\frac{I_{\max}}{I_{\min}} \qquad (8)$$

Average gradient (AG) measure is used to explore the relative clarity of the enhanced output image using Eqn. (9)

$$AG = \frac{1}{(M-1)(N-1)}\sum_x\sum_y\frac{|G(b,c)|}{\sqrt{2}} \qquad (9)$$

Where |G (b,c) | denotes absolute value of the Magnitude of Gradient.

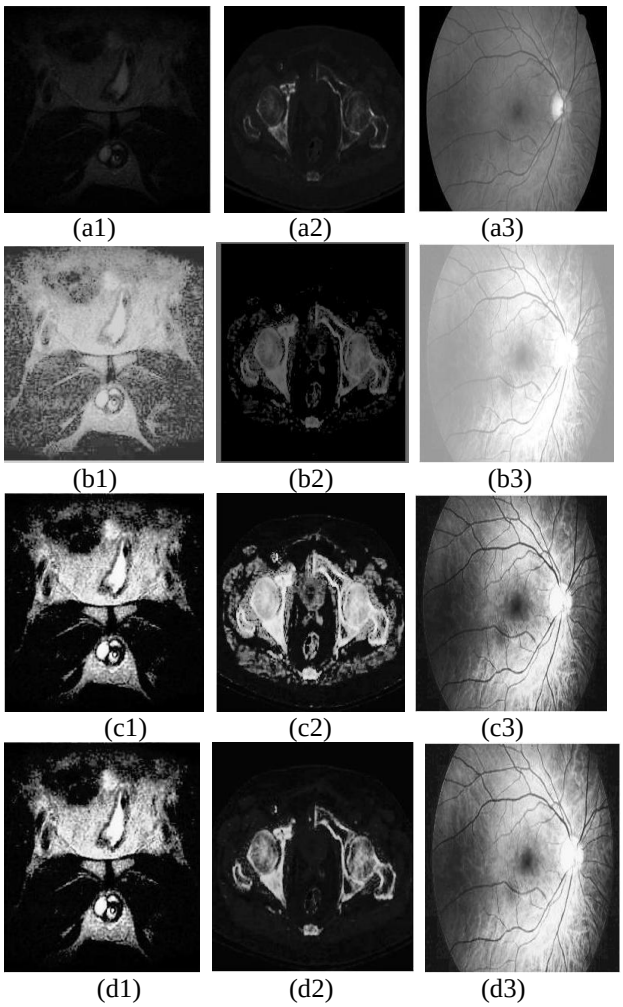
Finally Contrast improvement index (CII) is calculated to quantify contrast enhancement of an output image using Eqn. (10)

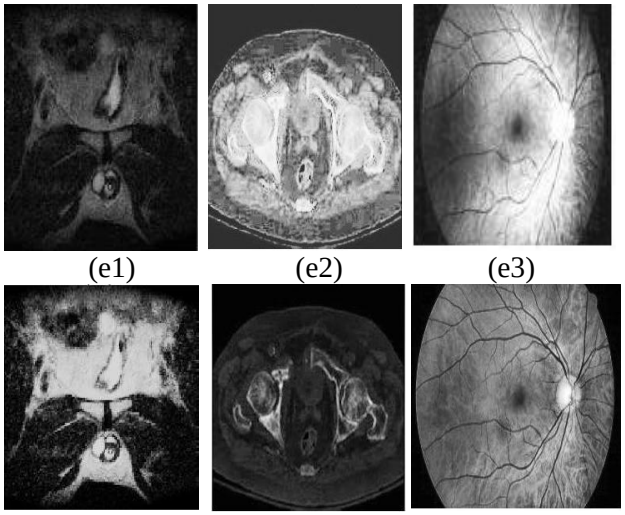
$$CII = \frac{C_{input}}{C_{0utput}} \qquad (10)$$

Where C_{output} and C_{input} denote the enhanced and input images contrast values respectively.

A. Performance Analysis

Simulation outcome of the GTDDHE approach and the four other methods are shown in figure.3 and they are investigated both quantitatively and qualitatively. The proposed GTDDHE offers highly normalized histogram distribution as shown in Figure 4.

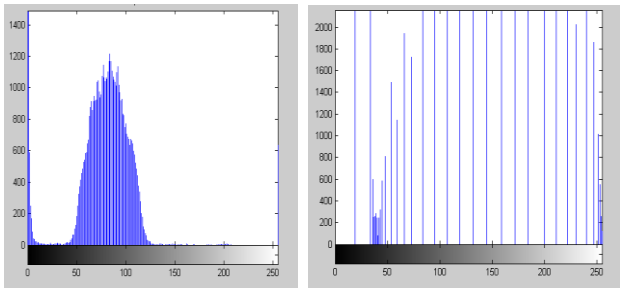




(e1) (e2) (e3)

(f1) (f2) (f3)

Figure.3 Qualitative valuation of low-contrast image enhancement: (a1-a3) Original low contrast images, (b1-b3) enhanced output image using HE [18], (c1-c3) enhanced output image using DSIHE [19], (d1-d3) enhanced output image based on RMSHE [20], (e1-e3) enhanced output image based on CLAHE [21], and (f1-f3) enhanced output image based on proposed GTDDHE method.



(a)Input Histogram (b) Output Histogram

Figure.4. Histogram comparison plot.

The proposed GTDDHE offers peak PSNR over all classes of given image sets as shows in table.1. It is also proves that inclusion of texture interpolation and gradient measures increase the PSNR and SSIM measure with improved spatial details. The proposed method outperformance both in terms of quality as well as spatial and boundary information retention. As shown in figure.5 and figure.6, the GTDDHE method shows better results compared to DWT-SVD[17], HE[18], DSIHE[19], RMSHE[20] AND CLAHE[21].

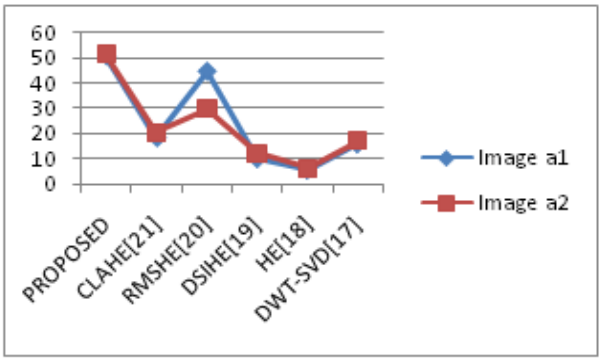


Figure.5 Comparison of PSNR

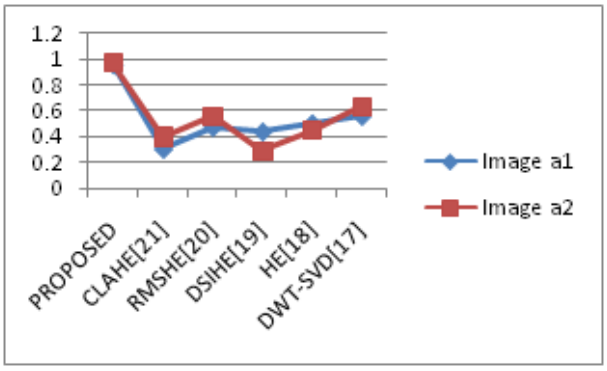


Figure.6 Comparison of SSIM

Table 1. Average metric values for 60 images

Methods/ Measures	HE	DSIHE	RMSHE	CLAHE	GTDDHE
PSNR	12.90	17.230	29.724	19.365	51.345
SSIM	0.475	0.6185	0.6206	0.5666	0.9905
MSE	3333	1060.8	500.553	859.58	0.0680
ENTROPY	4.967	5.6764	5.9957	6.0665	6.72
EMME	0.0414	0.0583	0.07407	0.0801	0.0684
MC	0.42	0.67	0.81	0.79	1.0415
CII	0.6162	0.7578	0.87419	1.2196	1.96205
AG	50.184	45.749	48.4786	54.296	61.2167

V. Conclusion

This paper investigates the quantitative effect of histogram equalization based image enhancement on various medical image modalities. The proposed color space conversion-based image enhancement technique shows significant quality improvements and overcome the distortions caused by over enhancement and associated problems in pattern classification. Here spatial texture pattern and gradient threshold bounds are applied not only for improving visual appearance of enhanced output image, but also for significantly restore the basic details required for accurate medical pattern analyzes.

References

[1]Shome, Saikat Kumar, and Siva Ram Krishna Vadali, "Enhancement of diabetic retinopathy imagery using contrast limited adaptive histogram equalization", *International Journal of Computer Science and Information Technologies*, Vol.2, No. 6, pp.2694-2699, 2011.

[2]Huang, Yu-Len, Kao-Lun Wang, and Dar-Ren Chen, "Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines", *Neural Computing & Applications*, Vol.15, No. 2, pp.164-169, 2006.

[3]El-Gamal, Fatma El-Zahraa A., Mohammed M. Elmogy, Mohammed Ghazal, Ahmed Atwan, Gregory N. Barnes, Manuel F. Casanova, Robert Keynton, and Ayman S. El-Baz, "A novel CAD system for local and global early diagnosis of Alzheimer's disease based on PIB-PET scans", *IEEE International Conference on Image Processing (ICIP)*, pp. 3270-3274, 2017.

[4]Qi, Qunfen, Paul J. Scott, Xiangqian Jiang, and Wenlong Lu, "Design and implementation of an integrated surface texture information system for design, manufacture and measurement", *Computer-Aided Design*, Vol. 57, pp.41-53, 2014.

[5]Bhateja, Vikrant, Mukul Misra, Shabana Urooj, and Aime Lay-Ekuakille, "A robust polynomial filtering framework for mammographic image enhancement from biomedical sensors", *IEEE Sensors Journal*, Vol.13, No.11, pp.4147-4156, 2013.

[6]Singh, Kuldeep, and Rajiv Kapoor, "Image enhancement using exposure based sub image histogram equalization", *Pattern Recognition Letters, Vol.*36 , pp.10-14,2014.

[7]Wong, Chin Yeow, Shilong Liu, San Chi Liu, Md Arifur Rahman, Stephen Ching-Feng Lin, Guannan Jiang, Ngaiming Kwok, and Haiyan Shi, "Image contrast enhancement using histogram equalization with maximum intensity coverage", *Journal of Modern Optics*, Vol.63, No.16,pp.1618-1629,2016.

[8]Lee, Su-Ling, and Chien-Cheng Tseng, "Color image enhancement using histogram equalization method without changing hue and saturation", *IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pp. 305-306, 2017.

[9]Wan, Minjie, Guohua Gu, Weixian Qian, Kan Ren, Qian Chen, and Xavier Maldague, "Infrared image enhancement using adaptive histogram partition and brightness correction", *Remote Sensing*, No. 10,682, 2018.

[10]Ying, Qichao, Zhenxing Qian, Xinpeng Zhang, and Dengpan Ye, "Reversible data hiding with image enhancement using histogram shifting", *IEEE Access*, pp.46506-46521, 2019.

[11]Srinivas, Kankanala, Ashish Kumar Bhandari, and Anurag Singh. "Low-contrast image enhancement using spatial contextual similarity histogram computation and color reconstruction." *Journal of the Franklin Institute*, Vol.357, No. 18, pp.13941-13963, 2020.

[12] Bai, Linfeng, Weidong Zhang, Xipeng Pan, and Chenping Zhao, "Underwater image enhancement based on global and local equalization of histogram and dual-image multi-scale fusion", *IEEE Access*, pp. 128973-128990,2020.

[13]Kandhway, Pankaj, Ashish Kumar Bhandari, and Anurag Singh, "A novel reformed histogram equalization based medical image contrast enhancement using krill herd optimization", *Biomedical Signal Processing and Control*, pp.101677, 2020.

[14]Subramani, Bharath, and Magudeeswaran Veluchamy, "Quadrant dynamic clipped histogram equalization with gamma correction for color image enhancement", *Color Research & Application* , 2020.

[15] Bhandari, Ashish Kumar, Pankaj Kandhway, and Shubham Maurya", Salp swarm algorithm based optimally weighted histogram framework for image enhancement", *IEEE Transactions on Instrumentation and Measurement*, 2020.

[16]Subramani, Bharath, and Magudeeswaran Veluchamy. "Fuzzy contextual inference system for medical image enhancement", *Measurement* ,148 pp.106967-106976, 2019.

[17]Fathi Kallel, Ahmed Ben Hamida,"A new adaptive gamma correction-based algorithm using DWT-SVD for non-contrast CT image enhancement", *IEEE Trans. Nano Biosci*, Vol.16, No.8, pp.666–675, 2017.

[18]Gonzalez RC, Woods RE, Eddins SL Digital image processing using MATLAB Gatesmark Publishing, USA, 2009.

[19]Yu Wang, Qian Chen, Baomin Zhang, "Image Enhancement Based On Equal Area Dualistic Sub-Image Histogram Equalization Method", *IEEE Transactions on Consumer Electronics*, Vol. 45, No. 1,pp.68-75, 1999.

[20]Soong-Der Chen, Abd. Rahman Ramli, "Contrast Enhancement using Recursive Mean-Separate Histogram Equalization for Scalable Brightness Preservation", *IEEE*, pp.1301-1309, 2003.

[21] Himanshu Singh, Vivek Singh, "A Comparative Analysis on Histogram Equalization Techniques for Medical Image Enhancement", *International Journals of Advanced Research in Computer Science and Software Engineering*, Vol.7, No.6, pp.364-370,2017.

J. Inst. Eng. India Ser. B
https://doi.org/10.1007/s40031-021-00693-4

ORIGINAL CONTRIBUTION

Epileptic Seizure Detection using Spectral Transformation and Convolutional Neural Networks

T. Saneesh Cleatus[1,2] · **M. Thungamani**[3]

Received: 30 November 2020 / Accepted: 17 October 2021

Abstract Automatic seizure detection and classification of seizures, as well as identification of pre-ictal activity in the electroencephalogram (EEG), are extremely important in clinical research. This decreases the time it takes to identify seizures and, as a result, improves seizure activity prediction. We propose a computer-aided method to detect the pre-ictal and ictal activity from a multichannel EEG signal. Three pre-processing techniques that are applied to EEG time-domain signals to generate an image database are proposed here. This image database is given as input to the machine learning algorithm for classification. Conversion of a time domain EEG signal to an image is accomplished by extracting EEG signal features such as correlation coefficient, short-time Fourier transform (spectrogram), and mutual information. The processed EEG waveform, which is represented as images, is used to train a convolutional neural network (CNN). The CNN classifies input signals into three classes—Seizure, Normal and Pre-ictal. We used the transfer learning method, which uses Alexnet, a pre-trained CNN architecture, for image training and classification. After training on the Spectrogram, Mutual Information, and Correlation coefficient image representations of the EEG signals, we have obtained a validation accuracy of 99.33%, 95.33%, and 97.5%, respectively.

✉ T. Saneesh Cleatus
saneesh@bmsit.in

[1] Department of Electronics and Communication Engineering, BMS Institute of Technology and Management, Bangalore, India

[2] Visvesvaraya Technological University, Bangalore, India

[3] Department of Computer Science, College of Horticulture, University of Horticultural Sciences, GKVK, Bangalore, India

Keywords Alexnet · CNN · EEG · Epilepsy

Introduction

Epilepsy refers to brain disorders that are characterized by epileptic seizures. A seizure is a brief occurrence of involuntary movements due to synchronized discharges of electrical activities in the brain. More than 50 million people worldwide suffer from epilepsy. 70% of the epileptic patients can live seizure-free if it is diagnosed properly [1]. Understanding EEG signals from an epileptic person serves as a good diagnostic tool in epilepsy research. This is due to the temporal information which is available in the EEG signal [2]. Ongoing research in experimental treatments for epilepsy is broad and comprehensive. Non-medicinal therapies such as electrical brain stimulation are being investigated in addition to clinical research [3]. The future looks bright for epilepsy treatment. Therefore, unequivocally there exists a need for accurate and timely diagnosis of this disorder. In clinical research, automatic seizure detection and classification of seizures, as well as the identification of pre-ictal activity in the electroencephalogram (EEG), are particularly significant. Thus, the time required to identify seizures can be reduced, and seizure activity prediction can be improved.

Electroencephalogram (EEG) measures and records the brain activity. It is commonly used in the diagnosis of medical conditions such as sleep disorders, detection of epileptic seizures, etc. Being a non-invasive neuroimaging modality, EEG is a popular choice for this purpose due to many reasons. The epileptic and non-epileptic paroxysmal attacks are differentiated with the help of EEG using interictal epileptiform discharges (IEDs), revealing

Published online: 05 February 2022

characteristic findings in several epilepsy syndromes [4]. Compared to other imaging techniques like Magnetoencephalography (MEG), EEG can pick up and detect electrical signals caused by the electric potential due to various neural activities and possess higher signal-to-noise ratios (SNR) for radially oriented spikes [5]. Hence, in this study, EEG data set which belongs to seizure and non-seizure classes that are made available by the University of Bonn is used.

Inspecting the EEG patterns manually is one method medical practitioners adopt in detecting abnormalities. This method is mostly tedious, time-consuming, and subjective and this might lead to the wrong diagnosis. Clinicians sometimes make errors and differ among themselves in evaluation. With the help of an automatic software-based seizure detection system, such issues can be eliminated. In this work, we propose a neural network model that can detect normal, pre-ictal and ictal stages of EEG signal. The proposed method is based on converting EEG time-domain signal to image data and classifying it using deep learning techniques.

Deep learning is a subclass of machine learning in which features are chosen by the algorithm itself. The algorithm uses representation learning wherein it automatically learn from the data itself and discovers the features that are best for the classification of the data. To learn the features we need to train the system with sufficient training data set. With enough training data, feature learning outperforms the features that are identified by engineering methods [6–9]. In the work, it has been used a class of networks called convolutional neural network (CNN). The selected network classifies EEG signals into a three-class problem. These classes are seizure EEG, normal EEG and pre-ictal EEG. It has been used AlexNet, a popular CNN architecture proposed by Krizhevsky et al. to the ImageNet challenge in December 2012 [10]. AlexNet was a breakthrough in image classification. Since then, many researchers have used this algorithm and tried on different domains apart from image classification.

In convolutional neural networks, the input to the network is a two-dimensional image whereas the EEG data used in our work is a time-series signal. Hence it was represented the EEG signal into an image by extracting the following properties from the EEG signals: Correlation coefficient, Mutual information, and short-time Fourier transform (STFT).

Methodology

The method comprises three major stages. In the first stage, it was obtained the EEG data from the BONN University's online repository. The EEG data is further pre-processed to ensure that it meets the requirements of the deep learning stage that follows. In this stage, the time domain signal is converted into a set of images. The details of transformations used to convert the EEG signal to an image and the steps involved in each case are explained. In the third stage, transformed images are given as the input to a convolutional neural network. The output of this stage classifies the signal into three classes (Healthy, Pre-ictal and Seizure). The proposed method is outlined in Fig. 1. The selection of CNN architecture and modification of the architecture to meet the requirement of transfer learning is explained in the work.

EEG Database

In this work, use of a database from the BONN university repository was made. This database consists of five sets of data namely **Z, O, N, F,** and **S**. The five sets of data fall into three classes of EEG. The folder **Z** and **O** contains the signal from healthy subjects whereas folder **F** and **N** contains the intra-cranial EEG recordings of epileptic patients during seizure-free intervals. Folder **S** contains data from epileptic patients during the occurrence of seizures. Each set contains 100 files of single-channel EEG segments. The duration of each file is 23.6 seconds. The EEG recordings in this database are sampled at a rate of 173.61 Hz. [11]. A description of each set is given in Table 1. Figure 2 shows the time-domain representation of the EEG signal from each category. It is evident from the figure that the seizure category signal has a higher amplitude in the order of 500 microvolts whereas the normal and pre-ictal category of signals has amplitude in the order of 200 microvolts.

EEG Time Series to Image Data

In this stage, the time-domain EEG signals are converted to image format. It has been proposed three algorithms for converting the EEG time series data. They are, using Spectrogram, Correlation Coefficient and Mutual Information. Spectrogram is obtained for the complete samples from each channel whereas the correlation coefficient and mutual information is evaluated between selected samples. In order to obtain these samples, we have time to slice the entire EEG sequence into small data sequences has been explained.

Time Slicing

Assume the EEG time series data be $g(n)$ and the total length of the sequence $g(n)$ is M. It has been divided the sequence $g(n)$ into the number of slices of length K, where M is an integer multiple of K.

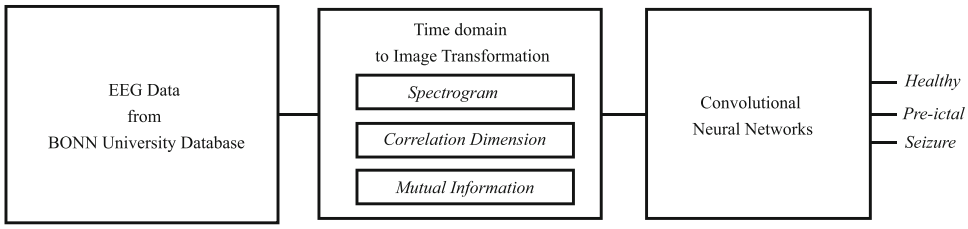


Fig. 1 Block diagram of the proposed method

Table 1 Bonn university EEG database

Database	Class	Description
A (Set Z)	Healthy	Scalp EEG recordings of healthy people with eyes open
B (Set O)	Healthy	Scalp EEG recordings of healthy people with eyes closed
C (Set N)	Pre-ictal	Intra-cranial EEG recordings of epileptic patients during seizure free intervals
D (Set F)	Pre-ictal	Epileptogenic zone EEG recordings during seizure-free intervals
E (Set S)	Seizure(Ictal)	EEG recordings of epileptic patients during seizures

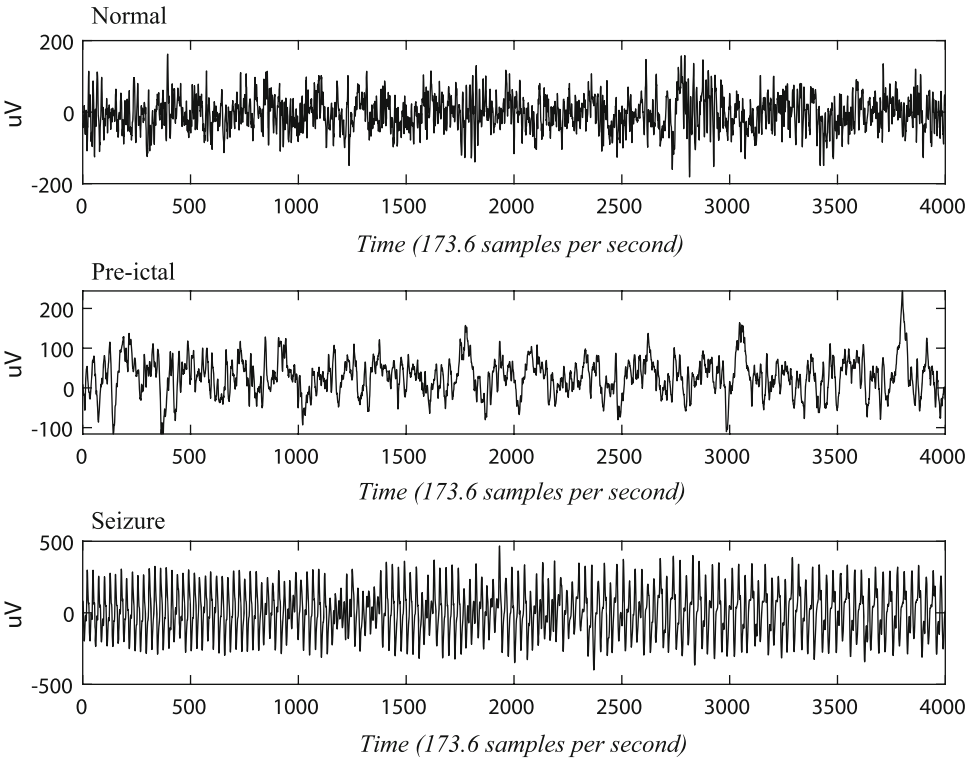


Fig. 2 Typical waveform of selected EEG Signal from each category, i.e., Normal (Z, O), Pre-ictal (F, N) and Seizure (S). In the *x*-axis, time is represented in samples. The sampling rate is 173.6 samples per second

$$g(n) = \sum_{m=0}^{K-1} g(iK + m); i = 0, 1, 2, \ldots, L \quad (1)$$

where K is the length of each slice and L is the total number of slices. For calculating the discrete time Fourier transform in order to obtain spectrogram images, the entire sequence was used as indicated in Eq. 1. For evaluating the cross correlation and mutual information, it has been derived two sequences out of the sliced series. The n^{th} sample from each slice is taken to form input sequence to the correlation and mutual information algorithm. The sequence created out of n^{th} sample can be written as

$$x_i'(n) = \sum_{m=0}^{L} g(mK) \quad (2)$$

Spectrogram

The time-domain signal is represented as an image in a spectrogram, and the time and frequency information are not lost in the process. Preserving the time-frequency information in an EEG signal is extremely beneficial. When it was understood the pattern of the signal in both the time domain and the frequency domain, we gain a better understanding of the nature of the signal. Once a pattern has been identified, it is straightforward to analyse and categorise the pattern. Another important advantage of using time-frequency analysis is that it can achieve a greater reduction in the random noises in noise-corrupted signals. Due to the large number of benefits, in spectrogram, time-frequency analysis has been implemented using STFT (Short-Time Fourier Transform) [12–14]. In this approach, it has been considered a windowed signal and apply Fourier transform to the windowed signal. Hence, the time-localized frequency information can be easily obtained from this method. In STFT, we will get the information related to the frequency contents at each time point (time window) in the signal. This is very advantageous since tasks which are depending on the time stamp can be easily controlled or labelled in various applications such as prediction of medical abnormalities, brain-computer interface systems, etc. In the STFT, the input signal $x(t)$ is multiplied using a window function $w(t)$. The window function $w(t)$ has nonzero values for a short time duration. The Fourier transform of the resulting signal (product of $x(t)$ and $w(t)$) is evaluated. The window signal is swept along the time axis. This results in a signal which can be considered as the two-dimensional representation of the input which contains both time and frequency information. For discrete-time signals, STFT is calculated as in Eq. 3.

$$X_m(\omega) = \sum_{n=-\infty}^{+\infty} x(n)w(n - mR)e^{-j\omega n} \tag{3}$$

where,

$X_m(\omega)$—DTFT of windowed sequence centered at mR.
$x(n)$ —Input signal
$w(n)$ —Window function
R—Hop size between successive DTFTs.

In this work, input signal $x(n)$ is obtained using the method described in Sect. 2.2.1 and is given as

$$x(n) = \sum_{m=0}^{K-1} g(iK + m); i = 0, 1, 2, \ldots, L \tag{4}$$

While applying the STFT to the EEG signal, we have ensured to keep the window length varying to get optimal results. It is to be noted that the total number of samples has to be integer multiples of the window length. After calculating the DTFT, we have generated the spectrogram images and upon visual inspection of the spectrogram images, it has been observed that there was a significance difference in the visual appearance of the spectrogram images for normal, pre-ictal and seizure category of signals. This turns out to be a reasonable explanation for applying convolutional neural networks for the classification purpose. The spectrogram of selected three category of signals is shown in Fig. 3. In the spectrogram, the default color-map is used. The ratio of power in dB to the frequency in Hz is mapped from indigo to yellow color with diminishing values of the ratio. The indigo color refers to the ratio of more than 140 dB/Hz where and yellow refers to less than 40dB/Hz.

Correlation Coefficient

Correlation coefficient is used to quantitatively analyze how two signals are correlated. In this work, it has been measured the correlation coefficient between two data sequences which are obtained from the samples of sliced data. The correlation coefficient values ranges from -1 to $+1$ where -1 represents a perfect negative correlation and $+1$ represents a perfect positive correlation. If the value is zero, this shows no relationship between the two signals (in this case the two EEG signals from two different channels which are not correlated) [15]. Equation 5 calculates the correlation coefficient between two input signals which are represented as time domain sequences x_i and y_i

$$r = \frac{n(\sum x_i y_j) - (\sum x_i)(\sum y_j)}{\sqrt{[n(\sum x_i^2) - (\sum x_i)^2][n(\sum y_j^2) - (\sum y_j)^2]}} \tag{5}$$

where

r—correlation coefficient
n—number of observations
x_i, y_j—input variables which is defined as instances of x_i' in Eq. 2

Figure 4 depicts the image representation of correlation coefficient values between the input sequences x_i and y_j for three distinct classes of signals.

Mutual Information

Mutual Information is a metric that quantifies the signals' mutual dependency. This is a quantitative measurement of the "amount of information" obtained about one set of data by observing the other data [16, 17]. Mutual Information between two variables X and Y can be written as

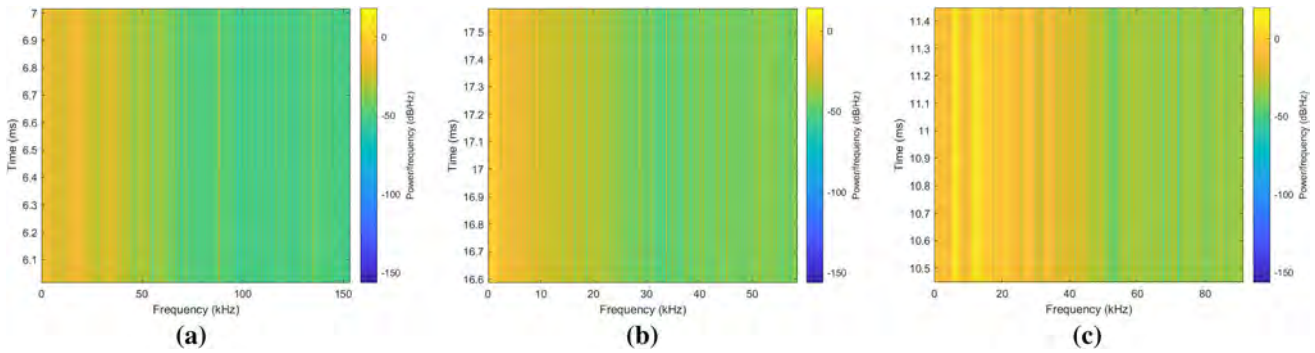


Fig. 3 Spectrogram Images of three classes **a** Healthy, **b** Preictal, **c** Seizure

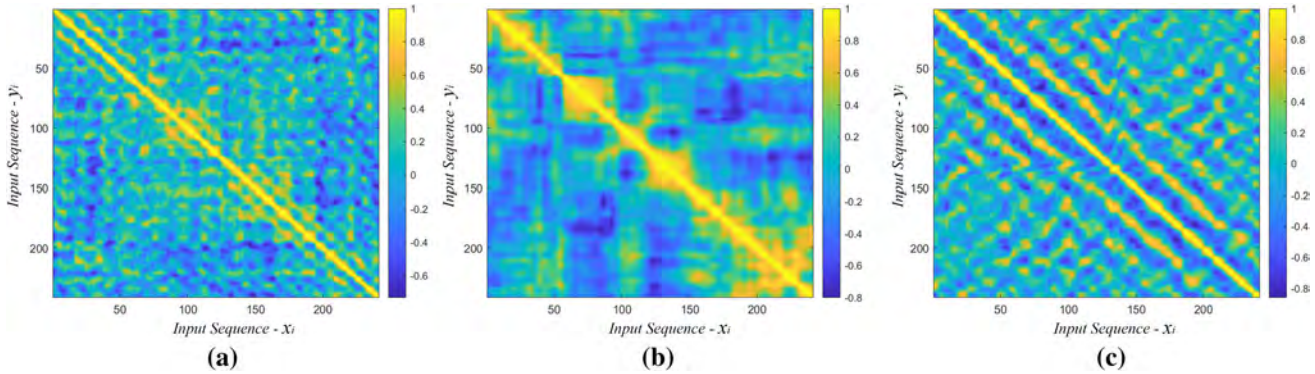


Fig. 4 Correlation Coefficient Images of three classes **a** Healthy, **b** Preictal, **c** Seizure

$$I(X;Y) = \sum_{x,y} p(x,y) log(\frac{p(x,y)}{p(x)p(y)}) \quad (6)$$

where

$I(X; Y)$ —mutual information
$p(x)$ —marginal probability mass function of X
$p(y)$ —marginal probability mass function of Y
$p(x, y)$ —joint probability mass function of X and Y

Mutual information can also be expressed in terms of entropy of the variable using Eq. 7

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \quad (7)$$

where

$H(X)$ —marginal entropy of X
$H(Y)$ —marginal entropy of Y
$H(X, Y)$ —joint entropy of X and Y

The inputs X and Y are the sequences created from the n^{th} sample of time slices as explained in the Eq. 1 and Eq. 2 respectively. The input $X = x_i$ for $i = 0, 1, 2, \ldots, L$ and $Y = x_j$ for $j = 0, 1, 2, \ldots, L$. The useful information between two inputs is obtained when $i \neq j$. The image obtained after calculating mutual information of selected three category of signals are shown in Fig. 5. The x-axis of the figure is the input sequence x_i and y-axis is the input sequence y_i which are obtained from n^{th} sample of time slices.

Procedure for Generating Images

The steps involved in generating image under each category is explained here. The process is explained in detail for one EEG signal of duration of 23.6 seconds. All three algorithms were implemented on 500 EEG files each of 23.6 seconds duration. Each signal is sampled at a rate of 173.61 samples per second, hence, the total number of samples available is 4097 per EEG signal.

For the generation of short-time Fourier transforms the standard approach is to use a window of a specific type (Rectangular, Hamming, Hanning, etc.) and time duration. It has been used the window size which varies from 17 samples (17 being a perfect divisor to get an integer quotient while dividing the value 4097) to 4097 samples and the best results in classification were achieved when the rectangular window is of size 4097. The sampling frequency was obtained by calculating the highest frequency component in the EEG signal selected and this varies for all 500 signals. Based on this method, it has been obtained one image for each EEG signal and the total number of images generated under the spectrogram process is 500 (Healthy—200, Pre-Ictal—200 and Seizure—100).

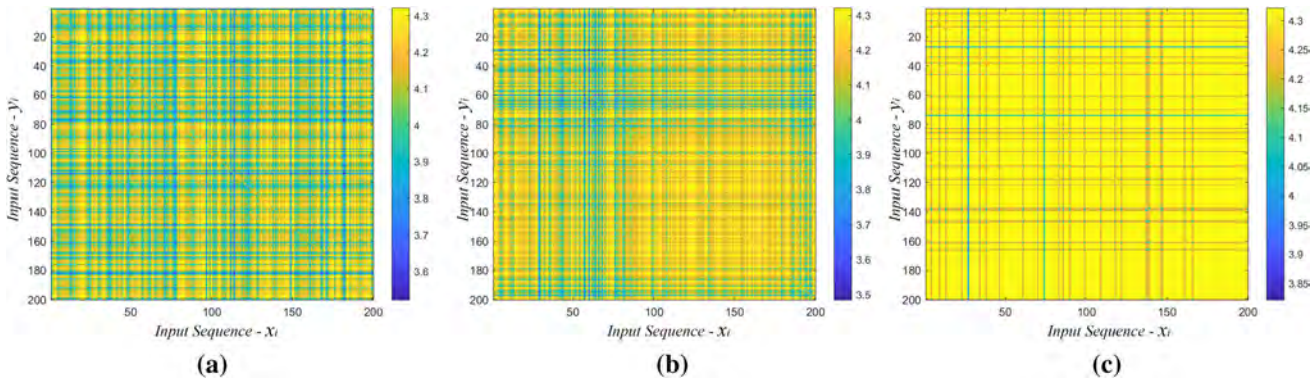


Fig. 5 Mutual Information Images of three classes **a** Healthy, **b** Preictal, **c** Seizure

To generate the images using correlation coefficients, each EEG signal of length 4097 samples are truncated to 4000 samples. This was required to create a perfect integer multiple of K (number of slices) and L (length of window) values to obtain M (length of input sequence). The highest classification accuracy was obtained when window length (L) and the number of slices (K) were 200 and 20 respectively. Similar to the spectrogram, for every individual EEG signal, it has been obtained a correlation coefficient matrix of dimension 200 × 200. This matrix was converted to an image as shown in Fig. 4. The color of the image varies from indigo to yellow as the correlation coefficient varies from − 1 to + 1. The yellow diagonal stripe in the image corresponds to correlation coefficient 1. The default file size of the image created is 875 x 656 pixels and the resolution is 150 pixels per inch.

The mutual information between every slice of 200 sample length in one EEG signal is used for generating one image as shown in Fig. 5. Like in the previous algorithm to construct the images of each EEG signal, the samples of length 4097 is truncated to 4000 samples to maintain an integer ratio between the total number of samples (M) and samples in each slice (L). It has been generated a matrix of dimension 200 × 200 with each value corresponding to the mutual information between each slice in one EEG signal. Table 2 summarizes the length of input, window length and number of slices used for generating images under each algorithm. The total number of images generated using this algorithm is 500 (Healthy—200, Pre-Ictal—200 and Seizure—100). However, for a two-class problem (Normal and Seizure), the number of images were reduced from 500 to 400 by combining the healthy and pre-ictal images into normal category and then eliminating the first 100 images under this category which carried less information. This was done manually by examining the file size of each image. With this process, the classificiation accuracy achieved was 97.5%.

The images created using the spectrogram, correlation coefficient, and mutual information for the entire EEG data set were stored in separate folders with the same name as the class type.

Convolutional Neural Network Architecture

For the classification of the images created using above mentioned methods, it has been used deep learning methods which use convolutional neural network (CNN) architecture. These architectures are generally used for image classification [18]. A standard CNN will have many layers such as input layer, convolution layer, pooling layer, dropout layer, flatten layer, fully connected layer, and output layer. Convolution, pooling, dropout, and flatten layers correspond to convolution operation whereas a fully connected layer is responsible for the classification process. The number of output nodes in the fully connected layer determines the number of classes [10]. The concept of transfer learning is widely used in which a CNN architecture, which is pre-trained on a large image data set is readily used and further it is tuned for application of our choice [19]. Several CNN architectures are proposed which

Table 2 Experiment settings for creating the images

Experiment parameters	Spectrogram	Correlation coefficient	Mutual information
Input file size (M)	4097	4000	4000
Window length (L)	Rectangular window of length 4097	200	200
Number of slices (K)	1	20	20

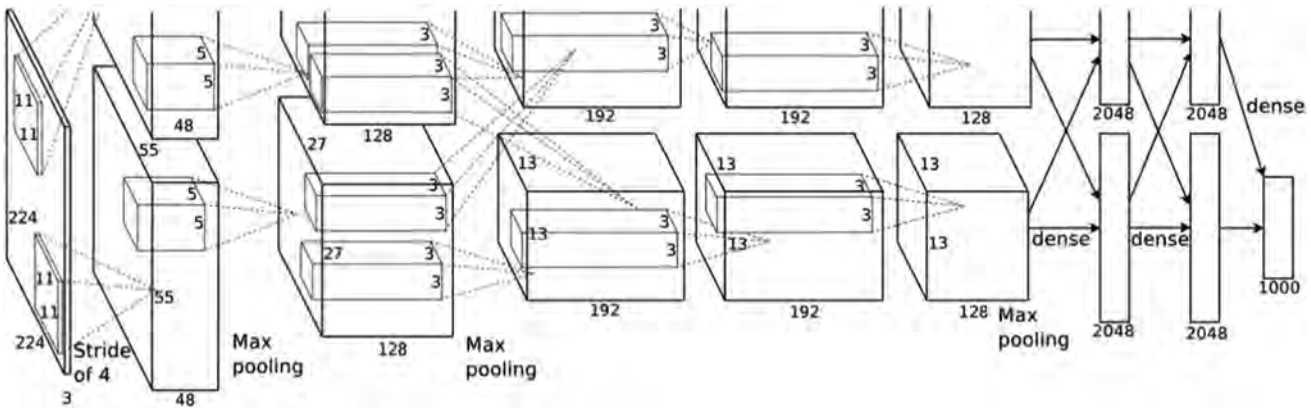


Fig. 6 Illustration of AlexNet's architecture. Image credits to Krizhevsky et al.

are successful in the image classification tasks. Some of the notable architectures are Alexnet [10], VGGnet [20], GoogleNet [21], Resnet [22] etc.

The pre-trained network was used, Alexnet for the classification purpose. AlexNet is trained on images from ImageNet database (http://www.image-net.org). AlexNet consists of eight layers and has the ability to classify images into 1000 categories [10]. The Alexnet architecture has five convolution layers of which three of them are followed by max-pooling layers. the last three layers are fully connected layers. It uses the non-saturating ReLU (rectified linear units) activation function. Due to the linear nature in comparison with *tanh* function of the sigmoid function, ReLU shows improved training performance [23]. The size of the input image for the network is 227 × 227 pixels.

The architecture and layer information of Alexnet is shown in Fig. 6. It was successfully modified the architecture for a three-class problem by changing the properties of fully connected layer in the final stage. The default number classes of typical Alexnet architecture is 1000 and it is modified to three to meet our requirement. The input images created by the methods were of dimension 875 × 656 pixels. This was scaled to 227 × 227 pixels to match the requirement of the input stage. The output stage, which is the classification layer is also replaced to suit the three-class problem.

Training

The images corresponding to each class are saved in respective folders. Totally, it has been created 1500 images using three methods. Under each method, 500 images were created which falls in any of the three classes. Under healthy and pre-ictal class, there are 200 images each and for seizure class, there are 100 images. Out of these 500 images, 70% of data is used for training the network and 30% for testing and validating the network. The selection of the training options for each class of transformation is summarized in Table 3. The training is carried out separately for spectrogram images, correlation coefficient images, and mutual information images. In each case, the training is carried out for 100 iterations where one epoch contains two iterations. The maximum training time taken was less than 3 minutes 47 seconds with the Dell workstation with Xeon Processor and Quadro P1000 GPU. Matlab R 2021a was used to perform the pre-processing and classification algorithms. The transfer learning procedure is carried out using the MATLAB deep network designer tool. The training progress graph is shown in Fig. 7.

Table 3 Selection of hyperparameters for each class of transformation

Training options	Spectrogram	Correlation coefficient	Mutual information
Solver	sgdm	sgdm	sgdm
Initial learning rate	0.0007	0.0006	0.0002
Maximum epochs	50	50	50
Minimum Batch Size	128	64	128
L2Regularizatio	0.0001	0.0001	0.0001
Momentum	0.9	0.9	0.9

Fig. 7 Training progress plot shows the final validation accuracy of 99.33% for spectrogram images, 95.33% for correlation coefficient images, and 97.5% for mutual information images

Results and Discussion

The spectrogram and correlation coefficient images were trained for classifying into healthy class, pre-ictal class and seizure class. Validation accuracy of 99.33% was obtained for the spectrogram images. For images using a correlation coefficient matrix, validation accuracy obtained was 95.33%. However, for images created with mutual information, three-class validation provided a low value of validation accuracy. Hence, it has been converted the image output to fit into two classes, healthy/normal and seizure, this was done by combining the images of healthy and pre-ictal class into normal class keeping seizure class untouched. A validation accuracy of 97.5 percent was obtained for this binary classification. Figure 7 illustrates the training progress and overall accuracy for each training example.

The performance of the algorithm was also evaluated using F1-Score. F1-Score is the harmonic mean of precision and recall. The general formula to calculate precision, recall and F1-Score, as well as accuracy is provided in Eq. 8. Since this is not a binary classification problem, we have calculated the overall F1-Score from the arithmetic mean of F1-Scores obtained for each class against other classes [24].

$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \tag{8a}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \tag{8b}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}} \tag{8c}$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8d}$$

Figure 8 shows the confusion matrix obtained for all three techniques proposed. The performance parameters such as precision, recall, accuracy and F-Score are given in Table 4.

The validation accuracy is compared with the techniques proposed by other authors. Acharya et.al performed the seizure detection using BONN University database and acheived an accuracy of 88.7 % using CNN architecture [25]. The chrononet architecture proposed by Roy et al., used a mixed architecture of CNN and RNN. This architecture was tested in abnormal EEG data from TUH database and they have achieved an accuracy of 86.57% [26]. Ye Yuan et.al., used $STFT - mConvA_{semi}$ model and achieved an ccuracy of 93.97%. They have used CHB-MIT dataset and for a two-class (ictal and non-ictal) problem [27]. Omerhodzic et al., used Wavelet-Neural Network Classifier for the BONN university data and has got an

Fig. 8 Confusion matrix generated for all three classification

Table 4 Performace parameters for each method

Performance parameter	Spectrogram	Correlation coefficient	Mutual information
Precision	99.45	95.71	95.45
Recall	98.89	96.11	98.33
Accuracy	99.33	95.33	97.50
F1-Score	99.16	95.81	96.77

Table 5 Comparison of performance of our network in terms of accuracy

Author	Classifier	Accuracy in %
Acharya et.al (2017)	CNN	88.7
Roy et al (2018) (Chua et al. 2009)	Chrononet	86.7
Ye Yuan et.al.(2018)	STFT-$mConvA_{semi}$	93.97
I. Omerhodzic et al	Wavelet NN	94
Ahmedt-Aristizabal(2018)	SVM	95.54
Proposed method with spectrogram images	**CNN**	**99.33**
Proposed method with correlation images	**CNN**	**95.33**
Proposed method with mutual information images and for two classes	**CNN**	**97.5**

Bold rows are the results of the proposed methods

overall accuracy of 94% [28]. Ahmedt et.al., for two-class problem using BONN university database, could achieve an average accuracy of 95.54% with a minimum accuracy of 90.25% between set A and D and a maximum accuracy of 99.5% between set A and E [29]. Table 5 compares the performance of our approach in terms of accuracy with similar works carried out. It has been seen that there is a significant improvement in the validation accuracy for our approach as compared to the different approaches mentioned in Table 5.

Conclusion

In this work, a method for automatically detecting epileptic seizures using electroencephalogram (EEG) signals is proposed. It was demonstrated how to pre-process EEG time-domain signals to build an image database and then use a deep learning architecture to classify the signals. This method was used to automatically classify EEG signals into three categories: normal, pre-ictal, and seizure. The one-dimensional time series data were converted to images by processing and obtaining features using the spectrogram, correlation coefficients, and mutual information from the seizure, pre-ictal, and normal EEG datasets. For training and classifying the dataset, we employed the Alexnet architecture. The spectrogram images produced an accuracy of 99.33% whereas mutual information and correlation images produced an accuracy of 95.33% and 97.5%, respectively.

Funding None.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. World Health Organization, Epilepsy: a public health imperative (2019)
2. S. Sakhavi, C. Guan, S. Yan, Learning temporal information for brain-computer interface using convolutional neural networks. IEEE Trans. Neural Netw. Learn. Syst. **29**(11), 5619–5629 (2018). https://doi.org/10.1109/TNNLS.2018.2789927
3. A. Puce, M.S. Hämäläinen, A review of issues related to data acquisition and analysis in EEG/MEG studies. Brain Sci. **7**(6), 58 (2017)
4. S. Noachtar, J. Rémi, The role of EEG in epilepsy: a critical review. Epilep. Behav. **15**(1), 22–33 (2009)
5. Piastra, M. Carla et al., A comprehensive study on electroencephalography and magnetoencephalography sensitivity to cortical and subcortical sources. Human Brain Map. **42.4**, 978–992 (2021)
6. P. Bashivan et al., Learning representations from EEG with deep recurrent-convolutional neural networks. arXiv preprint arXiv:1511.06448 (2015)
7. S. Sengupta et al., A review of deep learning with special emphasis on architectures, applications and recent trends. Knowl. Based Syst. **194**, 105596 (2020)
8. G. Dong, H. Liu (eds.), *Feature Engineering for Machine Learning and Data Analytics* (CRC Press, 2018)
9. O. Koteluk et al., How do machines learn? artificial intelligence as a new era in medicine. J. Person. Med. **11**(1), 32 (2021)
10. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks. Adv. Neural Inform. Process. Syst. **25**, 1097–1105 (2012)
11. R.G. Andrzejak et al., Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys. Rev. E **64**(6), 061907 (2001)
12. H.K. Kwok, D.L. Jones, Improved instantaneous frequency estimation using an adaptive short-time Fourier transform. IEEE Trans. Sig. Process. **48**(10), 2964–2972 (2000)
13. V. Gerla et al., P01-Comparison of short-time Fourier transform and continuous wavelet transform for frequency analysis of sleep EEG. Clin. Neurophysiol. **129**(4), e14 (2018)
14. H. Chen et al., Deep convolutional neural network with Scalogram for audio scene modeling. Interspeech. (2018)
15. H. Díaz, F. Maureira, F. Cordova, Time series of closed and open eyes EEG conditions reveal differential characteristics in the temporality of linear and non-linear analysis domain. Proc. Comput. Sci. **139**, 570–577 (2018)

16. A. Schlögl, C. Neuper, G. Pfurtscheller, Estimating the mutual information of an EEG-based brain-computer interface **3–8** (2002)
17. U. Melia et al., Mutual information measures applied to EEG signals for sleepiness characterization. Med. Eng. Phys. **37**(3), 297–308 (2015)
18. J. Schmidhuber, Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
19. A. Vedaldi, L. Karel, in *Proceedings of the 23rd ACM international conference on Multimedia*. Matconvnet: convolutional neural networks for matlab (2015)
20. K. Simonyan, Z. Andrew, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
21. C. Szegedy, et al., in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Going Deeper with Convolutions (2015)
22. K. He, et al., in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Deep Residual Learning for Image Recognition (2016)
23. R. Tibor Schirrmeister, et al., Deep learning with convolutional neural networks for decoding and visualization of eeg pathology. arXiv e-prints: arXiv-1708 (2017)
24. C. Goutte, E. Gaussier, *A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation* (European conference on information retrieval, Springer, Berlin, Heidelberg, 2005)
25. R.U. Acharya et al., Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput. Biol. Med. **100**, 270–278 (2018)
26. S. Roy, I. Kiral-Kornek, S. Harrer, in *Conference on Artificial Intelligence in Medicine in Europe*. ChronoNet: A Deep Recurrent Neural Network for Abnormal EEG Identification. (Springer, Cham, 2019)
27. Y. Yuan et al., A multi-view deep learning framework for EEG seizure detection. IEEE J. Biomed. Health Inform. **23**(1), 83–94 (2018)
28. S. Wang, EEG detection and de-noising based on convolution neural network and Hilbert-Huang transform. et al., *10th International Congress on Image and Signal Processing* (BioMedical Engineering and Informatics (CISP-BMEI), IEEE, 2017), pp. 2017
29. Y. Yuan, et al., A multi-view deep learning framework for EEG seizure detection. IEEE J. Biomed. Health Inform. **23.1**, 83–94 (2018)

Wireless Personal Communications
https://doi.org/10.1007/s11277-021-09235-z

Parametric Analysis of Compact UWB-MIMO Antenna with Improved Isolation Using Parasitic Reflectors and Protruded Ground Strips

Tathababu Addepalli[1] · V. R. Anitha[2]

Accepted: 14 October 2021

Abstract
A Compact Multiple-Input Multiple-Output (MIMO) antenna of size 16×26 mm^2 ($0.144\ \lambda_0 \times 0.234\ \lambda_0$) with improved isolation is presented in this paper. It covers the bandwidth of 2.7–14.9 GHz (FBW is 138.2%), which includes the entire UWB, X band, and K_u band applications. The elements in the structure are placed at a distance of $0.018\ \lambda_0$ (2 mm). Optimized values of two parallel strips are placed between the elements for isolation, giving $|S_{21}| > 15$ dB in most of the band. For further enhancement, four small rectangular strips with different dimensions are protruded to the ground plane, resulting in improved isolation $|S_{21}| > 20$ dB in most of the band. The radiation efficiency and peak gain values are 86–96% and 0.8–6.6 dBi respectively. Diversity performance is evaluated with the Envelope Correlation Coefficient (ECC < 0.06), Diversity Gain (DG > 9.975 dB) and Total Active Reflection Coefficient (TARC < − 10 dB). The Parametric study has been done for optimization values of the proposed model. The proposed model is simulated & fabricated on the FR4 substrate, tested and verified practically. The measured results almost follow the simulated results.

Keywords Elliptical patch · MIMO · Isolation · ECC · DG and TARC

1 Introduction

Multiple Input Multiple Output (MIMO) technology plays a key role in present wireless communication owing to the high data rate and reliability. High data rates are achieved through spatial multiplexing and reliability with space–time coding. But the major limitation of MIMO technology is coupling between antenna elements when they are arranged in a closed manner. If the separation is high, less coupling occurs but the size of the antenna housing will increase. The number of antennas connected tightly together for compactness cause more coupling between them. Coupling is introduced mainly due to currents flowing

✉ Tathababu Addepalli
babu.478@gmail.com

[1] Department of ECE, JNTUA, Anantapur, Andra Pradesh, India

[2] Department of ECE, BMS Institute of Technology & Management, Bangalore, Karnataka, India

on the metal surface, surface waves in common substrate and collision of radiation in free space. Coupling due to free space radiation can be overcome by pattern diversity technique. Effect of surface waves and currents on element to element can be reduced by placing parasitic patches/stubs, EBG structures, neutralization lines and DGS [1–4]. Ultra-wideband technology is being developed rapidly due to low power consumption and less interference. But it is facing a problem with multipath fading. This problem is overcome by MIMO technology using diversity techniques. The combination of UWB with MIMO technology is leading efficient role in wireless systems. Isolation is improved between rectangular shaped MIMO antennas of size 70×120 mm^2 and separated with a distance of 4 mm using printed stubs [5]. Circular shaped MIMO antenna of size 47×93 mm^2.

($0.49\ \lambda_0\times0.95\ \lambda_0$) with improved isolation using T- shaped stub is mentioned in [6]. Two elements hexagonal-shaped MIMO antenna of dimension 50×60 mm^2 with isolation enhancement using double-layer EBG structures is discussed in [7]. Orthogonal arrangement of four-element fractal antenna for UWB applications with 45×45 mm^2 substrate dimensions is discussed in [8]. 35×50 mm^2 ($0.333\lambda_0\times0.233\ \lambda_0$) size UWB MIMO antenna with correlation coefficient (ECC < 0.0036) is studied in [9]. Reconfigurable MIMO antenna for ultra-wideband applications with the port-to-port isolation above 20 dB is presented in [10]. Offset fed microstrip patch antenna operating from 3.1 to 12 GHz for MIMO systems is discussed in [11].

Circular shaped MIMO antenna with improved isolation using a resistive loading technique is discussed in [12]. Compact MIMO antennas structure of size of 35×35 mm^2 with improved isolation using diversity technique for UWB applications [13]. Quasi circular-shaped MIMO antenna of size 24×42mm^2 is arranged in an orthogonal manner for isolation improvement [14]. Ultra-wideband MIMO antenna, achieve high isolation using L shaped stubs as discussed in [15]. UWB-MIMO antenna of size 30×30 mm^2 with isolation enhancement using Y shaped DGS structure is discussed in [16]. Rectangular shaped UWB antenna with isolation is $S_{21} > 15$ dB for MIMO systems is discussed in [17]. Decoupling structures are used for the reduction of coupling between tightly coupled antennas [18]. Isolation is improved between antenna elements using the neutralization line technique, the size of PCB is 80×135 mm^2 as discussed in [19]. Owl shaped MIMO antenna with isolation improvement using protruded strip and the dimensions of the antenna are 26×31 mm^2 as discussed in [20]. The overall size of MIMO antennas is 22×36 mm^2 and an isolation enhancement of $S_{21} > 15$ dB for UWB-MIMO applications [21]. A compact half hexagonal-shaped UWB-MIMO antenna of size 20×34 mm^2 with high isolation ($S_{21} > 20$ dB) in most of the band using a double circular ring resonator is discussed in [22]. Eye shaped MIMO antenna with low mutual coupling and ECC values less than 0.02 for ultra-wideband applications is discussed in [23]. 18×34 mm^2 sized MIMO antenna with high isolation ($S_{21} > 22$ dB) using stubs and operating from 2 to 20 GHz is presented in [24]. Two nine shaped antenna fed with an orthogonal arrangement of microstrip lines, whose isolation is greater than 22.5 dB is discussed in [25]. Coupling is reduced between antennas to -22 dB when connected side by side with spacing of 2 mm using neutralization line and the size of the substrate is 16×35 mm^2 as mentioned in [26]. The compact elliptical-shaped MIMO antenna of dimensions 18×23 mm^2 with isolation 15 dB from 3 to 4 GHz and 20 dB from 4 to 12.4 GHz using modified T shaped stub is presented in [27].

In this communication, a compact MIMO antenna is designed, fabricated on low cost & widely available FR-4 substrate of size 16×26 mm^2 ($0.144\ \lambda_0\times0.234\ \lambda_0$), whose dielectric constant is 4.4. The antennas are separated with a distance of 2 mm ($0.018\lambda_0$), which is very small and automatically leads to high coupling ($|S_{21}| < 15$ dB) between them. To enhance the isolation, optimized values of two parallel strips are placed between the

antennas, however not sufficient for MIMO antennas. Four rectangular strips of varying dimensions are protruded to the ground plane, which further improves the isolation of $|S_{21}| > 20$ dB in the entire band. The proposed model simulated using ANSYS Electronics desktop 16.2 and checked the S-parameter values with an Agilent N5230A vector network analyzer. It was tested and verified practically in an anechoic chamber.

2 Antenna Theory

2.1 Design Process

The schematic of the proposed model with optimized values is shown in Fig. 1. It consists of two elliptical-shaped antennas, two parallel strips and DGS with protruded strips. The elements in the model are fed by a microstrip line, whose impedance is 50 Ω. The evolution of proposed model is in seven stages and it is depicted in Fig. 2. The basic single radiating antenna is designed from the basic elliptical shaped radiator equations [28] and the equations are represented in (1). To get the better impedance matching, truncation ground has been taken to the basic antenna. The parametric study has been done on various parameters like ground length, truncated ground and major axis and axial ratios of elliptical shaped patch antenna for getting optimized values. After optimizing the parameters, it provides a wide impedance bandwidth which is from 3.9 to 12.5 GHz. Later, the same radiator placed at a distance of 2 mm, 4 mm, 6 mm and 8 mm is shown in Fig. 3.

If the two radiators are closer, then more coupling is introduced. The second radiator placed at a distance of 2 mm ($0.018\lambda_0$, λ_0 measured at the lower frequency of 2.7 GHz) the obtained isolation is below 15 dB in most of the band and improvement in lower frequency is observed. To improve the isolation between elements two parallel strips with optimized dimensions are placed between them. Due to strips, isolation is improved to 15 dB in most of the band from 10 dB, except in the band 2.7–4.1 GHz and lower frequency is also improved to 2.8 GHz from 3.9 GHz. Coupling is reduced at the lower band by adding rectangular strips to the ground plane. After adding three strips to the ground plane one after the other, the isolation is improved at the lower band by 3 dB. Finally, the fourth strip of

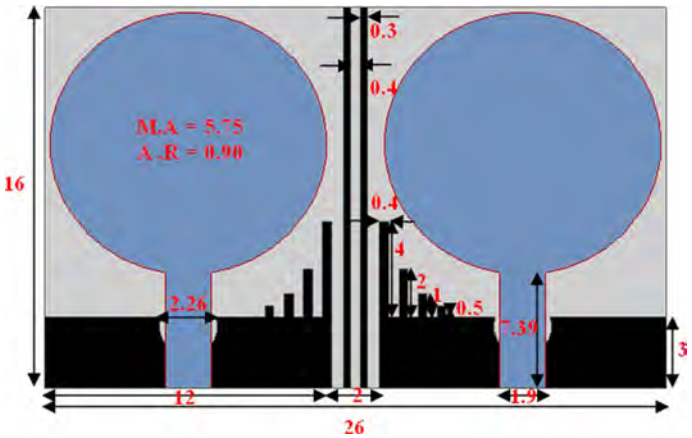


Fig. 1 Proposed model with geometry

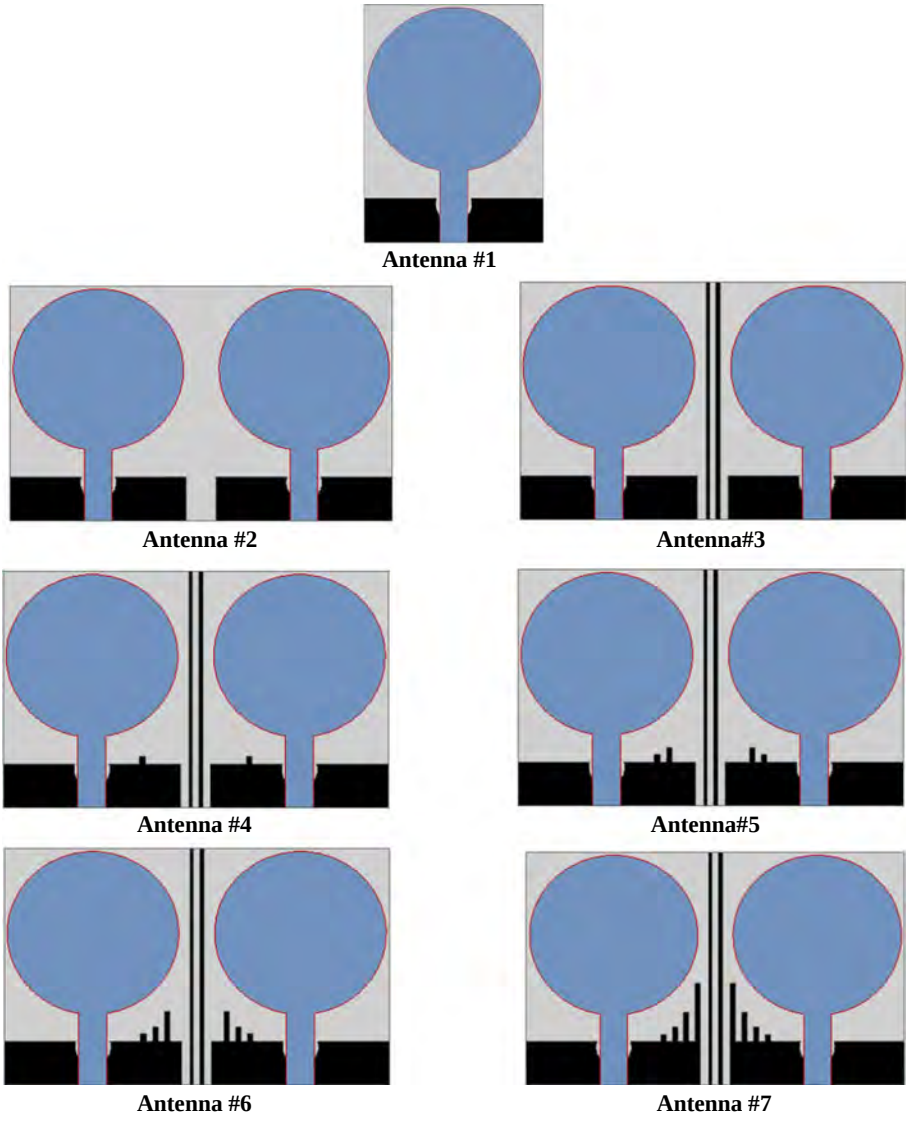


Fig. 2 Evolution stages of proposed model

size 0.4×4 mm^2 is added to the edge of the ground plane, and the isolation is improved to 20 dB in most of the band. The resultant simulated S-parameters are shown in Fig. 4.

$$f_l = \frac{7.2}{\{(L + r + p) \times k\}} \text{ GHz} \tag{1}$$

where, 'L' is the height of the planar monopole antenna in cm, 'r' is the effective radius of the equivalent monopole antenna in cm. L = 2B, r = A/4; B = semiminor axis, A = semimajor axis, 'P' = feed gap with respect to the partial ground in cm, 'K' = 1.15; (Proportionality

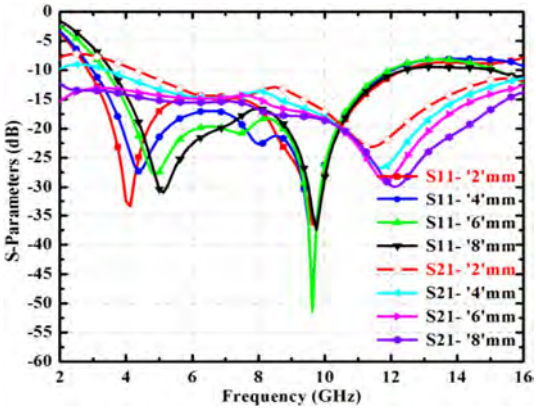


Fig. 3 Simulated S-parameters of antenna #2 with various distance separations

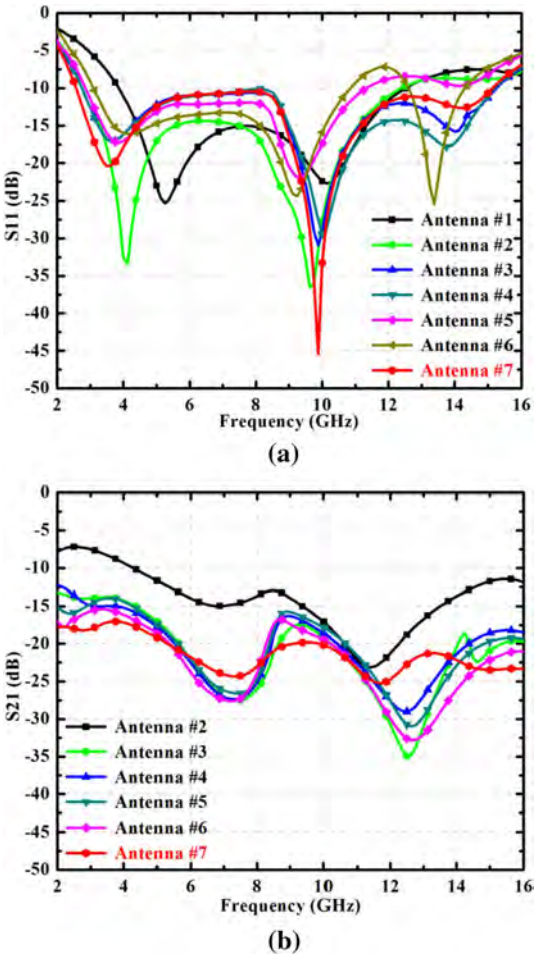


Fig. 4 **a** Simulated S_{11} values; **b** Simulated S_{21} values

constant for the 4.4 dielectric constant of FR4 substrate) estimates lower band-edge frequency within 10%.

From the basic design (antenna I):
A = 0.575 cm, B = 0.517 cm, L = 1.034 cm,
r = 0.143 cm, P = 0.52 cm, k = 1.15.
`According to the equation:
$f_l = 7.2/((1.034 + 0.143 + 0.36) \times 1.15)$
$= 4.07$ GHz.

2.2 Parametric Analysis

Parametric analysis of major & minor axis of the radiator, the width of the feed element, partial ground plane and the truncated ground plane is performed to determine the impedance performance of antenna. The major axis (M.A = 5.75 mm) & axial ratio (A.R = 0.90) and feed width 1.9 mm of the radiator are observed to give good impedance performance. The resultant graphs are shown in Fig. 5a and b and the overall impedance bandwidth is from 2.7 to 14.9 GHz. Partial ground plane and the truncated ground plane will always give wider bandwidth. The parametric study is carried with optimized values of 3 mm ground length and the axial ratio of 1 & a major axis of 1.2 mm. Figure 5c and d are represents the respective graphs. To enhance the isolation between radiators, two parallel strips are placed between them. Due to these strips of optimized values of 16 mm length give isolation above 15 dB in most of the band except at small band. Figure 5e Shows the parametric study of the length of the two strips. The effect of truncated ground on impedance matching is shown in Fig. 5f, enhancing isolation to some extent.

2.3 Surface Current Distribution

To understand the effect of one radiator on the other, the plots of surface current distributions are required. Figure 6a–h shows the surface current distributions with 3-D polar plots at 3.5 GHz, 5.5 GHz, 9.8 GHz, and 14.2 GHz. Here, parallel strips and protruded strips to the ground act as reflectors and storage elements. Due to reflections from those elements, a little amount of directivity of the proposed model increases thus increasing the gain.

3 Simulated and Measured Results

The comparison of simulated and measured S- Parameter values is shown in Fig. 7. The impedance bandwidths of simulated and measured results are obtained in the range 2.7–14.9 GHz and 2.6–14.7 GHz respectively. The simulated $|S_{21}|$ values are above 20 dB in most of the band except at the small band. The measured $|S_{21}|$ values are above 25 dB in most of the band. The small discrepancy is observed between simulated and measured values due to cable losses and tolerances in fabrication & soldering. Figure 7b and c are the front and rear view photographs of the fabricated model.

The radiation efficiency and peak gain values of the proposed model are shown in Fig. 8 and is above 95% in most of the band. The range of peak gain values of the proposed model is 0.8–6.6 dBi and the maximum value is 6.6 dB at 11.2 GHz. The radiation performance is studied at 3.5 GHz, 5.5 GHz, 9.8 GHz, and 14.2 GHz. The measured and simulated E and H filed patterns are shown in Fig. 9. The shapes of E and H fields are dumbbell and

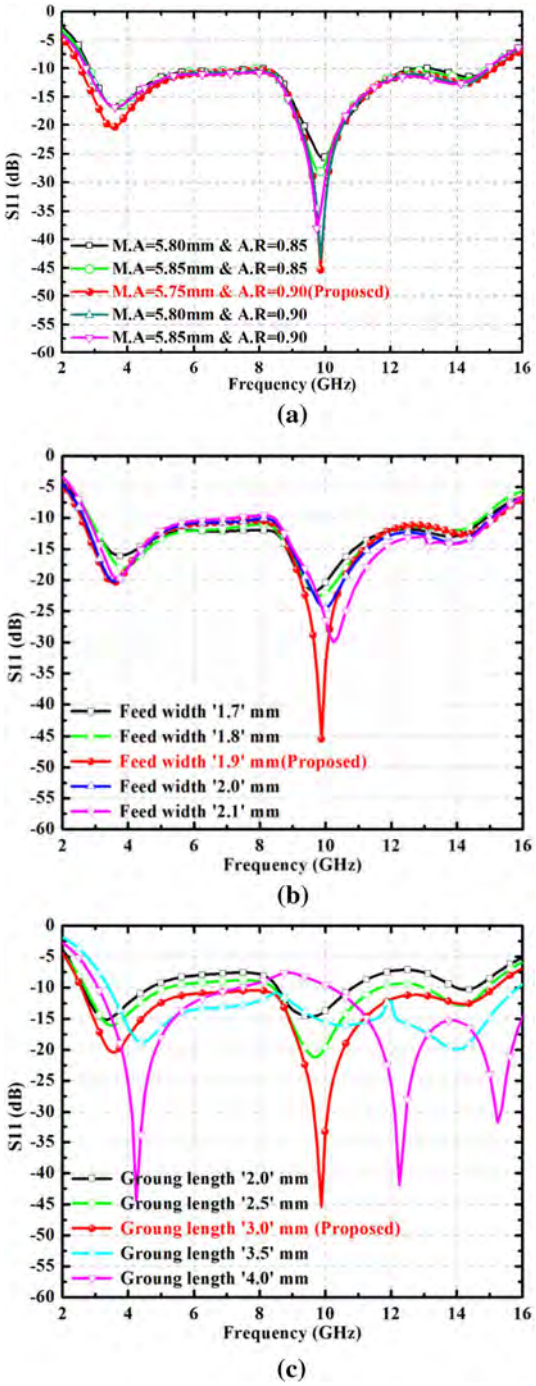
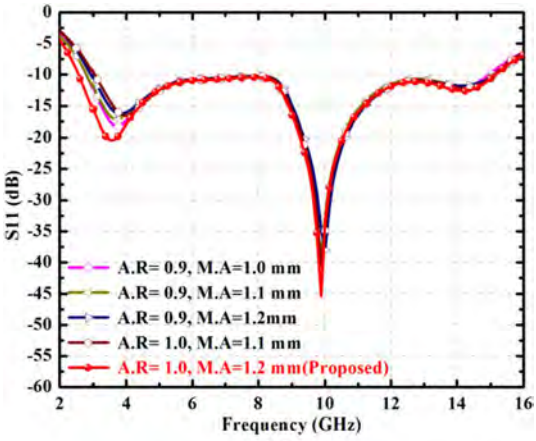
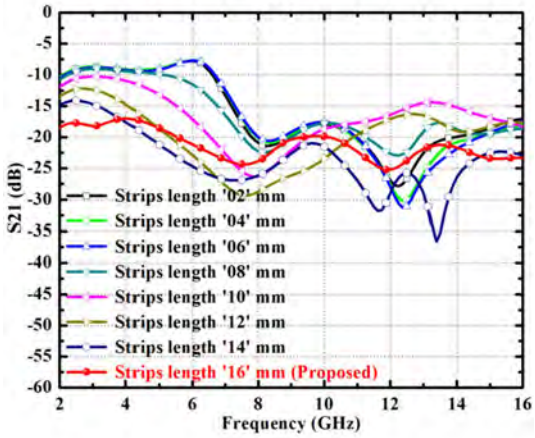


Fig. 5 Parametric analysis of **a** Major axis and Axial ratio of Radiator; **b** Feed Width of Radiator; **c** Ground Length; **d** Truncated Ground Plane; **e** Length of Parallel Strip Lines and **f** Simulated S—parameter values of partial ground with & without elliptical cut

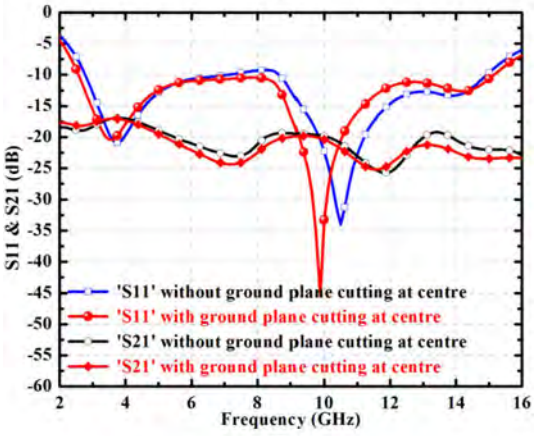
Fig. 5 (continued)



(d)



(e)



(f)

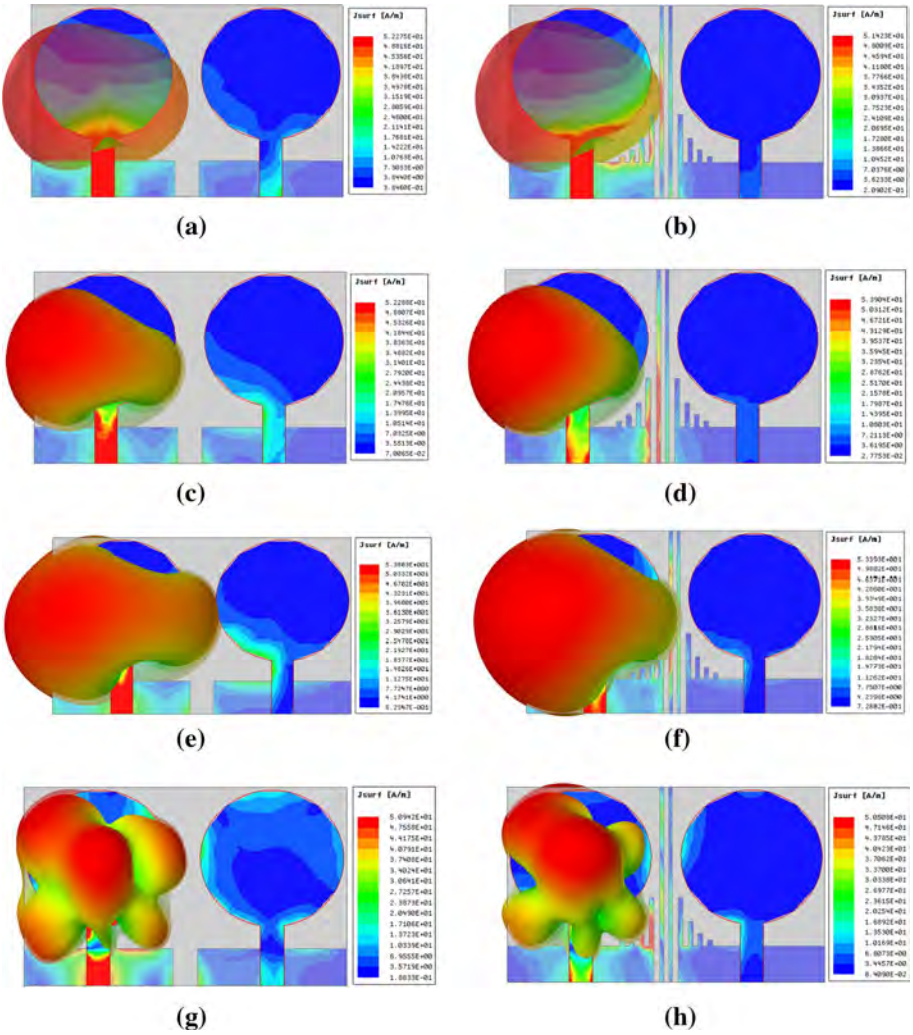


Fig. 6 Surface Current Distributions: (**a**) and (**b**) at 3.5 GHz; (**c**) and (**d**) at 5.5 GHz; (**e**) and (**f**) at 9.8 GHz and (**g**) and (**h**) at 14.2 GHz

Omnidirectional patterns at low frequencies. But, due to higher modes at higher frequencies, the patterns are disturbed. At these frequencies, the energy will be distributed to all the modes, hence patterns are not in shape of dumbbell and Omnidirectional.

4 Diversity Performance

The diversity performance of MIMO antennas is studied through three metrics namely the envelope correlation coefficient (ECC), diversity gain (DG) and total active reflection coefficient (TARC). The isolation between the two radiators can be studied easily

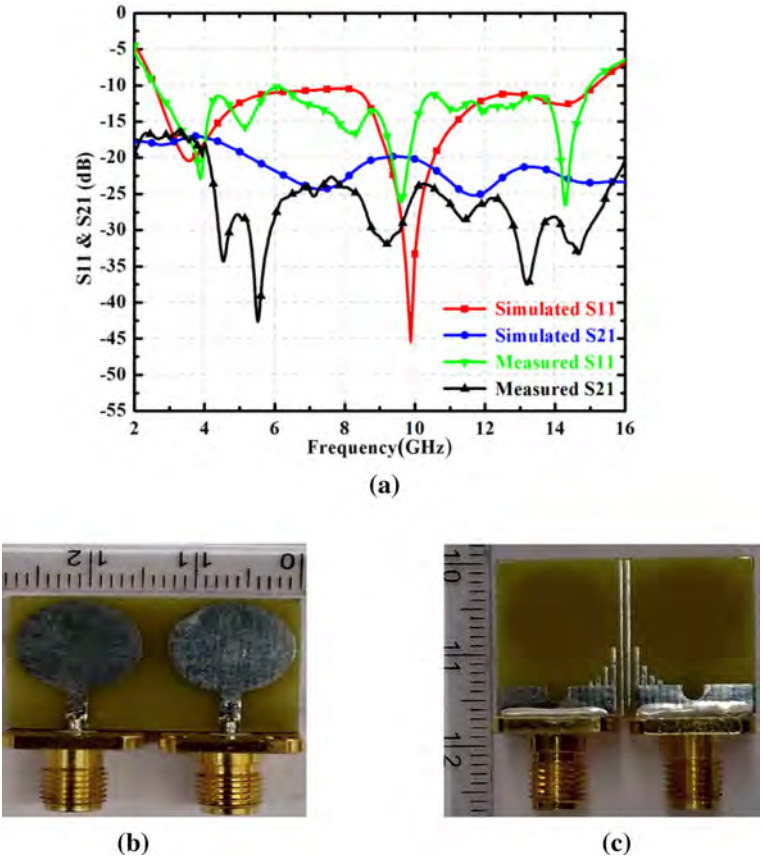


Fig. 7 (**a**) Simulated and Measured S- parameters of proposed model; (**b**) & (**c**) are photographs of front and rear views of proposed model

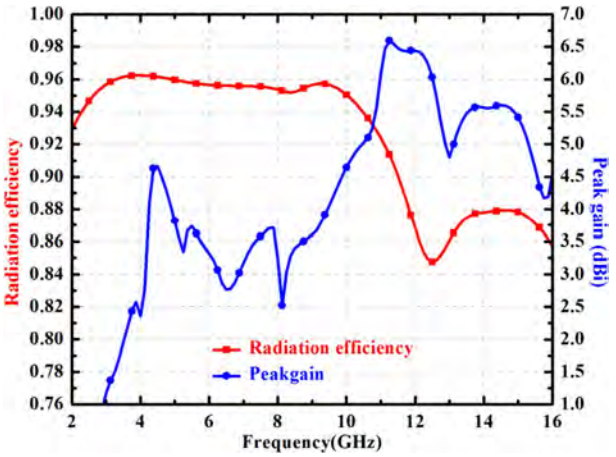


Fig. 8 Simulated radiation efficiency and peak gain values of proposed model

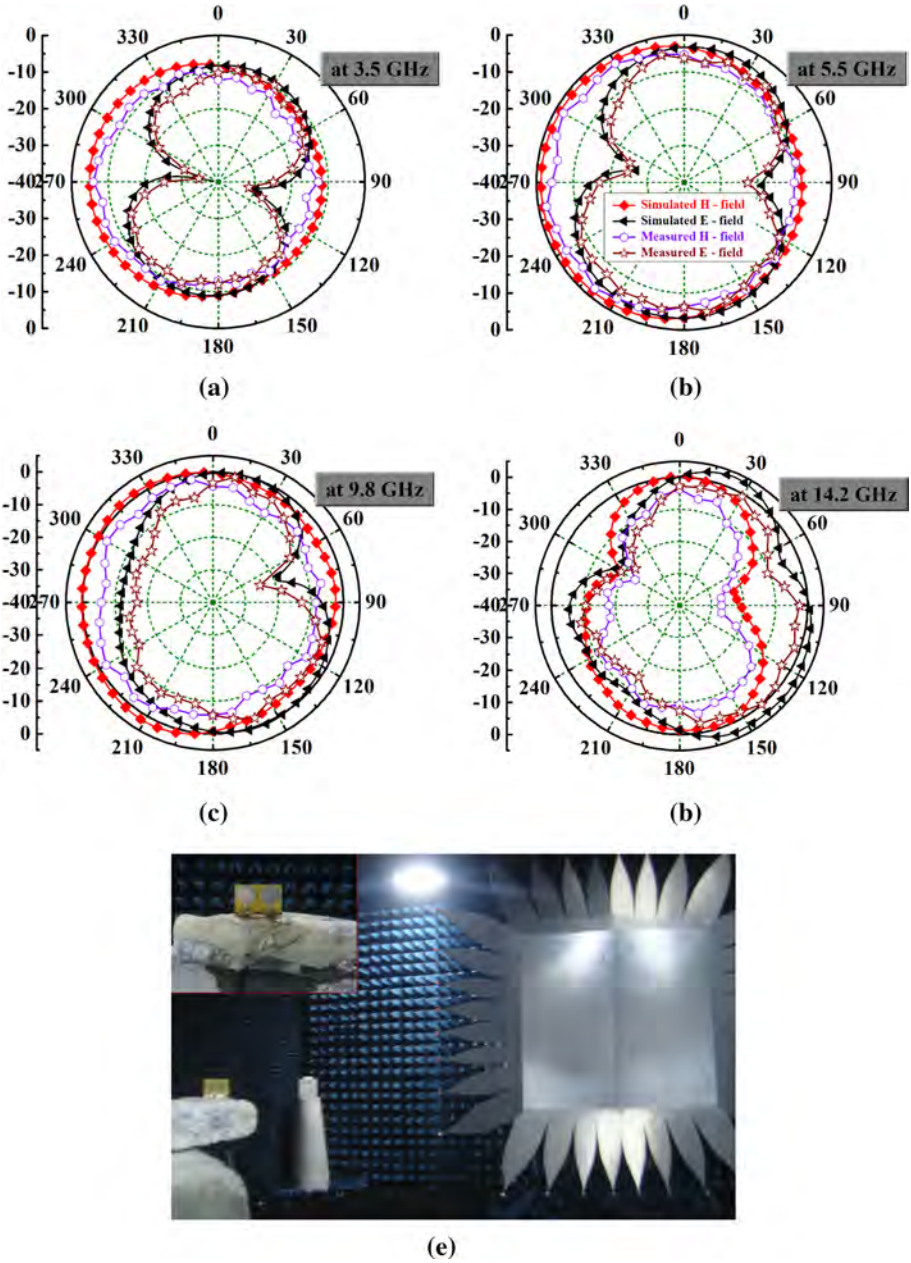


Fig. 9 E and H field radiation patterns: **a** at 3.5 GHz; **b** at 5.5 GHz; **c** at 9.8 GHz; **d** at 14.2 GHz and **e** Photograph of fabricated model in an anechoic chamber while measuring radiation patterns

through ECC value. If two radiators are more isolated, the ECC value is low (Ideally zero) and if more coupling is present between the antennas, then ECC value is high (Ideally one). But, practical acceptable value is 0.5 and it is a real number as it is a

magnitude square of signal correlation [29, 30]. The ECC value can be measured easily with S- parameter values and it is represented in Eq. (2). Diversity gain is also an important metric to evaluate the diversity performance of MIMO antenna. The relation between ECC and DG is represented in Eq. (3). Figure 10a, shows simulated values of ECC and DG parameters. The proposed model's ECC and DG values are lower than 0.06 and higher than 9.975 dB respectively.

$$ECC = \frac{\left|S_{11}^{*}S_{12} + S_{21}^{*}S_{22}\right|^{2}}{\left(1 - |S_{11}|^{2} - |S_{21}|^{2}\right)\left(1 - |S_{22}|^{2} - |S_{12}|^{2}\right)} \tag{2}$$

$$DG = 10\sqrt{1 - ECC^{2}} \tag{3}$$

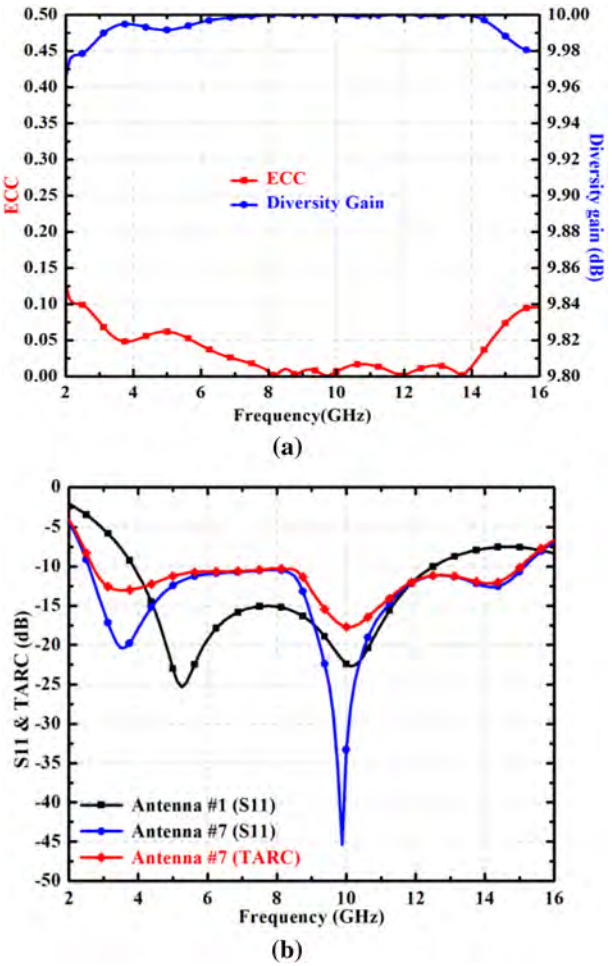


Fig. 10 **a** Simulated ECC and DG Values; **b** Simulated S_{11} and TARC values

In MIMO systems, when two radiators are operated at a time, the radiation of one radiator will affect the performance of the other radiator. This is due to changes in self impedances and mutual impedances. So, the S-parameter values of the individual radiator are not sufficient to evaluate the system performance. The entire system performance can be evaluated by considering changes in self and mutual impedances one term introduced named as total active reflection coefficient (TARC). It is defined as the square root of the difference between all the available power at ports and applied power, divide by the applied power [31] and it is represented in Eq. (4). TARC in terms of S-parameters for the two-element MIMO antenna is represented in Eq. (5). Figure 10b, describes the comparison between S_{11} values of single radiator and S_{11} & TARC values of the proposed model. The acceptable TARC values for MIMO systems is below 0 dB (TARC < 0 dB). The proposed model gives good TARC values for the entire impedance bandwidth, which is below − 10 dB.

$$TARC = \sqrt{\frac{\text{Available power} - \text{Radiated power}}{\text{Available power}}} \tag{4}$$

$$TARC = \sqrt{\frac{\left|\left(S_{11} + S_{12}e^{j\theta}\right)^2\right| + \left|\left(S_{21} + S_{22}e^{j\theta}\right)^2\right|}{2}} \tag{5}$$

The performance comparison of the proposed model with other models is listed in Table 1 in terms of size of the antenna, dielectric constant (εr), the spacing between antenna elements, impedance bandwidth, isolation, peak gain, radiation efficiency, ECC, and TARC with references [5–8, 10–15, 17, 20–23, 25 and 27] for performance comparison. The proposed model has more advantages like compactness, wider impedance bandwidth, high isolation, high-efficiency, low ECC and high DG compared with the latest literature. In most of the references, TARC values are not represented. But, TARC is an important parameter to evaluate the overall MIMO system performance. It considers the changes in individual antenna's impedances due to adjacent elements and changes in mutual impedances. So, S- parameters of individual antennas are not sufficient for analyze the overall system performance. The proposed model gives good TARC values, which are below − 10 dB (Acceptable TARC values are below 0 dB). Hence, the proposed model is suitable choice for the portable wireless device applications.

5 Conclusion

A Compact MIMO antenna of size 16 × 26 mm2 (0.144 λ0 × 0.234 λ0) is presented in this paper. The elements in the model are separated with a distance of 2 mm (0.018λ0). The proposed model gives an impedance bandwidth ($|S_{11}|$ >10 dB) of 2.7 to 14.9 GHz. Two parallel and four strips are protruded to the ground plane for the isolation enhancement. Due to these elements, the isolation ($|S_{21}|$) is improved to above 20 dB. The radiation efficiency and peak gain values are above 86% for the entire band and 0.8 to 6.6 dBi respectively. The diversity performance is checked with metrics ECC, DG and TARC values. These are below 0.06, above 9.975 dB and below − 10 dB respectively. The proposed model covers the entire UWB, X- band and Ku band applications.

Table 1 Performance comparison of proposed model with other models

Ref. no	Antenna size (mm^2)	εr	Elements separation (λ_0)	Impedance bandwidth (GHz)	Isolation $\lvert S_{21} \rvert$ (dB)	Peak gain (dBi)	R.E (%)	ECC	TARC (dB)
#05	70 × 120	3.02	0.02	1.6–4	≥ 20	3.8	–	–	–
#06	47 × 93	4.4	0.35	3.1–10.6	≥ 31	3.5	≥ 75	− 29 dB	–
#07	50 × 60	4.5	0.50	3–6	≥ 21	–	–	–	–
#08	45 × 45	4.4	–	2–10.6	≥ 17	3	–	0.005	–
#10	37.5 × 40	4.4		2–11	≥ 20	–	–	0.1	≤ − 4
#11	38.5 × 38.5	4.4	–	3.08–11.8	≥ 15	1.4–3.6	≥ 75	-	–
#12	35 × 36	2.2	0.12	3–9	≥17	1–5.5	≥ 87	0.02	–
#13	35 × 35	4.4	–	3–12	≥ 20	–	–	0.40	–
#14	30 × 40	4.4	–	3.1–10.6	≥ 16	–	≥ 70	0.15	–
#15	24 × 42	4.3	–	3.1–10.9	≥ 17	1.8–3.5	≥ 75	0.10	–
#17	20 × 40	4.4	–	3–11	≥ 15	1.5–5	≥ 80	0.8	–
#20	26 × 31	4.3	–	3.1–11.12	≥ 20	2.5–5.54	–	0.002	–
#21	22 × 36	3.5	–	3.1–11	≥ 15	1–5	≥ 70	0.1	–
#22	20 × 34	4.4	–	3.1–11	≥ 20	3–4	≥ 75	0.2	–
#23	18 × 36	4.4	–	2.8–20	≥ 20	1.6–6	–	0.02	≤ − 10
#25	21× 27	4.4	0.15	5.1–5.4 & 7.3–7.6	≥ 22.5	> 9.38	≥ 79.8	0.1	≤ − 10
#27	18 × 23	2.65	–	3–12.4	≥ 20	4	≥ 70	0.1	–
Proposed	16 × 26	4.4	0.018	2.7–14.9	≥ 20	0.8–6.6	≥ 86	0.06	≤ − 10

Acknowledgements Authors feel a need to convey heartfelt thanks to Rakesh Kumar Singh, Scientist 'F' and Head, Stealth technologies and test ranges, Directorate of radar seekers and systems (DRSS), RCI, DRDO, Hyderabad, India for providing lab facilities for testing and measure the radiation patterns of the fabricated antenna as a part of research work.

Authors' contributions All authors read and approved the final manuscript.

Funding For this manuscript there is no funding availability.

Availability of data and material In the manuscript must include an 'Availability of data and materials' statement.

Code availability For this manuscript there is no Code availability.

Declarations

Conflicts of interest The authors declare that they have no conflicts of interest /competing interests" in this section.

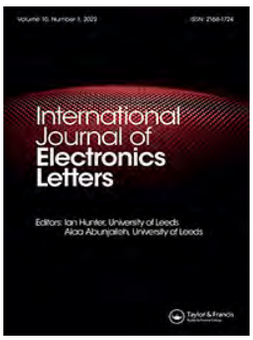
References

1. Jayanthi, K., & Kalpana, A. M. (2017). Mutual coupling reduction techniques between MIMO antennas for UWB applications. *International Journal of Recent Innovation Trends Computer Communications* 5(9).
2. Addepalli, T., & Anitha, V. R. (2020). Compact two-port MIMO antenna with high isolation using parasitic reflectors for UWB, X and Ku band applications. *Progress In Electromagnetics Research C, 102*, 63–77.
3. Nadeem, I., & Choi, D.-Y. (2018). Study on mutual coupling reduction technique for MIMO antennas. *IEEE Access, 7*, 563–586.
4. Addepalli, T., & Anitha, V. R. (2020). A very compact and closely spaced circular shaped UWB MIMO antenna with improved isolation. *AEU-International Journal of Electronics and Communications, 114*, 153016.
5. Isaac, A. A., Al-Rizzo, H., Yahya, S., Al-Wahhamy, A., & Abushamleh, S. (2018). Decoupling of two closely-spaced planar monopole antennas using two novel printed-circuit structures. *Microwave and Optical Technology Letters, 60*(12), 2954–2963.
6. Radhi, A. H., Nilavalan, R., Wang, Y., Al-Raweshidy, H. S., Eltokhy, A. A., & Ab Aziz, N. (2018). Mutual coupling reduction with a wideband planar decoupling structure for UWB–MIMO antennas. *International Journal of Microwave and Wireless Technologies, 10*(10), 1143–1154.
7. Li, Q., Feresidis, A. P., Mavridou, M., & Hall, P. S. (2015). Miniaturized double-layer EBG structures for broadband mutual coupling reduction between UWB monopoles. *IEEE Transactions on Antennas and Propagation, 63*(3), 1168–1171.
8. Tripathi, S., Mohan, A., & Yadav, S. (2015). A compact Koch fractal UWB MIMO antenna with WLAN band-rejection. *IEEE Antennas and wireless propagation letters, 14*, 1565–1568.
9. Chithradevi, R., & Sreeja, B. S. (2017). A compact UWB MIMO antenna with high isolation and low correlation for wireless applications. In: *2017 IEEE International Conference on Antenna Innovations & Modern Technologies for Ground, Aircraft and Satellite Applications (IAIM)*. IEEE.
10 Jafri, S. I., Saleem, R., Shafique, M. F., & Brown, A. K. (2016). Compact reconfigurable multiple-input-multiple-output antenna for ultra-wideband applications. *IET Microwaves, Antennas & Propagation, 10*(4), 413–419.
11. Kang, L., Li, H., Wang, X., & Shi, X. (2015). Compact offset microstrip-fed MIMO antenna for band-notched UWB applications. *IEEE Antennas and Wireless Propagation Letters, 14*, 1754–1757.
12. Park, J.-D., Rahman, M., & Chen, H. N. (2019). Isolation enhancement of wide-band MIMO array antennas utilizing resistive loading. *IEEE Access, 7*, 81020–81026.
13. Zhu, J., Li, S., Feng, B., Deng, L., & Yin, S. (2015). Compact dual-polarized UWB quasi-self-complementary MIMO/diversity antenna with band-rejection capability. *IEEE Antennas and Wireless Propagation Letters, 15*, 905–908.

14. Deng, J.-Y., Guo, L.-X., & Liu, X.-L. (2015). An ultra-wideband MIMO antenna with a high isolation. *IEEE antennas and wireless propagation letters, 15*, 182–185.
15 Alsath, M. G. N., & Kanagasabai, M. (2015). Compact UWB monopole antenna for automotive communications. *IEEE Transactions on Antennas and Propagation, 63*(9), 4204–4208.
16. Zhu, J., Feng, B., Peng, B., Deng, L., & Li, S. (2016). A dual notched band MIMO slot antenna system with Y-shaped defected ground structure for UWB applications. *Microwave and Optical Technology Letters, 58*(3), 626–630.
17. Mathur, R., & Dwari, S. (2019). Compact planar reconfigurable UWB-MIMO antenna with on-demand worldwide interoperability for microwave access/wireless local area network rejection. *IET Microwaves, Antennas & Propagation, 13*(10), 1684–1689.
18. Zhao, L., & Ke-Li, Wu. (2015). A dual-band coupled resonator decoupling network for two coupled antennas. *IEEE Transactions on Antennas and Propagation, 63*(7), 2843–2850.
19. Wang, S., & Zhengwei, Du. (2014). Decoupled dual-antenna system using crossed neutralization lines for LTE/WWAN smartphone applications. *IEEE Antennas and Wireless Propagation Letters, 14*, 523–526.
20. Mchbal, A., Amar Touhami, N., Elftouh, H., & Dkiouak, A. (2018). Mutual coupling reduction using a protruded ground branch structure in a compact UWB owl-shaped MIMO antenna. *International Journal of Antennas and Propagation, 2018*.
21 Liu, Li., Cheung, S. W., & Yuk, T. I. (2015). Compact MIMO antenna for portable UWB applications with band-notched characteristic. *IEEE Transactions on Antennas and Propagation, 63*(5), 1917–1924.
22. Mathur, R., & Dwari, S. (2019). A compact UWB-MIMO with dual grounded CRR for isolation improvement. *International Journal of RF and Microwave Computer-Aided Engineering, 29*(1), e21500.
23. Chandel, R., Gautam, A. K., & Rambabu, K. (2018). Design and packaging of an eye-shaped multiple-input–multiple-output antenna with high isolation for wireless UWB applications. *IEEE Transactions on Components, Packaging and Manufacturing Technology, 8*(4), 635–642.
24. Chandel, R., Gautam, A. K., & Rambabu, K. (2018). Tapered fed compact UWB MIMO-diversity antenna with dual band-notched characteristics. *IEEE Transactions on Antennas and Propagation, 66*(4), 1677–1684.
25. Dkiouak, A., Zakriti, A., & El Ouahabi, M. (2020). Design of a compact dual-band MIMO antenna with high isolation for WLAN and X-band satellite by using orthogonal polarization. *Journal of Electromagnetic Waves and Applications, 34*(9), 1254–1267.
26. Zhang, S., & Pedersen, G. F. (2015). Mutual coupling reduction for UWB MIMO antennas with a wideband neutralization line. *IEEE antennas and wireless propagation letters, 15*, 166–169.
27. Tao, J., & Feng, Q. (2016). Compact ultra-wideband MIMO antenna with half-slot structure. *IEEE Antennas and wireless Propagation letters, 16*, 792–795.
28 Ray, K. P. (2008). Design aspects of printed monopole antennas for ultra-wide band applications. *International journal of antennas and propagation, 2008*, 1–8.
29. Sharawi, M. S., Hassan, A. T., & Khan, M. U. (2017). Correlation coefficient calculations for MIMO antenna systems: A comparative study. *International Journal of Microwave and Wireless Technologies, 9*(10), 1991–2004.
30. Addepalli, T., & Anitha, V. R. (2019). Design and parametric analysis of hexagonal shaped MIMO patch antenna for S-band, WLAN, UWB and X-band applications. *Progress In Electromagnetics Research C, 97*, 227–240.
31. Fritz-Andrade, E., Jardon-Aguilar, H., & Tirado-Mendez, J. A. (2020). The correct application of total active reflection coefficient to evaluate MIMO antenna systems and its generalization to N ports. *International Journal of RF and Microwave Computer-Aided Engineering, 30*(4), e22113.

Tathababu Addepalli was born in 1986. He received his B. Tech degree from JNTUH, Hyderabad in 2007 and M. Tech degree from JNTUK, Kakinada in 2010. Currently, he is a full-time research scholar in JNTUA, Anantapuramu and registered in 2017. He has published more than five research papers in various reputed international journals and conferences like ELSEVIER (AEU), PIER, Wiley (IJCS), MDPI (Electronics), IEEE and SPRINGER conferences, etc. His areas of interest are microstrip patch antennas, MIMO antennas, UWB antennas, Characteristic Mode Analysis(CMA), THz antennas and metamaterial antennas.

V.R. Anitha has been working as a Professor, Department of ECE in BMS Institute of Technology & Management, Bangalore, Karnataka, INDIA. She obtained Bachelor of Technology in the Department of Electronics and Telecommunication Engineering from Institute of Electronics and Tele Communications Engineering (IETE), New Delhi and Post Graduate in Microwave Engineering from Nagarjuna University Guntur, and Ph.D. in Microwave Engineering from Sri Venkateswara University, Tirupati. She is a member in IEEE, IETE and ISSS and published 38 papers in national and international journals and conferences and presented many technical lectures in various scientific forums (International & National). She was invited to deliver lectures in various Institutions. She received Travel Grant award from URSI General Assembly-2011, Istanbul, Turkey. She was session chair in International Conference on ICEAA-IEEE APWC-EMS'13, Torino, Italy. She received "Young Scientist" grant from DST for her presentation held at Kuala Lumpur, Malaysia in October 2013. Presently she completed couple of research projects under INUP and NPMASS and she is holding research projects from various funding agencies. She is doing collaborative research with Yeungnam University, Republic of Korea and CSIR-NAL. Dr. Anitha is continuously upgrading her expertise in the areas of Electrically Small Antennas (ESA), Microstrip Antennas, Fractal Antennas and MEMS. Dr.Anitha is instrumental in establishing a National MEMS Design Centre at Sree Vidyanikethan Engineering College under the aegis of NPMASS. Competence in handling COMSOL, CoventorWare and Intellisuite will provide an ideal environment for conducting simulation studies in the area of MEMS, HFSS, ADS & IE3D will provide platform for conducting simulation studies in the area of Antennas & Microwaves. She played a key role to obtain Fund of Rs. 20 laks for Improvement of S&T Infrastructure (FIST) from DST during 2012–13. Also, she is holding few funded projects from government agencies. She is serving as Ph.D guide, BOS member, reviewer, editor and DC member.



International Journal of Electronics Letters

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tetl20

X-band Miniaturized frequency reconfigurable travelling wave antenna

Narjala Sri Pravallika, V R Anitha & K. Rama Naidu

To cite this article: Narjala Sri Pravallika, V R Anitha & K. Rama Naidu (2022): X-band Miniaturized frequency reconfigurable travelling wave antenna, International Journal of Electronics Letters, DOI: 10.1080/21681724.2022.2068194

To link to this article: https://doi.org/10.1080/21681724.2022.2068194

Published online: 27 Apr 2022.

INTERNATIONAL JOURNAL OF ELECTRONICS LETTERS
https://doi.org/10.1080/21681724.2022.2068194

Taylor & Francis
Taylor & Francis Group

X-band Miniaturized frequency reconfigurable travelling wave antenna

Narjala Sri Pravallika[a], V R Anitha[b] and K. Rama Naidu[c]

[a]Research Scholar, Department of ECE, JNTUA, Anantapur, India; [b]Department of ECE, BMS Institute of Technology & Management, Bangalore, India; [c]Department of ECE, JNTUA, Anantapur, India

ABSTRACT

The paper presents a novel miniaturized ($1.2\ \lambda_0 \times 0.5\ \lambda_0 \times 0.026\ \lambda_0$) frequency reconfigurable travelling wave antenna (TWA) operating at 8–12 GHz band. The antenna uses PIN diodes arranged in a novel configuration to produce frequency reconfigurability and operates at single frequency bands of 8.5, 9, 9.35, 9.9, 10.45 and 11 GHz with an average bandwidth of 1 GHz based on different switch configurations. The proposed antenna demonstrates an efficiency>80% for the operating frequency in different stations while maintaining a stable average gain of 8.5 dBi. The antenna's compact nature and high gain make it suitable for several applications within modern satellite communication devices and radar systems. The proposed antenna is designed and tested for its conformality using CST Microwave-Studio. The proposed design is further verified with a fabricated prototype measured in an anechoic chamber, agreeing well with the simulated results.

ARTICLE HISTORY
Received 4 October 2021
Accepted 6 March 2022

KEYWORDS
TWA; frequency reconfigurability; leakywave; miniaturization; conformality

1. Introduction

To support the recent technical advancements, miniaturized reconfigurable antennas have been in great demand for remote sensing applications. Frequency reconfigurable antennas allow the user to resonate at different frequency bands, allowing several functions to be operated within the same device. Furthermore, frequency reconfigurable antennas allow the development of compact and flexible systems.

Different frequency reconfigurability techniques have been explored in the literature. One of the most common techniques is changing the antenna's electrical length, resulting in a change in the resonant frequency. Planar antennas such as microstrips (Bansal & Gupta, 2020) are demonstrated to be easily reconfigured using PIN diodes and micro-electro-mechanical system (MEMS) switches. Several antennas have been presented in the literature (Abdulraheem et al., 2017; Fadamiro et al., 2019; Ismail et al., 2021; Sahu & Sharma, 2017; Valkonen et al., 2010) where such RF switches are used in radiating elements and a ground plane to control the surface current and, hence, affect the band of operation. Such RF switches allow fast response time and ease of fabrication, making them very popular for several reconfigurable antennas. Another interesting method of

CONTACT Narjala Sri Pravallika sripravallika37@gmail.com Electronics and Communication Engineering Jawaharlal Nehru Technological University - Anantapur 515002 India

using PIN diodes for frequency reconfigurability was presented in (Nikolaou et al., 2009), where three frequency bands were switched by truncating the antenna using closed stub matching networks. Another standard method for introducing frequency reconfigurability involves varying electrical properties of the resonator using varactor diodes. Varactors introduce an external capacitance that can be varied with the change in voltage, which further introduces frequency reconfigurability (Atallah et al., 2016; Shirazi et al., 2018; Young et al., 2015).

Apart from the reconfigurable characteristics of an antenna, miniaturization is another essential factor. Compact antennas allow smaller form-factor for the overall device (Iftikhar et al., 2020). Several ultra-wideband (UWB) and multi-band antennas with some sort of reconfigurability have been presented in the literature (Arif et al., 2019; Gupta et al., 2020; Luadang et al., 2018; Mohamadzade et al., 2020; Paracha et al., 2018; Simorangkir et al., 2018), but such antennas tend to have limited applications. Modern-day communication systems require compact solutions with multi-band operation and easy switching mechanisms (Ahmad et al., 2018).

This paper presents a compact travelling wave antenna operating at X-band with frequency reconfigurability. Antenna design methodology and the novel switching technique are described in Section 2. Antenna's performance with the help of simulation and further verification with measured results are presented in Section 3. Parametric analysis for the antenna and results for its conformality is described in Section 4, followed by a conclusion in Section 5.

2. Proposed design

This section provides the geometrical description of the proposed antenna and proposed switched array configuration. The antenna is demonstrated to be reconfigurable with the help of PIN diodes used to control the electrical parameters of the antenna.

2.1. Antenna geometry

The proposed antenna geometry is presented in Figure 1. The proposed antenna is designed on an RT-Duroid 5880 substrate (relative permittivity, $\varepsilon_r = 2.2$; loss tangent, $tan\ \delta = 0.0009$) of height, $h = 0.787$ mm. The unit cell is designed to be a mutually coupled S-shaped microstrip line with an open stub fed using a microstrip line. The proposed structure is an array of 8 S-shaped radiating elements connected using PIN-diodes. The S-shaped radiating elements are bent at an angle of 30° to achieve the maximum radiation characteristics for the antenna. The radiating elements are backed with a full conducting ground plane of thickness $t = 0.035$ mm. The geometrical parameters of the proposed antenna are as follows: $w_f = 1.6$ mm, $l_1 = 9$ mm, $l_2 = 4.3$ mm, $d = 0.4$ mm, $g = 1$ mm, $w_u = 0.5$ mm, $l = 35$ mm ($\approx 1.2\lambda_0$) and $w = 15$ mm ($\approx 0.5\lambda_0$).

To further facilitate the design process and understand the unit-cell better, the unit cell was simulated using CST-Microwave Studio. The simulated S-parameters can then be used to calculate the phase constant for the unit cell using the equation (Sarkar et al., 2018),

$$\beta p = \cos^{-1}\left(\frac{1 - S_{11}S_{22} + S_{12}S_{21}}{2S_{21}}\right) \tag{1}$$

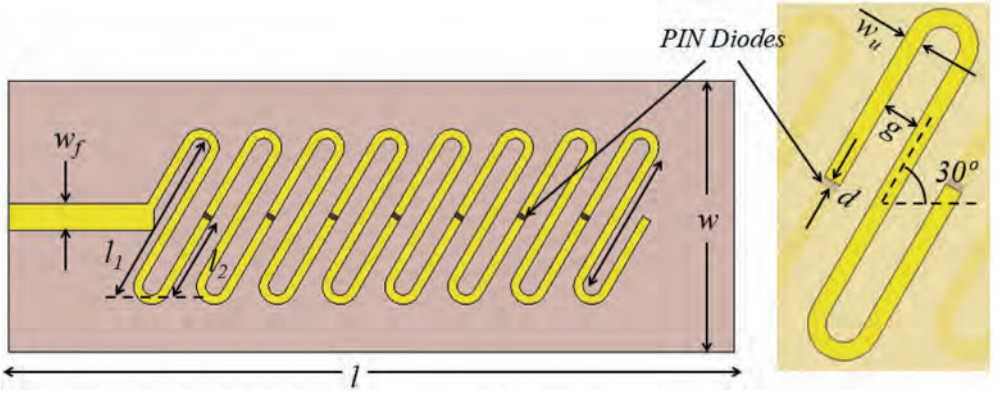


Figure 1. Top view of the schematic for the proposed antenna and zoomed-in view of the radiating element.

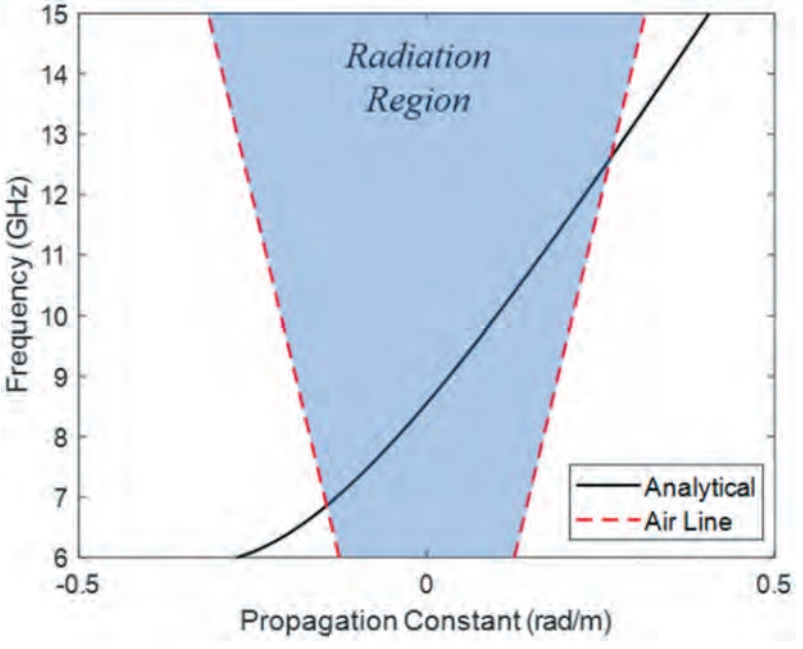


Figure 2. Propagation constant for the proposed antenna.

Here, β is the phase constant for the proposed antenna, p is the periodicity of the unit cell, S_{11}& S_{22} are the simulated reflection coefficients, and S_{12} and S_{21} are the simulated transmission coefficients for the unit cell. The calculated dispersion curve is presented in Figure 2, which demonstrates the frequency band of 7–12 GHz falling under the radiating region.

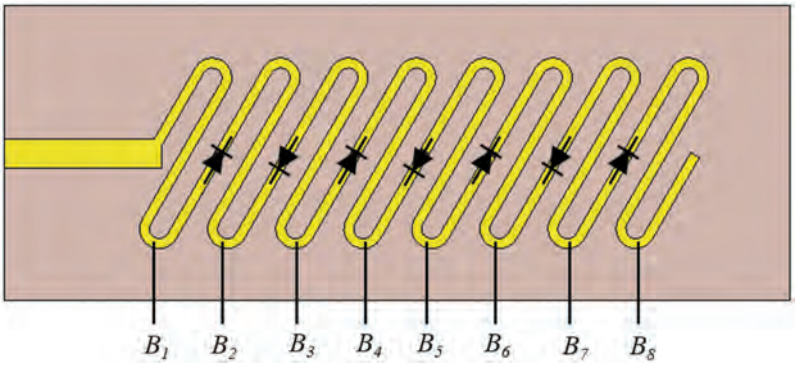


Figure 3. Proposed PIN diodes-based switching configuration and biasing lines.

2.2. Switching technique

PIN diodes behave as RF switches in the radiating elements, and they create an RLC circuit for both ON and OFF states. The proposed design uses Skyworks' SMP1340 PIN diodes. It offers a resistance of 0.81 Ω in the ON state and a very high resistance for the OFF state, making it suitable for switching applications at X-band. The proposed design uses seven PIN diodes, connecting the radiating elements. The switches are arranged alternatively for ease in biasing and make the switching method robust, as shown in Figure 3. The switch diodes being arranged alternatively can be controlled using alternative biasing lines. If the diode is being maintained at the same potential at both arms, it will behave in the OFF state unless the potential difference between the two arms is 3.3 V.

3. Simulated and measured results

3.1. Antenna performance

To evaluate the performance of the proposed leaky-wave antenna, it was designed and simulated using CST Microwave Studio. The design uses a waveguide port to feed the designed model and touchstone files for the PIN diodes to include their effects on the antenna performance within the simulation. The final model was fabricated on an RT-Duroid 5880 substrate (relative permittivity, $\varepsilon_r = 2.2$; loss tangent, $tan\ \delta = 0.0009$) of height, $h = 0.787$ mm, and an SMA connector is used to feed the proposed antenna. The fabricated antenna has been presented in Figure 4 and is measured on an Anritsu MS46522B vector network analyser (VNA).

The resonant frequency for the proposed antenna is reconfigured using different switch configurations. The switches are biased (for ON state) by maintaining a 3.3 V potential difference on the corresponding biasing lines and are kept in OFF state by keeping the potential difference at 0 V (by either grounding or providing the same voltage for both biasing lines of the switch). Table 1 gives a truth table specifying different resonant bands for different switch configurations. Simulated and measured S-parameter results for the proposed antenna shows a perfect agreement and further verify the frequency reconfigurability characteristics of the proposed antenna (see, Figure 5). The antenna offers an average gain of 8.4 dBi with an efficiency greater than 80% for the resonant frequency at different frequency bands, as shown in Figure 6.

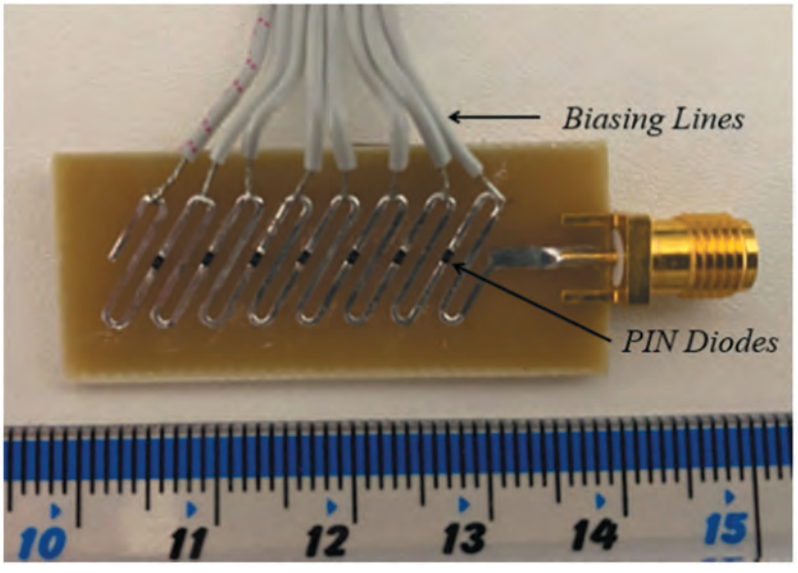


Figure 4. Fabricated TWA design with highlighted PIN diodes and biasing lines.

Table 1. Frequency truth table.

Case	SW1	SW2	SW3	SW4	SW5	SW6	SW7	Freq (GHz)	BW (MHz)
A	0	0	0	0	0	0	0	8.50	1200
B	1	0	0	0	0	0	0	9.05	1000
C	1	0	0	1	0	0	1	9.35	1200
D	1	0	1	0	1	0	1	9.90	1150
E	1	1	0	0	1	1	0	10.45	1200
F	1	1	1	1	1	1	1	11.00	1150

The measured radiation pattern for the proposed antenna is presented in Figure 7. The antenna presents an average gain of 8.52 dBi for all six switch configurations. The slight change in beam-angle is due to a change in phase delay between each radiating element with the change in frequency and can simply be calculated using the general principle of leaky-wave antennas (Balanis, 2005; Elliott, 2003). However, the antenna presents an average half-power beamwidth of 60° in the elevation plane and 90° in the azimuth plane, allowing almost constant radiation in one direction. The side-lobe level (SLL) is maintained below 12 dB for all cases.

3.2. Parametric analysis

To present the robustness of the design and further verify the antenna's performance for its conformality, some of the crucial parameters of the proposed design have been tested in simulation. The S-shaped radiating element is bent at an angle of 30° for a suitable impedance match between the feedline and the radiating element, allowing maximum radiation. This is further verified with the S-parameter plot for the unit cell presented in Figure 8.

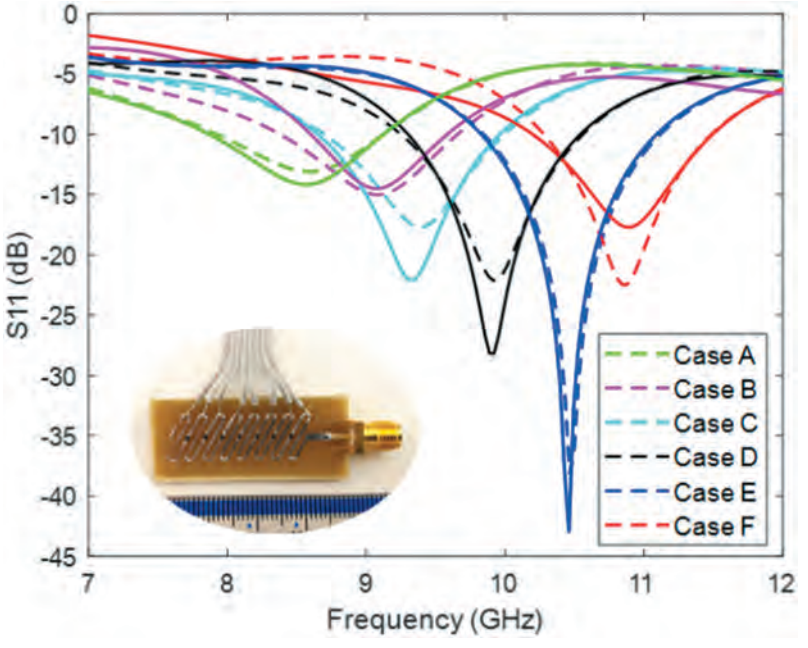


Figure 5. Simulated (dashed) and measured (solid) S11 for the proposed antenna.

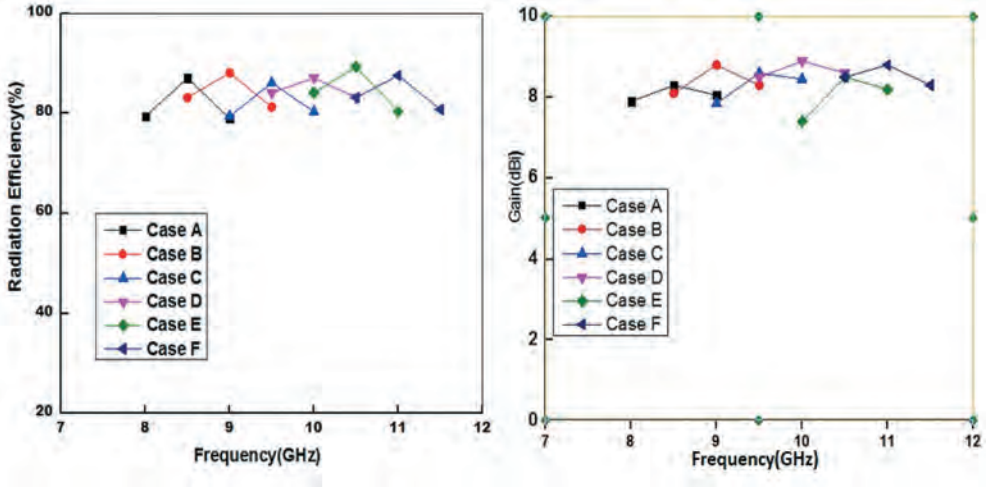


Figure 6. Measured radiation efficiency (dB) measured gain (dBi) for the proposed antenna.

The antenna is a travelling wave structure, and each radiating element receives a reduced power from the last element. The amount of power left after the eight elements have radiated and reflected is negligible, hence the number of radiating elements. This is further verified in simulation by defining a waveguide port on either side of the radiating element and studying the S-parameters. The amount of power reflected can be calculated using S_{11}, and the amount of power sent forward can be calculated using S_{21}. Figure 9

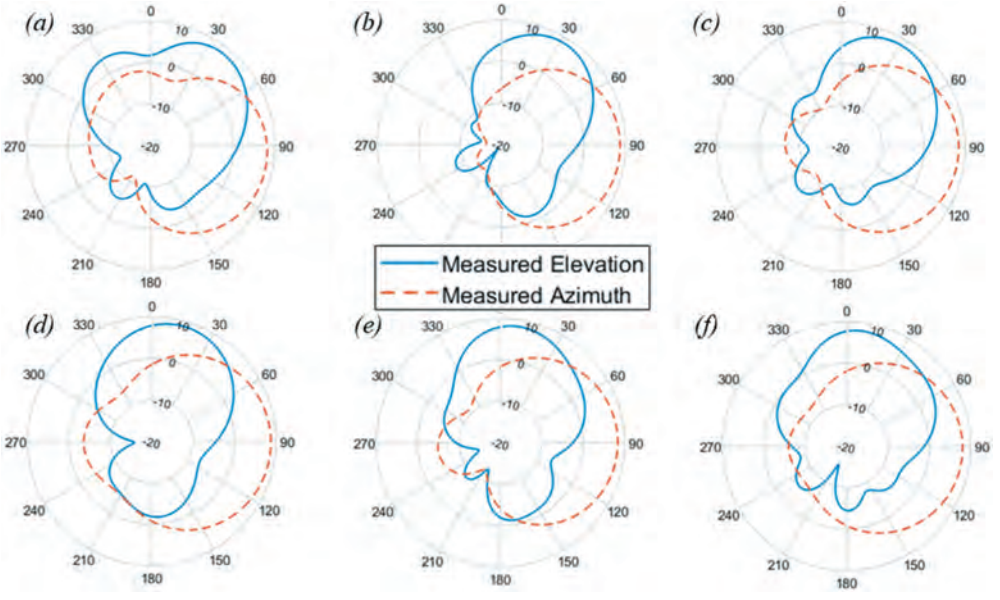


Figure 7. Measured radiation pattern for different cases in Azimuth and Elevation plane at (a) 8.5 GHz, (b) 9 GHz, (c) 9.35 GHz, (d) 9.9 GHz, (e) 10.45 GHz and (f) 11 GHz.

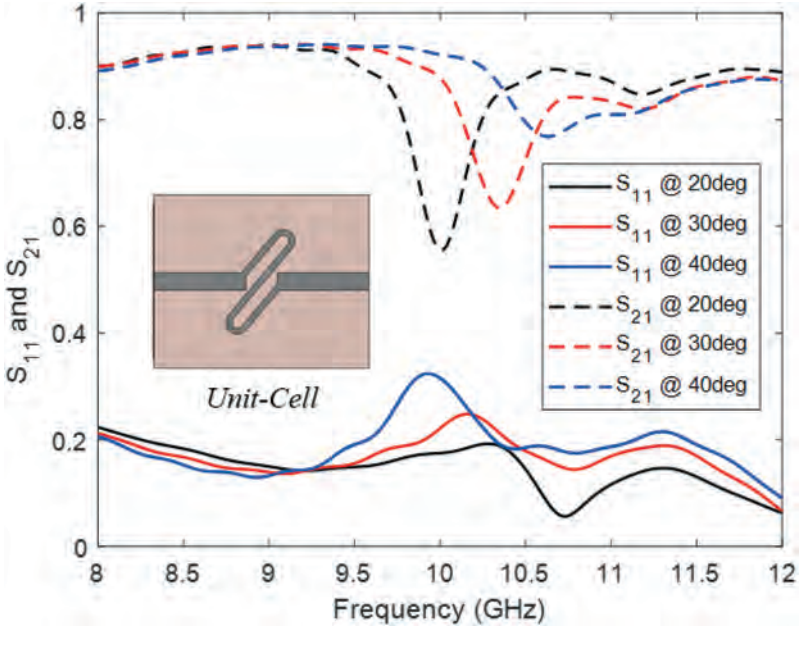


Figure 8. Simulated linear S_{11} and S_{21} for the unit cell at different tilt angles.

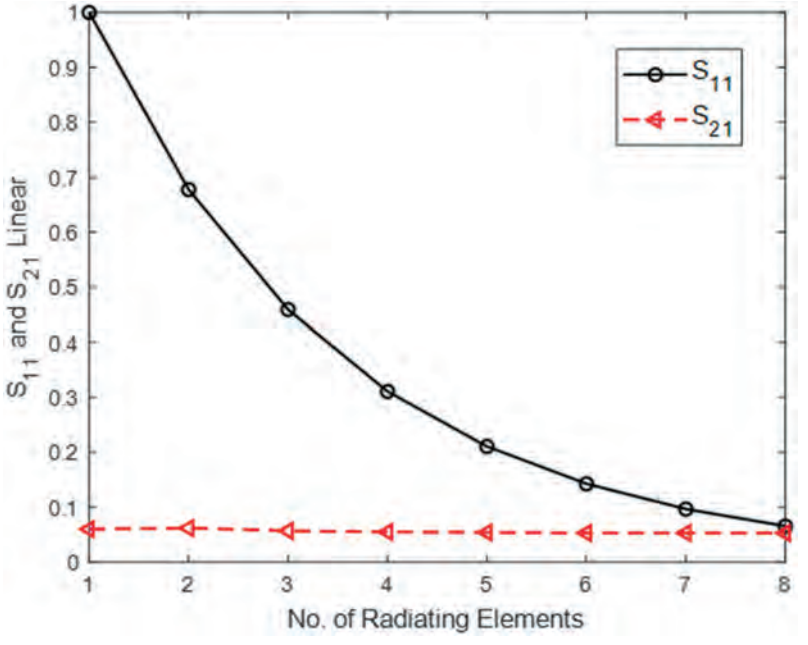


Figure 9. Simulated S11 and S21 demonstrating the amount of power left after 8 elements of the proposed array have radiated out.

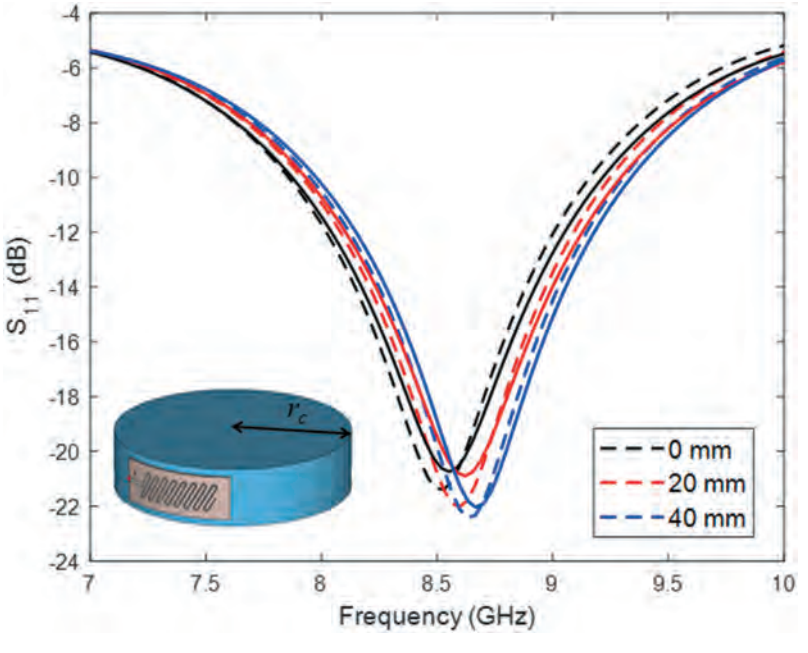


Figure 10. Measured (solid lines) and simulated (dashed lines) S-parameter plot for conformality of the proposed antenna for a radius of 0, 20 and 40 mm.

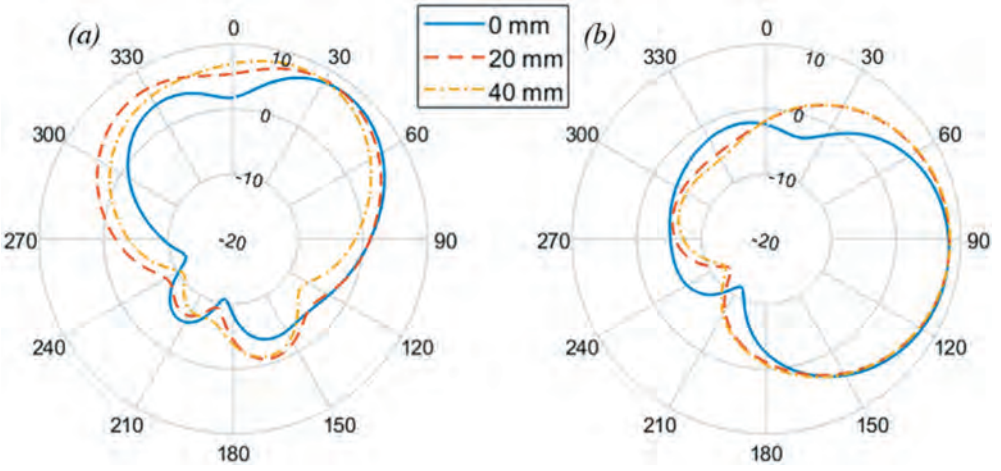


Figure 11. Measured radiation pattern in (a) Elevation and (b) Azimuth plane for the conformed antenna for radius, r_c = 0, 20 and 40 mm.

Table 2. Comparison of the proposed work with the literature.

Ref	Freq. Band	Size	No. of Reconf. Freq. Bands	Avg. Bandwidth	Gain (dBi)
(Abdulraheem et al., 2017)	2.2–6 GHz	50 mm × 45 mm	10	240 MHz	0.8–5.8
(Fadamiro et al., 2019)	1–3 GHz	63 mm × 69.75 mm	4	1 GHz	1.2–4
(Shirazi et al., 2018)	1.8–5.7 GHz	80 mm × 80 mm	2	300 MHz	0.1–4.2
(Young et al., 2015)	0.5–3.5 GHz	47.7 mm × 45 mm	5	500 MHz	2
(Atallah et al., 2016)	1.7–5.7 GHz	40 mm × 40 mm	3	500 MHz	0.3–0.8
(Ahmad et al., 2018)	4.2–7 GHz	24 mm × 20 mm	5	400 MHz	2.5–3.1
This Work	8–12 GHz	35 mm × 15 mm	7	1 GHz	8.1–8.6

presents simulated S_{11} and S_{21} for the antenna when tested for one to ten elements. As seen from the plotted graph, the amount of power left after the eighth element is radiated out is negligible, and hence, verifying the initial theory.

Conformal antennas have been in massive demand for weather radar and similar applications. The proposed antenna is printed on a thin substrate, allowing for minimal conformality. However, the antenna being tiny in the Y-axis (w = 15 mm), it is impossible to bend the antenna in this direction. The antenna was measured for conformality in X-axis for a bend radius of 0 mm, 20 mm, and 40 mm. The variation in S-parameters for the conformed shape is presented in Figure 10.

The measured radiation pattern for the conformality is shown in Figure 11. The antenna was measured in an NSI2000 anechoic chamber and was conformed on a metallic pipe of radius, r_c = 20, and 40 mm. The antenna's gain is found to have a minimal change of ±0.12 dBi which can be considered negligible. However, the direction of radiation had a significant change. The beam direction changed from 31° in elevation

plane for no conformality to 27° for r_c = 20 mm and 22° for r_c = 40 mm. This can be explained by the change in the radiator's direction with conformality. The antenna maintains a similar efficiency with and without conformality, and hence, the antenna can be considered suitable for several weather radar applications and for mounting on curved surfaces.

The performance and miniaturised nature of the proposed antenna are compared with the literature in Table 2 to further appreciate the presented work's importance.

4. Conclusion

The article presented a miniaturized travelling-wave antenna with frequency reconfigurability, operating at X-band (8–12 GHz). The proposed antenna used PIN diodes in a novel configuration to control different radiating elements of the array. The antenna is presented to be operating at 8.5, 9.05, 9.35, 9.9, 10.45 and 11 GHz for different switch configurations. The frequency reconfigurability is attained by connecting different radiating elements of the proposed array using PIN diodes. Each operating mode is shown to have a minimum bandwidth of 1 GHz. The antenna was measured in an NSI2000 anechoic chamber for its radiation characteristics, and it presented an average gain of 8.5 dBi throughout the frequency band for different configurations with an efficiency >80%. The antenna is further tested for its conformality, where it was bent across its length on a metallic pipe of radius 20 and 40 mm, respectively. The antenna presented a stable gain with a slight change in the beam direction. Each antenna configuration demonstrates a slight change in the beam declination, which is believed to be because of the antenna's travelling-wave nature. Each element being fed sequentially introduces a phase delay, resulting in a beam declination. This beam-declination changes with the phase delay, which is further controlled with the operating frequency. The fabricated prototype presents a good agreement with the simulated results. Furthermore, the directional behaviour of the antenna, along with frequency reconfigurability, conformality and high gain, maintains a small form-factor of 1.2 $\lambda_0 \times 0.5\ \lambda_0 \times 0.026\ \lambda_0$, which makes it a suitable candidate for several communications and radar applications.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Abdulraheem, Y. I., Oguntala, G. A., Abdullah, A. S., Mohammed, H. J., Ali, R. A., Abd-Alhameed, R. A., & Noras, J. M. (2017). Design of frequency reconfigurable multiband compact antenna using two PIN diodes for WLAN/WiMAX applications. *IET Microwaves, Antennas & Propagation*, *11*(8), 1098–1105. https://doi.org/10.1049/iet-map.2016.0814

Ahmad, A., Arshad, F., Naqvi, S. I., Amin, Y., Tenhunen, H., & Loo, J. (2018). Flexible and compact spiral-shaped frequency reconfigurable antenna for wireless applications. *IETE Journal of Research*, *66*(1), 22–29. https://doi.org/10.1080/03772063.2018.1477629

Arif, A., Zubair, M., Ali, M., Khan, M. U., & Mehmood, M. Q. (2019). A compact, low-profile fractal antenna for wearable on-body WBAN applications. *IEEE Antennas Wireless Propagation Letter*, *18* (5), 981–985. https://doi.org/10.1109/LAWP.2019.2906829

Atallah, H. A., Abdel-Rahman, A. B., Yoshitomi, K., & Pokharel, R. K. (2016). Compact frequency reconfigurable filtennas using varactor loaded T-shaped and H-shaped resonators for cognitive radio applications. *IET Microwaves, Antennas & Propagation*, *10*(9), 991–1001. https://doi.org/10.1049/iet-map.2015.0700

Balanis, C. A. (2005). *"antenna theory: analysis and design"* (3rd ed.). Wiley.

Bansal, A., & Gupta, R. A. (2020). Review on microstrip patch antenna and feeding techniques. *International Journal of Information Technology*, *12*(1), 149–154. https://doi.org/10.1007/s41870-018-0121-4

Elliott, R. S. (2003). *Antenna Theory And Design* (Revised ed.). IEEE Press Series on Electromagnetic Wave Theory.

Fadamiro, A. O., Famoriji, O. J., Zakariyya, R. S., Zhang, Z., & Lin, F. (2019). Design of H-Tree fractal slots frequency reconfigurable hexagonal patch antenna using PIN diodes. *Journal of Electromagnetic Waves and Applications*, *33*(12), 1591–1604. https://doi.org/10.1080/09205071.2019.1618740

Gupta, R., Bakshi, G., & Bansal, A. (2020). Dual-band circularly polarized stacked sapphire and TMM13i rectangular DRA. *Progress In Electromagnetics Research*, *91*, 143–153. https://doi.org/10.2528/PIERM20012701

Iftikhar, A., Parrow, J., Asif, S., Fida, A., Allen, J., Allen, M., Braaten, B., & Anagnostou, D. (2020). Characterization of novel structures consisting of micron-sized conductive particles that respond to static magnetic field lines for 4G/5G (Sub-6 GHz) reconfigurable antennas. *Electronics*, *9*(6), 903–910. https://doi.org/10.3390/electronics9060903

Ismail, M. F., R, M. K. A., & Majid, H. A. (2021). Wideband Frequency Reconfiguration using PIN diode. *Microwave Optical Technology Letter*, *54*(6), 1407–1412. https://doi.org/10.1002/mop.26826

Luadang, B., Sakonkanapong, A., Dentri, S., Pansomboon, R., & Phongcharoenpanich, C. (2018). NFC-enabled far-field antenna on PET flexible substrate for 3G/4G/LTE mobile devices. *IEEE Access*, *7*, 71966–171973.

Mohamadzade, B., Simorangkir, R. B. V. B., Hashmi, R. M., Chao-Oger, Y., Zhadobov, M., & Sauleau, R. (2020). A conformal band-notched ultrawideband antenna with monopole-like radiation characteristics. *IEEE Antennas Wireless Propagation Letter*, *19*(1), 203–207. https://doi.org/10.1109/LAWP.2019.2958036

Nikolaou, S., Kim, B. K. B., & Vryonides, P.(2009). "Reconfiguring antenna characteristics using PIN diodes," *Proceedings of EuCAP*, pp 3748–3752.

Paracha, K. N., Rahim, S. K. A., Chattha, H. T., Aljaafreh, S. S., Rehman, S. U., & Lo, Y. C. (2018). Low-cost printed flexible antenna by using an office printer for conformal applications. *International Journal of Antennas and Propagation*, *2018*, 1–7. https://doi.org/10.1155/2018/3241581

Sahu, N. K., & Sharma, A. K. (2017). "An investigation of pattern and frequency reconfigurable microstrip slot antenna using PIN diodes", *Progress in electromagnetics research symposium-spring (PIERS)*, pp. 971–976.

Sarkar, A., Adhikary, M., Sharma, A., Biswas, A., Akhtar, M. J., & Hu, Z. (2018). "Composite right/left-handed based compact and high gain leaky-wave antenna using complementary spiral resonator on HMSIW for Ku band applications. *IET Microw Antennas Propag*, *12*(8), 1310–1315. https://doi.org/10.1049/iet-map.2017.0478

Shirazi, M., Huang, J., Li, T., & Gong, X. (2018). A switchable-frequency slot-ring antenna element for designing a reconfigurable array. *IEEE Antennas and Wireless Propagation Letters*, *17*(2), 229–233. https://doi.org/10.1109/LAWP.2017.2781463

Simorangkir, R. B., Kiourti, A., & Esselle, K. P. (2018). UWB wearable antenna with a full ground plane based on PDMS-embedded conductive fabric. *IEEE Antennas and Wireless Propagation Letters*, *17*(3), 493–496. https://doi.org/10.1109/LAWP.2018.2797251

Valkonen, R., Luxey, C., Holopainen, J., Icheln, C., & Vainikainen, P. (2010). "Frequency-reconfigurable mobile terminal antenna with MEMS switches", *Proceedings of the fourth european conference on antennas and propagation* (pp. 1–5).

Young, M. W., Yong, S., & Bernhard, J. T. (2015). A miniaturized frequency reconfigurable antenna with single bias, dual varactor tuning. *IEEE Transactions on Antennas and Propagation*, *63*(3), 946–951. https://doi.org/10.1109/TAP.2015.2388776

Int J Imaging Syst Technol. 2021 Mar; 31(1): 28–46.
Published online 2020 Nov 23. doi: 10.1002/ima.22525
PMCID: PMC7753711
PMID: 33362346

An integrated feature frame work for automated segmentation of COVID-19 infection from lung CT images

Deepika Selvaraj, [1] Arunachalam Venkatesan, [1] Vijayalakshmi G. V. Mahesh, [2] and Alex Noel Joseph Raj [3]

Abstract

The novel coronavirus disease (SARS-CoV-2 or COVID-19) is spreading across the world and is affecting public health and the world economy. Artificial Intelligence (AI) can play a key role in enhancing COVID-19 detection. However, lung infection by COVID-19 is not quantifiable due to a lack of studies and the difficulty involved in the collection of large datasets. Segmentation is a preferred technique to quantify and contour the COVID-19 region on the lungs using computed tomography (CT) scan images. To address the dataset problem, we propose a deep neural network (DNN) model trained on a limited dataset where features are selected using a region-specific approach. Specifically, we apply the Zernike moment (ZM) and gray level co-occurrence matrix (GLCM) to extract the unique shape and texture features. The feature vectors computed from these techniques enable segmentation that illustrates the severity of the COVID-19 infection. The proposed algorithm was compared with other existing state-of-the-art deep neural networks using the Radiopedia and COVID-19 CT Segmentation datasets presented specificity, sensitivity, sensitivity, mean absolute error (MAE), enhance-alignment measure (EM_{φ}), and structure measure (S_m) of 0.942, 0.701, 0.082, 0.867, and 0.783, respectively. The metrics demonstrate the performance of the model in quantifying the COVID-19 infection with limited datasets.

Keywords: artificial intelligence, computed tomography image, deep neural network, feature extraction, limited training points, segmentation, Zernike moment

1. INTRODUCTION

The novel coronavirus disease-2019 (COVID-19) is a lung infection caused by Severe Acute Respiratory Syndrome (SARS) and is transmitted from person to person easily. The first human COVID-19 case has been reported in Wuhan city, China in December 2019.[1] This disease was

traced during the sewage water study from Milan and Turin in Italy before China identified the first case and the related information was shared in a recent article.[2] The World Health Organization (WHO) has reported 10.59 million COVID cases across 213 countries and territories, as of June 2020. A large percentage of the population at present is affected by this disease. The recovery and mortality rates are 54.22% and 4.91% approximately. All over the world, researchers from biomedical departments are involved in an effort to find the effective vaccine for the COVID-19.[1] However, early detection is very important for choosing the right type of treatment to control the disease and avoid the spread. AI has been crucial in helping and serving the automated diagnostics and public health surveillance to analyse the severity of the COVID-19 as well as in protecting the people from the virus spread.[3]

In the human body, the lungs are the major organs of the respiratory system. The disease in the lung will affect the entire air circulation system and may even result in death. The other major diseases that affect the lungs are lung cancer, tuberculosis, pneumonia.[4], [5] Also, some microorganisms such as fungus, bacteria, and virus may cause illness which requires proper medical treatment to cure the infection/disease. Likewise, the infection caused due to the COVID-19 virus requires the same. The most commonly used COVID- 19 tests are RT-PCR and antigen testing. Reverse transcription polymerase chain reaction (RT-PCR) is the standard COVID-19 test used to detect the viral nucleic acid that provides the genetic information of the virus. The test requires nasopharyngeal and throat swabs collected from the infected person. The duration of the test is too long and the result of RT-PCR is affected by sampling errors and low viral load.[6] The antigen test is designed to detect the spike protein that is responsible for facilitating the entry of virus into the human cell. Though the test is fast, its sensitivity is poor.

An alternate detection method for COVID-19 is based on medical imaging techniques that identify the infected location and infection rate. Medical imaging records the information in the lung region using chest radiography (X-rays) and Computed-Tomography (CT) scan.[7], [8], [9], [10] Compared to the X-rays, a CT scan is mostly preferred for detecting the radiographic features of COVID-19 from the three-dimensional view of the lung's region.[11] Further, the CT scans can be used to give two-dimension views: axial view, coronal view, and sagittal view for the proper COVID infection detection. CT scan imaging of lungs gives good sensitivity to examine the COVID-19 infected region even before the medical symptoms occur.[12]

The spread of the infection due to COVID-19 in the lungs are identified through Ground glass opacification (GGO) and pulmonary consolidation phases. GGO is identified in the early stage of the infection whereas pulmonary consolidation indicates the final stage of the disease. These phases are observed from the CT scan images of the COVID cases.[13], [15] The qualitative rate of the disease indicated in CT scans provides important information to guide against COVID-19.

Rajinikanth et al.[16] suggested that infection visibility analysis using CT scans gives better capability and reliability in the detection of the disease using the two-dimensional (coronal) view. The dataset used is obtained from the Radiopedia database.[17] The experimental analysis and the simulation of the model were performed in a MATLAB environment.

Accurate segmentation of radiographic features is important to quantify the disease in CT scan images. Segmentation of medical images needs marking manually by experts. As the number of COVID cases is increasing rapidly, it is difficult to label the infected region manually. Thus, automatic segmentation is needed from the CT scans to detect the infected region.[18] From the above consideration, our contributions are as following:

1. We propose a Deep Neural Network (DNN) model to detect and segment the COVID-19 from the axial view CT scan image from Radiopedia and COVID-19 CT Segmentation database. Our model provides better performance with limited number of training points.
2. Using the region-specific approach, instead of taking all the training points or images,[19] limited training points are chosen from the infected region as well as the background region of the CT scan image. Before selecting the training points, for each image, the shape and texture descriptors are extracted using Zernike moments and GLCM, respectively.
3. The performance of the proposed method is evaluated using the standard metrics for the different test images and evaluated metrics are compared with the existing models.[19], [20], [21], [22], [23], [24]

The methodology needs enhancement techniques for CT scans of the COVID-19 dataset, feature extraction and appropriate classifier model to detect and segment the infected region. The rest of the paper is arranged as follows. Section 2 discusses the related works on the COVID-19 using AI techniques. Section 3 gives a detailed explanation and workflow of methodology and implementation. Sections 4 and 5 outline the experimental design, results and discussion. Section 6 provides the conclusion.

2. RELATED WORKS ON THE COVID-19 USING AI

A comprehensive list of work for COVID-19 image-based AI techniques are found in Maga et al.[18] and Fu et al.[25] Developing a deep network gives more benefits for automatic and fast segmentation of the medical images.[26] Camouflaged object detection (COD) is introduced to identify the embedded object with their surroundings.[27] COD is beneficial in medical image applications such as lung infection segmentation. Ali et al.[28] proposed the integration approach consisting of unsupervised machine learning (self-organizing map), dimensionality reduction (principal component analysis) and computational classification (Adam Deep Learning) to present a better classification performance. Particularly, the U-Net model works well in many of the segmentation tasks for medical images.[29] There are a few applications that have adopted U-Net for liver, heart, and multi-organ segmentation.[30], [31], [32] Chen et al.[33] has proposed a deep learning model for multiple regions auto segmentation for COVID-19 and used the aggregated residual transformations to extract the features from the CT image. Contrastive learning is proposed to train the encoder module, which provides the expressive feature information from the publicly available CT image COVID-19 dataset.[34] The Deep Learning model requires a lesser number of image samples for training the dataset to provide an automatic classification of COVID-19 infected images. Also, the authors concluded that the contrast learning method achieves better performance than ResNet-50. Narin et al.[35] have analysed the ROC result of three Convolutional Neural Network (CNN) models such as ResNet-50, InceptionV3, and Inception-ResNetV2 using a chest X-ray image. Shi et al.[36] have discussed pre-processing the COVID CT images using location-specific feature extraction and infection Size Aware Random Forest method (iSARF) to distinguish the infection level and then classification using random forest. The iSARF performed well by providing Sensitivity, Specificity and Accuracy of 0.907, 0.833 and 0.879, respectively. Yujin et al.[37] proposed, the patch-based CNN using a novel probabilistic Grad-CAM saliency map with a limited number of training points from the chest X-rays. Wang et al.[38] proposed an open-source CNN model called COVIDNet, which is trained and tested by the COVIDx dataset. It is useful to identify the COVID-19 infection from chest X-rays and has achieved a sensitivity of 80%.

Lung Infection Segmentation Deep Network (Inf-Net) used to detect the COVID-19 infection in the lungs, from the CT image slices where a parallel partial decoder is adapted to extract the high-level features.[19] Later implicit reverse attention and explicit edge attention are used to enhance the features. Charmaine et al.[39] have presented the deep learning model using the location-attention oriented approach to calculate the probability of COVID-19 infection region. Chest CT dataset is used for training the deep learning system.

A novel semi-supervised shallow learning model including Parallel Quantum-Inspired Self-supervised Network (PQIS-Net) with Fully-Connected (FC) layers for automatic segmentation of COVID-19 CT image is proposed in Konar et al.[40] The patch-based classification was applied to the segmented images for the diagnosis of COVID-19 using the two publicly available datasets. The efficiency (F1-score and AUC) of the PQIS-Net was compared with pre-trained convolutional based models.

The truncated VGG-19 model was proposed to analyse the COVID-19 CT scans. TheVGG-16 model was used to extract features from the CT images using fine-tuning.[41] The feature selection was achieved through Principal Component Analysis (PCA) and the classification was done using four different classifier methods such as deep convolutional neural network (CNN), Bagging Ensemble with support vector machine (SVM), Extreme Learning Machine (ELM) and Online sequential ELM on 208 test images. From the different classifiers, Bagging Ensemble with SVM have achieved the following: an accuracy of 95.7%, AUC0.958 and F1-score 95.3%.

The three-phase COVID-19 CT image detection model is introduced in Ahuja et al.[42] The modules are as follows: (a) Augmentation using stationary wavelets; (b) COVID-19 detection using pre-trained CNN model such as ResNet18, ResNet50, ResNet101, and SqueezeNet; (c) Abnormality localization in CT scan images. The experimental analysis showed that the pre-trained ResNet18 transfer learning model has given a better classification accuracy of 99.82% for training and 99.4% for testing.

A pipeline model is presented in Dey et al.[43] with several sub-modules to classify the segmented region from the COVID-19 images. First, the COVID-19 CT images are segmented using the using Social-Group-Optimization and Kapur's Entropy thresholding, followed by K-means clustering and morphology-based segmentation. Next, a classification module is introduced to classify the segmented region. Here PCA based fusion technique is used to fuse the features and then fused features are trained with different classifiers such as Random Forest, k-Nearest Neighbours (KNN), Support Vector Machine with Radial Basis Function, and Decision Tree. Experimental results indicated an accuracy of 91% and 87% with Morphology-based segmentation and kNN classifier. A summary of the various techniques available for COVID-19 screening is presented in Table 1.

TABLE 1

Literature summary of related techniques for COVID-19 screening

Reference	Technique	Dataset	Findings
33	Aggregated residual transformations to learn a robust and expressive feature representation.	110 COVID-19 CT image of size 512 × 512 are collected from 60 patients.	The model gives 0.95 precision and 0.89 accuracy. 12.7% of accuracy and 14.5% of precision are improvement with U-Net model.
34	Contrastive learning method.	MedSeg dataset:110 CT COVID-19 images. COVID-19 CT dataset: 349 CT COVID-19 images 397 Non-COVID images.	ResNet-50 performance gives accuracy of 0.868, recall of 0.872, AUC of 0.931.
35	ResNet50, InceptionV3 and InceptionResNetV2.	Chest X-ray 100 images: 50 non-infected image and 50 infected images. All images are 224 × 224 pixel size.	ResNet50 model provides good classification performance with 98% accuracy than other two proposed models (97% accuracy for InceptionV3 and 87% accuracy for Inception-ResNetV2).
36	An infection Size Aware Random Forest method (iSARF).	COVID-19 images are taken from 1658 patients.	iSARF yielded sensitivity of 0.907, specificity of 0.833, and accuracy of 0.879.
37	A patch-based convolutional neural network approach.	COVID-19 dataset: 180 infected image of size 224 × 224.	Global approach gives 70.7% of accuracy, 60.6% of precision, 60.1% of recall, 59% of F1-score. Local approach gives 88.9% of accuracy, 83.4% of precision, 85.9% of recall, 84.4% of F1-score.
19	COVID-19 Lung Infection Segmentation Deep Network (Inf-Net) with a parallel partial decoder method is to aggregate the high-level features.	Dataset consists of 100 axial CT images from different COVID-19 patients. 45 CT images randomly	Inf-Net achieves 0.692 of sensitivity, 0.943 of specificity, 0.725 of sensitivity, 0.960 of specificity.

Wang et al.[44] have proposed the weakly-supervised deep learning (UNet and DeCoVNet) using 3-Dimensional CT volumes to identify COVID-19.The literature indicates the successful evaluation of COVID-19 infection using deep learning and traditional machine learning methods by various researchers. But the study also clarifies that some of the regions are not evaluated in the classifier layer due to lack of training data samples and inefficient feature representation. The detailed design flow of the proposed framework is explained in the next section.

3. METHODOLOGY

The Datasets used in the proposed framework consists of 80 images of size 512 × 512 pixels (3 channels). The segmentation process of each CT image involves (a) pre-processing, (b) extracting the robust and sensitive features, (c) training the region-specific region, and (d) segmenting the region of interest (infected region) from the background. The block diagram of the proposed methodology is as shown in Figure 1.

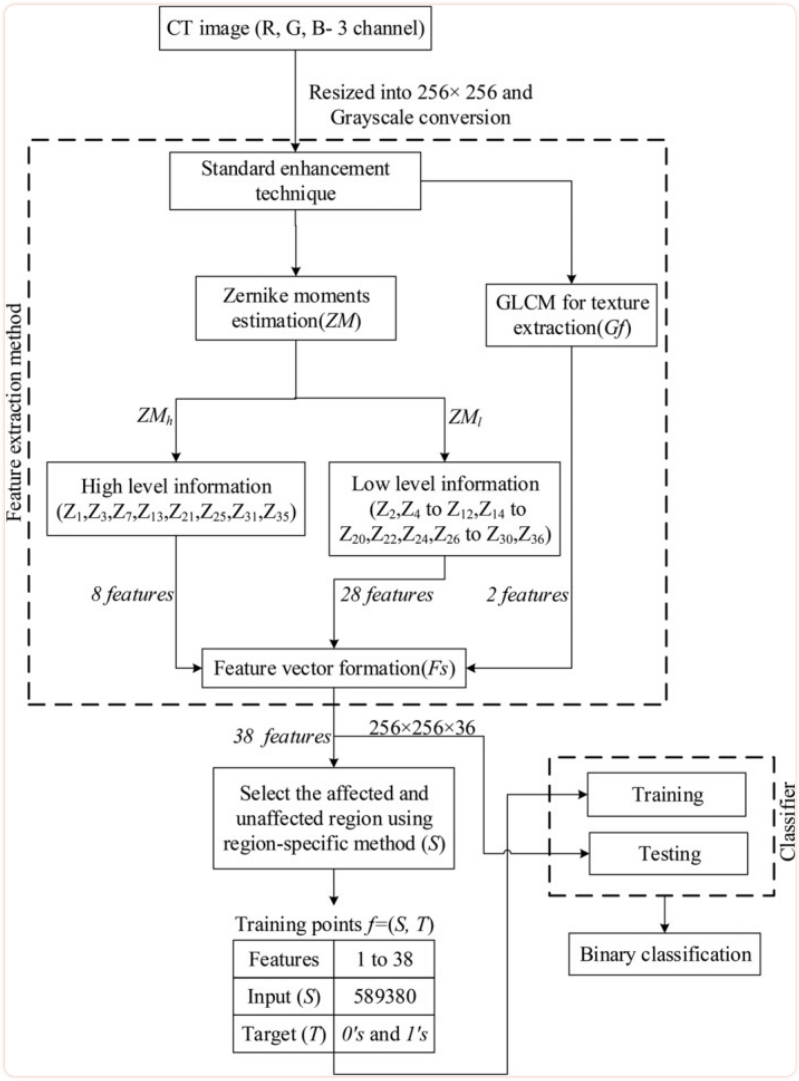


FIGURE 1

Block diagram for the flow of binary classification for computed tomography image

3.1. Pre-processing and feature extraction methods

The pre-processing is based on the standard enhancement approach [45] called adjust image intensity values (imadjust). This approach describes the expressive information from the COVID-19 CT image by adjusting the contrast of the GGO and pulmonary consolidation region. The infected pixels are brightened visually as shown in Figure 2. Also, the Contrast Improvement Index (CII) [46] metrics are analysed to highlight the importance of pre-processing on the COVID dataset. The image quality (pixel intensity) is improved by 30% to 40% when compared to the original CT image as shown in Table 2. CII is computed based on Equations (1) and (2).

$$\mathrm{CII} = \frac{C_{\mathrm{enhanced}}}{C_{\mathrm{original}}} \tag{1}$$

$$C = \frac{I_f - I_b}{I_f + I_b} \tag{2}$$

where C_{enhanced} and C_{original} are the contrast of the enhanced and original images respectively, C is the average value of the contrast in the enhanced or original image. I_f and I_b are the mean value for foreground and background contrast of the image. A higher CII value indicates improvement in image quality.

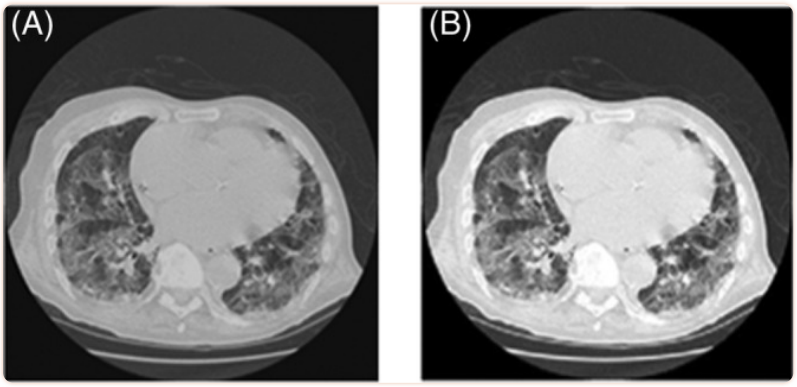


FIGURE 2

A, Original computed tomography image; B, Image after enhancement

TABLE 2

CII values for the pre-processed COVID-19 CT images

Number of images	Contrast for the original CT image	Contrast for the pre-processed image	CII for the pre-processed image
1	0.5984	0.8166	1.3646
2	0.4264	0.7348	1.7234
3	0.1650	0.3279	1.9872
4	0.4808	0.6698	1.3931
5	0.4846	0.9497	1.9599

Note: Default CII value for the original image is 1.000.

The enhanced images present significant details for extracting the features. The extraction of features is a description of each pixel of the CT image that contains descriptive information in the form of a stack of the vectors (feature vector). The feature vector helps to find out the output class (infected COVID-19 region and background region). The feature vector considered for the work is the integration of shape and texture features. The texture features are derived from the gray level co-occurrence matrix and shape descriptor features are extracted from the Zernike moment.

3.2. Zernike moment (ZM) based shape descriptor features and gray level co-occurrence matrix (GLCM) for texture features

Based on the infection stage/level, the infected region of COVID-19 CT images has dissimilarity in shapes, sizes and structures of the geometry. The shape descriptor is good to define the infected pixel information range from its background region. The shape parameters can be extracted using descriptor based on Zernike and Hu moments.[47], [48] For better representation, the shape descriptor should possess the properties such as (a) provide low redundancy and higher discrimination ability. (b) Rotation, scale and translation invariance, and (c) present coarse to finer detail representation.

Hu moments require extensive computation power to calculate higher order moments. Therefore, we choose the ZM based shape descriptor for representing shape features of the COVID-19 lungs image. The ZM is calculated for every pixel of the pre-processed image by sliding a window of size 5 × 5 with a stride of 1 along the column and row. For the computation of ZMs, different polynomial orders (n) and the corresponding repetition factor (m) were considered for the function $I(x, y)$. Next, we compute the Zernike moments, by projecting $I(x, y)$ onto the set of complex Zernike polynomials as given in Equation (3).

$$Z_{nm} = \frac{m+1}{\pi} \sum_{x} \sum_{y} I(x,y)\, P^{*}_{n,m}(x,y), x^2 + y^2 \leq 1 \tag{3}$$

where, $P_{nm}(x, y)$ is Zernike polynomial, $I(x, y)$ is the image of size 5 × 5 pixels, n is a non-negative integer, m is an integer such that $0 \le |m| \le n$, and $n - |m|$ is even. The function $P^{*}_{nm}(x, y)$ is the complex conjugate of the orthogonal basis function $P_{nm}(x, y)$ defined as:

$$P_{nm}(x,y) = P_{n,m}(r,\theta) = R_{n,m}(r)e^{jm\theta} \tag{4}$$

where, $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}\frac{y}{x}$, $with\ \ 0 \le \theta \le 2\pi, j = \sqrt{-1}$.

The radial polynomial $R_{nm}(r)$ is defined as:

$$R_{nm}(r) = \sum_{q=0}^{n-|m|} (-1)^q \frac{(2n+1-q)!}{q!(n-|m|-s)!(n+|m|+1-s)!} r^{n-q} \tag{5}$$

The Zernike moment obtained from the Equation (3) is a complex number represented as shown in Equation (6):

$$Z_{nm} = R_{\mathrm{ZM}} + jI_{\mathrm{ZM}} \tag{6}$$

$$\left|Z_{nm}\right| = \sqrt{R_{\mathrm{ZM}}^2 + I_{\mathrm{ZM}}^2} \tag{7}$$

Equation (7) is the magnitude of the Zernike moment $|Z_{nm}|$ that denotes the ZM shape descriptor for specified *n* and *m*. Thus, a total number of 36$|Z_{nm}|$were obtained with *n* ranging over 1 to 10 and matching combinations of *n*, *m* with *m* such that $0 \le |m| \le n$, and $n - |m|$ is even for every pixel. The magnitude of ZM calculated for every pixel to find the ZM feature maps are as displayed in Figure 3.

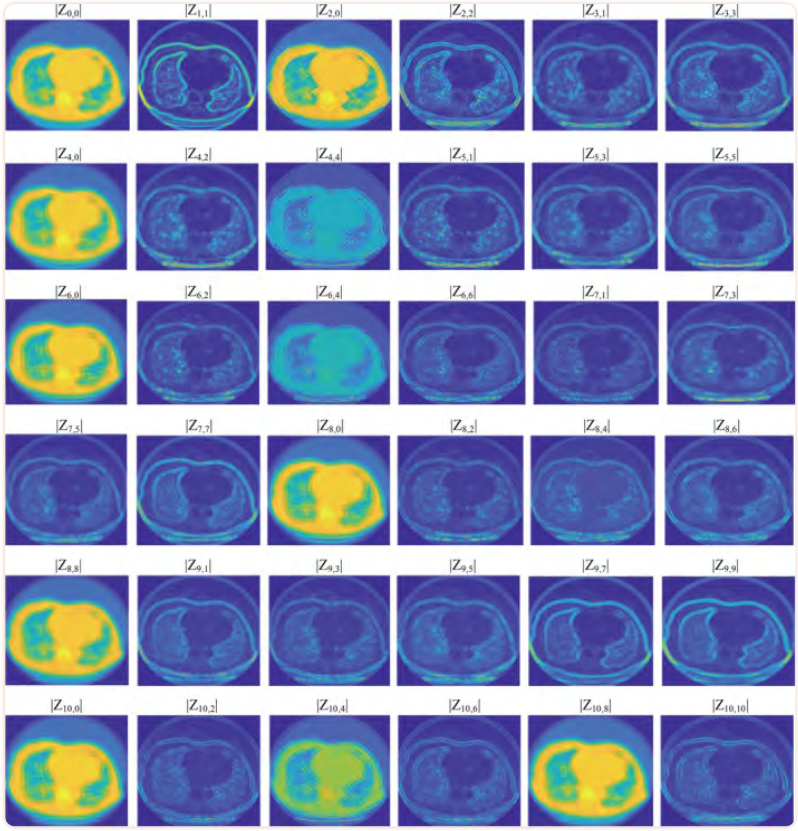


FIGURE 3

Zernike moment features with different combination of n and m [Color figure can be viewed at wileyonlinelibrary.com]

From the observation of Zernike features, we find that all the 36 different $|Z_{nm}|$ provide unique information about the infected and background region. To analyse the Zernike robust feature (ZM_h) and Zernike sensitive feature (ZM_l), each point is taken from both the regions with 36 ZM for different *n*, *m* and are plotted as indicated in Figure 4 and the separation of the ZM_h and ZM_l is indicated in Equations (8) and (9).

$$\mathrm{ZM}_h = \{|Z_{0,0}|, |Z_{2,0}|, |Z_{4,0}|, |Z_{6,0}|, |Z_{8,8}|, |Z_{8,8}|, |Z_{10,0}|, |Z_{10,8}|\} \tag{8}$$

$$\mathrm{ZM}_l = \{|Z_{1,1}|, |Z_{2,12}| \text{ to } |Z_{5,5}|, |Z_{6,2}| \text{ to } |Z_{7,7}|, |Z_{8,2}|, |Z_{8,6}|, |Z_{9,1}| \text{ to } |Z_{9,9}|, |Z_{10,10}|\} \tag{9}$$

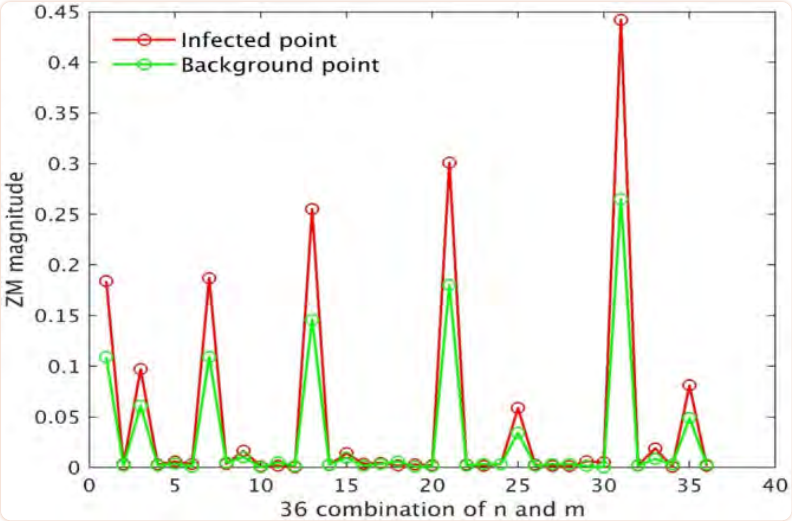


FIGURE 4

Experimental analysis of low-level and high-level feature from Zernike moment (total 36 Zernike feature, red – infected point, green – background point) [Color figure can be viewed at wileyonlinelibrary.com]

From the different combinations of ZM feature extraction image, we plot the intensity range for the set of pixels for the 1st image of the dataset. The graphical plot presented in Figure 5 represents number of pixels in *x*-axis and ZM features in y-axis. The above observation indicates that the Zernike features from background information are dense at the bottom region and the infected region information mostly occurs above a specific threshold.

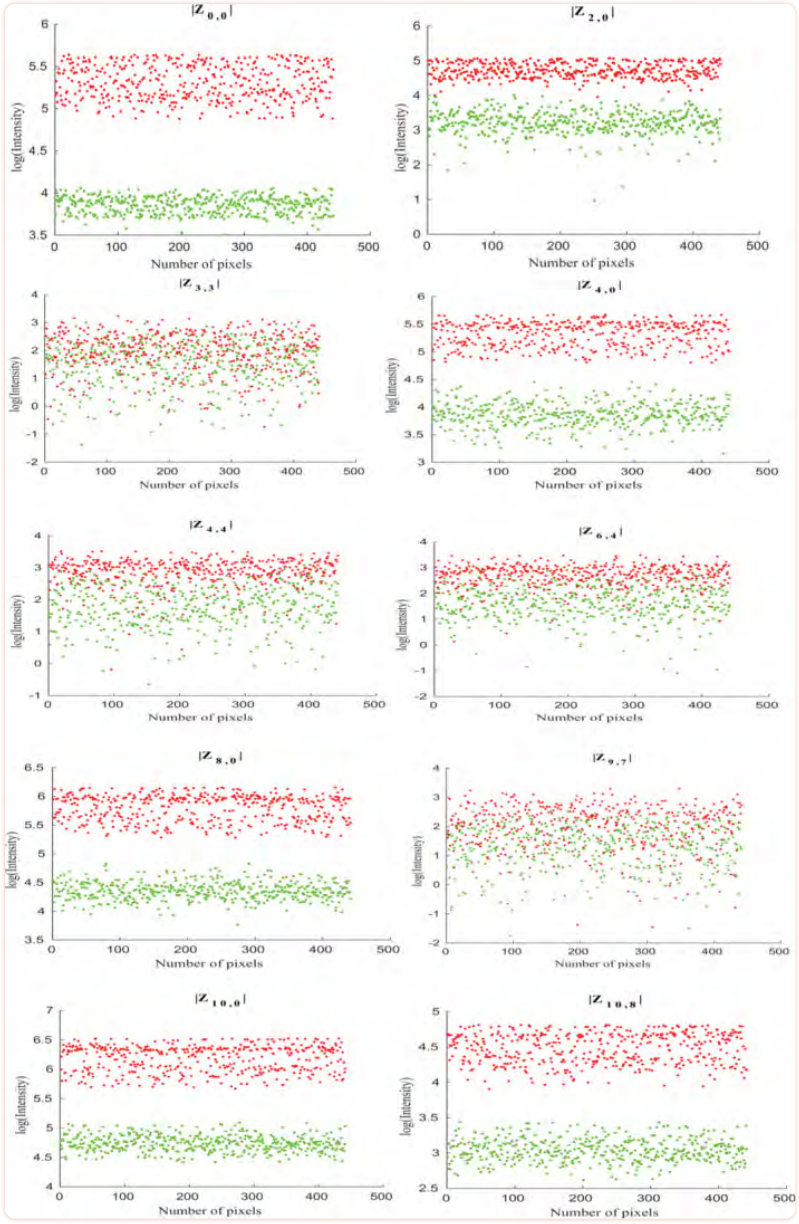


FIGURE 5

Plots for Zernike moment features from specific region of the training image (Red points represent background and Green points represents infected pixels) [Color figure can be viewed at wileyonlinelibrary.com]

To make the binary classification more accurate, we add an extra texture feature computed using the GLCM. The texture feature is more important in medical imaging (CT image) for extracting the GGO and pulmonary consolidation features.[49] GLCM consists of the relationship of the different angles between image pixels. If the gray level co-occurrence matrix obtained from an image be denoted as $q = [q(r, s\,|d, \theta)]$. In this point, GLCM is used to estimate the features of r^{th} pixel frequency with the feature of sth pixel frequency by the length ($d = 1$) and direction ($\theta = 0$). The computation of the texture feature follows the same procedure as that of ZM, that is, computing the texture features for every pixel by moving a window of 5 × 5 with a stride of 1 over the pre-processed image.

From GLCM, texture features such as entropy, variance, contrast, correlation, dissimilarity, energy and homogeneity are calculated. Among all these features, we considered the GLCM variance and contrast as shown in Equations (10) and (11). These two features are combined to form GLCM features *Gf* that gives proper discrimination of infected region and background region from the CT image.

$$\text{Contrast} = \sum_{r=0}^{N-1} \sum_{s=0}^{N-1} q(r,s)(r-s)^2 \tag{10}$$

$$\text{Variance} = \sum_{r=0}^{N-1} \sum_{S=0}^{N-1} q(r,s)(r-\mu)^2 \tag{11}$$

$$\text{Gf}_i = [\text{variance}_i, \text{contrast}_i] \tag{12}$$

where, *i* refers to the number of training images, ' μ ' denotes mean and '*N*' denotes the image size, Gf_i refers the GLCM features (2 features). The intensity for the GLCM feature images (contrast and variance) is plotted as shown in Figure 6. It represents number of pixels in *x*-axis and GLCM features in y-axis. Here the red points denote infected region and green points denote background respectively. The observation gives clear distinction between the background region and infected region.

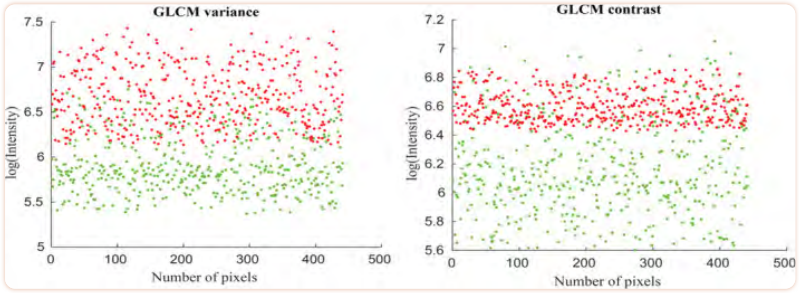


FIGURE 6

Plots for GLCM features from specific region of the training image (Red points represents background and Green points represents infected pixels) [Color figure can be viewed at wileyonlinelibrary.com]

Now the shape features computed from the ZM and the texture features obtained from the GLCM are integrated to form a feature vector of 38 features as indicated in Equation (13).

$$Fs_i = \{\text{ZM}_{h,i}, \text{ZM}_{l,i}, \text{Gf}_i\} \tag{13}$$

The next section discusses about the formation of training dataset from Fs_i features.

4. EXPERIMENTAL DESIGN

4.1. COVID-19 dataset

COVID-19 dataset (axial view) is collected from References 17 and 50 to evaluate our implementation. The previous implementations have used X-rays and CT image datasets for segmentation and classification of the infected COVID regions.[50], [51] The COVID-19 CT dataset used in this work contains 80 images of size 512 × 512 pixels (3 channel) along with the ground truth image (binary image). Ground truth images are useful for the identification of targets from the input dataset. From the dataset, 30 images are selected for training the network, the remaining 50 images are used for testing the trained network for performance evaluation.

4.2. Training and testing dataset

COVID-19 CT images contain additional background pixels (without infected region) other than infected pixels. Using the whole image as the training data gives testing results with less accuracy that reduce the overall performance. It should be able to classify the background region from the infected region during the testing. Hence, a balanced training dataset is required for the both infected region *If* and background *Bf* pixels along with all the features from the training images.

The works[19], [33], [35] have used the whole image as training data. The training result has delivered a performance greater than 90% as compared to the overall testing accuracy. However, as explained in Fan et al.[19] the overall sensitivity of the test dataset is reduced by not selecting the true positive target properly (infected pixels). Hence, we chose properly infected pixels and background pixels from the extracted images.

4.3. Region-specific for formation of training points

The dataset of training points is manually selected from the infected and background pixels from the 30 training images (after pre-processing and Fs_i) by using the region- specific approach. The set of features (S_i) is formed by marking the specific region (infection region (*If*) and background region (*Bf*)) for each feature from the training image dataset as shown in Figure 7.

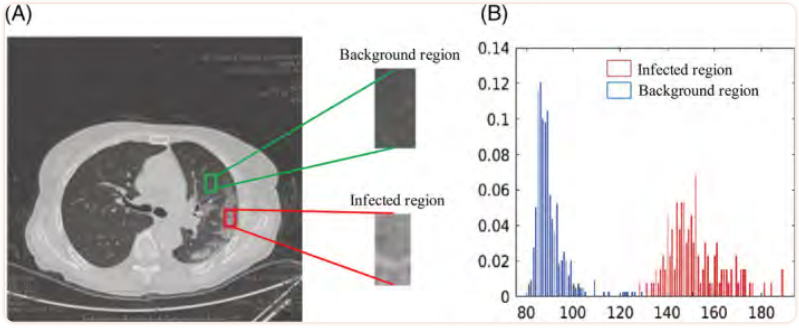


FIGURE 7

A, Specific region of background (*Bf*) and infection features (*If*) are separated out for training. B, Histogram plot for *Bf* (blue colour) and *If* (brown colour) [Color figure can be viewed at wileyonlinelibrary.com]

We chose the *If* region with 62 pixels and Bf region as 456 pixels. This feature helps to classify the image pixels as infected or not from COVID-19. The algorithm for region- specific approach is explained below:

Input:
Fs_i= {$Z_{h,i}$, $Z_{l,i}$, Gf_i}; //i refers to number of training images

Output: If; //Infection region feature.

Bf; // Background region features from Fs_i.

S_i; // Input feature set for training.

Start

Initialize $S_{i,k}$ = [ ];

Initialize j = 518; // Number of training point in each features

for i = 1 to 30 // Number of training images

for m = 1 to 38 // Number of features

If = Fs_i[j-517:j-456] // Vectorized format of Infected region point

Bf = Fs_i[j-517:j-62] // Vectorized format of Background region poin

M_i = [If, Bf];

S_i = [S_i, M_i] // Training points features from 30 training image

end

end

Stop

Region selection is helpful in getting balanced image information between background and infected pixels. The selected pixels are rearranged into the vector points for training. The total training points consist of 589 380 pixels with 519 371 background pixels and 70 009 infected pixels. In the total training points, each image contains 19 646 points of 38 features. A large number of points are taken into COVID infected region to reduce the misclassification of testing image features.

During the training process, the feature vector is labelled using the ground truth as shown with T_i as the target vector represented in binary form.

$$f_{i,k} = [S_i, T_i] \tag{14}$$

$$T_i = [\text{class } 1_i, \quad \text{class } 2_i] \tag{15}$$

here $f_{i,k}$ is the formed training dataset, *S* is the input features, *T* is target for the corresponding input features, *i* refers to the number of training images, *k* refers to the number of training points (*k = 1* to *589 380* points from all features of Fs_i). From the *T*, *class 1* = '*1*' and *class 2* = '*0*' represents an infected pixel, *class 1* = '*0*' and *class 2* = '*1*' represents the background pixel.

Furthermore, to select a suitable classifier for the proposed work, the performance of different classifiers such as Support Vector Machine (SVM), Deep Neural Network (DNN), Decision Tree (DT), Logistic Regression (LR) and Gaussian Naive Bayes (GNB) are compared using the training data points. The performance of classifiers is evaluated from the confusion matrix using the standard metrics: Area Under Curve (AUC), Specificity, Sensitivity and Accuracy as presented in Table 3. From Table 3, it is observed that the DNN classifier outperforms the other classifier in terms of Accuracy, AUC and Sensitivity for the given training points and corresponding confusion matrix for the $f_{i,k}$ for DNN has been presented in Table 4. AUC is estimated from the ROC graph for the different classifiers that are as plotted in Figure 8.

TABLE 3

Evaluation of training points metrics for different classifiers

Different classifiers	AUC	Sensitivity	Specificity	Accuracy (%)
SV M	0.8723	0.5549	0.8988	88.6
GNB	0.9107	0.4025	0.9735	83.3
LR	0.8187	0.5211	0.8950	88.2
DT	0.94	0.756	0.9593	93
DNN	**0.9427**	**0.7678**	**0.9285**	**93.8**

Note: The significance of the bold numbers are highlighted results of the proposed algorithm.

TABLE 4

Confusion matrix of training points for DNN classifier

		T_i		
		Infected pixels	**Background pixels**	**Accuracy**
Training output	Infected pixels	46 469 (7.9%)	12 684 (2.2%)	93.8%
	Background pixels	23 540 (4.0%)	506 687 (85.9%)	

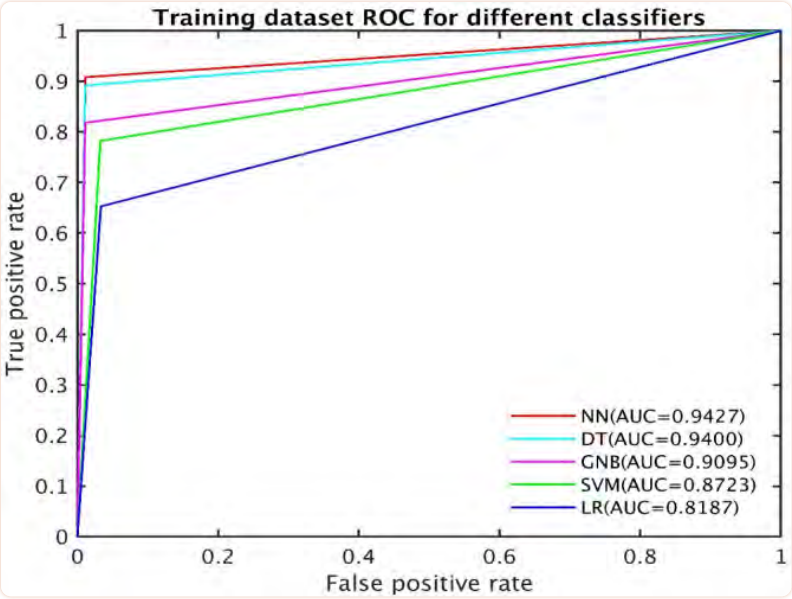


FIGURE 8

ROC for different classifiers [Color figure can be viewed at wileyonlinelibrary.com]

4.4. Binary classification using DNN

The set of ZM and GLCM features extracted using the methodology as explained in Section 3 form the feature set. The corresponding ground truth data is considered as a target. A supervised learning method is implemented by training a Deep Neural Network (DNN). Our architecture is identical to the basic feed-forward network with multiple hidden layers. This work considers the architecture where the size of the input layer is 38 neurons (38 features), three hidden layers with 58 neurons per layer and binary classification output layer as shown in Figure 9.

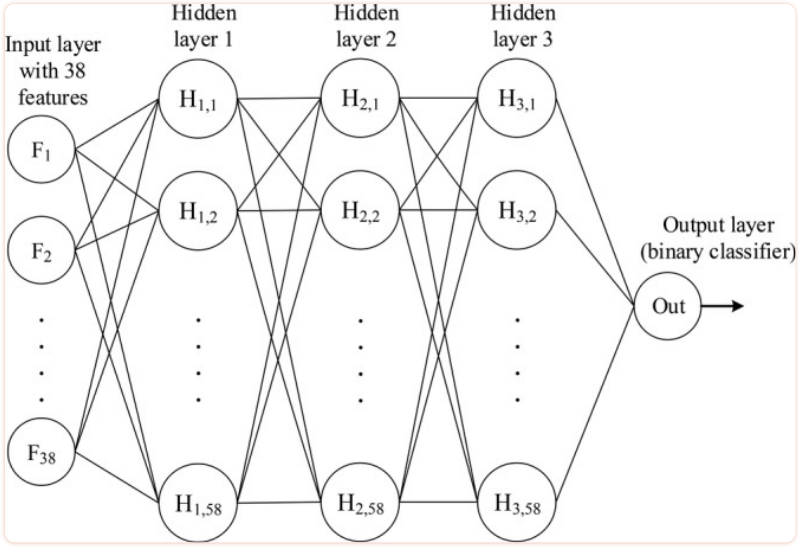


FIGURE 9

Deep neural network (DNN) architecture

The total number of learnable parameters (weights and biases) in the neural network is 9224. The first layer has 2262 (38 × 58 + 58), the second layer 3422 (58 × 58 + 58), the third layer 3422 (58 × 58 + 58) and the final layer with 118 (58 × 2 + 1). The loss function (cross-entropy loss) converges by modifying the parameters using Scaled Conjugate Gradient (SCG) training function. The three hidden layers are followed with the activation function called tan-sigmoid and the output layer involves the softmax operation that gives the probability of two outputs. The probability of the first neuron in the output layer represents the lungs infected region and the second neuron is complementary of the first probability called background region. The training required 8 minutes and 49 seconds to get the best error performance of 0.1367 at 100 epochs.

5. RESULTS AND DISCUSSION

The trained DNN classifier is evaluated for the test image (50 images) from the COVID-19 dataset. After the pre-processing and Fs_i, the test vector is formed for each test image of size 38 × 256 × 256. All the experimental analysis is done in MATLAB R2020a run on Intel Corei3-2330 M CPU. The feature generation process takes around 250 minutes (about 4 hours) for the test set of 50 images and testing time for each test image is around 4 to 7 minutes.

The challenges of automatic segmentation from the CT image are.

1. CT Images are varying in terms of texture, shape, and position across different cases and consolidation is small, which may result in a false negative output from a whole CT image.
2. Due to the low contrast and blurriness within the GGO of the infection area, there is a significant difficulty in the identification of the contours in the CT images.
3. Artificial intelligence (AI) model requires a greater number of training data to train the model and more human/ specialist resources are required to collect and annotate the ground truths of the COVID-19 dataset.

As explained in Section 3, the feature vector of 38 features is given to the trained DNN model as input. The output of the classifier gives the probability score for two classes (infected and background). This probability score decides and falls on particular class 1 or class 0. The probability vector output is reshaped into the image size 256 × 256. Further, the image of size 256 × 256 is compared with the ground truth image to detect and segment the COVID-19 regions and also evaluate the performance of the model.

To make the output binary image effective, the morphological operation is applied to the output binary image. Different morphological structuring element produces different test image results. Hence, we find the suitable structuring element with a proper size for all the test images to achieve the overall average accuracy and selecting the structuring element that gives high accuracy. The average accuracy vs morphological structuring element graph for COVID-19 CT image test data is shown in Figure 10. From the plot, it is determined that "imclose" disk structuring element of size '2' reaches the maximum accuracy. The test dataset results are displayed in Figure 11. In Figure 11d column, GT contours are overlaid on the output image which gives visual understanding of the non-distinguishable pixels. This arises due to the poor lightening between infected pixels and background pixels from the original image as shown in Figure 11 (3rd-row image) and also the thin occurrence of GGO and consolidation in Figure 11 (6th-row image). From the ablation study Table 5, 36 ZM shape descriptor training feature points alone do not achieve better results in Sensitivity and Mean IOU. Shape descriptor features do not capture the proper infected region (True Positive) information from the test image. so that, 2 GLCM texture features are added with 36 ZM features and trained the DNN model with 38 features. Evaluated metrics achieve better results with the combination of ZM and GLCM features (38 features) compared to the 36 features results. We have analysed each test image performance by standard metrics[39] such as Accuracy (A), Sensitivity or Recall (R), Specificity, F1 score, Dice index, Precision (P), Mean IOU and score using the confusion matrix as represented in Table 6.

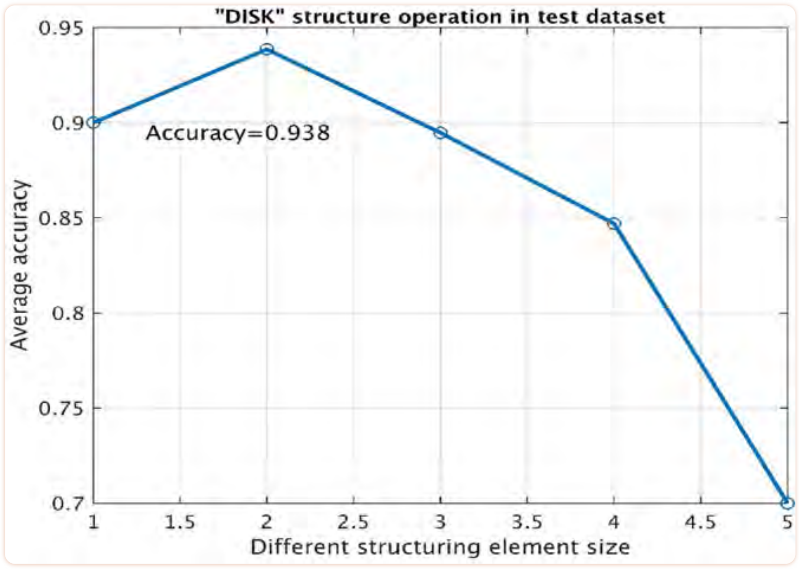


FIGURE 10

Test dataset plot for morphological structuring element vs average accuracy [Color figure can be viewed at wileyonlinelibrary.com]

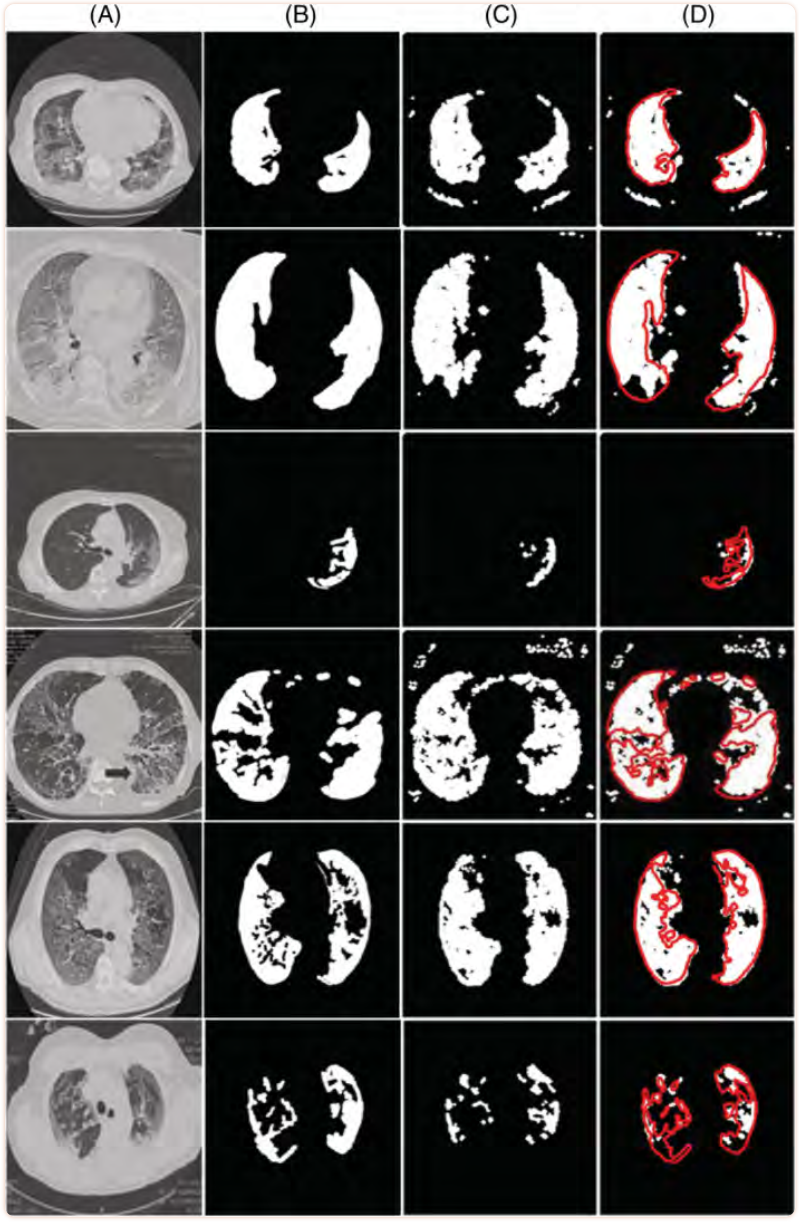


FIGURE 11

Segmentation results for the COVID-19 dataset. A, Column shows computed tomography axial view input image; B, Column shows ground truth image; C, Column shows the output from deep neural network model and D, Column shows overlay the GT contour on output image for visual understanding [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5

Performance evaluation for the test dataset (36 ZM features)

Number of test images	Accuracy (A) = $\frac{TP+TN}{TP+TN+FP+FN}$	IOU score		Mean IOU = $\frac{C_0+C_1}{2}$	Sensitivity or Recall (R) = $\frac{TP}{TP+FN}$	Specificity = $\frac{TN}{TN+FP}$	F1 score = $2 \times \frac{P \times R}{P+R}$	Precision (P) = $\frac{TP}{TP+FP}$
		Class 0 (C_0) = $\frac{TN}{TN+FP+FN}$	Class 1 (C_1) = $\frac{TP}{TP+FP+FN}$					
1	83.5	0.393	0.815	0.604	0.414	0.98	0.564	0.884
2	85.6	0.337	0.843	0.59	0.381	0.969	0.505	0.747
3	88.8	0.459	0.936	0.697	0.490	0.992	0.629	0.879
4	91.5	0.166	0.928	0.547	0.177	0.994	0.285	0.736
5	89.3	0.333	0.925	0.629	0.371	0.987	0.5	0.765
6	82.2	0.382	0.881	0.631	0.453	0.967	0.552	0.708
7	97.7	0.058	0.895	0.476	0.060	0.99	0.110	0.583
8	86.3	0.086	0.872	0.479	0.101	0.977	0.16	0.375
9	83	0.66	0.922	0.791	0.711	0.984	0.600	0.915
10	83.3	0.238	0.824	0.531	0.285	0.955	0.385	0.590

TABLE 6

Confusion matrix for test image 1

		GT		
		Infected pixels	Background pixels	Accuracy
Testing output	Infected pixels	14 025 (21.4%)	2228 (3.4%)	92.30%
	Background pixels	2818 (4.3%)	46 465 (70.9%)	

Additionally, three more important metrics are Enhance-alignment Measure (EM_φ), Mean Absolute Error (MAE)[38] and Structure measure (S_m)[52] are used for performance evaluation. Enhance-alignment Measure (EM_φ) is a newly proposed metric. It is useful to estimate the global and local similarities of two binary images as presented in Equation (16). where, W represents width, H represents the height of GT, F represent the final predicted output, (x, y) represents each pixel coordinate of GT and F. φ denotes the enhanced alignment matrix.

$$\mathrm{EM}_\varphi = \frac{1}{W \times H} \sum_{x}^{W} \sum_{y}^{H} \varphi(F(x,y), \mathrm{GT}(x,y)) \tag{16}$$

Mean Absolute Error (MAE), measures the error between the F and GT of each pixel in the image as indicated in Equation (17). Structure measure (S_m), measures the structural similarity between the GT mask and prediction map as represented in Equation (18). Where $\alpha \in [0, 1]$, the default setting $\alpha = 0.5$ represents the balance factor between region-aware similarity (S_r) and object-aware similarity (S_o). The performance of each test image is evaluated by metrics as listed in Table 7.

$$\mathrm{MAE} = \frac{1}{W \times H} \sum_{x}^{W} \sum_{y}^{H} \left| F(x,y) - \mathrm{GT}(x,y) \right| \tag{17}$$

$$S_m = \alpha * S_o + (1-\alpha) * S_r \tag{18}$$

TABLE 7

Performance evaluation for the random 10 test image from COVID-19 dataset (36 ZM and 2 GLCM features)

Number of test images	Accuracy (A) = $\frac{TP+TN}{TP+TN+FP+FN}$	IOU score		Mean IOU = $\frac{C_0+C_1}{2}$	Sensitivity or Recall (R) = $\frac{TP}{TP+FN}$	Specificity = $\frac{TN}{TN+FP}$	F1 score = $2 \times \frac{P \times R}{P+R}$	Precision (P) = $\frac{TP}{TP+FP}$	Dice index	M
		Class 0 (C_0) = $\frac{TN}{TN+FP+FN}$	Class 1 (C_1) = $\frac{TP}{TP+FP+FN}$							
1	0.9230	0.90	0.73	0.815	0.832	0.95	0.847	0.86	0.8028	0
2	0.9237	0.907	0.68	0.793	0.876	0.93	0.815	0.76	0.7369	0
3	0.9610	0.95	0.66	0.805	0.723	0.98	0.795	0.88	0.8121	0
4	0.9447	0.94	0.54	0.74	0.835	0.95	0.705	0.61	0.7923	0
5	0.9553	0.95	0.56	0.75	0.600	0.99	0.720	0.906	0.6479	0
6	0.9579	0.95	0.74	0.845	0.882	0.96	0.851	0.82	0.7488	0
7	0.9788	0.97	0.3	0.62	0.330	0.98	0.432	0.66	0.673	0
8	0.9394	0.938	0.31	0.624	0.346	0.98	0.473	0.81	0.6366	0
9	0.8771	0.848	0.61	0.729	0.828	0.89	0.759	0.70	0.6665	0
10	0.9248	0.910	0.67	0.79	0.852	0.94	0.704	0.76	0.6945	0

We have compared our method with some of the existing works that employ deep network for the segmentation of the COVID-19 infected region from the CT image dataset as presented in Table 8. In Ronneberger et al,[20] standard U-Net model is evaluated using 45 training images. The performance of the outcome is low in terms of sensitivity, specificity, MAE, Dice and EM_{φ}.

The modernized version of the U-Net model is described in References 21, 22, 23, 24 which resulted in a poor evaluation of the Dice index, sensitivity compared to our developed method. In our method, MAE is quite higher than the Semi-Inf-Net and Inf-Net model reported in Fan et al. [19] EM_{φ}, S_m and Sensitivity are moderately 3%, 2.12%, 3.3% lesser compared to Semi-Inf-Net model. However, the overall average comparison of our test image result is much improved and superior with a limited number of training points as shown in Table 9. Also, Table 9 shows a performance improvement with our method in each of the metrics compared to other existing works. The proposed model improved the sensitivity by 31.2%, 10%, 6.5%, 18%, 4.3% and 1.3% compared with U-Net model, Attention-UNet, Gated-UNet, Dense-UNet, U-Net++, and Inf-Net respectively. Likewise, the percentage improvements of other metrics in relation to other state-of-the-art is presented in Table 9.

TABLE 8

Comparative performance evaluation of other existing methods

Existing reference	Methods	Sensitivity	Specificity	Dice	MAE	EM_{φ}	S_m
Ronneberger et al [20]	U-Net [a]	0.534	0.858	0.439	0.186	0.625	0.622
Oktay et al [21]	Attention-UNet [a]	0.637	0.744	0.583	0.112	0.739	0.744
Schlemper et al [22]	Gated-UNet [a]	0.658	0.725	0.623	0.102	0.814	0.725
Li et al [23]	Dense-UNet [a]	0.594	0.655	0.515	0.184	0.662	0.655
Zhou et al [24]	U-Net++ [a]	0.672	0.722	0.581	0.120	0.720	0.722
Deng-Ping Fan et al [19]	Inf-Net	0.692	0.781	0.682	0.082	0.838	0.781
Deng-Ping Fan et al [19]	Semi-Inf-Net	0.725	0.800	0.739	0.064	0.894	0.800
Our method	DNN (specific background and infected region) [b]	0.701	0.942	0.757	0.082	0.867	0.783

[a] Models are analysed in Reference 19 using 45 training images and 50 test images. [50]

[b] 589 824 training point from 30 images and 20 test images, [17] 30 test images. [19]

TABLE 9

Percentage improvement compared between our model with existing work

Models	Sensitivity	Specificity	Dice	MAE	EM_{φ}	S_m
U-Net	31.2%	9.79%	72.43%	44%	38.7%	25.8%
Attention-UNet	10.04%	26.61%	29%	73%	17.3%	5.2%
Gated-UNet	6.53%	29.93%	21%	80%	6.5%	8%
Dense-UNet	18.01%	43.8%	46%	44.4%	30.9%	19%
U-Net++	4.31%	30.47%	30%	68%	20.4%	6%
Inf-Net	1.30%	20.6%	10.9%	—	3.4%	0.25%
Semi-Inf-Net	—	17.75%	2.4%	—	—	—
Avg. improvement with our method	10.19%	25.5%	30.2%	44.2%	16.7%	9.17%

Note: —, Unavailability of data.

6. CONCLUSION

In this paper, we propose a DNN model for COVID-19 detection from the lung's CT axial view image. The proposed method has adopted the ZM for shape features, GLCM for texture feature and specific- region for selecting the training points, precisely to extract the unique information from the CT image. Our design gives an effective detection of the COVID-19 infection in the lungs and an appropriate tool for radiologists to define the infection stage/percentage. Our model provided better results with limited number of training points. The average test dataset performance reached 70%, 94%, 86% and 78% in terms of Sensitivity, Specificity, EM_{φ} and S_m, respectively. With the proposed method, the average performance is improved in terms of Sensitivity, Specificity, Dice and EM_{φ} by 10.9%, 26%, 24.7%, and 16.5%, respectively in comparison to other popular deep networks such as U-Net, Gated-UNet, Dense-UNet, U-Net++, Inf-Net, Semi-UNet. Moreover, these popular deep networks need a larger number of images for training dataset to maintain the model performance. From the results and higher evaluation metrics, it is evident that the proposed model performed significantly well with much smaller number of training points in the training dataset. The limitation of our model is the difficulty in detecting the GGO from the poor contrast CT images which require additional features and enhancement methods to provide more detailed information. In future, the inner structure information can be extracted using unique texture features and added to the training dataset to improve the performance of the DNN model and also to overcome the limitations.

ACKNOWLEDGEMENTS

This research was financially supported by The Research Start-Up Fund Subsidized Project of Shantou University, China, Grant No: NTF17016. The authors would like to thank Vellore Institute of Technology, Vellore for providing lab facilities.

Notes

Selvaraj D, Venkatesan A, Mahesh VGV, Joseph Raj AN. An integrated feature frame work for automated segmentation of COVID-19 infection from lung CT images . *Int J Imaging Syst Technol*. 2021;31:28–46. 10.1002/ima.22525 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Funding information The Research Start-Up Fund Subsidized Project of Shantou University, China, Grant/Award Number: NTF17016

DATA AVAILABILITY STATEMENT

COVID-19 CT image dataset that support the findings of this study are openly available in https://radiopaedia.org/articles/COVID-19-3?lang=us.Codes are deposited into the public github database https://github.com/deepika-s517/Auto-segment-COVID-19-CT-images.git. Remaining materials and supporting data of this article are included within the article.

REFERENCES

1. Shi F, Wang J, Shi J, et al. Review of Artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. *IEEE Reviews in Biomedical Engineering*. 2020;1-1. 10.1109/rbme.2020.2987975. [PubMed] [CrossRef] [Google Scholar]

2. La Rosa G, Iaconelli M, Mancini P, Bonanno Ferraro G, Veneri C, Bonadonna L, Lucentini L, Suffredini E. First detection of SARS-CoV-2 in untreated wastewaters in Italy. *Science of The Total Environment*. 2020;736:139652. 10.1016/j.scitotenv.2020.139652. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

3. Fong SJ, Dey N, Chaki J. *Artificial Intelligence for Coronavirus Outbreak*. Singapore: Springer; 2020. 10.1007/978-981-15-5936-5 [CrossRef] [Google Scholar]

4. Syrjala H, Broas M, Ohtonen P, Jartti A, Pääkkö E. Chest magnetic resonance imaging for pneumonia diagnosis in outpatients with lower respiratory tract infection. *European Respiratory Journal*. 2017;49(1):1601303. 10.1183/13993003.01303-2016. [PubMed] [CrossRef] [Google Scholar]

5. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine*. 2018;15(11):e1002683. 10.1371/journal.pmed.1002683. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

6. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for Typical Coronavirus Disease 2019 (COVID-19) Pneumonia: Relationship to Negative RT-PCR Testing. *Radiology*. 2020;296(2):E41-E45. 10.1148/radiol.2020200343. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

7. Lee JH, Kim DI, Cho MK. Computed tomography apparatus and method of controlling X-ray by using the same, ed: Google Patents, 2017. https://patents.google.com/patent/US9655584B2/en

8. Forthmann P, Pfleiderer G. Augmented display device for use in a medical imaging laboratory. Google Patents, 2019.

9. Jensen VT. Method and system of acquiring images with a medical imaging device, ed: Google Patents, 2009. https://patents.google.com/patent/US20060293582

10. Scheib S. Dosimetric end-to-end verification devices, systems, and methods. U.S. Patent 9,643,029, issued May 9, 2017. https://patents.google.com/patent/US20150085993

11. Li Y, Xia L. Coronavirus Disease 2019 (COVID-19): Role of Chest CT in Diagnosis and Management. *American Journal of Roentgenology*. 2020;214(6):1280-1286. 10.2214/ajr.20.22954. [PubMed] [CrossRef] [Google Scholar]

12. Salehi S, Abedi A, Balakrishnan S, Gholamrezanezhad A. Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients. *American Journal of Roentgenology*. 2020;215 (1):87-93. 10.2214/ajr.20.23034. [PubMed] [CrossRef] [Google Scholar]

13. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. 2020;395(10223):497-506. 10.1016/s0140-6736(20)30183-5. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

14. Wang L, Wang Y, Ye D, Liu Q. Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence. *International Journal of Antimicrobial Agents*. 2020;55(6):105948. 10.1016/j.ijantimicag.2020.105948. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

15. Li K, Fang Y, Li W, Pan C, Qin P, Zhong Y, Liu X, Huang M, Liao Y, Li S. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). *European Radiology*. 2020;30(8):4407-4416. 10.1007/s00330-020-06817-6. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

16. Rajinikanth V, Dey N, Raj ANJ, Hassanien AE, Santosh KC, Raja N. Harmony-search and otsu based system for coronavirus disease (COVID-19) detection using lung CT scan images. *arXiv preprint arXiv:2004.03431*;2020. [Google Scholar]

17. https://radiopaedia.org/articles/COVID-19-3?lang=us. Last Accessed date April 5, 2020.

18. Gaál G, Maga B, Lukács A. Attention u-net based adversarial architectures for chest x-ray lung segmentation. *arXiv preprint arXiv:2003.10304* ; 2020. https://arxiv.org/pdf/2003.10304.pdf

19. Fan DP, Zhou T, Ji GP, et al. Inf-net: automatic COVID-19 lung infection segmentation from CT images. *IEEE Trans Med Imaging*. 2020;39:2626-2637. 10.1101/2020.04.22.20074948 [PubMed] [CrossRef] [Google Scholar]

20. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer; 2015, October:234-241. 10.1007/978-3-319-24574-4_28 [CrossRef] [Google Scholar]

21. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Glocker B. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* ; 2018. https://arxiv.org/pdf/1804.03999.pdf

22. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, Rueckert D. Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*. 2019;53:197-207. 10.1016/j.media.2019.01.012. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

23. Li X, Chen H, Qi X, Dou Q, Fu C-W, Heng P-A. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes. *IEEE Transactions on Medical Imaging*. 2018;37(12):2663-2674. 10.1109/tmi.2018.2845918. [PubMed] [CrossRef] [Google Scholar]

24. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer; 2018:3-11. 10.1007/978-3-030-00889-5_1. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

25. Fu H, Fan DP, Chen G, Zhou T . COVID-19 imaging-based AI research collection. 2020. https://github.com/HzFu/COVID19_imaging_AI_paper_list.

26. Angulakshmi M, Lakshmi Priya GG. Automated brain tumour segmentation techniques- A review. *International Journal of Imaging Systems and Technology*. 2017;27(1):66-77. 10.1002/ima.22211. [CrossRef] [Google Scholar]

27. Fan DP, Ji GP, Sun G, Cheng MM, Shen J, Shao L. Camouflaged object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA, USA: IEEE; 2020:2777-2787. 10.1109/CVPR42600.2020.00285 [CrossRef] [Google Scholar]

28. Ali MNY, Sarowar MG, Rahman ML, Chaki J, Dey N, Tavares JMRS. Adam Deep Learning With SOM for Human Sentiment Classification. *International Journal of Ambient Computing and Intelligence*. 2019;10(3):92-116. 10.4018/ijaci.2019070106. [CrossRef] [Google Scholar]

29. Shan F, Gao Y, Wang J., Shi W, Shi N, Han M, Shi Y. Lung infection quantification of covid-19 in CT images with deep learning. *arXiv preprint arXiv:2003.04655* ; 2020. https://arxiv.org/ftp/arxiv/papers/2003/2003.04655.pdf

30. Ye C, Wang W, Zhang S, Wang K. Multi-Depth Fusion Network for Whole-Heart CT Image Segmentation. *IEEE Access*. 2019;7:23421-23429. 10.1109/access.2019.2899635. [CrossRef] [Google Scholar]

31. Liu Z, Song Y-Q, Sheng VS, Wang L, Jiang R, Zhang X, Yuan D. Liver CT sequence segmentation based with improved U-Net and graph cut. *Expert Systems with Applications*. 2019;126:54-63. 10.1016/j.eswa.2019.01.055. [CrossRef] [Google Scholar]

32. Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, Liu T, Yang X. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. *Medical Physics*. 2019;46(5):2157-2168. 10.1002/mp.13458. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

33. Chen X, Yao L, Zhang Y. Residual attention U-net for automated multi-class segmentation of COVID-19 chest CT images. *arXiv preprint arXiv:2004.05645* ; 2020. https://arxiv.org/pdf/2004.05645.pdf

34. Chen X, Yao L, Zhou T, Dong J, Zhang Y. Momentum contrastive learning for few-shot covid-19 diagnosis from chest CT images. *arXiv preprint arXiv:2006.13276* ; 2020. https://arxiv.org/pdf/2006.13276.pdf [PMC free article] [PubMed]

35. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *arXiv preprint arXiv:2003.10849* ; 2020. https://arxiv.org/ftp/arxiv/papers/2003/2003.10849.pdf [PMC free article] [PubMed]

36. Shi F, Xia L, Shan F, Wu D, Wei Y, Yuan H, Shen D. Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification. *arXiv preprint arXiv:2003.09860* ; 2020. https://arxiv.org/ftp/arxiv/papers/2003/2003.09860.pdf [PubMed]

37. Oh Y, Park S, Ye JC. Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets. *IEEE Transactions on Medical Imaging*. 2020;39(8):2688-2700. 10.1109/tmi.2020.2993291. [PubMed] [CrossRef] [Google Scholar]

38. Wang L, Wong A. COVID-Net: a tailored deep convolutional neural network Design for Detection of COVID-19 cases from chest X-ray images. *arXiv preprint arXiv:2003.09871* ; 2020. https://arxiv.org/pdf/2003.09871.pdf [PMC free article] [PubMed]

39. Butt C, Gill J, Chun D, Babu BA. RETRACTED ARTICLE: Deep learning system to screen coronavirus disease 2019 pneumonia. *Applied Intelligence*. 2020;1-7. 10.1007/s10489-020-01714-3. [CrossRef] [Google Scholar]

40. Konar D, Panigrahi BK, Bhattacharyya S, Dey N. Auto-Diagnosis of COVID-19 Using Lung CT Images with Semi-Supervised Shallow Learning Network. *Research Square*. 2020. 10.21203/rs.3.rs-34596/v1. [CrossRef] [Google Scholar]

41. Singh M, Bansal S, Ahuja S, Dubey RK, Panigrahi BK, & Dey N. Transfer Learning Based Ensemble Support Vector Machine Model for Automated COVID-19 Detection Using Lung Computerized Tomography Scan Data. *Research Square*. 2020. 10.21203/rs.3.rs-32493/v1. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

42. Ahuja S, Panigrahi BK, Dey N, Rajinikanth V, Gandhi TK. Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices. *Applied Intelligence*. 2020. 10.1007/s10489-020-01826-w. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

43. Dey N, Rajinikanth V, Fong SJ, Kaiser MS, Mahmud M. Social group optimization–assisted Kapur's entropy and morphological segmentation for automated detection of COVID-19 infection from computed tomography images. *Cognitive Computation*. 2020;12(5):1011-1023. 10.1007/s12559-020-09751-3. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

44. Wang X, Deng X, Fu Q, Zhou, Feng J, Ma H, Liu W, Zheng C. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. *IEEE Transactions on Medical Imaging*. 2020;39(8):2615–2625. 10.1109/tmi.2020.2995965. [PubMed] [CrossRef] [Google Scholar]

45. Chaki J, Dey N. *A beginner's guide to image preprocessing techniques*. Boca Raton: CRC Press; 2018. 10.1201/9780429441134 [CrossRef] [Google Scholar]

46. Subramani B, Veluchamy M. MRI brain image enhancement using brightness preserving adaptive fuzzy histogram equalization. *International Journal of Imaging Systems and Technology*. 2018;28(3):217-222. 10.1002/ima.22272. [CrossRef] [Google Scholar]

47. Raj ANJ, Mahesh VG. Zernike-moments-based shape descriptors for pattern recognition and classification applications. *Advanced Image Processing Techniques and Applications*. Hershey: IGI Global; 2017:90-120. 10.4018/978-1-5225-2053-5.ch004 [CrossRef] [Google Scholar]

48. Adapa D, Joseph Raj AN, Alisetti SN, Zhuang Z, Ganesan K, Naik G. A supervised blood vessel segmentation technique for digital Fundus images using Zernike Moment based features. *PLOS ONE*. 2020;15(3):e0229831. 10.1371/journal.pone.0229831. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

49. Barstugan M, Ozkaya U, Ozturk S. Coronavirus (covid-19) classification using ct images by machine learning methods. *arXiv preprint arXiv:2003.09424* ; 2020. https://arxiv.org/ftp/arxiv/papers/2003/2003.09424.pdf

50. COVID-19 CT segmentation dataset . https://medicalsegmentation.com/covid19/. Accessed April 11, 2020.

51. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. Covid-19 image data collection: Prospective predictions are the future. *arXiv preprint arXiv:2006.11988* ; 2020. https://arxiv.org/pdf/2006.11988.pdf

52. Fan DP, Cheng MM, Liu Y, Li T, Borji A. Structure-measure: a new way to evaluate foreground maps. *Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy: IEEE; 2017:4548-4557. [Google Scholar]

Bilingual text detection from natural scene images using faster R-CNN and extended histogram of oriented gradients

Industrial and Commercial Application Published: 06 April 2022

Volume 25, pages 1001–1013, (2022) Cite this article

Pattern Analysis and Applications

Aims and scope

Submit manuscript

Alex Noel Joseph Raj, Chen Junmin, Ruban Nersisson, Vijayalakshmi G. V. Mahesh & Zhemin Zhuang

Abstract

In today's world, scene text detection is important for a wide range of scientific and industrial processes. Compared with text detection in documents, text detection in natural scenes is challenging since they are subjected to different orientations, scaling, brightness variations, and complex backgrounds. Further scenes can contain multiple scripts which limits the performance of detection algorithms. In this paper, we propose a state-of-the-art algorithm for text detection for a bilingual natural scene dataset. The framework consists of (a) Faster R-CNN employed to extract probable text regions within the scene images, (b)

rearrangement of the text region as consecutive frames along the time axis and extraction of global and local shape features from the three orthogonal planes and (c) use of simple and effective classifier to predict the features extracted from regions as text or non-text region. The proposed frame when compared to other text detection techniques improves the overall text detection accuracy. The validity of the algorithm is verified on the bilingual text detection dataset MSRA-TD500, and a promising F1 score of 0.70 is reported.

References

1. Yan C, Xie H, Liu S, Yin J, Zhang Y, Dai Q (2017) Effective Uyghur language text detection in complex background images for traffic prompt identification. IEEE Trans Intell Transp Syst 19(1):220–229

Article Google Scholar

INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Spectrum Sensing in Non-Gaussian Noise

Deepa N Reddy[1]*, **Yerram Ravinder**[2]

1 Assistant Professor, Department of Electronics and Communication, BMS Institute of Technology & Management, Bengaluru, India
2 Professor, Department of Electronics and Telecommunication, Pune Institute of Computer Technology, Pune, India

OPEN ACCESS

Received: 06.06.2021
Accepted: 29.08.2021
Published: 27.09.2021

Citation: Reddy DN, Ravinder Y (2021) Spectrum Sensing in Non-Gaussian Noise. Indian Journal of Science and Technology 14(32): 2596-2606. https://doi.org/10.17485/IJST/v14i32.1034

* **Corresponding author**.

reddydeepa2680@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background: Spectrum sensing is a crucial step to realize the Cognitive Radio technology. The spectrum sensing schemes at low signal-to-noise ratio, noise uncertainty and especially under the background of non-Gaussian noise, provide low detection of the primary user. This results in missed detection or false alarm and increases higher interference to the primary user. **Objectives:** Detection schemes designed for additive Gaussian noise exhibit poor performance in the non-Gaussian environment. This study considers the problem of spectrum sensing with the assumption that the noise follows a non-Gaussian distribution with heavier tails. **Methods/findings:** A fuzzy logic-based method is proposed for primary user detection under non Gaussian Noise. The results are highlighted for the Laplacian noise. Through Monte Carlo simulations it is observed that Laplacian noise noticeably affects the performance of energy detector. Also, a fractional change in noise uncertainty degrades the performance of energy detector. The performance of the proposed scheme is presented through receiver operating characteristic (ROC) and plot of the detection probability versus signal-to-noise ratio (SNR) using simulations. It is shown that by appropriately choosing the membership functions and the fuzzy rule base in the fuzzy inference system the proposed fuzzy logic method for spectrum sensing provides reliable detection.

Keywords: NonGaussian noise; Fuzzy logic; Spectrum sensing; noise uncertainty

1 Introduction

Majority of spectrum sensing schemes are focused on primary user detection in Gaussian noise. The Gaussian distribution fails to satisfactorily describe some noise in practice. For instance, the Gaussian distribution cannot be used to model Radio Frequency (RF) noise and low frequency atmospheric noise. This is primarily due to the fact that the noise in practice is likely to generate observations of high magnitude than what can be produced from the Gaussian distribution. High magnitude observations are also referred to as impulsive noise. This conveys that the probability distribution function (pdf) of such noise has heavy tails. It can be stated that the pdf of the

non-Gaussian impulsive noise decays at a lower rate than the Gaussian. Therefore, a prime feature of the non-Gaussian model is that it has heavy tails than the Gaussian distribution. Also in practical wireless communication scenarios the noise usually has a heavy tailed nature. The Laplacian noise is an important non-Gaussian noise distribution and hence it is frequently used in engineering studies.

For accurate detection of primary user (PU) signal under noise uncertainty many detection schemes are designed for additive Gaussian noise. The effect of threshold selection over the performance of spectrum sensing in cognitive radio network (CRN) using energy detector (ED) is studied in [1–4]. Through simulations it is shown that it provides better detection performance through suitable selection of the dynamic threshold factor. The effect of noise uncertainty on the energy detection approach is analyzed in [5,6]. In [6] a two-step cooperative stochastic resonance energy detection algorithm is proposed to address the problem, where the traditional energy detection (ED) performance is susceptible to noise uncertainty.

Recently, under the category of non-parametric detection, some detection schemes based on the Goodness-of-fit test (GoFT) have been proposed. They include the Anderson-Darling (AD) detection, the Kolmogorov-Smirnov (KS) test, the detection based on ordered statistics and so on. These schemes perform better than energy detection in a Gaussian environment [7]. A Frequency domain Goodness of Fit Test based spectrum sensing method is proposed under dynamically varying noise. Through analytical and simulated results, it is shown that this method is robust to dynamically varying noise [8]. In [9–11] fuzzy logic is used to address the decision on the PU activity. This scheme showed better detection performance in low signal to noise ratio (SNR) regions. It is also shown through simulation results that the fuzzy based detection could achieve accurate sensing results better than conventional schemes in noisy conditions. Several studies in [12–15] have focused on signal detection under a NGN environment. Schemes like energy detection, sequential detection and absolute value cumulation detection are investigated in a Laplacian environment. However, there still exist some issues in PU detection in Cognitive Radio (CR). Two primary issues are 1) Improving the detector's performance under non-Gaussian noise, 2) need for computationally efficient method to meet the real-time detection requirement.

This work considers the problem of spectrum sensing with the assumption that the noise follows a non-Gaussian distribution with heavier tails. The main contributions in this paper are summarized as follows:

- The effect of non-Gaussian noise on the energy detector's performance in noise uncertain environment is investigated.
- To overcome the effect of noise uncertainty, the spectrum sensing algorithm using adaptive fuzzy threshold proposed in [16] primarily studied under Gaussian noise is investigated under non-Gaussian noise environment.
- The effect on the probability of detection due to the characterization of the statistical moments in the noise distribution is investigated.
- Performance evaluation of the conventional and fuzzy threshold methods are carried out through simulations under different noise uncertainty conditions.
- The best method out of these techniques under varying noise uncertainty conditions is identified.

The rest of this paper is organized as follows. Section 2 gives the outline on the non-Gaussian noises. Section 3 and 4 detail the system model for the spectrum sensing and the proposed method. The results are presented and discussed in Section 5 and concluding remarks are provided in Section 6.

2 Non-Gaussian Noise

Pearson introduced kurtosis to understand whether a given distribution is Gaussian or not. Three conditions proposed by Pearson for a distribution to be called as Gaussian are (i) kurtosis, (ii) skewness, (iii) the distance between mean and mode (modal divergence). Skewness and kurtosis are the higher order moments that are used to determine the difference between other distributions and the Gaussian distribution [17,18].

Since the Laplace distribution is the most popular distribution and is frequently used in engineering studies this work focuses on primary user signal detection in Laplacian noise. The Laplace distribution is within the class of generalized Gaussian distributions.

The Gaussian pdf is given as

$$p(n) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{n^2}{2\sigma^2}\right) \tag{1}$$

The Laplacian pdf is given by

$$p(n) = \frac{1}{\sqrt{2\sigma^2}} exp\left(-\sqrt{\frac{2}{\sigma^2}}|n|\right), -\infty < n < \infty \tag{2}$$

Where, σ^2 is the noise variance

Normally, a heavier tailed noise has a larger degree of non-Gaussianity, also known as kurtosis, defined as

$$k = \frac{E[n^4]}{E^2[n^2]} \tag{3}$$

Equivalently the kurtosis coefficient minus 3 is referred to as excess kurtosis. As the kurtosis increases the tails are heavier and k as 6 (excess kurtosis is 3) corresponds to Laplacian noise, while when the pdf is Gaussian the excess kurtosis is 0. Figure 1 gives the pdf plot of Eqn (1) and (2). It is observed that the non-Gaussian model has heavier tail than the Gaussian distribution.

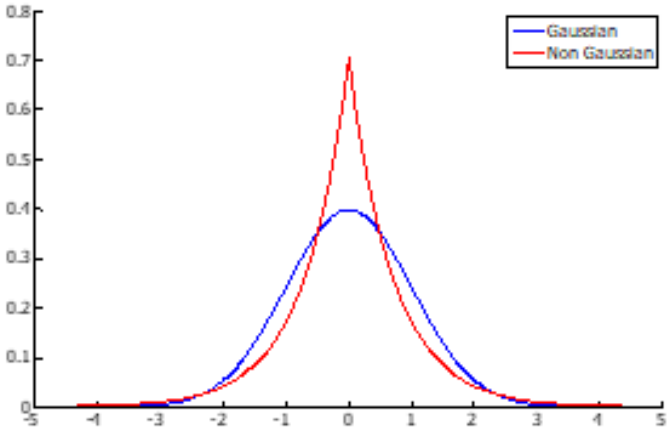


Fig 1. Gaussian and non Gaussian distributions

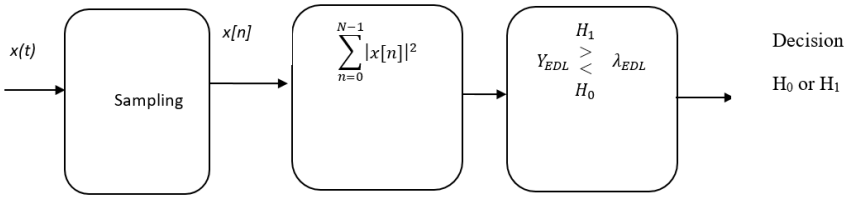


Fig 2. Energy Detector under Laplacian noise

3 System Model

The detection of the PU activity in the presence of noise by the secondary user (SU) is modeled as a binary hypothesis testing problem as given below:

$$\left\{ \begin{array}{c} H_0 : x[n] = v[n] \\ H_1 : x[n] = s[n] + v[n] \end{array} \right\} \tag{4}$$

where v[n]'s are independent and identically distributed (i.i.d.) noises, with non-Gaussian distribution of zero mean and variance σ_w^2 given as $L(0, \sigma_w^2)$ and s[n] is the signal transmitted by the PU.

3.1 Energy Detector under Laplacian Noise

The energy detector (ED) is the most commonly used method for PU detection as it has very low computational complexity. The energy detector under Laplacian noise (EDL) is discussed in [19]. The block diagram of energy detector under Laplacian noise is given in Figure 2.

The test statistic of energy detector under Laplacian noise is given as

$$Y_{EDL} = \sum_{n=1}^{N} |x[n]|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{EDL} \tag{5}$$

The energy measurement Y_{EDL} is calculated from N samples of the received signal at the CR receiver. The decision static Y_{EDL} is compared against a fixed threshold λ_{EDL}. The simple hypothesis testing problem is formulated in Eqn (5), where λ_{EDL} is the detection threshold which is set assuming constant false alarm rate. The sensing performance of the energy detector in Laplacian noise is measured by two metrics called as the detection probability (P_D) and false alarm probability (P_F) as formulated in [19].

$$P_F = Q\left(\frac{\lambda_{EDL} - N\sigma_w^2}{\sqrt{5N}\sigma_w^2}\right) \tag{6}$$

$$P_D = Q\left(\frac{\lambda_{EDL} - N(P + \sigma_w^2)}{\sqrt{5N\sigma_w^4 + 4NP\sigma_w^2)}}\right) \tag{7}$$

Where, P gives the power of the PU signal, N gives the number of samples, variance σ_w^2, λ_{EDL} is the detection threshold and Q(.) is the Gaussian complementary CDF.

3.2 Effect of Noise Uncertainty on Energy Detector under Laplacian Noise

In practical scenarios, it is not possible to accurately know the statistics of the noise v[n]. This creates the uncertainty in noise in the detection that may vary with time. The noise uncertainty is modelled such that the noise power is distributed in a single interval as given below.

$$\sigma_w^2 \, \varepsilon[\left(\frac{1}{\rho}\right)\sigma^2{}_n, \rho\sigma^2{}_n] \tag{8}$$

Where, $\sigma^2{}_n$ is the nominal noise power and ρ>1 indicates the uncertainty quantification.

The SNR wall for an energy detector under Laplacian noise is given in [19,20]. It is the SNR beneath which reliable signal detection cannot be achieved.

$$SNR_{wall}^{LED} = \frac{\rho^2 - 1}{\rho} \tag{9}$$

Considering the worst-case noise uncertainty Eqn (6) and (7) which gives false alarm and the probability of detection respectively are modified.

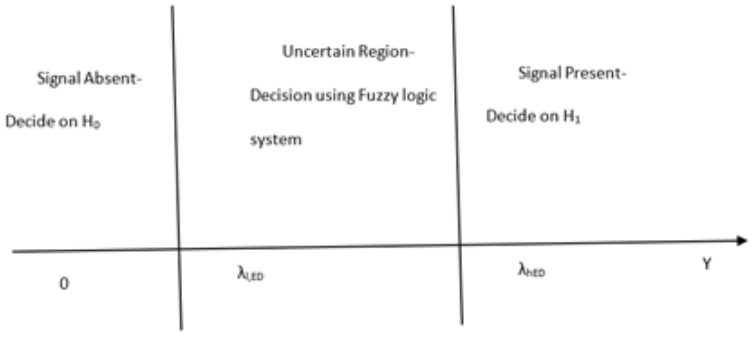


Fig 3. Two thresholds of the energy detector under Laplacian Noise

$$P'_F = Q\left(\frac{\lambda_{EDL} - N\rho\sigma_w^2}{\sqrt{5N}\rho\sigma_w^2}\right) \tag{10}$$

$$P'_D = Q\left(\frac{\lambda_{EDL} - N(P + \frac{(\sigma_w^2)}{\rho})}{\sqrt{5N\sigma_w^4(1/\rho) + 4NP(1/\rho)\sigma_w^2)}}\right) \tag{11}$$

The decision on PU activity by the CR device is usually made by a conventional ED using a single threshold. But for accurate detection it is necessary to fix an optimum threshold level, as the detection performance is dependent on the threshold level. The optimum threshold cannot be easily defined due to the problem of noise uncertainty. The situation worsens with low SNR, since the noise uncertainty leads to SNR Wall which is defined as the SNR below which the detector cannot provide reliable detection. In addition, the detection performance is dissatisfactory under the non-Gaussian noise in comparison to the Gaussian noise [19,20].

4 Proposed Method

To overcome the effect of noise uncertainty in CR systems several studies have been proposed as discussed in Section 1 of the paper. In this study a fuzzy thresholding scheme is investigated under non-Gaussian scenarios. This study is the first of its kind which uses a fuzzy thresholding approach for PU detection in a non-Gaussian environment. The two thresholds are therefore formulated from Eqn (6) as

$$\lambda_{l,EDL} = \left(\sqrt{5N}Q^{-1}(P_F) + N\right)\left(\frac{1}{\rho}\right)\sigma_w^2 \tag{12}$$

$$\lambda_{h,EDL} = \left(\sqrt{5N}Q^{-1}(P_F) + N\right)\rho\sigma_w^2 \tag{13}$$

Where, $\lambda_{l,EDL}$ is the lower threshold and $\lambda_{h,EDL}$ is the upper threshold. These thresholds are pictorially represented in Figure 3. Using these two thresholds the whole range of the observed test statistic Y_{EDL} is divided into three zones 1) Signal present zone -H_1 declared 2) uncertainty zone and 3) Signal absent zone -H_0 declared. If the test statistic exceeds the upper threshold $\lambda_{h,EDL}$ then the SU decides on H_1 which means the presence of primary user. While if the test statistic is less than $\lambda_{l,EDL}$ the SU decides on H_0. The lower and upper thresholds are selected in accordance with the lower and the upper limit of noise variance respectively.

The decision process using the two thresholding scheme is given below:

$$D = \left\{ \begin{array}{llll} 1 & \text{if} & Y > \lambda_{h,EDL} & H_1 \text{ is declared} \\ F & \text{if} & \lambda_{l,EDL} < Y < \lambda_{h,EDL} & \text{Fuzzy decision} \\ 0 & \text{if} & Y > \lambda_{l,EDL} & H_0 \text{ is declared} \end{array} \right\}$$

The fuzzy decision F is defined as

$$F = \left\{ \begin{array}{ll} 1 \text{ if } Z > \lambda_F, & H_1 \text{ is declared} \\ 0 \text{ if } Z < \lambda_F, & H_0 \text{ is declared} \end{array} \right\}$$

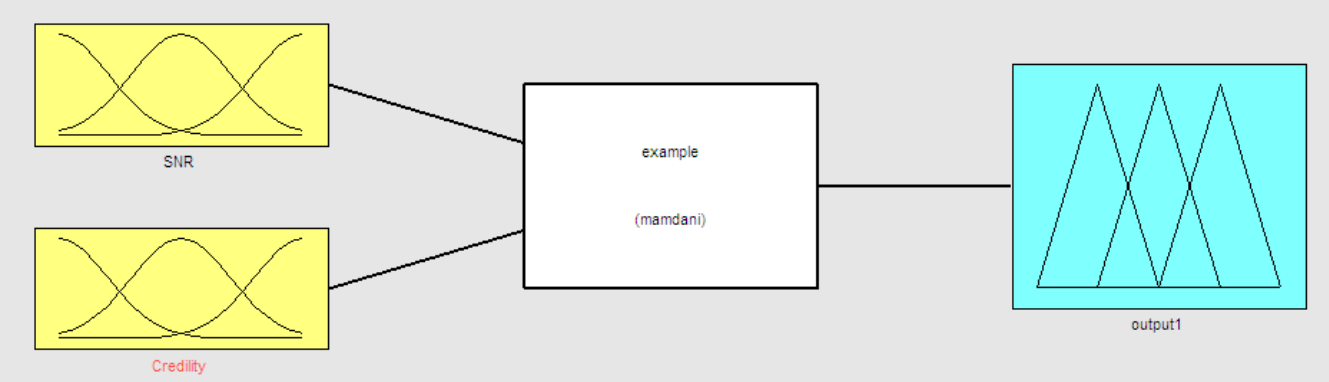


Fig 4. Membership functions in the input and output of the fuzzy inference system

Where, Z is the output of the fuzzy logic system, λ_F is the threshold set for taking fuzzy logic decision.
The final decision X=D+F

$$X = \left\{ \begin{array}{cc} D+F>0, & PU \text{ is present} \\ \text{else}, & \text{PU is absent} \end{array} \right\} \tag{16}$$

The decision in the confused region is taken using two antecedents called 1) Credibility 2) SNR available at SU. The output of the fuzzy logic system is the consequent which gives the possibility of the primary user activity.

The credibility C is defined as

$$C = \frac{(Y_{EDL} - \lambda_{l,EDL})}{(\lambda_{h,EDL} - \lambda_{l,EDL})} \tag{17}$$

If the value of C is high, then H_1 is favoured and if the value of C is low H_0 is favoured.

The fuzzy rule base and the membership for the antecedents and consequents are detailed in Figure 4 and Table 1 respectively.

Table 1. Rule base for decision on the PU activity

Antecedent 1 (Credibility)	Antecedent 2 (SNR)	Consequent (Likelihood of the presence of PU)
Low	Low	Very Low
Low	Medium	Low
Low	High	Medium
Medium	Low	Low
Medium	Medium	Medium
Medium	High	High
High	Low	Medium
High	Medium	High
High	High	Very High

5 Results and Discussion

This section details the results obtained through MATLAB simulations to analyze the performance of fuzzy two threshold scheme under non-Gaussian noise.

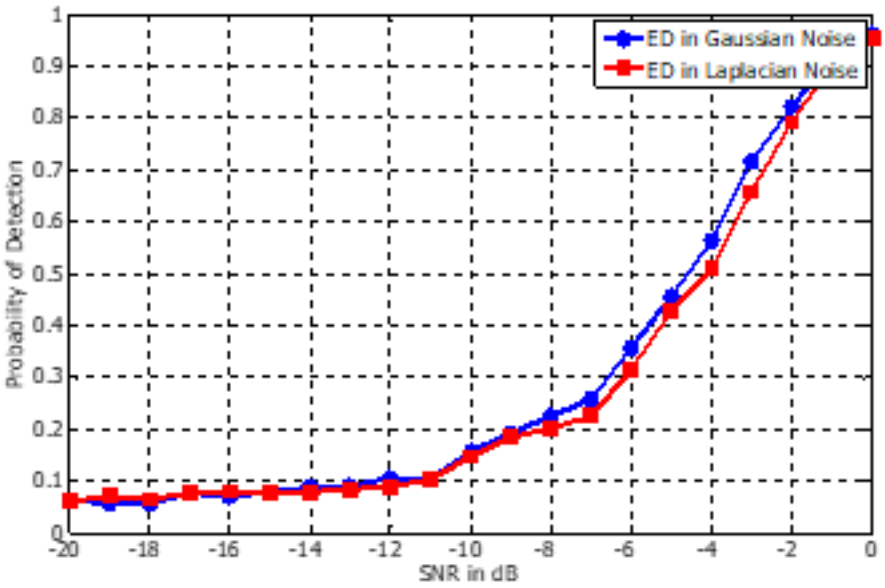


Fig 5. Performance of energy detector under Gaussian and non-Gaussian noise

The spectrum sensing is carried out in a CR environment using energy detection. The detection probability is used as a standard of measurement to determine the sensing accuracy. The primary transmitter signal is considered as a sinusoidal pilot signal of known frequency. The noise is modeled as non-Gaussian. In the simulations the Laplacian noise is considered, as it is a special case of non-Gaussian noise. It has pdf with excess kurtosis of 3 while the pdf of the Gaussian has the excess kurtosis as 0 [17]. The results provided are averaged over 10000 Monte Carlo simulations. Table 2 outlines the parameters used in the simulation study.

Figure 5 gives the performance of ED under the Gaussian and non-Gaussian noise. From the plot it is evident that the energy detector performs better under Gaussian noise than Laplacian noise. The energy detector provides optimal performance under Gaussian noise even when no prior information on the PU signal is available. Hence it is necessary to investigate methods to boost the energy detector's performance under Laplacian noise.

Table 2. Values used in Simulations

Parameters		Values
Target false alarm probability $\bar{P_F}$		0.05
Total numbers of samples N		100
SNR range of interest		-20 to -5dB
Excess Kurtosis	Gaussian	0
	Non-Gaussian	3

Figure 6 illustrates the impact of noise uncertainty on the energy detector's performance under Laplacian noise. Energy detection using single threshold in the detection process is sensitive to noise uncertainty. A marginal change in the noise uncertainty factor ρ deteriorates the performance quickly.

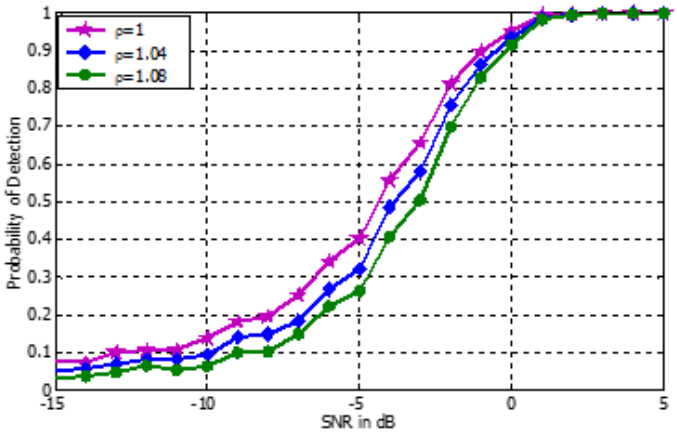


Fig 6. Impact of noise uncertainty on detection probability under non-Gaussian noise

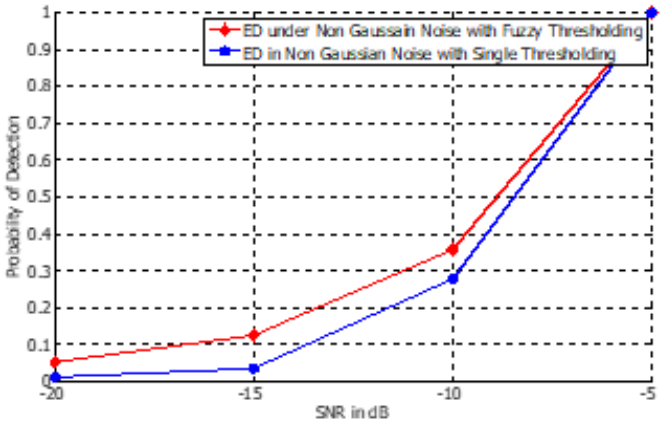


Fig 7. Effect of Energy detector under non-Gaussian Noise for ρ=1.023, P_F=0.01, N=1000 Samples

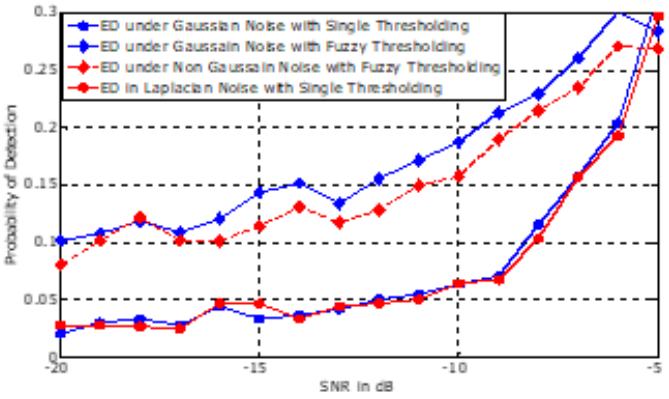


Fig 8. Thresholding methods to combat noise uncertainty in detection performance for ρ=1.088

The method proposed in[16] to address the problem of noise uncertainty are simulated and compared with the single threshold using the constant false alarm rate. From Figures 7 and 8 , it is observed that the PU detection using single threshold fails as the noise uncertainty increases, because the threshold fixed is independent of the noise variance.

Table 3. Results obtained using Single and Fuzzy thresholding schemes

SNR	PD using Single Thresholding under Gaussian Noise	PD using Single Thresholding under Non Gaussian Noise	PD using Fuzzy Thresholding under Gaussian Noise	PD using Fuzzy Thresholding under Non-Gaussian Noise
-15	0.04	0.05	0.15	0.12
-10	0.06	0.06	0.18	0.15

Realizing Eqn (10-17) in MATLAB and by appropriately choosing the membership functions and the fuzzy rule base in the fuzzy inference system the proposed adaptive thresholding using fuzzy logic method for ED under Laplacian noise, provides reliable detection compared to the other methods when the noise uncertainty increases. Another interesting observation drawn is on the improvement in signal detection in the low SNR regimes.

From Figure 7 it is observed that for the conditions in the existing literature on sensing under Laplacian noise[19], the fuzzy thresholding provides 20% increase in the detection, at false alarm rate 0.01,1000 samples and ρ of 0.1dB. Table 3 summarizes the value of probability of detection obtained in Gaussian and non-Gaussian environments using both the thresholding schemes. It is observed that the fuzzy thresholding method provides increased detection accuracy of around 20% compared to the single thresholding scheme.

Kurtosis refers to the size of the tails on a distribution. The tails of a distribution quantify the number of events which have appeared that are outside the standard range. If the distributions of event outcomes have lots of occurrences of outlier results, this causes heavy tails on the bell-shaped distribution curve. This is referred to as excess kurtosis. A kurtosis of 3 is termed as mesokurtic. Standard normal distributions are mesokurtic. If the kurtosis is >3 it can be visualized as a thin bell-shaped curve with a high peak termed as leptokurtic. One of the most well-known leptokurtic distributions is Student's t distribution, Laplace distribution. When kurtosis <3 is recognized as platykurtic with broad peak and thick tails. All uniform distributions are platykurtic.

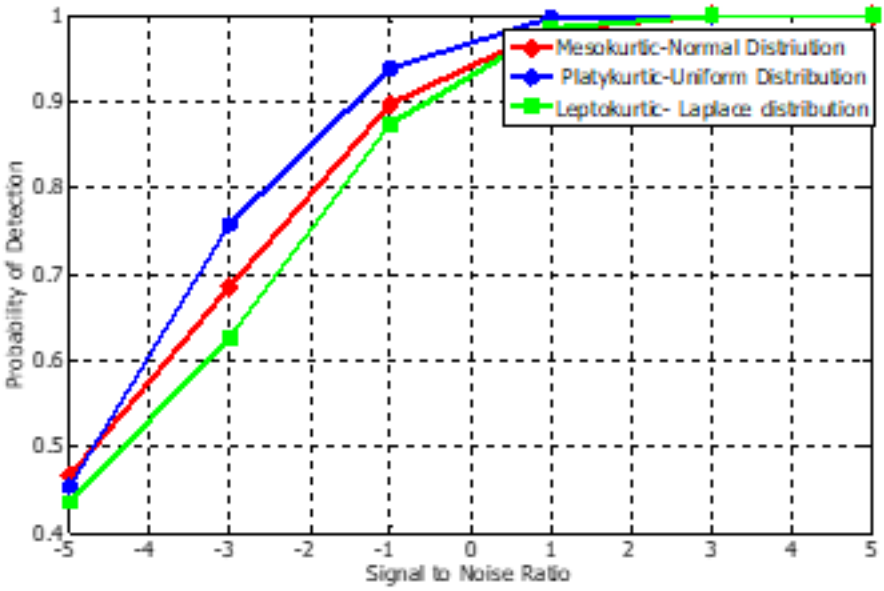


Fig 9. Effect of Kurtos is on primary user detection

The detection probability achieved for the mesokurtic, platykurtic and leptokurtic distributions are plotted in Figure 9. As the kurtosis increases, the detection probability decreases. This is because the platykurtic distribution produces fewer outliers and less extremes than the normal distribution.

In the leptokurtic distribution, like the Laplace distribution, the tails approach asymptotically to zero more slowly than a Gaussian. Hence the number of outliers are more than the normal distribution. Thus, the uniform distribution with fewer numbers of outlier's shows higher PU signal detection compared to normal and Laplace distributions.

6 Conclusion

In this paper, some typical impairments in PU detection, such as the effect of a non-Gaussian noise and noise uncertainty are investigated. Majority of spectrum sensing schemes are focused on primary user detection in Gaussian noise. But in practical wireless communication scenarios the noise usually has a heavy tailed nature. The Laplacian noise is an important non-Gaussian noise distributions and hence it is frequently used in engineering studies. This work focuses on primary user signal detection in Laplacian noise. Through Monte Carlo simulations it is observed that a non-Gaussian noise noticeably affects the performance of ED. Also a fractional change in noise uncertainty degrades the performance of energy detector. Hence to combat the above degradation in energy detector's performance the spectrum sensing algorithm using adaptive fuzzy thresholding which was primarily studied under Gaussian noise is extended under a non-Gaussian noise environment. Through simulations it is shown that by appropriately choosing the membership functions and the fuzzy rule base in the fuzzy inference system the proposed adaptive thresholding using fuzzy logic method for Laplacian energy detector provides reliable detection. The effect of the probability of detection due to the characterization of the skewness and kurtosis are also studied. Results show that as the kurtosis increases, the tails are heavier and it degrades the PU detection. The study limits to the use of single antenna in the sensing process. Hence the effect of using multiple antennas in the sensing process and in fading scenarios can be considered for future investigation in a non-Gaussian noise environment.

References

1) Kumar A, Thakur P, Pandit S, Singh G. Threshold selection and cooperation in fading environment of cognitive radio network: Consequences on spectrum sensing and throughput. *AEU - International Journal of Electronics and Communications*. 2020;117:153101. Available from: https://dx.doi.org/10.1016/j.aeue.2020.153101.
2) Sarala B, Devi SR, Sheela JJJ. Spectrum energy detection in cognitive radio networks based on a novel adaptive threshold energy detection method. *Computer Communications*. 2020;152:1–7. Available from: https://dx.doi.org/10.1016/j.comcom.2019.12.058.
3) Yu S, Liu J, Wang J, Ullah I. Adaptive Double-Threshold Cooperative Spectrum Sensing Algorithm Based on History Energy Detection. *Wireless Communications and Mobile Computing*. 2020;2020:1–12. Available from: https://dx.doi.org/10.1155/2020/4794136.
4) Mahendru G, Shukla AK, Banerjee P, Patnaik LM. Adaptive Double Threshold Based Spectrum Sensing to Overcome Sensing Failure in Presence of Noise Uncertainty. *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*. 2019;p. 466–471. Available from: https://doi.org/10.1109/SPIN.2019.8711570.
5) Salahdine F, Kaabouch N, Ghazi HE. Techniques for dealing with uncertainty in cognitive radio networks. *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*. 2017;p. 1–6. Available from: https://doi.org/10.1109/CCWC.2017.7868352.
6) Yang T, Wu Y, Li L, Xu W, Tan W. A Two-Step Cooperative Energy Detection Algorithm Robust to Noise Uncertainty. *Wireless Communications and Mobile Computing*. 2019;2019:1–10. Available from: https://dx.doi.org/10.1155/2019/3912784.
7) Reddy DN, Ravinder Y. Novel receiver diversity combining methods for spectrum sensing using meta-analytic approach based on p-values. *J Eng Sc Tech*. 2018;13(9):2883–2897.
8) Gao R, Qi P, Zhang Z. Frequency domain goodness of fit test based spectrum sensing method with dynamically varying noise. *China Communications*. 2020;17(12):172–179. Available from: https://dx.doi.org/10.23919/jcc.2020.12.012.
9) Ahuja B, Kaur G. Two-Stage Spectrum Sensing Using Fuzzy Logic for Cognitive Radio Networks. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*. 2020;90(3):515–525. Available from: https://dx.doi.org/10.1007/s40010-019-00595-7.
10) Bharatula S, Murthy BS. Innovative Fuzzy-Based Spectrum Sensing Technique for Noisy Conditions. In: Lecture Notes in Electrical Engineering. Springer Singapore. 2021;p. 735–750. doi:10.1007/978-981-15-9019-1_64.
11) Mabrook MM, Taha HA, Hussein AI. Cooperative spectrum sensing optimization based adaptive neuro-fuzzy inference system (ANFIS) in cognitive radio networks. *Journal of Ambient Intelligence and Humanized Computing*. 2020;5:1–2. Available from: https://dx.doi.org/10.1007/s12652-020-02121-9.
12) Sinha K, Trivedi YN. Spectrum Sensing Based on Dynamic Primary User with Additive Laplacian Noise in Cognitive Radio. In: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer International Publishing. 2021;p. 16–28. Available from: http://dx.doi.org/10.1007/978-3-030-73423-7_2.
13) Gao R, Jing F, Wang J, Zhang J, Zhang Y. A New Spectrum Sensing Method with Low SNR under Laplace Noise. *2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI)*. 2020;p. 291–297. Available from: https://doi.org/10.1109/IICSPI51290.2020.9332346.
14) Gao R, Qi P, Zhang Z. Performance analysis of spectrum sensing schemes based on energy detector in generalized Gaussian noise. *Signal Processing*. 2021;181:107893. Available from: https://dx.doi.org/10.1016/j.sigpro.2020.107893.
15) Tan F, Song X, Leung C, Cheng J. Collaborative Spectrum Sensing in a Cognitive Radio System with Laplacian Noise. *IEEE Communications Letters*. 2012;16(10):1691–1694. Available from: https://dx.doi.org/10.1109/lcomm.2012.080312.120517.
16) Ahuja B, Kaur G. Two-stage spectrum sensing using fuzzy logic for cognitive radio networks. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*. 2020;90:515–540.
17) Celikoglu A, Tirnakli U. Skewness and kurtosis analysis for non-Gaussian distributions. *Physica A: Statistical Mechanics and its Applications*. 2018;499:325–334. Available from: https://dx.doi.org/10.1016/j.physa.2018.02.035.
18) Aryal GR. Study of laplace and related probability distributions and their applications. . Available from: https://scholarcommons.usf.edu/etd/2443/.

19) Gao R, Li Z, Li H, Ai B. Absolute Value Cumulating Based Spectrum Sensing with Laplacian Noise in Cognitive Radio Networks. *Wireless Personal Communications*. 2015;83(2):1387–1404. Available from: https://dx.doi.org/10.1007/s11277-015-2457-4.
20) Tandra R, Sahai A. SNR Walls for Signal Detection. *IEEE Journal of Selected Topics in Signal Processing*. 2008;2(1):4–17. Available from: https://dx.doi.org/10.1109/jstsp.2007.914879.

PLACEMENT MANAGEMENT SYSTEM

Akash P. Galagali
Dept. of Electronics and Communication
BMS Institute of Technology & Management
(Affiliated to VTU)
Bengaluru, India
Email: akashgalagali4@gmail.com

Chethan
Dept. of Electronics and Communication
BMS Institute of Technology & Management
(Affiliated to VTU)
Bengaluru, India
Email: chethanjkulkarni@gmail.com

Mrs. Mamatha K.R
(Assistant professor)
Dept. of Electronics and Communication
BMS Institute of Technology & Management
(Affiliated to VTU)
Bengaluru, India
Email: mamathakr@bmsit.in

Debanshu Biswas
Dept. of Electronics and Communication
BMS Institute of Technology & Management
(Affiliated to VTU)
Bengaluru, India
Email: debanshubiswas111@gmail.com

Siddesh S
Dept. of Electronics and Communication
BMS Institute of Technology & Management
(Affiliated to VTU)
Bengaluru, India
Email: 1by17ec164@bmsit.in

***Abstract*— As we are moving forward in life, we are facing challenges everyday and till date we are trying to have a better solution to it. With the increasing number of human population, the crave for jobs and recruitment is also increasing in heavy demand day by day and this won't stop ever. Rather it would get worse. So, looking into this factor the placement offices are facing problems to conduct the recruitment process smoothly. They are looking forward to some automated and online process so that the officers can take a break and cut down their stress of work duty. In the view of this we bring you "Placement Management System."**

***Keywords*— ◎ Django ◎ HTML ◎ CSS ◎ JS**

◎ GIT ◎ Docker ◎ Bash Scripting ◎ Nginx

I. INTRODUCTION

The project "PLACEMENT MANAGEMENT SYSTEM" is a web-based application developed for the placement department of the college in order to provide the details of its students in a database for the companies to their process of recruitment. The main purpose of the System is to avoid manual process because the manual work makes the process slow and other problems such as inconsistency & ambiguity on operations. In order to avoid this web-based placement managed system is proposed, where the student information in the college with regard to placement is managed efficiently. It intends to help fast in fast access procedures in placement related activities and ensures to maintain the details of the student. Students logging should be able to upload their personal and educational information. The key feature of this project is that it is onetime registration enabled. The placement cell calls the companies to select their students for jobs via the campus interview. The placement cell allows the companies to view the student resumes in selective manner. They can filter the student profile as per their requirement. The job details of the placed students will be provided by the administrator. The administrator plays an important role in our project. Our project provides the facility of maintaining the details of the students and gets the requested list of candidates for the company who would like to recruit the students based on given criteria.

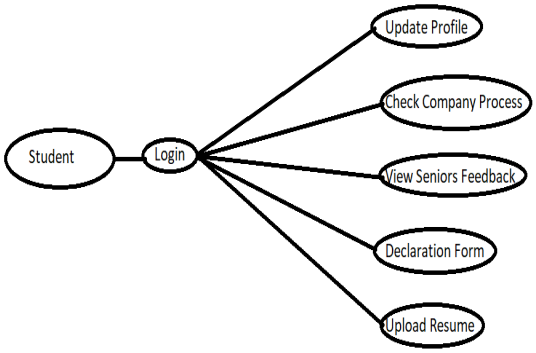


Figure 1: Student Registration Process

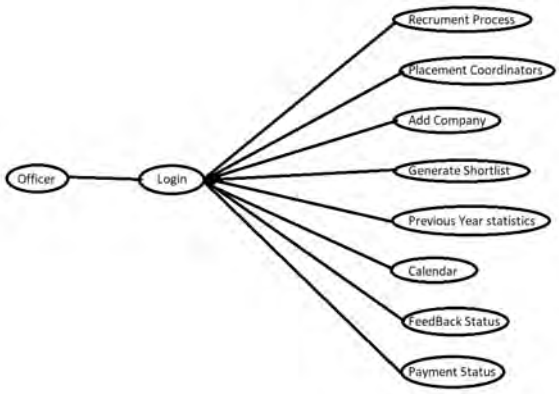


Figure 2: Working at the Admin Panel

II. PROBLEM ANALYSIS

Nowadays campus placements are conducted in all colleges. Various software and other sector companies are conducting campus selections for selecting merit candidates. When campus selections are conducted the students should provide their curriculum vitae to the concern officer for attending the campus interviews. This routine process is maintained manually, like maintenance of their resumes in papers. This can be automated by designing software.

The old manual system was suffering from many drawbacks. The PMS System is to maintain the data and make the process smooth, maintaining the data and retrieving information is very tedious and lengthy. there used to be lots of difficulties in associating any particular transaction with particular context. It is very difficult to find errors while entering the records. Once the records were entered it was very difficult to update records.

Main reason behind it is a lot of information to be maintained and managed. For this reason we provide features in PMS that can potentially automate the whole system and reduce the errors and manual work.

In Colleges, Training and Placement Officers(TPO) have to manage the student profile and data for training and placement process manually. Placement officers have to collect the information from companies and notify the students who are clearing the company criteria and this manual process is tedious and erroneous. Any modification in the details of a student or updates required in the profile of the student has to be searched and get done manually.

III. PROPOSED SYSTEM

Proposed system is inexpensive and no manual workload is needed. Everything, however smaller the details for the registration process are, they are going to get into the system through this Online Placement Management System.

The software used in this regards are as listed below:

- Django
- HTML
- CSS
- JS
- GIT
- Docker
- Bash Scripting
- Nginx

All these softwares are easily available in the market. If need be any college can develop it for the better conduction of placement process.

Below given are the basic functionalities of the softwares used in this project:

- Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design.

- Hypertext Markup Language (HTML) is the standard markup language for documents designed to be displayed in a web browser.

- Cascading Style Sheets(CSS) is designed to enable the separation of presentation and content, including layout, colors, and fonts.

- JavaScript(JS) conforms to the ECMAScript specification and is one of the core technologies of the World Wide Web.

- GIT is a distributed version-control system for tracking changes in any set of files, originally designed for coordinating work among programmers cooperating on source code during software development.

- Docker is a set of platform as a service (PaaS) products that use OS-level virtualization to deliver software in packages called containers.

- Bash Scripting helps in executing a shell command, running multiple commands together, customizing administrative tasks, performing task automation etc.

- Nginx is used as a web server designed for maximum performance and stability.

IV. FLOW OF THE PROCESS

In this project we will be using Django and our main framework backing up most of the authentication, communication with the frontend and the database, and effectively handling traffic, sending emails, sending SMS and much more. Front end will be handled by HTML, CSS and JS. JavaScript will be handling all the robust frontend actions like making XHR requests, visibility of elements, interactions with the users, etc. MySQL will be the database holding all the information about the students and PTO.

EXISTING SYSTEM	PROPOSED SYSTEM
1. Manual SGPA CGPA calculation	1. Automatic CGPA SGPA calculation
2. Manual shortlisting	2. Can generate Shortlist with Criteria
3.Notifying students individually	3. Notifying students on one click
4. No option for students to update Resume	4. Can Update resume at any time
5. Collect Feedback Manually	5. Can write Feedback on site
6. Manual Reminding about Training and Placement Fees	6. Automatic Reminding about Training and Placement Fees

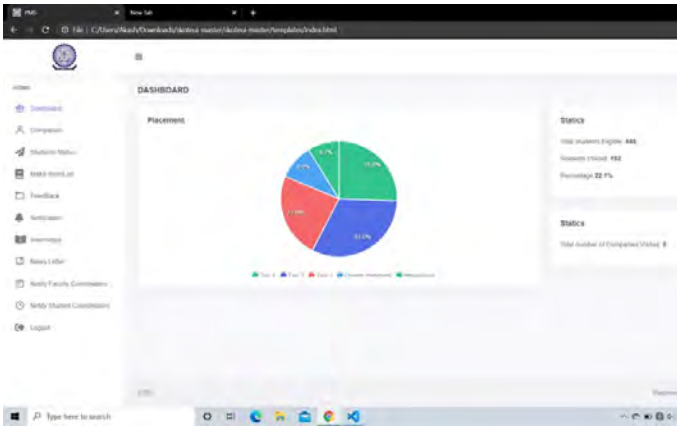


Figure 4: Allover Statistical Display of Registrations

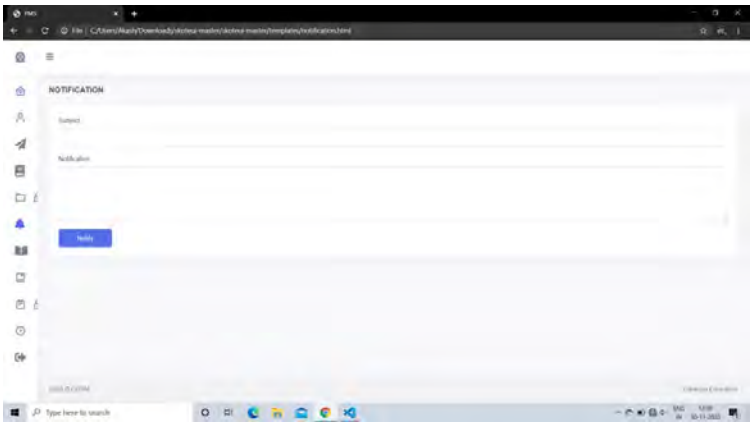


Figure 5: Notification box

Applications: -

1. It can be used in colleges to ease the way of the placement campus drive.
2. On a large scale it will help in the recruitment process more efficiently.

Advantages: -

1. Easy to Install
2. No Manual workload
3. Easy to use
4. More accuracy

V. RESULTS

This paper depicts how the system will work and how smooth the conduction process is going to take place from the very beginning of registration till the recruitment of the individuals and how stress-free the process is this for the placement officers.

VI. CONCLUSION

According to various surveys done, this turns out really well for both students as well as the officers and the companies' HRs too. We hope to look forward to implementation of this system not only in our college but in all the colleges of this region as well as the country. This way the process of recruitment is going to be hassle-free.The main advantage of this paper is that it suggests the solution to a very existing

problem at a very low cost. The accuracy of the system is no doubt of high rate for good possible outputs.

VII. ACKNOWLEDGMENT

The authors of this paper would like to thank the management of BMS Institute Of Technology for providing support to execute this project.

VIII. REFERENCES

- https://www.slideshare.net/
- https://acadpubl.eu/
- http://ijcsit.com/
- http://data.conferenceworld.in/

The peak-to-average power ratio reduction using hybrid scheme with companding and discrete Hartley transform for orthogonal frequency division multiplexing system

N. Rashmi*

Department of Electronics and Communication Engineering,
BMS Institute of Technology and Management,
Doddaballapur Main Road, Avalahalli,
Yelahanka, Bengaluru-560064, India
Email: rashmiswamy@bmsit.in
*Corresponding author

Mrinal Sarvagya

Department of Electronics and Communication Engineering,
REVA University, India
Email: mrinalsarvagya@reva.edu.in

Abstract: Orthogonal frequency division multiplexing is a multicarrier and high data rate system. Increasing data rate through such modulation techniques increases amplitude variation with large dynamic range usually referred as peak-to-average power ratio (PAPR) in OFDM systems. As a consequence, power amplifier need to operate in linear region to avoid nonlinear distortion. In this article, to combat the nonlinearity effect in communication systems we propose a novel hybrid scheme, combined with μ-law companding and discrete Hartley transform (DHT) spread. We present the detailed simulation results for PAPR reduction for OFDM and superposition coded modulation-OFDM. It is observed that SCM-OFDM performance is much superior to OFDM. In this article we proposed a scheme for SCM-OFDM system that reduces PAPR to 1.7949 dB. The article also compares the obtained results with clipping technique, partial transmit sequence, selective mapping. The computational complexity of the system using SCM-OFDM is reduced by 50% by applying hybrid scheme.

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 commulative distribution function; CCDF.

Reference to this paper should be made as follows: Rashmi, N. and Sarvagya, M. (2022) 'The peak-to-average power ratio reduction using hybrid scheme with companding and discrete Hartley transform for orthogonal frequency division multiplexing system', *Int. J. Intelligent Enterprise*, Vol. 9, No. 3, pp.303–317.

Biographical notes: N. Rashmi is working as an Assistant Professor in the Department of Electronics and Communication Engineering at the BMS Institute of Technology and Management, Doddaballapur Main Road, Avalahalli, Yelahanka, Bengaluru-560064, India.

Mrinal Sarvagya is a Professor in the School of ECE at the REVA University. Her research interests are wireless communication systems, communication networks, communication theory and systems.

1 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 ratio (PAPR) and carrier frequency offset (CFO). The complexity of digital-to-analogue 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 categorised 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) (Vittal and Naidu, 2017; Liang et al., 2014; El-Hamed et al., 2018; Hassan, 2016). In Satyavathi and Rao (2019), a low-complexity PAPR reduction with modified linear SM scheme was suggested for the OFDM system. Lee and Chung (2018) proposed a scheme to reduce hardware complexity by analysing loglikelihood ratio for channel coding there by reducing PAPR. The performance for

modified linear SM scheme for PAPR reduction in Satyavathi and Rao (2019) was insignificant compared with conventional schemes. Additive schemes include tone reservation (TR) and clipping and filtering techniques. Nagaraju et al. (2018) analysed the SM and PTS technique for PAPR reduction and performance analysis of modified SM technique based on BER and complexity of the system for PAPR were reported. TR technique with low-complexity tones with null subcarriers was proposed in Wang et al. (2016) to decrease the PAPR in the filter bank multi carrier systems combined with OFDM system. In Azim et al. (2017), decision directed method proposed for PAPR reduction for optical OFDM. In Nadal et al. (2014), clipping and quantisation 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 Chen and Liang (2007) implemented Reed Muller codes are implemented with PTS scheme. All these schemes reduces 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 discrete 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 organised as follows. In Section 2, the SCM-OFDM system is discussed, and the PAPR and cumulative complementary distribution function (CCDF) are formulated. In Section 3, the clipping technique for PAPR reduction for SCM-OFDM system, PTS and SM techniques were discussed. Section 4 is the simulation results were discussed and Section 5 is the conclusion.

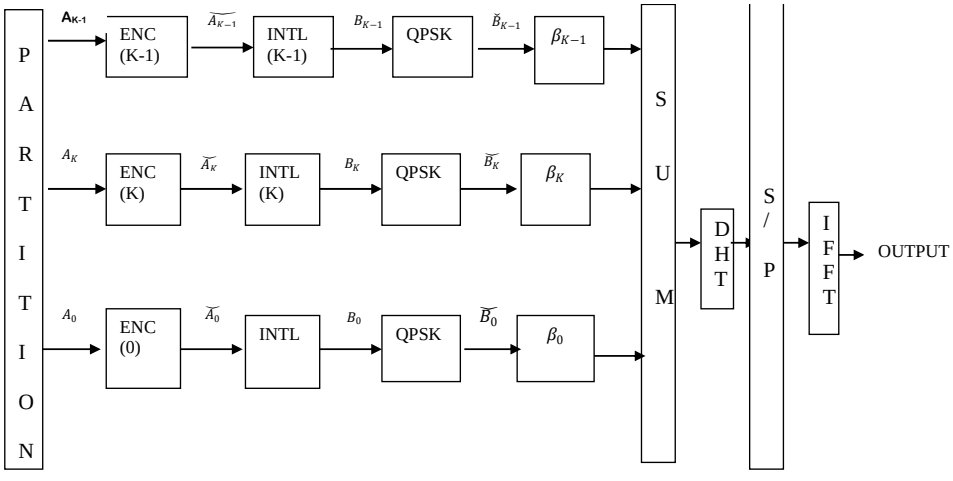
2 SCM-OFDM system

2.1 Characterisation 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 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. SCM 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.

Figure 1 The proposed SCM-OFDM system for PAPR reduction



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 $\{P_k\}$. Rashmi and Sarvagya (2016) proposed a transmitter for SCM-OFDM system. 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 $x_k(j)$.

$$x_k^K = x_{re,j}^K + ix_{im,j}^K \tag{1}$$

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:

$$x(j) = \sum_{k=0}^{K-1} \rho_k x_k(j),\ j = 0, 1, \ldots, J-1 \tag{2}$$

where $\{\rho_k\}$ are constant weighting colour k. In this study, $\{\rho_k\}$ was obtained using the simulation-based power allocation method projected in Chen and Liang (2007).

The superimposed symbols are fed to a customary OFDM modulator unit that consists of an N-point inverse 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 (ISI). 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

$$\begin{aligned} &S_i(n) = S_i(n+N) \text{ for } -N_g \le n \le N-1 \\ &S_i(k) = \left\{ \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_i(k) e^{j2\pi nk/N} \right\}, \end{aligned} \tag{3}$$

where $X_i(k) = x(j)$ from equation (2) is the superimposed modulated data symbol assigned to the sub carriers and N_g represents the length of CP ($-N_g < n < N-1$).

2.2 Peak-to-average power ratio

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:

$$S_i(k) = \left\{ \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_j(k)^{ej2\pi nk/LN} \right\}, 0 < n < LN \tag{4}$$

The PAPR is termed as the ratio of the maximum instantaneous power to the average power and is formulated mathematically as follows:

$$PAPR[S[n]] = \frac{\max\left[|s[K]|^2\right] 0 \le n \le LN-1}{E\left[|s[K]|^2\right]} \tag{5}$$

where $E[.]$ specifies the expectation operator.

2.3 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:

$$CCDF = \Pr\left(PAPR_{SCM-OFMD} > PAPR_o\right), \tag{6}$$

where $PAPR_{SCM-OFDM}$ and $PAPR_o$ are the output of the SCM-OFDM system and the threshold value respectively.

3 PAPR reduction techniques

3.1 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_k's are equal. To reduce the PAPR, clipping of X_j to form $\bar{X}_j$ is followed based on the rule presented below:

$$\bar{X}_j = \begin{cases} s_j, & |s_j| \le A \\ \dfrac{As_j}{|s_j|}, & |s_j| > A \end{cases} \tag{7}$$

where $A > 0$ is the clipping threshold.

From Liang (2015), we outline the clipping ratio as follows:

$$\gamma = A^2 \Big/ E\left(\left|X_j^2\right|\right) \tag{8}$$

The PAPR of the transmitted signal with clipping is given as follows

$$PAPR = A^2 \Big/ E\left(\left|\bar{X}_j^2\right|\right).$$

The received signal can be written as follows:

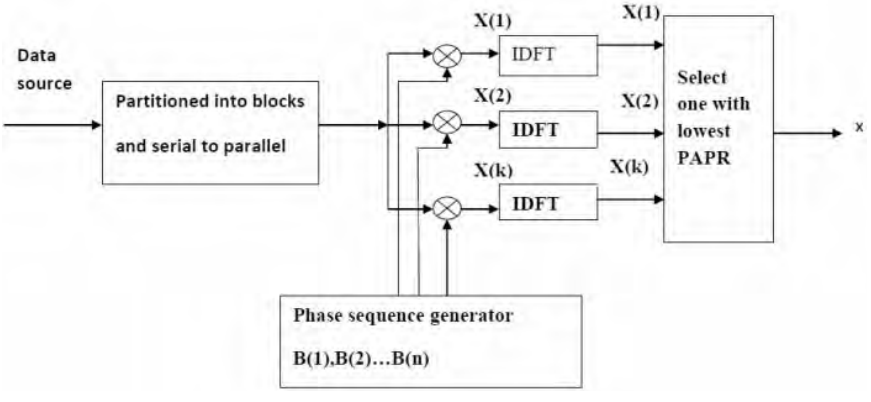
$$Y_j = X_j + w_j, \tag{9}$$

where w_j is the complex Gaussian noise with a zero mean and σ^2 variance.

3.2 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 DFT 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 Figure 2. This technique also requires side information. The side information is an extra bit that carries the phase sequences to receiver increasing the bandwidth. SM technique is not only spectral inefficient but also complex system.

Figure 2 Block diagram of SM



3.3 Partial transmit sequence

PTS 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

$$X^{(u)} = \left[X_1^{(u)}, X_2^{(u)}, \ldots, X_{M-1}^{(u)}\right] \tag{10}$$

$$X = \sum_{u=1}^{U} X^u \tag{11}$$

The phase rotation factors are given by

$$C_u = e^{j\theta u},\ u = 1, 2, \ldots, U \tag{12}$$

4 Proposed MMC-DHT PAPR reduction technique for SCM-OFDM system

The objective of this investigation is to realise 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

$$F_x(\cdot) = 1 - \exp\left(\frac{x_0^2}{\sigma_x^2}\right),\ x_o > 0 \tag{13}$$

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 Ghahremani and Shayesteh (2014)

$$F_{EC}\left(x(n)\right) = \text{sgn}(x) \cdot \sqrt{3\sigma_x^2} \cdot erf\left(\frac{|x|}{\sqrt{2\sigma_x^2}}\right),\ 0 \le x \le 1 \tag{14}$$

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

$$\begin{aligned} y_c(n) &= F\left(x_j(n)\right) \\ &= \text{sgn}\left(x_j(n)\right) \frac{A_s \times \log\left(1 + \mu\left|\frac{x_j(n)}{A_s}\right|\right)}{\log(1+\mu)} \end{aligned} \tag{15}$$

where $\text{sgn}(x(n)) = \dfrac{x(n)}{|x(n)|}$ 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 Ghahremani and Shayesteh (2014), 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.,

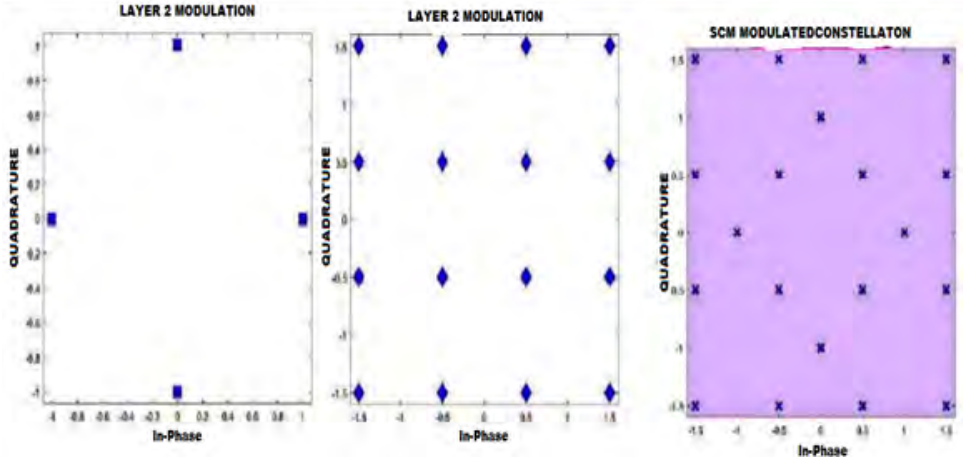
$$K = \frac{Peak\ amplitude\ of\ compressor}{Peak\ of\ actual\ signal} = \frac{A}{x_{jpeak}} \tag{16}$$

The transfer characteristics of the modified μ compander including new parameter peak ratio '*K*' is expressed by substituting the above equation in equation (15).

$$y = K \times x_{jpeak} \times \frac{\log\left(1+\mu\frac{|x_{jpeak}|}{K \times x_{jpeak}}\right)}{\log(1+\mu)}\text{sgn}(x) \tag{17}$$

The above companding profile permits the *μ-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.

Figure 3 Simulation result of SCM of two signals QPSK and 16 QAM modulated signal (see online version for colours)



OFDM signal is precoded using DHT precoder before IFFT block. The Hartley transform of $S_i(n)$ is expressed as

$$H_s(\omega) = \int_{-\infty}^{\infty} f(t)Cas(\omega t)dt. \tag{18}$$

$$Cas(\omega t) = \cos(\omega t) + \sin(\omega t) \tag{19}$$

$$f(t) = \int_{-\infty}^{\infty} H_g(\omega)Cas(\omega t)d\omega \tag{20}$$

To reduce the number of functional computation $Cas(\omega t)$ can be expressed as:

$$\sqrt{2}\cos\left(\omega t - \frac{\pi}{4}\right) = \sqrt{2}\sin\left(\omega t + \frac{\pi}{4}\right) \tag{21}$$

For discrete signal Hartley transform is expressed as:

$$H_{s(k)} = \frac{1}{\sqrt{2}} \sum_{n=0}^{NL-1} y_i(n) Cas\left(\frac{2\pi nk}{N}\right) \tag{22}$$

For discrete signal Hartley transform is expressed as:

$$H_{s(k)} = \frac{\sqrt{2}}{\sqrt{N}} \sum_{n=0}^{NL-1} y_i(n) \cos\left(\frac{2\pi nk}{N} \frac{\pi}{-4}\right) \tag{23}$$

$k = 0, 1, \ldots, N-1$.

The DHT 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:

$$Y_j(n) = \left\{ \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} H_s(k) e^{\frac{j2\pi nk}{LN}} \right\} + w(n) \tag{24}$$

$0 < n < LN - 1$.

5 Simulation results

In this study, computer simulations were performed to evaluate the performance of the SCM-OFDM system. CP is added to the time domain signal to avoid the ISI. Besides PAPR drawback in OFDM system, OFDM is additionally sensitive to spectral null problem over a frequency selective fading channel. By using DHT 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 1/2. 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, layer 1 adopted QPSK and for layer 2 16 QAM and resultant superimposed signal is shown in Figure 3. For the higher modulation scheme when SCM is simulated with 64 QAM PAPR obtained is 6.4089 dB, OFDM with 64 QAM is 9.31 dB. Compared to OFDM PAPR in SCM improved by 31.2%. However, PAPR in the hybrid system SCM-OFDM system is 13.53 dB. 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 2.87 dB. Even PAPR reduction using clipping and filtering technique PAPR is reduced to 2.87 dB, but BER performance of the same reduction scheme is poor as shown in Figure 4 compared to BER performance of the proposed method. PAPR reduction using DHT for SCM-OFDM signal is 5.8971 dB, thereby reducing PAPR by 4.7565 dB compared to clipping and without reduction technique by 7.6329 dB respectively, thereby improving performance of PAPR by 56.41%. With proposed 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, with 256 QAM PAPR in OFDM is as high as 30 dB 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 12.5 dB. Figure 3 shows the simulation of SCM symbols, considering $K = 2$ layers the number of symbols in SCM is 2. Figure 4 show that the PAPR reduction in SCM-OFDM system with DHT precoder and clipping as a reduction PAPR technique. DHT precoder outperforms clipping technique by 7.01 dB. BER performance is better compared to clipping technique for SCM-OFDM system. Figure 5 present that when modified μ-law compander with different value of μ, with $\mu = 30$ 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 Figure 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. Figure 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 out performance with the other PAPR reduction techniques is shown in Figure 7. 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. Figure 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 12.5 dB 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 1 Simulation parameters

Bandwidth	20 MHz
Carrier frequency	2.5 GHz
Number of subcarriers	256
Subcarrier spacing	15 KHz
Number of cyclic prefix	64
Convolution encoder rate	1/2
Interleaver	Random interleaver
Modulation for layer 1	256 QAM
Modulation for layer 2	256 QAM

Table 2 Comparison of DHT and DFT for computational complexity

Schemes	*Real multiplications*	*Complex additions*
DHT	$2L^2$	$L(L-1)$
DFT	$4L^2$	$L(L-1)$

Notes: Where L is order of DHT matrix and DFT matrix.

Table 3 Comparison of DHT for spectral efficiency

Scheme	*Sideband information (bits/OFDM symbol)*
DHT	None
PTS	S $\log_2 \varnothing$
SLM	$\log_2 \varnothing$

Notes: Where S is the number of subblocks and ∅: is the number of phase factors.

Figure 4 BER performance of clipping and DHT PAPR reduction techniques in SCM-OFDM system (see online version for colours)

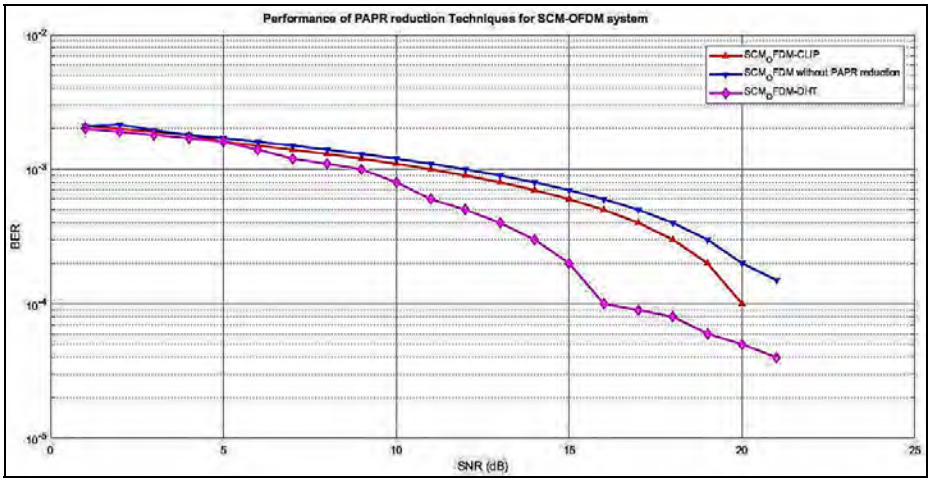


Figure 5 CCDF vs. PAPR for different μ companding profiles (u_1, u_2) and modified μ companding techniques (see online version for colours)

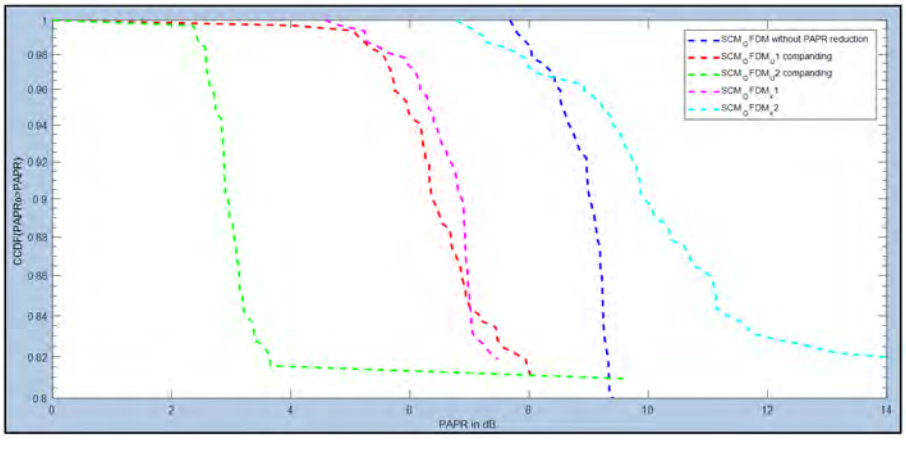


Table 4 Comparison of PAPR reduction techniques for SCM-OFDM system

PAPR reduction technique	*SCM*	*OFDM*	*SCM-OFDM without PAPR reduction*	*SCM-OFDM with clipping*
PAPR (in dB)	6.4089	9.31	13.53	10.6536
PAPR reduction technique	*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)	5.8971	3.64	1.7949	12.5

Figure 6 BER performance of modified μ-law companding technique on SCM-OFDM system with different companding levels U_1, U_2 and U_3 (see online version for colours)

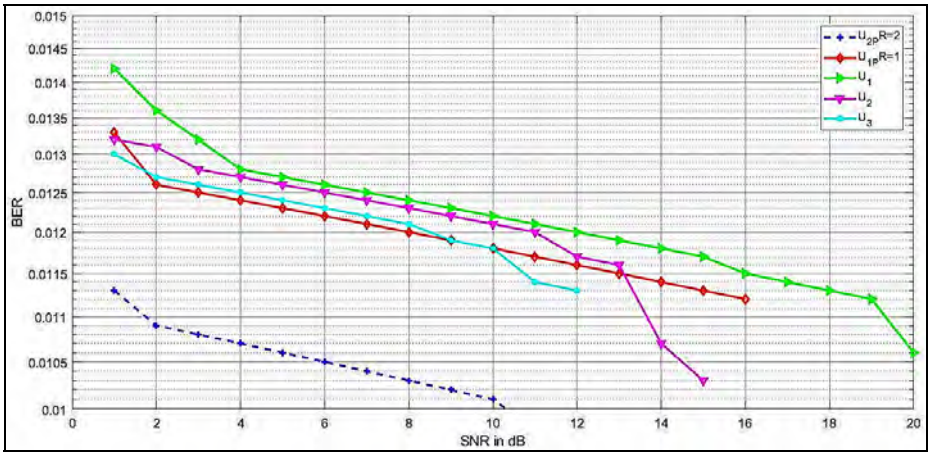


Figure 7 CCDF vs. PAPR for different PAPR reduction techniques on SCM-OFDM system (see online version for colours)

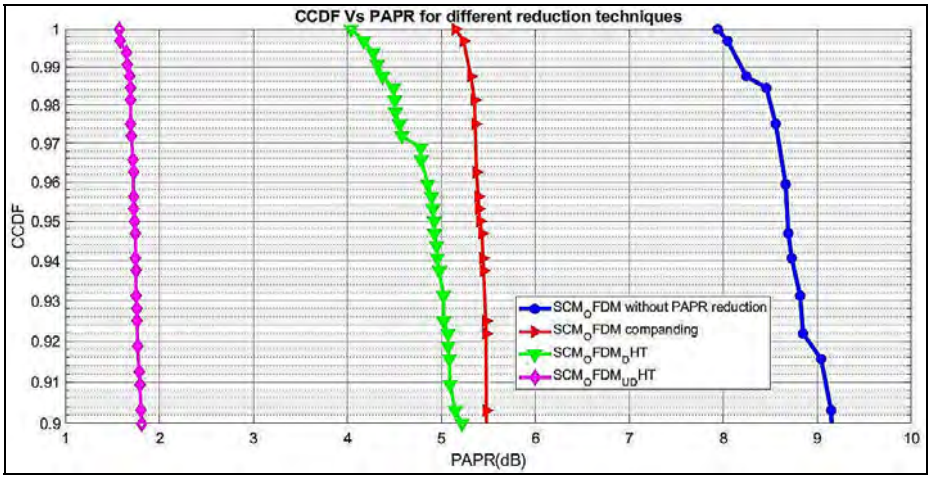


Figure 8 BER performance of different PAPR reduction technique on SCM-OFDM system (see online version for colours)

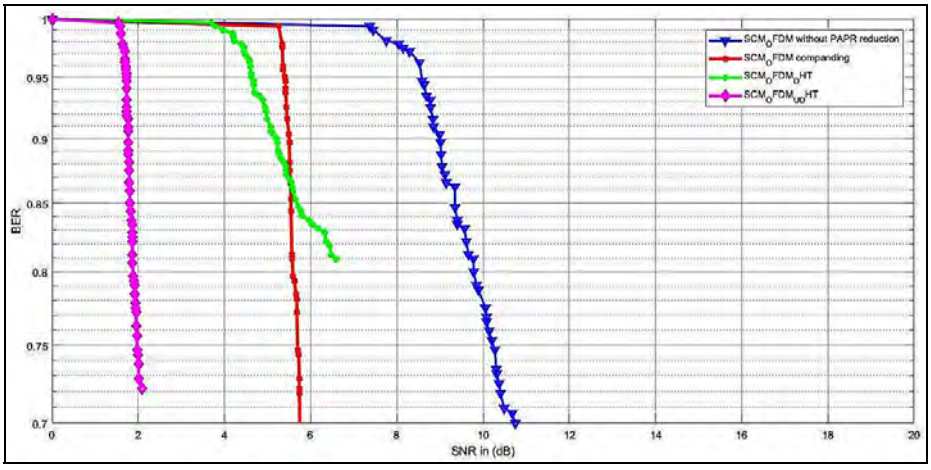
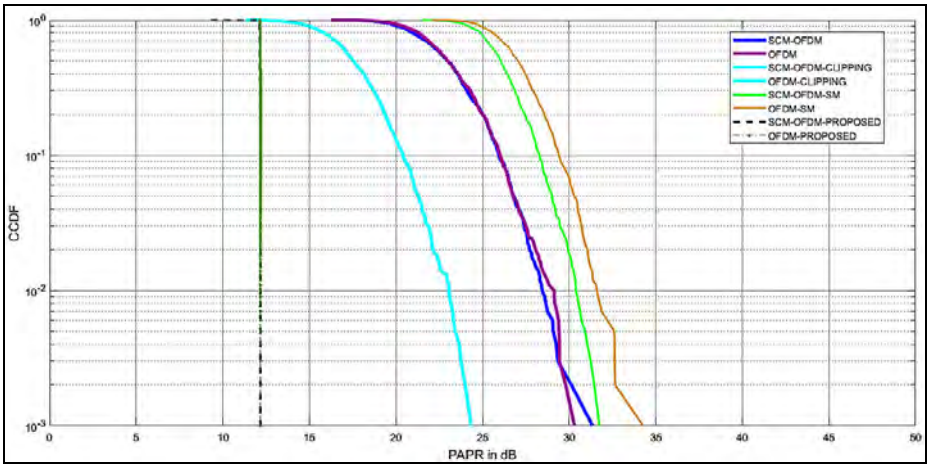


Figure 9 Comparative study of PAPR on SCM-OFDM systems for 256 QAM modulation with different reduction techniques (see online version for colours)



The SCM OFDM system is simulated for 64 QAM and 256 QAM respectively, PAPR for SCM signal and OFDM signal PAPR are 6.4089 dB and 9.31 dB respectively. With the combination of SCM-OFDM system PAPR increased to 13.53 dB due to Gaussian like signal of SCM-OFDM signal and with equal power allotted for two layers of SCM. Clipping technique reduces PAPR by 2.87 dB, where as DHT achieves PAPR reduction by 7.63 dB. PAPR is further reduced by 9.89 dB by μ companding technique. As observed from the simulation results DHT technique provides better BER performance, μ companding technique provides better PAPR performance. Combining the advantages of both the technique we achieved reduction in PAPR by 11.73 dB and improvement in PAPR by 86.7%. SCM-OFDM system is simulated with 256 QAM to demonstrate the proposed method works better compared to well known PAPR technique such as PTS and SM. Figure 9 shows that PAPR in SCM-OFDM without PAPR reduction is 33 dB,

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 12.5 dB, PAPR performance is improved by 62.1%.

6 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 64 QAM SCM-OFDM system, 12.5 dB with 256 QAM 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.

References

Azim, A.W., Le Guennec, Y. and Maury, G. (2017) 'Decision-directed iterative methods for PAPR reduction in optical wireless OFDM systems', *Optics Communications*, Vol. 389, pp.318–330.

Chen, H. and Liang, H. (2007) 'PAPR reduction of OFDM signals using partial transmit sequences and Reed-Muller codes', *IEEE Communications Letters*, Vol. 11, No. 6, pp.528–530.

El-Hamed, M.A., Zekry, A., Elagooz, S.S. and El-Samie, F.E.A. (2018) 'Blind selective mapping for single-carrier frequency division multiple access system', *Digital Signal Processing*, Vol. 75, pp.25–37.

Ghahremani, R. and Shayesteh, M.G. (2014) '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)*, IEEE, May, pp.1483–1487.

Hassan, E. (2016) 'PAPR reduction using selective mapping scheme', in *Multi-Carrier Communication Systems with Examples in MATLAB®*, pp.82–109, CRC Press.

Lee, J.H. and Chung, H. (2018) 'Exact and approximate log-likelihood ratio of M-ary QAM with two-time dimensions', *ICT Express*, Vol. 5, No. 3, pp.173–177.

Liang, H.Y. (2015) 'Integrating CE and modified SLM to reduce the PAPR of OFDM systems', *Wireless Personal Communications*, Vol. 80, No. 2, pp.709–722.

Liang, H.Y., Chu, H.C., Lin, C.B. and Lin, K.H. (2014) 'A partial transmit sequence technique with error correction capability and low computation', *International Journal of Communication Systems*, Vol. 27, No. 12, pp.4014–4027.

Nadal, L., Moreolo, M.S., Fàbrega, J.M. and Junyent, G. (2014) 'Low complexity PAPR reduction techniques for clipping and quantization noise mitigation in direct-detection O-OFDM systems', *Optical Fiber Technology*, Vol. 20, No. 3, pp.208–216.

Nagaraju, C.H., Sharma, A.K. and Subramanyam, M.V. (2018) 'Reduction of PAPR in MIMO-OFDM using adaptive SLM and PTS technique', *International Journal of Pure and Applied Mathematics*, Vol. 118, No. 17, pp.355–373.

Rashmi, N. and 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*, Vol. 8, No. 2, p.1017.

Satyavathi, K. and 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.

Vittal, M.V.R. and Naidu, K.R. (2017) 'A novel reduced complexity optimized PTS technique for PAPR reduction in wireless OFDM systems', *Egyptian Informatics Journal*, Vol. 18, No. 2, pp.123–131.

Wang, H., Wang, X., Xu, L. and Du, W. (2016) 'Hybrid PAPR reduction scheme for FBMC/OQAM systems based on multi data block PTS and TR methods', *IEEE Access*, Vol. 4, pp.4761–4768.

Journal of Non-Crystalline Solids xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Non-Crystalline Solids

journal homepage: www.elsevier.com/locate/jnoncrysol

A composition-dependent thermal behavior of $Si_{20}Te_{80-x}Sn_x$ glasses: Observation of Boolchand intermediate phase

Jagannatha K B [a,b], B. Tanujit [c], Diptoshi Roy [b], S. Asokan [c], Chandasree Das [b,*]

[a] Department of Electronics and Communication Engineering, BMS Institute of Technology and Management, Bangalore-64, India
[b] Department of Electrical and Electronics Engineering, Center for Nanomaterials and Display, BMS College of Engineering, Bangalore-19, India
[c] Department of Instrumentation and Applied Physics, Indian Institute of Science, Bangalore-12, India

ARTICLE INFO

Keywords:
Thermal stability
Glass-forming ability
Boolchand's intermediate phase
Network connectivity and rigidity

ABSTRACT

Thermal properties have been investigated using alternating differential scanning calorimetry (ADSC) for determining the phase transitions of $Si_{20}Te_{80-x}Sn_x$ glass samples ($1 \leq x \leq 6$). Experiments are performed to assess the variations of thermal characteristics such as glass transition temperature (T_g) and crystallization temperature (T_c) etc. as a function of Sn variation. The glass-forming ability (GFA) and thermal stability (ΔT) are analyzed for understanding the suitability of as-prepared material in terms of network connectivity and rigidity. The studies have revealed minimum non-reversing enthalpy (ΔH_{NR}) value between compositions ($2 \leq x \leq 5$) of $Si_{20}Te_{80-x}Sn_x$ glasses, which serve as a basis of the existence of Boolchand's intermediate phase (BIP). It is observed that an increase in Sn dopant resulted in an increase in the density and decrease in the molecular volume of $Si_{20}Te_{80-x}Sn_x$ samples. These studies reveal the correlation with the memory switching behavior displayed by $Si_{20}Te_{80-x}Sn_x$ glasses.

1. Introduction

Data storage and related energy consumption are major challenges in the era of artificial intelligence (AI), big data and smartphones. The development of a new technique for data storage that operates at higher speed, lower operating voltage, and long data retention is very much essential for the next generation of digital devices. Researchers are exploring novel memory technologies towards realizing next-generation memory. These include ferroelectric RAM, magnetic RAM, phase-change memory (PCM), etc. [1–3]. Among these, PCM technology which has phase transformation capability appears to be particularly promising due to its manufacturing technology, endurance, multibit ability, higher scalability and lower cost [4,5]. PCM materials are a class of semiconductor materials that do not have a long-range order. These materials include at least one chemical element from the VI main group of the periodic table, such as sulfur, selenium, and tellurium. Extensive study has been carried out on chalcogenide materials to investigate their properties for various applications [6,7]. Amorphous glasses mostly find their applications in photonics and electronics [8]. Over the last decade, companies like Intel, IBM, Micron, etc., have been commercially developing PCM-based nonvolatile random-access memories (NV-RAM). Meanwhile, considerable research has been carried out to understand the fundamentals of PCM [9–13]. The information stored in the PCM materials is prompted by an applied electric field. Ovshinsky observed the electrical switching phenomenon nearly five decades ago, which forms the basis for PCM [14]. The electrical switching can be classified into two groups namely memory and threshold switching. The switching happens when a voltage denoted by threshold voltage (V_{th}), is applied and the PCM material switches from high resistance OFF state (amorphous) considered as logic '0′ to low resistance ON state (crystalline) logic '1′. In threshold switching, withdrawal of applied external electric field prompts glasses to return to the amorphous state, whereas in memory switching, after removal of the field, the glass remains latched on to the crystalline state. The switching processes are electronic and thermal in nature, and the electronic switching befalls when the charge defect states in the sample are bursting by the applied electric field carriers. Further, the thermal switching leads to the development of a channel at the contact region in the crystalline state [15].

The research on electrical switching characteristics is used to determine the appropriate chalcogenide glasses (CG's) suitable for PCM. In addition to switching studies, the general understanding of thermal crystalline kinetics studies is essential for preparing appropriate memory materials [16]. T_g is a significant thermal parameter that specifies the connectivity of the glassy state in CG's. Alternating Differential

* Corresponding author.
E-mail address: chandasreedas.eee@bmsce.ac.in (C. Das).

https://doi.org/10.1016/j.jnoncrysol.2021.121311
Received 27 April 2021; Received in revised form 12 November 2021; Accepted 13 November 2021

Scanning Calorimetry (ADSC) is a thermal analysis tool in which a sine wave is superimposed on the linear temperature curve of differential scanning calorimetry (DSC) [17]. The composition dependency of thermal characteristics of CG's determined by ADSC, reveals several intriguing elements of the microscopic structure and local network connection. The advantages of the ADSC technique, include increased sensitivity, increased resolution without sacrificing sensitivity, the separation of total heat flow (THF) curve into thermally reversing heat flow (RHF) and non-reversing heat flow (NHF), from which non-reversing enthalpy change (ΔH_{NR}) and heat capacity (C_p) may be measured in a single experiment. Over the last two decades, researchers have been studying the effects of network connectivity and stiffness on the characteristics of CG's, particularly their thermal properties. Phillips [18] postulated that optimal network connectivity arises at the critical average coordination, $<r> = 2.4$, when the number of floppy modes equals zero, resulting in percolation in-network rigidity, which is known as rigidity percolation of the glassy system. Thorpe [19] states that the $<r> < 2.4$ glassy systems are in floppy phase and $< r > > 2.4$ are rigid but stress-free within the spectrum of glass-forming composition. At a typical coordination number, 2.4, the glassy samples undergo a percolation transition from floppy to rigid. These phenomena are seen in Si-Te [20], Si-Te-In [21], Ge-Se-Te [22], etc. Tanaka has also testified that the stiffness threshold changes to a different average coordination number for medium-range interacted systems [23]. Also, research groups have shown that the rigidity threshold compasses over a broad range of compositions in some glassy systems [24,25]. The glassy samples exhibit two different stiffness transitions, and the first transition occurs from a floppy phase to a rigid, and the second moves from a rigid to a stressed rigid. This range of composition represents the intermediate rigid phase (IP), which constitutes a thermally reversing glass window. The IP includes the self-organized, nearly perfect, and stress-free structure of the glass [24]. Theoretical calculations using graph theory, cluster approximations, and constraint counting indicate the presence of an intermediate phase known as BIP in a variety of glasses [26]. Si-based chalcogenide glasses have been proven to be suitable for optical and electronics applications [27,28]. The switching characteristics displayed by these glasses make them ideal for PCM applications. Previous studies on electrical switching and crystalline peaks of $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 7$) glasses, prompted us to clarify its thermal properties more clearly [29]. In this work, the thermal crystalline kinetics parameters such as glass transition temperature (T_g), crystalline temperature (T_c), thermal stability (ΔT), and non-reversing enthalpy (ΔH_{NR}) of $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 6$) chalcogenide glasses are presented.

2. Experimental approach

$Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 6$) bulk CG's are prepared via a melt quenching process. The appropriate quantity of Tin (Sn), Silicon (Si) and Tellurium (Te) elements with a purity of 99.9999% (Sigma-Aldrich), is weighed and transferred to quartz ampoules, which are pre-evacuated at 100 °C in the hot-air oven and sealed using a flash evaporation unit at ~7.3×10^{-6} mbar. To prepare homogenize melt, the ampoules are positioned at the center of the rotary furnace at 1100 °C for 24 h after reaching an optimum temperature which is decided by the melting point of the constituent elements. The samples are quenched in a bucket of freezing water (< 0 °C), which is mixed with sodium hydroxide (NaOH) to get the bulk ingots. Using X-ray diffraction (XRD) (PANalytical 'X'Pert3 Powder) the non-crystalline behavior of as-prepared bulk samples is documented.

Thermal studies on $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 6$) bulk glasses are carried out by ADSC (model ADSC822^e, Mettler Toledo). Two 40 ml aluminum cups are used for ADSC wherein one cup is held blank and treated as a reference while the other cup is sealed after placing a sample. Thermal diagnostics are carried out for a temperature range of 40 °C–300 °C at a heat flux of 3 °C / min and 1 °C/ min variation rate. Argon gas (Ar) with a stream rate of 75 ml/min is used as the cleansing agent [30]. Reversing heat flow (RHF) is equivalent to the thermodynamic specific heat variations on heating, while non-reversing heat flow (NHF) refers to the kinetic process and enthalpy differences that convoys structural changes in the glassy network [31]. The thermal parameters T_g, T_c, ΔT, ΔH_{NR} and ΔC_p are assessed from the RHF and NHF graphs derived from total heat flow (THF) curves of ADSC results. Error in thermal parameters measurement is within ± 5 °C, the variation in ΔH_{NR}, measured for heating scans are found to be within ±0.05 J/g. Errors are determined by the standard deviation process, with the resulting error bars depicted in the figures. The sample densities are measured with the ethanol as a reference fluid employing density measurement equipment (METTLER TOLEDO, MODEL ME 204). The density calculation error is below 4%.

3. Results and discussion

Fig. 1 displays the assessment of RHF and NHF graphs of a typical $Si_{20}Te_{75}Sn_5$ glass for numerous thermal factors. Fig. 2 depicts the THF curve of the representative $Si_{20}Te_{80-x}Sn_x$ glass samples ($1 \leq x \leq 6$), acquired via ADSC. Three noticeable exothermic crystalline peaks (T_{c1}, T_{c2} and T_{c3}) and one endothermic T_g have been observed in $Si_{20}Te_{80-x}Sn_x$ samples where x is 2–5. The reason for the appearance of two or three crystallization peaks may be due to the phase separation in the samples due to the disengagement of initial similar compounds into two or more crystalline phases [32]. Sn may crystallize at a lower temperature (T_{c1}) followed by Te at (T_{c2}) into cubic Sn_1Te_2 and hexagonal Te_1 phases. The remaining amorphous matrix may crystallize into hexagonal Si_1Te_2, hexagonal Si_2Te_3 and cubic Sn_1Te_2 [29]. Similar phases formed during crystallization are observed in Si-Te-Bi and Si-Te [29,33].

It is observed that almost all chalcogenide glasses follow the Mott 8-N principle [34], consequently, the coordination numbers (CN) of Si and Te can also be reasonably presumed to be four and two in Si-Te-Sn glassy samples. In particular, metal elements are classified as a 4-fold or higher coordination feature in the glassy chalcogen network samples that improve network connectivity. M. Stevens et al. reported that Sn atoms are coordinated by 4-fold in a Ge–Se(S)–Sn glassy systems [35]. Hence in the current Si-Te-Sn samples, the coordinating numbers are assumed to be four, two, and four for Si, Te and Sn, respectively. The average CN ($< r >$) can be found using the following formula for the above glassy system.

$$< r >= \frac{r1_{Si}(20) + r2_{Te}(80 - x) + r3_{Sn}(x)}{100} \quad (1)$$

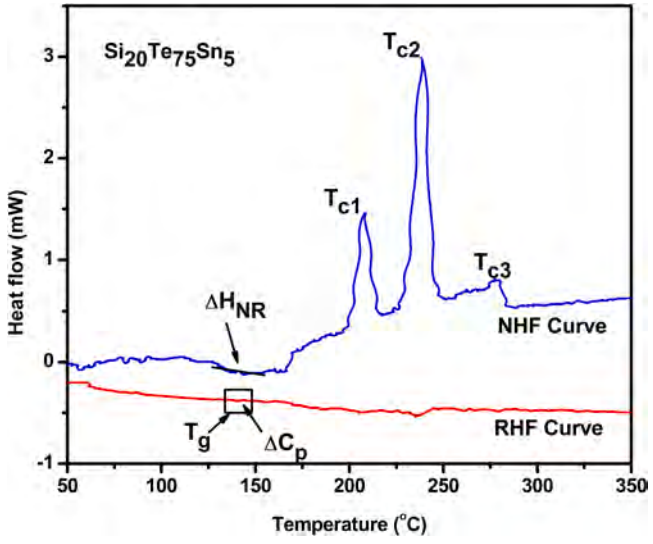


Fig. 1. Assessment of ADSC RHF and NHF graphs of a typical $Si_{20}Te_{75}Sn_5$ sample for thermal parameters.

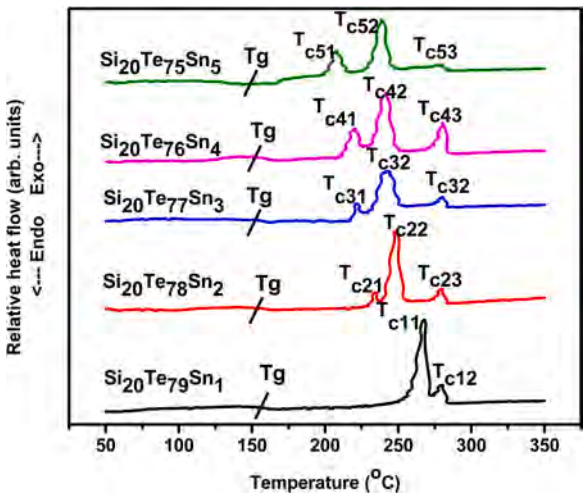


Fig. 2. Total heat flow graph of the $Si_{20}Te_{80-x}Sn_x$ glass samples ($1 \leq x \leq 6$).

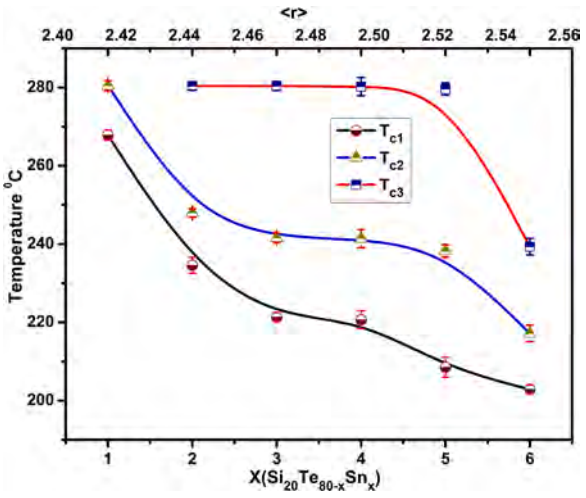


Fig. 4. The variation of T_c with x and $\langle r \rangle$ for $Si_{20}Te_{80-x}Sn_x$ glassy samples ($1 \leq x \leq 6$). The trend line obtained using B-Spline fit function.

where r is the CN of individual atoms. The reason for the variation of thermal parameters with respect to Sn variation in the Si-Te network is determined by network connectivity and rigidity of the samples. Consequently, topological limits may be correlated with the rapid change in the feature of thermal parameters with respect to composition.

Fig. 3 represents the variation of T_g with the addition of Sn dopants in the Si-Te matrix. To determine network connectivity and rigidity, the compositional dependency of T_g is used. A rise in T_g usually means network strengthening, while a decline in T_g indicates the phase separation due to the splitting of homopolar bonds [36]. From Fig. 3 it is seen that the T_g values decrease with the addition of Sn, which could be due to the semi-metallic nature of Sn, the disintegration of Te-Te homopolar bonds, and the bonding of Te-Si and Sn-Te atoms results in decreased network connectivity and stiffness. Similar studies revealed that metallic dopants, such as In, Sb, and Bi inflowing to the host matrix, partially disrupt the chain length and thus increase the number of weak Te rings [37,38]. A decrease in T_c values is observed in $Si_{20}Te_{80-x}Sn_x$ glassy alloys ($1 \leq x \leq 6$). Fig. 4 displays the decrease in first and second crystallization temperature (T_{c1} and T_{c2}) of $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 6$) samples. However, there is no significant change in the third crystallization temperature, T_{c3}. Similar crystalline temperatures are observed in Si-Te-In glasses [39]. The decrease in T_c values indicates the existence of loosely connected heteropolar and homopolar bonds in as-prepared samples [27].

The difference between T_c and T_g values ($\Delta T = T_{c1}\text{-}T_g$) is directly related to glassy system's thermal stability and glass-forming ability (GFA). ΔT values help in determining the tendency of the glassy material towards crystallization [40]. Fig. 5 shows the thermal stability of $Si_{20}Te_{80-x}Sn_x$ glassy samples ($1 \leq x \leq 6$). It is observed from Fig. 5 that ΔT decreases and T_c values also fall with Sn doping (Fig. 4), which indicates that Si-Te glasses can easily be devitrified. These findings support the notion that the switching voltages, network connectivity, and network rigidity of $Si_{20}Te_{80-x}Sn_x$ glass samples ($1 \leq x \leq 6$) decrease with the addition of Sn in the Si-Te matrix.

Fig. 6 depicts the number of Si-Te, Te-Te, and Te-Sn bonds, as well as the average bond energy with Sn variation. The number of bonds have been computed using bond formation rules, and the bond energies have been obtained using bond energy values [41]. From Fig. 6, it is observed that Te-Te bonds decrease, no change in Si-Te bonds, and there is an increase in Te-Sn bonds with an increase in Sn dopant. Further, average bond energy is found to increase with Sn addition, as Te-Te bond energy is 257.6 ± 4.1 kJ/mol and Te-Sn is 359.8 kJ/mol. When Sn is added to the host matrix Te-Si, Sn atoms progressively replace lower energy Te-Te bonds with higher energy Te-Sn, which results in overall increase in the

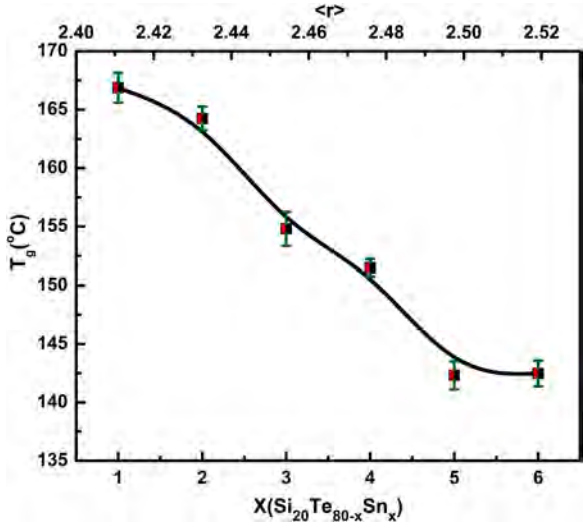


Fig. 3. The variation T_g of $Si_{20}Te_{80-x}Sn_x$ glass samples ($1 \leq x \leq 6$) with Sn (x) and average coordination number $\langle r \rangle$. The trend line obtained using B-Spline fit function.

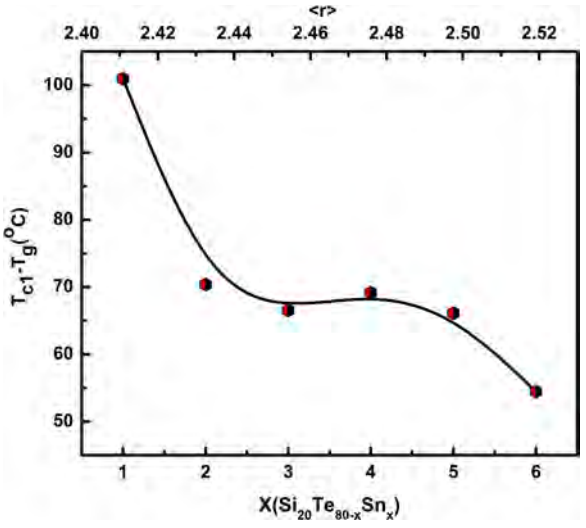


Fig. 5. The compositional dependence of ΔT for $Si_{20}Te_{80-x}Sn_x$ glasses ($1 \leq x \leq 6$). The trend line obtained using B-Spline fit function.

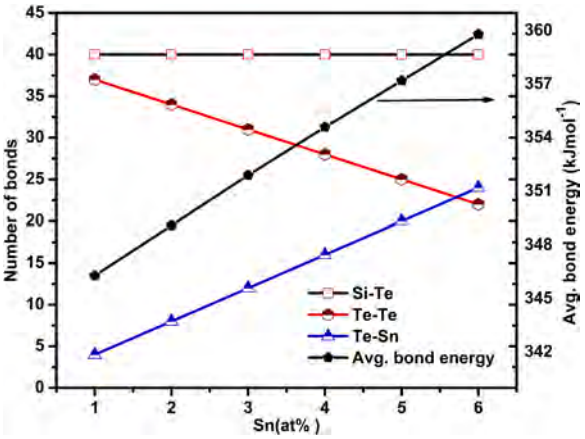


Fig. 6. The variation of average bond energy and number of bonds, of $Si_{20}Te_{80-x}Sn_x$ glassy samples ($1 \leq x \leq 6$) With Sn composition (x).

average bond energy of Si-Te-Sn system, similar increase in average bond energy is observed in GeSeTe glasses also [42]. The decrement in T_g and T_c values may be due to the influence of network connectivity, the replacement of 2-fold coordinated Te with 4-fold coordinated Sn atoms [41], and replacement of lower energy bonds of Te-Sn with higher energy bonds of Te-Si. This decrease in bond numbers correlates to a decrease in values of T_g and T_c, which are directly related to a decrease in rigidity and network connectivity of samples.

The ΔC_p calculated at T_g is a significant parameter to describe a specified amorphous material, as it indicates the heat accumulated in molecular motion. The T_g represents the temperature above which the base glass matrix reaches different structural phases and below which the host matrix attains a structurally stable state [43]. A variation in the ΔC_p during glass transition recommends the existence of a prolonged stiffness transformation. Glasses that exhibit significant variation in ΔC_p values are weak and those exhibit minor variations are strong liquids [44]. The weak liquids exhibit greater GFA tendency, whereas the strong liquids have lower GFA tendency [45]. The value of ΔC_p for strong liquids is between 0.09 and 0.35 J/g-K, while, for weak liquids, the ΔC_p value is above 0.35 J/g-K [46]. From Fig. 7, it is seen that the addition of Sn to the Si-Te matrix creates ternary alloys derived from strong liquids. Comparable studies have been published for Sn doping with Se-Sb and Ge-Te glasses [38,47].

ΔH_{NR} is used to quantify internal stress in a glass by providing heat energy between the glass and its melt. The decrease in ΔH_{NR} indicates that the composition range correlates to BIP during the glass transition phase. For glass composition range in the intermediate phase, ΔH_{NR} is found close to zero [24,48], and these studies recommend that glass and liquid structure in IP are close to each other and are stress-free. The present study on ΔH_{NR} versus compositional variation with coordination number $< r >$ (Fig.7) displays that ΔH_{NR} starts decreasing at $x = 2$ and reaches close to zero at $x = 4$, IP range can be assumed $2 \leq x \leq 5$ in $Si_{20}Te_{80-x}Sn_x$ glassy alloys. The intermediate phase has been observed in Si-Te-In, Ge-Te-In-Ag, and Si-Te-Cu glasses [39,49,50] over different composition ranges. In particular, Si-Te-Cu forms the base for the present study, the IP region seen in the $< r > = 2.32$ to 2.50, the compositional width (Δr) and centroid (Δc) are 0.18 and 2.41, respectively, whereas, for Ge-Te-In-Ag system, IP region seen in the range $< r > =$ 2.37 to 2.62, the compositional Δr and Δc are 0.25 and 2.49, respectively. In random networks, the width of IP almost vanishes, which agrees with the extended constraint theory [34]. This is true with the present Si-Te-Sn system where $< r > = 2.42$ to 2.52. The Δr and the Δc of $Si_{20}Te_{80-x}Sn_x$ glassy samples ($1 \leq x \leq 6$) are 0.1 and 2.47, respectively (Fig. 8).

Fig. 9 displays the variation of density with composition. It is seen from Fig. 9 that the density of $Si_{20}Te_{80-x}Sn_x$ glassy samples ($1 \leq x \leq 6$) decreases with the addition of Sn. Since the atomic weight of Sn (118.69. AMU) is nearer to Te (127.6 AMU), the replacement of Te with Sn atoms could be a reason for the decrease in density of the host matrix. Also, in the host matrix Si-Te bond length is less, and atoms are tightly packed; hence density will be higher. When Sn dopant is added to the host matrix, the Te-Sn atoms form the bonds, the bond length of Te-Sn is high compared to Si-Te hence atoms are loosely packed, and there is a decrease in density with the addition of Sn dopant.

The variation of molar volume with tin addition is shown in Fig. 10. The molar volume increases with the addition of Sn dopant, as density decreases due to an increase in bond length to accommodate the increased bond length, the molar volume increases. A plateau is seen in both density and molar volume in the composition range $x = 2$ to 5, which corresponds to the reversibility window identified by ADSC. It is also supported by the present ΔH_{NR} studies. It has been realized previously that the IP produces densely packed space-filling networks with low molar volumes [51]. Similar results are seen in Ge-Te-Si and Ge-Te-In-Ag systems [49,52].

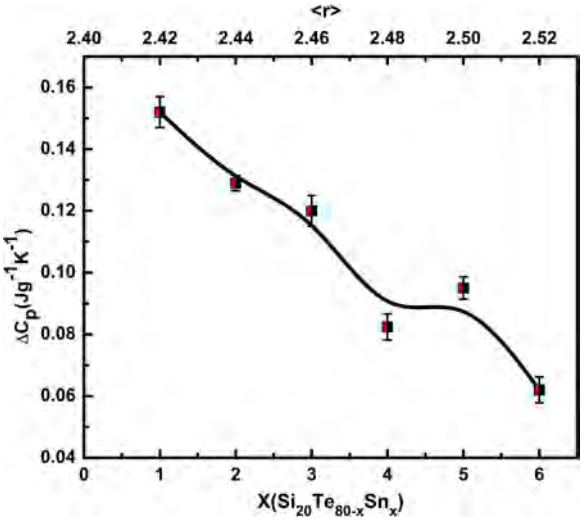


Fig. 7. Variation of ΔC_p measured at T_g for $Si_{20}Te_{80-x}Sn_x$ glassy samples ($1 \leq x \leq 6$). The trend line obtained using B-Spline fit function.

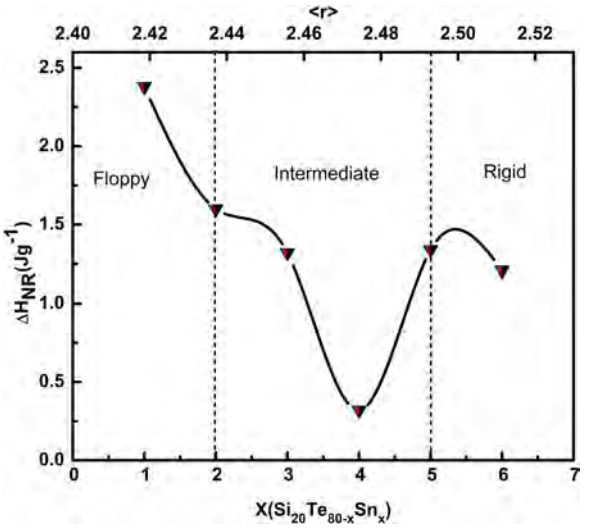


Fig. 8. ΔH_{NR} versus compositional variation with coordination numbers for $Si_{20}Te_{80-x}Sn_x$ glass samples ($1 \leq x \leq 6$).

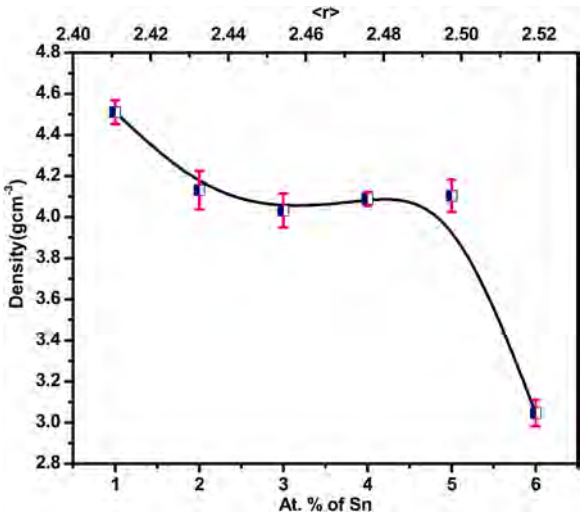


Fig. 9. Variation of density with composition in $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 6$) glassy samples. The trend line obtained using B-Spline fit function.

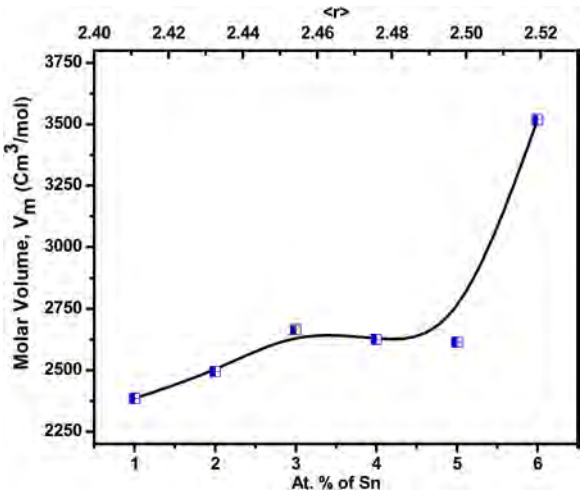


Fig. 10. Variation of molar volume with the composition in $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 6$) glassy samples. The trend line obtained using B-Spline fit function.

Conclusions

The studies on thermal parameters of melt quenched $Si_{20}Te_{80-x}Sn_x$ glassy samples ($1 \leq x \leq 6$) are carried out systematically using ADSC. These samples show one endothermic glass transition temperature (T_g), and three different crystalline temperature (T_c) peaks representing different stable phases percolating at different T_c. The decrease in T_g, T_c and density values with Sn addition indicates the decrease in network connectivity and rigidity of the glassy samples. Thermal stability (ΔT) and change in the specific heat capability (ΔC_p) determine that samples can be easily devitrified and has a lower glass-forming ability. The presence of the Boolchand Intermediate phase in the composition range $2 \leq x \leq 5$ is seen in ΔH_{NR} studies.

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.

Acknowledgment

Dr. Chandasree Das and Jagannatha KB acknowledge Science and Engineering Research Board (SERB) (File No.: SPG/2021/000359), and TEQIP (II), India for providing the funding support for their research work.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejps.2020.105216.

Reference

[1] J.G. Zhu, Magnetoresistive random access memory: the path to competitiveness and scalability, Proc. IEEE 96 (11) (2008) 1786–1798.
[2] H.S.P. Wong, et al., Phase change memory, Proc. IEEE 98 (12) (2010) 2201–2227.
[3] H.S.P. Wong, et al., Metal–oxide RRAM, Proc. IEEE 100 (6) (2012) 1951–1970.
[4] J. Maimon, E. Spall, R. Quinn, S. Schnur, Chalcogenide-based non-volatile memory technology, 2001 IEEE Aerosp. Conf. Proc. 5 (2001) 2289–2294 (Cat. No. 01TH8542).
[5] G.F. Zhou, Materials aspects in phase change optical recording, Mater. Sci. Eng. A 304 (2001) 73–80.
[6] 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 (5) (2016) 1644.
[7] Z. Han, et al., On-chip mid-infrared gas detection using chalcogenide glass waveguide, Appl. Phys. Lett. 108 (14) (2016) 1–4.
[8] H.S. Nalwa, Handbook of advanced electronic and photonic materials and devices, Handb. Adv. Electron. Photonic Mater. Devices 1 (2001) 321–350.
[9] M. Wuttig, H. Bhaskaran, T. Taubner, Phase-change materials for non-volatile photonic applications, Nat. Photonics 11 (8) (2017) 465–476.
[10] P. Fantini, Phase change memory applications: the history, the present and the future, J. Phys. D Appl. Phys. 53 (28) (2020), 283002.
[11] P. Noé, C. Vallée, F. Hippert, F. Fillot, J.Y. Raty, Phase-change materials for non-volatile memory devices: from technological challenges to materials science issues, Semicond. Sci. Technol. 33 (1) (2017) 13002.
[12] Z. Cheng, C. Rios, N. Youngblood, C.D. Wright, W.H.P. Pernice, H. Bhaskaran, Device-level photonic memories and logic applications using phase-change materials, Adv. Mater. 30 (32) (2018), 1802435.
[13] H. Flores-Ruiz, M. Micoulaut, From elemental tellurium to $Ge_2Sb_2Te_5$ melts: high temperature dynamic and relaxation properties in relationship with the possible fragile to strong transition, J. Chem. Phys. 148 (3) (2018) 34502.
[14] S.R. Ovshinsky, Reversible electrical switching phenomena in disordered structures, Phys. Rev. Lett. (1968).
[15] S. Asokan, Electrical switching in chalcogenide glasses- some newer insights, J. Optoelectron. Adv. Mater. 3 (3) (2001) 753–756.
[16] M. Abu El-Oyoun, A study of the crystallization kinetics of Ge_2QTe_{80} chalcogenide glass, J. Phys. D Appl. Phys. 33 (18) (2000) 2211–2217.
[17] S.L. Simon, Temperature-modulated differential scanning calorimetry: theory and application, Thermochim. Acta 374 (1) (2001) 55–71.
[18] J.C. Phillips, Topology of covalent non-crystalline solids I: short-range order in chacogenide alloys, J. Non. Cryst. Solids 34 (2–3) (1979) 153–181.
[19] M.F. Thorpe, Continous deformations in random netwokrs, J. Non. Cryst. Solids 57 (2) (1983) 355–370.
[20] S. Asokan, G. Parthasarathy, E.S.R. Gopal, Evidence for a critical composition in group-IV VI chalcogenide glasses, Phys. Rev. B 35 (15) (1987) 8269–8272.
[21] S. Rao Gunti, S. Asokan, Thermodynamic, Raman and electrical switching studies on $Si_{15}Te_{85-x}Ag_x$($4 \leq x \leq 20$) glasses, J. Appl. Phys. 111 (3) (2012).
[22] R Ganesan, A. Srinivasan, K.N. Madhusoodanan, Composition dependence of the glass transition in Ge-Se-Te glasses, Phys. Status Solidi 65 (1) (1994) 45–47.
[23] K. Tanaka, Structural phase transitions in chalcogenide glasses, Phys. Rev. B 39 (2) (1989) 1270–1279.
[24] D.G. Georgiev, P. Boolchand, M. Micoulaut, Rigidity transitions and molecular structure of As_xSe_{1-x} glasses, Phys. Rev. B Condens. Matter Mater. Phys. 62 (14) (2000) R9228–R9231.
[25] D. Roy, G.S. Varma, S. Asokan, C. Das, Electrical switching and optical bandgap studies on quaternary Ag-doped Ge-Te-In thin films, IEEE Trans. Electron Devices 66 (4) (2019) 1881–1886.
[26] Y. Wang, P. Boolchand, and M. Micoulaut, "Glass structure, rigidity transitions and the intermediate phase in the Ge -As-Se ternary," vol. 52, no. December, pp. 633–639, 2000.
[27] 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. December 2018.
[28] J.K. Babanna, D. Roy, S.G. Varma, S. Asokan, C. Das, Effect of Sn addition on glassy Si-Te bulk sample, AIP Conf. Proc. 1966 (2018) 20034.
[29] K.B. Jagannatha, D. Roy, C. Das, Electrical switching and crystalline peak studies on $Si_{20}Te_{80-x}Sn_x$ ($1 \leq x \leq 7$) chalcogenide bulk glasses, J. Non. Cryst. Solids 544 (2020). May.

[30] G. Sreevidya Varma, A. Chaturvedi, U. Ramamurty, S. Asokan, Kinetics based evidence for intermediate phase in $Ge_{15}Te_{85-x}In_x$ chalcogenide glasses, J. Non Cryst. Solids 471 (2017) 251–255. May.
[31] D.M. Price, Modulated-temperature thermomechanical analysis, Thermochim. Acta 358 (1998) 23–29, 1998.
[32] V. Modgil, V.S. Rangra, Effect of Sn addition on thermal and optical properties of glass, J. Mater. 2014 (2014). Hindawi Publishing Corporation.
[33] 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 (5) (1985) 502–504.
[34] N.F. Mott, Conduction in non-crystalline systems, Philos. Mag. 17 (150) (1968) 1259–1268.
[35] M. Stevens, J. Grothaust, P. Roolchand, and E. C. Devices, "Vol. 47, No. 3," vol. 47, no. 3, pp. 199–202, 1983.
[36] P. Boolchand, D.G. Georgiev, T. Qu, F. Wang, L. Cai, S. Chakravarty, Nanoscale phase separation effects near *r* glasses, Comput. Eng. 5 (2002) 713–724.
[37] D. Tonchev, S.O. Kasap, Thermal properties of Sb_xSe_{100-x} glasses studied by modulated temperature differential scanning calorimetry, J. Non Cryst. Solids 248 (1) (1999) 28–36.
[38] 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. Alloy. Compd. 721 (2017) 674–682.
[39] 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 (12) (2011) 3345–3352.
[40] A. Hrubý, Evaluation of glass-forming tendency by means of DTA, Czechoslov. J. Phys. 22 (11) (1972) 1187–1193.
[41] S. Asokan, PhD thesis, Department of physics, Indian Institute of Science, 1986.
[42] S.B. Bhanu Prashanth, S. Asokan, A composition dependent thermal behavior of $Ge_xSe_{35-x}Te_{65}$ glasses, J. Non Cryst. Solids 355 (22–23) (2009) 1227–1230.
[43] M.M.A. Imran, N.S. Saxena, D. Bhandari, M. Husain, Glass transition phenomena, crystallization kinetics and enthalpy released in binary $Se_{100-x}In_x$ ($x = 2$, 4 and 10) semiconducting glasses, Phys. Status Solidi Appl. Res. 181 (2) (2000) 357–368.
[44] C.A. Angell, Spectroscopy simulation and scattering, and the medium range order problem in glass, J. Non Cryst. Solids 73 (1–3) (1985) 1–17.
[45] P.Z. Saheb, B.H. Sharmila, S. Asokan, K.A. Gowda, Extended rigidity percolation and chemical thresholds in Ge–Te–Pb glasses as revealed by MDSC, Solid State Commun. 129 (12) (2004) 765–768.
[46] K. Chebli, J.M. Saiter, J. Grenet, A. Hamou, G. Saffarini, Strong-fragile glass forming liquid concept applied to GeTe chalcogenide glasses, Phys. B Condens. Matter 304 (1–4) (2001) 228–236.
[47] O.A. Lafi, M.M.A. Imran, N.I. Abu-Shaweesh, F.M. Al-Kurdi, I.K. Khatatbeh, Effect of chemical ordering on the crystallization behavior of $Se_{90}Te_{10-x}Sn_x$ (x=2, 4, 6, and 8) chalcogenide glasses, J. Phys. Chem. Solids 75 (6) (2014) 790–795.
[48] D. Selvanathan, W. Bresser, P. Boolchand, Stiffness transitions in glasses from Raman scattering and temperature-modulated differential scanning calorimetry, Phys. Rev. B Condens. Matter Mater. Phys. 61 (22) (2000) 15061–15076.
[49] G. Sreevidya Varma, D.V.S. Muthu, A.K. Sood, S. Asokan, Raman signatures of intermediate phase in quaternary $Ge_{15}Te_{80-x}In_5Ag_x$ glasses, J. Non Cryst. Solids 387 (2014) 143–147.
[50] D. Roy, B. Tanujit, J. K. B, G.S. Varma, S. Asokan, C. Das, Manifestation of intermediate phase in Cu doped Si-Te glasses, J. Non Cryst. Solids 531 (2020). Mar.
[51] P. Boolchand, D.G. Georgiev, T. Qu, F. Wang, L. Cai, S. Chakravarty, Nanoscale phase separation effects near r= 2.4 and 2.67, and rigidity transitions in chalcogenide glasses, C. R. Chim. 5 (11) (2002) 713–724.
[52] C. Das, M.S.R.N. Kiran, U. Ramamurty, S. Asokan, Manifestation of intermediate phase in mechanical properties: nano-indentation studies on Ge-Te-Si bulk chalcogenide glasses, Solid State Communications 152 (24) (2012) 2181–2184.

Influence of Cu Doping in Si–Te-Based Chalcogenide Glasses and Thin Films: Electrical Switching, Morphological and Raman Studies

Diptoshi Roy, B. Tanujit, K. B. Jagannatha, S. Asokan, and Chandasree Das, *Senior Member, IEEE*

***Abstract*— To understand the electrical switching behavior of $Si_{15}Te_{85-x}Cu_x$ ($1 \leq x \leq 10$) series, *I–V* characterization has been performed on bulk as well as amorphous thin films of the as-prepared samples. Both the bulk glasses and amorphous thin films are found to manifest memory-type switching behavior. The threshold voltages of thin-film devices are found to be much lower than the bulk counterparts and hence could find application for phase change memory (PCM). The composition analyses of both have divulged the existence of intermediate phase (IP) in the composition range of $2 \leq x \leq 6$. To examine the probability of the given glass for PCM application, Set–Reset studies have been performed on the glasses with a triangular pulse of 6 mA for set operation and rectangular pulse of 12 mA for the reset operation. The study has revealed a continuous repetition of few Set–Reset cycle by the Si–Te–Cu series. Raman studies carried out on the bulk glasses report the occurrence of blue shift over the composition in a regular manner. Further, SEM studies have been carried out on Si–Te–Cu samples to understand the morphological changes that would have occurred during switching. Additionally, thickness dependence of threshold voltage of representative $Si_{15}Te_{80}Cu_5$ and $Si_{15}Te_{76}Cu_9$ glasses has been carried out to reveal the relationship between threshold voltage and thickness.**

***Index Terms*— Chalcogenide glasses, electrical switching, intermediate phase (IP).**

I. Introduction

PHASE change memory (PCM) is enduring moderate renascence of interest as the faults it encounters can be easily deciphered through engineering solutions. The success of PCM depends on the development of engineering solutions and the corresponding sufficient reduction in costs. The amorphous and crystalline states have a substantial resistance variance which leads to the empowerment of PCM [1], [2]. The noncrystalline state possesses high resistivity whereas the crystalline state manifests resistivity of magnitude which is three to four orders of magnitude lower.

In the last couple of years, a wide variety of improvements like density improvement, aggressive scaling of devices, and reduction in programming power have been seen in the field of PCM. The main reason for reduction in switching power which stands as the rudimentary strength of PCM technology is the pattern of device scaling. The promising attributes of PCM technology are effective scalability (up to few nanometers), faster programming time (in the order of few nanoseconds), ameliorated endurance (up to 10^9 programming cycles), and bit alterability (the capability of PCM to directly write in the memory, without the requirement of prior erasing step). For the features like nonvolatility, reduced usage of power, and direct write, PCM gains to envelope an extensive area of applications like addressing wireless systems, embedded applications, solid-state subsystems, and computing platforms [3].

Metal-doped chalcogenide glasses possess electronic and optical properties and are dissimilar from the properties of the ordinary ones. Glasses containing transition metal like Cu, Ag, and Sn have been employed in optical memory [4] and laser materials [5]. Additionally it has been seen that glasses doped with Cu manifests photoconductivity and shrinkage in bandgap, illustrating their probable usage in economical mid-IR detection which in turn leads to the investigation of electrical properties [6]. The metallic impurities such as copper when added to Si–Te bring interesting variations in their properties. They enter the structural network of chalcogenide glasses in a special way and increase the network connectivity, crystallizing ability, and the electrical conductivity [7], [8]. That is the reason an attempt has been made to study systematically the properties of the Si–Te glassy system with added Cu impurity. The study on the switching behavior of a material is an important part to find the possibility of the material in PCM application.

II. Experimental Techniques

The bulk sample that is required for making thin-film devices are concocted by melt quenching technique. Pure elements of the order of 99.999% purity have been taken in proportional quantity in clean quartz ampoule which has been sealed maintaining a vacuum of 10^{-6} mbar. The furnace is

Manuscript received December 15, 2020; accepted January 10, 2021. Date of publication January 26, 2021; date of current version February 24, 2021. The work of Chandasree Das was supported in part by the Science and Engineering Research Board (SERB) through the "Fast Track Scheme for Young Scientist," India, under Grant SB/FTP/ETA-95/2013 and in part by the Technical Education Quality Improvement Programme (TEQIP)-II. The review of this article was arranged by Editor B. K. Kaushik. *(Corresponding author: Chandasree Das.)*

Diptoshi Roy and Chandasree Das are with the BMS College of Engineering, Bengaluru 560019, India (e-mail: diptoshi.pej16@bmsce.ac.in; chandasreedas.eee@bmsce.ac.in).

B. Tanujit and S. Asokan are with the Indian Institute of Science, Bengaluru 560012, India (e-mail: tanujitb@iisc.ac.in; sasokan@iisc.ac.in).

K. B. Jagannatha is with the BMS Institute of Technology, Bengaluru 560064, India (e-mail: jagan@bmsit.in).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TED.2021.3051925.

Digital Object Identifier 10.1109/TED.2021.3051925

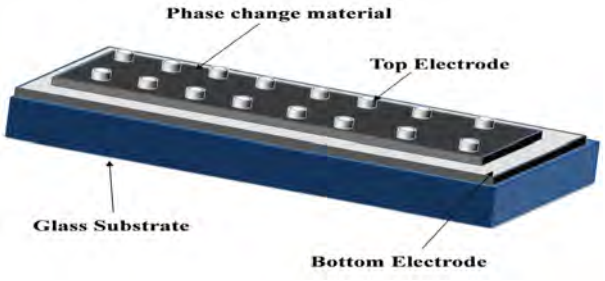


Fig. 1. 3-D schematic of sandwich geometry utilized for making amorphous Si–Te–Cu thin-film switching devices.

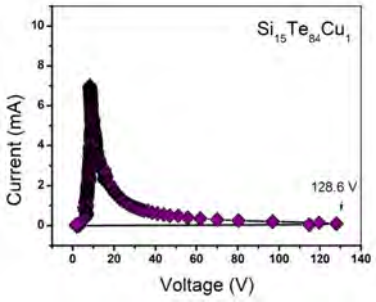


Fig. 2. *I–V* characteristic of a representative $Si_{15}Te_{84}Cu_1$ glass showing MS.

programmed to revolve at 10 r/min uninterruptedly with an increment of 100 °C per hour till it reaches 1100 °C. The homogeneity of the melt can be maintained by keeping the ampoule at a sustained temperature of 1100 °C for 24 h and finally quenching it in ice water mixed with NaOH. The studies on the bulk switching have been undertaken to understand whether the material undergoes switching or not. Once it has been confirmed that electrical switching occurs in the bulk samples further switching analysis has been carried out on thin films which are indeed a very important aspect for memory device research.

A clean biological glass slide of 25 mm × 75 mm dimension is used as a substrate on which the source material has been coated using the flash evaporation approach. To execute the switching studies, the devices have been fabricated in sandwich style. The glass substrate has been used as a base for coating and the top and bottom electrodes have been coated with aluminum. The 3-D schematic of sandwich geometry that has been adopted for making amorphous Si–Te–Cu thin-film switching devices is shown in Fig. 1.

The switching studies have been carried out using Keithley (2410^c) source measure unit (SMU) with Lab View 7 (National Instruments) which has a voltage and current limit of 1100 V and 1 A, respectively. To carry out the switching studies on bulk samples, the samples have been polished to obtain 0.3-mm thickness and placed between the contact points of the Keithley SMU. The purpose of testing the bulk sample is to know the property of the material and its feasibility in the application in the form of thin film. 0.3 mm has been maintained only for bulk samples whereas in the films the approximate thickness obtained is 10 μm. A probe station with an applicability of x-, y-, z-direction movement is employed to place the thin film for switching studies. Threshold switching (TS) is volatile which means once the applied field is removed the material reverts to its original state. As memory switching (MS) draws maximal importance in PCM application, in this work MS has been emphasized specifically for both bulk and film.

SEM studies are carried out using VEGA3 TESCAN. Horiba Jobin Yvon (LabRAM HR) in backscattered mode is used to perform Raman studies on the glassy samples. To analyze the backscattering light, the triple monochromator is used and discerned by a charge-coupled device (CCD) cooled at −70 °C. An argon ion laser of 514.5 nm line is utilized to illuminate the glassy sample wherein the laser is focused with a 50X objective. A superior signal-to-noise ratio is obtained by using an acquisition time of around 120 s.

III. Results and Discussion

A. Switching Analysis in $Si_{15}Te_{85-x}Cu_x (1 \leq x \leq 10)$ Bulk Glasses

The I–V characteristic of an exemplary bulk $Si_{15}Te_{85-x}Cu_x$ glass is shown in Fig. 2 where an ohmic behavior exhibited by the sample is observed initially followed by a negative resistance behavior. Specified glasses manifest a current-controlled negative resistance (CCNR) behavior at threshold voltage, leading to a high conducting "ON" state. Factors on which the switching by a glassy semiconductor depends are ON-state current, thermal stability of the glass, thermal diffusivity of the material, and the network connectivity. It has been seen here that upon restriction of ON-state current to low values, the glass exhibits TS. The glass is seen to display MS at a higher ON-state current and is being shown in Fig. 2. An analogous trend is also seen in other tellurides [9].

The MS is associated with structural change and it becomes difficult for the system to restructure itself with the increase in cross-linkages and rigidity. Consequently, a memory to TS is noticed in systems having more cross-linking elements [10]. The phase change attribute of the glasses helps in understanding the MS in chalcogenides.

The transition from amorphous to crystalline phase occurs during switching and the later phase is retained in MS. The temperature of the glassy material between the electrodes augments because of the joule heating and creates a filamentary path [11], [12].

The compositional dependence of switching voltage of $Si_{15}Te_{85-x}Cu_x$ glasses is shown in Fig. 3. The two governing features that determine the compositional variation of switching voltage of chalcogenides are dopant resistivity [13] and the network connectedness [14]. Typically the inclusion of metallic dopants is known to reduce the resistivity and consequently the switching voltage in chalcogenides [15]. Conversely, the strengthening of connectivity and rigidity of network is seen to increase the switching voltage. In many cases, high resistivity and network connectivity make the structural changes difficult during switching process resulting in TS as seen in Al–As–Te glasses [16].

Further, the increase or decrease in resistivity and network connectivity can be judged from the variation of threshold (or switching) voltage with composition, and the same has been interpreted in this work. The Si–Te–Cu system has exhibited good MS behavior and also the switching data are reproducible. In the present system, with the intensification of Cu, the threshold voltage (V_T) is seen to increase initially followed by the exhibition of a plateau in the region of $2 \leq x \leq 6$. A precipitous decline is exhibited by V_T above 6% of Cu. The consequence of both the network connectivity because of the incorporation of high coordinated Cu atoms and metallic nature of the dopant are seen to influence the contrast of V_T of bulk $Si_{15}Te_{85-x}Cu_x$ glasses with composition. The analysis of the extended X-ray absorption fine structure (EXAFS) spectrum in literature reveals the existence of coordination number in copper as four [17], a number which has also been observed in the other systems of copper like CuFeS [18], Cu_3AsS_4 [18], and CuAsI [19].

The movement of Cu^+ions also has a critical contribution in the switching behavior of Si–Te–Cu glasses. The OFF-state highlights the subsequent recombination processes: hole seizure by C_1^- centers ($C_1^- + e^+ \rightarrow C_1^0$) and electron seizure by C_3^+ centers ($C_3^+ + e^- \rightarrow C_3^0$). A passivation of C^- centers can be carried out by Cu^+ ions that exist in the sample by diffusion process [20], consequently minimizing the voltage needed to instigate the switching process, also additionally reducing the switching time of the sample. The initial increment of V_T with composition of $Si_{15}Te_{85-x}Cu_x$ glasses is due to the increase in network connectivity and the subsequent decrement is owing to the incorporation of higher metallic dopant and rapid passivation of charged defect states. Thermal studies on $Si_{15}Te_{85-x}Cu_x$ glasses [21] have shown the presence of a wide trough in the range $2 \leq x \leq 6$ in the compositional dependence of nonreversing enthalpy at glass transition (ΔH_{NR}), revealing the existence of Boolchand intermediate phase (IP). The present study also shows the presence of a wide plateau in the range $2 \leq x \leq 6$ in the compositional dependence of V_T which could be linked with the IP as observed in the electrical switching behavior of Ge–Te and Si–Te glasses [22]. The composition of the glass which resides in the IP is presumed to have self-organization and is independent of stress and so the width of the IP stands as a measurement of self-organization of the glassy system [23]. Additionally, the latent heat of melting of glasses in the IP is nearly absent and this can be a possible reason of low threshold field in the IP [24]. Fig. 3 shows the variation of V_T with respect to composition wherein the thickness of various samples has been kept constant as 0.3 mm.

B. Electrical Switching in $Si_{15}Te_{85-x}Cu_x$ ($1 \leq x \leq 10$) Amorphous Thin Films

The I–V characteristic of a typical thin-film device is depicted in Fig. 4. The thickness of the material is approximately 10 μm and the dimension of the top electrode is circular in shape with 6 mm dia.

The comparison of electrical switching response of bulk $Si_{15}Te_{85-x}Cu_x$ glasses with their thin-film counterparts is important for the applicability of the material in PCM.

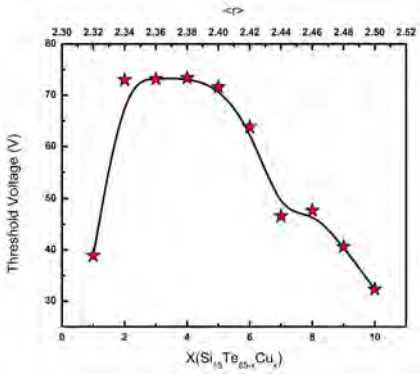


Fig. 3. Compositional dependence of switching voltage of $Si_{15}Te_{85-x}Cu_x$ glasses ($1 \leq x \leq 10$); $< r >$ represents the average coordination number of specific composition.

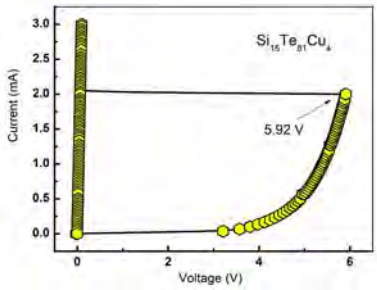


Fig. 4. I–V characteristic of a representative $Si_{15}Te_{81}Cu_4$ thin film showing MS.

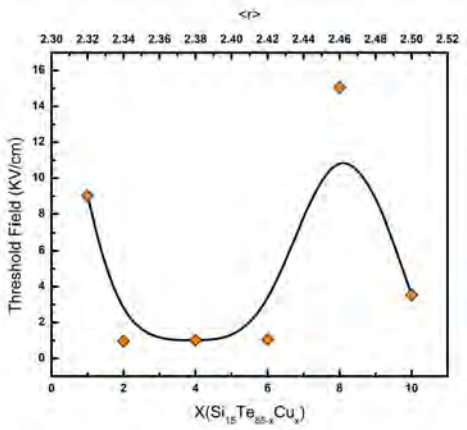


Fig. 5. Compositional dependence of switching fields of $Si_{15}Te_{85-x}Cu_x$ ($1 \leq x \leq 10$) thin films.

The studies on the bulk $Si_{15}Te_{85-x}Cu_x$ sample indicate that both MS and TS [25] have been exhibited by the glassy samples which depend on the ON-state current. Howbeit, the studies on amorphous thin films of $Si_{15}Te_{85-x}Cu_x$ samples indicate the exhibition of only MS. The memory form of the switching responses has been manifested by other binary, ternary, and quaternary amorphous thin films such as GeS_2 [26], Ge–Te–Si [27], and Ge–Te–In–Ag [28].

The compositional dependence of threshold field of amorphous $Si_{15}Te_{85-x}Cu_x$ thin films is shown in Fig. 5. To study the effect of incorporation of dopant on the switching

phenomena, the compositional dependence of threshold field has been investigated for the thin-film devices. In addition, as the usage of flash evaporation technique does not guarantee the uniformity in the thickness of deposition among the devices, the threshold voltage has been normalized as threshold field and used for studying the composition dependence. Further, the coordination number in glasses helps to define the rigidity percolation threshold (RPT) in glasses which manifests the percolative transition of glass from floppy polymeric glass to rigid amorphous solid and is included in Fig. 5. Interestingly, amorphous thin films do not show one to one correspondence of percolation effect when compared with their bulk counterparts. The change observed in the threshold field of thin-film samples as compared to bulk at the onset of percolation could be because of several factors like high field conductions, methods of preparations of thin films, and difference of electrodes in bulk and thin films. The semiconducting thin films which are developed in sandwich geometry with electrodes of metallic nature may have an occurrence of some specific high field conduction process. The several conduction processes are categorized as bulk-limited or electrode-limited.

The bulk-limited is further categorized into space-charge-limited conductivity (SCLC), Poole–Frenkel effect, and hopping. Electrode-limited is classified as tunneling and Schottky effect. These nonlinear conductions may lead to high electric field across the film [29].

The substantial types of variables experienced in thin-film deposition mechanism (e.g., substrate temperature, deposition rate, angle between vapor stream and substrate) enable deposition of different structured films under nominally homogenous environment, which can sequentially influence the optical and electrical properties [29]. Further in bulk the electrode used is brass whereas for thin-film aluminum is used. Even the shape of the top and bottom electrodes is different in thin film and bulk experiments. This difference could also be one of the reasons for change in the threshold field at the onset of percolation. As seen in Fig. 5, the thin films also display an IP in the region $2 \leq x \leq 6$, where it is seen that the samples have a low threshold field which could make them potential phase change materials.

C. Set–Reset Studies in $Si_{15}Te_{85-x}Cu_x$ ($1 \leq x \leq 10$) Bulk Glassy System

A study on Set–Reset operation has been carried out on bulk samples of 0.3 mm thickness. To carry out the operation, triangular pulse of 6 mA for Set and rectangular pulse of 12 mA for Reset has been used for bulk samples. The bulk samples of $Si_{15}Te_{85-x}Cu_x$ ($1 \leq x \leq 10$) glasses have been found to manifest 5–10 Set–Reset cycles repeatedly. The Set–Reset operation on the representative $Si_{15}Te_{79}Cu_6$ glass is shown in Fig. 6. In the SET operation the sample behaves as a semipermanent storage of information as the sample transits from amorphous to crystalline phase with the introduction of applied current or voltage pulse. The phase is reversed to amorphous by the Reset operation wherein on the application of a sharp current pulse of higher magnitude, local melting

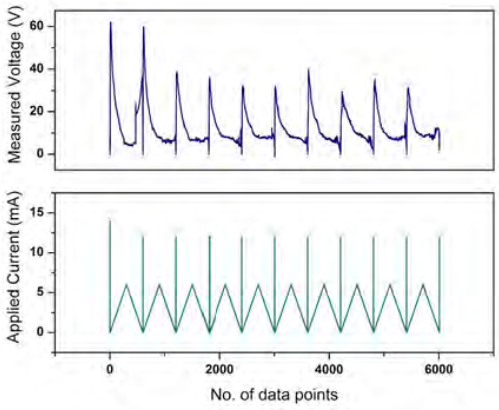


Fig. 6. Electrical switching response of $Si_{15}Te_{79}Cu_6$ glass for a triangular Set pulse of 6-mA amplitude and rectangular Reset pulse of 12-mA amplitude.

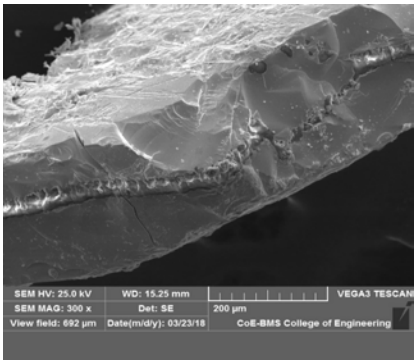


Fig. 7. SEM image of $Si_{15}Te_{77}Cu_8$ showing formation of channel-like structure between the electrodes.

of conducting crystal and amorphization occurs. In Fig. 6, the width of set pulse is 500 ms and that of reset pulse is 1 ms.

D. Morphological Studies on $Si_{15}Te_{85-x}Cu_x$ System Edge Switching

In order to understand the phenomena associated with the switching of $Si_{15}Te_{85-x}Cu_x$ ($1 \leq x \leq 10$) bulk glasses, switching on the edge of a representative bulk sample is carried out. There have been several models (both electronic and thermal) proposed to understand the mechanism of TS and MS in glassy chalcogenides.

For any qualitative analysis of switching behavior, both thermal and electronic effects must be considered, and the two can produce a coupled response called "electrothermal effect." Both the initiation of switching and the maintenance of the filamentary ON-state in chalcogenide threshold switches are fundamentally electronic in nature [30]. The most important role played by the thermal effects is the latching of the sample to the ON-state (MS) which involves a thermally induced amorphous-to-crystalline phase transition in the conducting channel [31]. Fig. 7 is the SEM image of $Si_{15}Te_{77}Cu_8$ bulk sample with a thickness of 0.3 mm, after conducting a switching operation at the edge of the sample. From Fig. 7, a channel-like structure is evident which supports the concept of

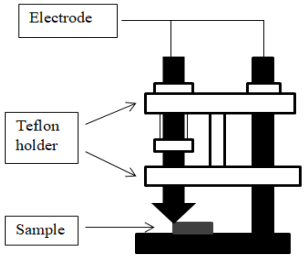


Fig. 8. Schematic of edge switching test device and the sample.

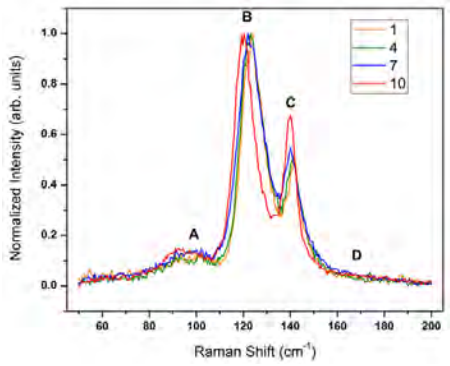


Fig. 9. Intensity normalized Raman spectra of selected as prepared Si–Te–Cu glasses. Band A ~ 97.9–92.3 cm^{-1}, B ~ 123.9–120.7 cm^{-1}, C ~ 140.9–139.5 cm^{-1}, and D ~ 165–175 cm^{-1}.

conducting channel like formation during switching. However, further studies on this channel-like structure are not carried out at this stage of research. The schematic of edge switching test device and the sample has been shown in Fig. 8.

E. Raman Studies on $Si_{15}Te_{85-x}Cu_x$ Bulk Glasses

Raman studies are performed on the glassy samples in the backscattering mode. The presented Raman spectra (Fig. 9) are intensity normalized. There are two dominating high-intensity modes (B and C) and two very low intense humps (A and D). The band positions have the following ranges of the wavenumber: band A ~ 97.9–92.3 cm^{-1}, B ~ 123.9–120.7 cm^{-1}, C ~ 140.9–139.5 cm^{-1}, and D ~ 165–175 cm^{-1}. The A and C bands are attributed to the E_{TO} modes of crystalline Te–Te chain. Band B is attributed to A_1 mode of crystalline Te–Te chain [32]. The ordered chains (Te_n) of $c-$Te are recognized to have analogous phonon frequencies at 123 and 143 cm^{-1}.

Band A features chains of Te which are distorted and can be either phase separated by photo-induced bond cleavage or can be attached at the end of each chain of structural elements existing in the glass. Band D can be expected from the stretching mode localized in edge-sharing tetrahedral $SiTe_{4/2}$ units [32]. Band D is the least intense among all. Shift in band A and D does not show much regularity. Band B and C exhibit a significant blue shift over composition, in a regular manner. Within the composition range $Si_{15}Te_{84}Cu_1$ and $Si_{15}Te_{75}Cu_{10}$, band B shifts from 123.95 to 120.69 cm^{-1}. The peak intensity remains almost same, within error, for all samples.

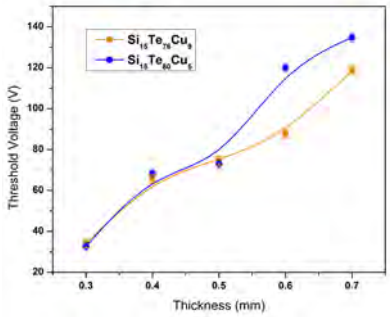


Fig. 10. Thickness dependence of threshold voltage of representative $Si_{15}Te_{80}Cu_5$ and $Si_{15}Te_{76}Cu_9$ glasses.

Within the composition range $Si_{15}Te_{84}Cu_1$ and $Si_{15}Te_{75}Cu_{10}$, band C shifts from 140.89 to 139.53 cm^{-1}. The peak intensity increases significantly with the inclusion of Cu. This blue shift with increasing Cu content in the sample suggests: 1) frequency of phonons interacting with the incident photon is increased; 2) tensile strain in the material lattice is increased; and 3) phonon in the material gains more energy from photons [33]. The blue shift can be anticipated to occur at RPT. The gradual addition of Cu atoms in the range of composition causes cleavage of Te_n chains which in turn blue-shifts the optical-absorption edge and leaving a bulk glass rich in Si inside which prevails metastable clusters of $SiTe_{4/2}$ units tetrahedral in shape [32].

It is worthy to mention here that, in this work, Raman is performed on as-prepared bulk glassy samples; these samples are neither switched nor annealed. The main purpose to perform Raman on as-prepared samples is to identify the signature of RPT. This does not involve crystalline phase formation but a distribution of atomic coordination, which in turn, modifies the mechanical, optical, and electrical behavior of these types of glasses. Thus, by performing Raman the samples are being classified, but how the electric pulse (switching) or temperature (annealing)-induced crystallization takes place in the microstructure of these glasses is not investigated.

F. Thickness Dependence of Threshold Voltage

The V_T of glassy samples manifesting MS is seen to exhibit proportionality to t or $t^{1/2}$, where t denotes the corresponding thickness of the glasses. Fig. 10 shows the trend of V_T with the increase in thickness for two representative $Si_{15}Te_{80}Cu_5$ and $Si_{15}Te_{76}Cu_9$ glasses.

Composition like Ge–Te–Al [34] and $Ge_{18}Te_{82-x}Bi_x$ ($1 \leq x \leq 4$) [35] exhibits a linear dependence of V_T against t whereas Ge–As–Te [12] is found to exhibit $t^{1/2}$ relation with thickness. The representative $Si_{15}Te_{80}Cu_5$ and $Si_{15}Te_{76}Cu_9$ is found to manifest a linear variation of V_T with t.

IV. Conclusion

The chalcogenide glasses prepared by the melt quenching technique have been found to exhibit MS behavior followed by the manifestation of same type of behavior in amorphous thin films which are designed in sandwich geometry. The threshold voltages of thin films are found to be much lower than the bulk counterparts and hence could be suitable for

PCM applications. Both the bulk and thin films have been seen to display IP in the composition range $2 \leq x \leq 6$. Amorphous thin films do not show one to one correspondence of percolation effect when compared with bulk, and the various factors responsible for such behavior are assumed to be high field conductions, methods of deposition of thin films, difference of electrode materials in bulk and thin films. Electrical switching operation conducted at the edge of bulk samples has shown the exhibition of MS behavior and also the formation of channel-like structure, which has helped to correlate the phenomena associated with the switching effect in $Si_{15}Te_{85-x}Cu_x$ ($1 \leq x \leq 10$) glasses. Set–Reset operation on a representative bulk samples of 0.3 mm thickness using a triangular pulse of 6 mA for Set and a rectangular pulse of 12 mA for Reset have shown a manifestation of 5–10 Set–Reset cycles repeatedly. Raman studies performed on the glassy samples have shown an exhibition of significant blue shift by Band B and C. The blue shift is exhibited by the glass at RPT. Linear variation of V_T with thickness is evident in these glasses.

REFERENCES

[1] G. W. Burr *et al.*, "Phase change memory technology," *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron., Mater., Process., Meas., Phenomena*, vol. 28, no. 2, pp. 223–262, Mar. 2010, doi: 10.1116/1.3301579.

[2] H. S. P. Wong *et al.*, "Phase change memory," *Proc. IEEE*, vol. 98, no. 12, pp. 2201–2227, Dec. 2010, doi: 10.1109/JPROC.2010.2070050.

[3] M. H. R. Lankhorst, B. W. Ketelaars, and R. A. M. Wolters, "Low-cost and nanoscale non-volatile memory concept for future silicon chips," *Nature Mater.*, vol. 4, no. 4, pp. 347–352, Mar. 2005, doi: 10.1038/nmat1350.

[4] T. J. Moore, T. L. Gilton, and K. A. Campbell "Methods to form a memory cell with metal-rich metal chalcogenide," U.S. Patent 7 056 762, B2, Jun. 6, 2006.

[5] T. J. Carrig, "Transition-metal-doped chalcogenide lasers," *J. Electron. Mater.*, vol. 31, no. 7, pp. 759–769, Jul. 2002, doi: 10.1007/s11664-002-0233-1.

[6] E. Mytilineou, "Chalcogenide amorphous semiconductors: Chemical modification or doping," *J. Optoelectron. Adv. Mater*, vol. 4, pp. 705–710, Sep. 2002.

[7] Z. U. Borisova, *Glassy Semiconductors*. New York, NY, USA: Springer, 1981, pp. 317–343, doi: 10.1007/978-1-4757-0851-6.

[8] J. Z. Liu and P. C. Taylor, "The formal valence shell model for structure of amorphous semiconductors," *J. Non-Crystalline Solids*, vol. 114, pp. 25–30, Dec. 1989, doi: 10.1016/0022-3093(89)90057-4.

[9] K. B. Jagannatha, D. Roy, and C. Das, "Electrical switching and crystalline peak studies on $Si_{20}Te_{80-X}Sn_X$ ($1 \leq X \geq 7$) chalcogenide bulk glasses," *J. Non-Cryst. Solids*, vol. 544, May 2020, Art. no. 120196, doi: 10.1006/j.jnoncrysol.2020.120196.

[10] G. Sreevidya Varma, D. V. S. Muthu, A. K. Sood, and S. Asokan, "Electrical switching, SET-RESET, and Raman scattering studies on $Ge_{15}Te_{80-X}In_5Ag_X$ glasses," *J. Appl. Phys.*, vol. 115, no. 16, Apr. 2014, Art. no. 164505, doi: 10.1063/1.4873237.

[11] A. C. Warren, "Thermal switching in semiconducting glasses," *J. Non-Crystalline Solids*, vol. 4, pp. 613–616, Apr. 1970, doi: 10.1016/0022-3093(70)90098-0.

[12] H. J. Stocker, C. A. Barlow, and D. F. Weirauch, "Mechanism of threshold switching in semiconducting glasses," *J. Non-Crystalline Solids*, vol. 4, pp. 523–535, Apr. 1970, doi: doi.org/10.1016/0022-3093(70)90088-8.

[13] B. J. Fernandes, K. Sridharan, P. Munga, K. Ramesh, and N. K. Udayashankar, "Memory type switching behavior of ternary $Ge_{20}Te_{80-X}Sn_X$ ($0 \leq X \leq 4$) chalcogenide compounds," *J. Phys. D, Appl. Phys.*, vol. 49, no. 29, Jun. 2016, Art. no. 295104, doi: 10.1088/0022-3727/49/29/295104.

[14] S. Asokan and K. P. Lakshmi, "Electrical switching and other properties of chalcogenide glasses," *J. Indian Inst. Sci.*, vol. 91, pp. 319–330, Apr./Jun. 2011.

[15] B. J. Fernandes, K. Ramesh, and N. K. Udayashankar, "Electrical switching in $Si_{20}Te_{80-X}Bi_X$ ($0 \leq X \leq 3$) chalcogenide glassy alloys," *J. Non-Crystalline Solids*, vol. 483, pp. 86–93, Mar. 2018, doi: 10.1016/j.jnoncrysol.2018.01.001.

[16] S. Prakash, S. Asokan, and D. B. Ghare, "Electrical switching behaviour of semiconducting aluminium telluride glasses," *Semicond. Sci. Technol.*, vol. 9, no. 8, pp. 1484–1488, Aug. 1994, doi: 10.1088/0268-1242/9/8/007.

[17] S. H. Hunter, A. Biennenstack, and T. M. Hayes, *The Structure of Non-Crystalline Materials*, P. H. Gaskell, Ed. London, U.K.: Taylor and Francis, 1977.

[18] L. Pauling, *Uniones Quimicas*. Kapelusz, Argentina: Buenos Aires, 1969.

[19] R.B. Heslop and K. Jones, *Inorganic Chemistry*. Amsterdam, The Netherlands: Elsevier, 1976.

[20] S. R. Gunti and S. Asokan, "Thermal and electrical switching studies on $Ge_{20}Se_{80-X}Bi_X$ ($1 \leq X \leq 13$) ternary chalcogenide glassy system," *J. Non-Crystalline Solids*, vol. 356, nos. 33–34, pp. 1637–1643, Jul. 2010, doi: 10.1016/j.jnoncrysol.2010.06.028.

[21] D. Roy, B. Tanujit, G. S. Varma, S. Asokan, and C. Das, "Manifestation of intermediate phase in Cu doped Si-Te glasses," *J. Non-Crystalline Solids*, vol. 531, Mar. 2020, Art. no. 119863, doi: 10.1016/j.jnoncrysol.2019.119863.

[22] C. N. Murthy, V. Ganesan, and S. Asokan, "Electrical switching and topological thresholds in Ge-Te and Si-Te glasses," *Appl. Phys. A, Solids Surf.*, vol. 81, no. 5, pp. 939–942, Oct. 2005, doi: 10.1007/s00339-005-3221-5.

[23] P. Boolchand, D. G. Georgiev, and B. Goodman, "Discovery of the intermediate phase in chalcogenide glasses," *J. Optoelectron. Adv. Mater.*, vol. 3, no. 3, pp. 703–720, 2001.

[24] X. Feng, W. J. Bresser, and P. Boolchand, "Direct evidence for stiffness threshold in chalcogenide glasses," *Phys. Rev. Lett.*, vol. 78, no. 23, pp. 4422–4425, Jun. 1997, doi: 10.1103/PhysRevLett.78.4422.

[25] D. Roy *et al.*, "Electrical switching studies on $Si_{15}Te_{85-X}Cu_X$ bulk ($1 \leq X \leq 5$) glasses," *Proc. AIP Conf.*, 1966, Art. no. 020033, doi: 10.1063/1.5038712.

[26] R. T. A. Kumar, C. Das, S. Asokan, C. Sanjeeviraja, and D. P. Padiyan, "Optical, photo-acoustic and electrical switching studies of amorphous GeS_2 thin films," *Appl. Phys. A, Solids Surf.*, vol. 115, pp. 1151–1158, Jun. 2014, doi: doi.org/10.1007/s00339-013-7976-9.

[27] C. Das, G. M. Rao, and S. Asokan, "Electrical switching behavior of amorphous $Ge_{15}Te_{85-X}Si_X$ thin films with phase change memory applications," *Mater. Res. Bull.*, vol. 49, pp. 388–392, Jan. 2014, doi: 10.1016/j.materresbull.2013.09.020.

[28] D. Roy, G. S. Varma, S. Asokan, and C. Das, "Electrical switching and optical bandgap studies on quaternary ag-doped Ge–Te–In thin films," *IEEE Trans. Electron Devices*, vol. 66, no. 4, pp. 1881–1886, Apr. 2019, doi: 10.1109/TED.2019.2898935.

[29] R. D. Gould, S. Kasap, and A. K. Ray, "Thin films," in *Springer Handbook of Electronic and Photonic Materials*, S. Kasap and P. Capper, Eds. Cham, Switzerland: Springer, 2017, pp. 645–702, doi: doi.org/10.1007/978-3-319-48933-9_28.

[30] D. M. Kroll, "Theory of electrical instabilities of mixed electronic and thermal origin. II. Switching as a nucleation process," *Phys. Rev. B, Condens. Matter*, vol. 11, no. 10, pp. 3814–3821, May 1975, doi: 10.1103/PhysRevB.11.3814.

[31] M. P. Shaw and A. Madan, *The Physics and Applications of Amorphous Semiconductors*. London, U.K.: Academic, 1988.

[32] S. Rao Gunti and S. Asokan, "Thermodynamic, Raman and electrical switching studies on $Si_{15}Te_{85-X}Ag_X$ ($4 \leq X \leq 20$) glasses," *J. Appl. Phys.*, vol. 111, no. 3, Feb. 2012, Art. no. 033518, doi: 10.1063/1.3682759.

[33] C. N. Banwell and E. M. McCash, "Raman spectroscopy," in *Fundamentals of Molecular Spectroscopy*, 4th ed. New York, NY, USA: McGraw-Hill, 2016, pp. 100–126.

[34] J. R. Bosnell and C. B. Thomas, "Preswitching electrical properties,'forming', and switching in amorphous chalcogenide alloy threshold and memory devices," *Solid-State Electron.*, vol. 15, pp. 1261–1271, Nov. 1972, doi: doi.org/10.1016/0038-1101(72)90047-0.

[35] C. Das, G. M. Rao, and S. Asokan, "Electrical switching and thermal studies on bulk Ge–Te–Bi glasses," *J. Non-Crystalline Solids*, vol. 357, no. 1, pp. 165–169, Jan. 2011, doi: 10.1016/j.jnoncrysol.2010.09.046.

Applied Physics A (2022) 128:36
https://doi.org/10.1007/s00339-021-05174-5

Applied Physics A
Materials Science & Processing

Nanoarchitectonics earth-abundant chalcogenide Cu_2SnS_3 thin film using ultrasonic spray pyrolysis for visible light-driven photocatalysis

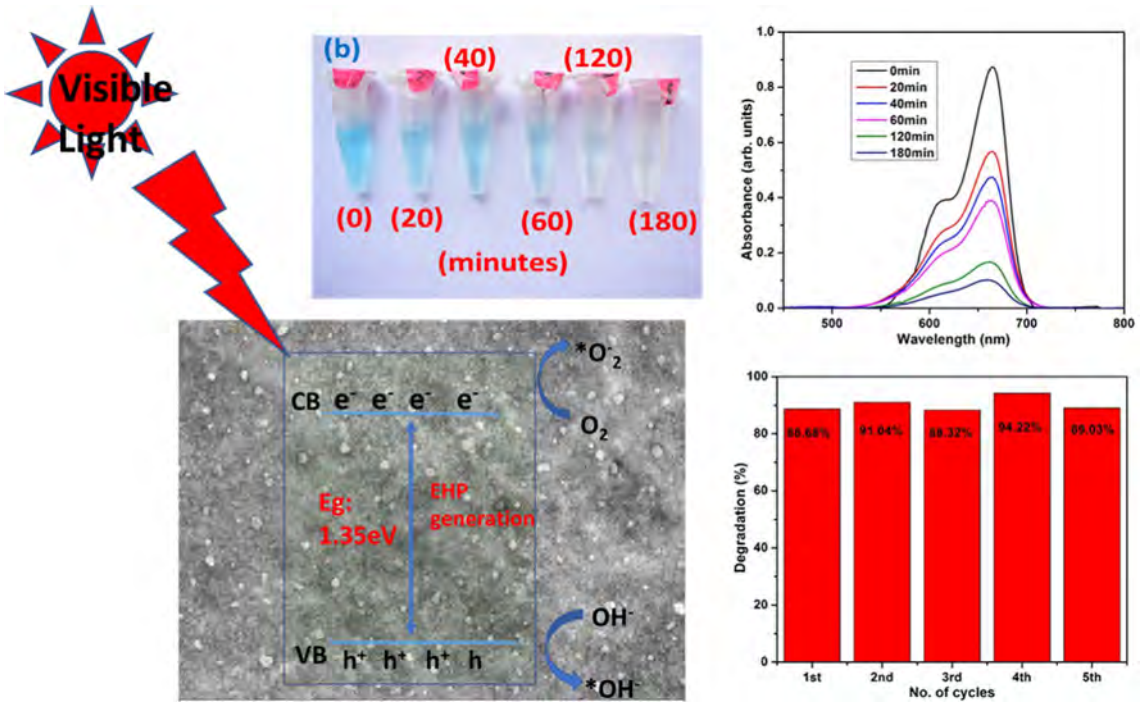
Sabina Rahaman[1] · **Monoj Kumar Singha**[2,3]

Received: 1 October 2021 / Accepted: 30 November 2021

Abstract

Photocatalysis is the process which used to remove the dye from the effluent. Here, we have used earth-abundant chalcogenide Cu_2SnS_3 (CTS) thin film as a photocatalyst to remove the dye (methylene blue) under visible light irradiation. A low cost ultrasonic spray pyrolysis was used to deposit the film on soda-lime glass substrate in a single step at optimized temperature 500 °C. The structural, morphological and optical properties of the CTS film have been studied using XRD, SEM and UV–Vis spectroscopy. Structural analysis confirms the formation of Cu_2SnS_3 tetragonal structure without any secondary impurities. SEM image indicates the surface of this film is smooth and uniform. Bandgap of film is found to be 1.35 eV. Photocatalysis activity of CTS thin film is studied by degrading methylene blue (MB) dye (1×10^{-5} M) in water solution under visible light irradiation. It degrades 90% MB in 3 h. To evaluate the industrial effluent, we have studied the photocatalytic activity in different pH (4, 7, 9) medium. It shows that the MB degradation is faster in the base medium compare to acidic or neutral medium. It takes 1 h to degrade 90% dye in base medium. Films are showing good repeatable performance of the photocatalytic activity.

Graphical abstract



Extended author information available on the last page of the article

Published online: 13 December 2021

Keywords Cu_2SnS_3 (CTS) thin film · Ultrasonic spray pyrolysis · Visible light photocatalyst · Repeatability · Different pH medium

1 Introduction

Harmful dyes produced from pharmaceutical effluent, textile, chemical sewages are polluting the aqua systems and environment. These toxic dyes are creating water pollution in our precious resource of fresh water. To prevent the water pollution from these dyes, there are various methods used and among them photocatalysis process is mostly used to degrade the dyes. Photocatalysis processes use photocatalyst (semiconductor materials) in the presence of light to degrade the dyes [1–3]. Mostly oxide-based semiconductor like titanium oxide (TiO_2), zinc oxide (ZnO) are used as photocatalyst [4–11]. Though oxide semiconductors are non-toxic, economical, earth-abundant and high chemical stability, they have shown photocatalytic activity under UV light due to a wide bandgap like 3.2 eV (TiO_2) and 3.37 (ZnO). It is very rare to use them as photocatalyst under natural visible light. Sunlight has around 3–5% UV spectrum, 42–43% visible spectrum and 52–55% infrared spectrum. So, most of the spectrum cannot be absorbed by theses oxide materials for photocatalytic application [12]. Therefore, it is very essential to develop new photocatalysts that can utilize UV, visible and infrared energy of the solar spectrum. In this regard, researchers have identified narrow bandgap semiconductor photocatalysts to degrade dyes under visible light irradiation. CdS, PbS, $ZnIn_2S_4$, In_2S_3, $CdIn_2S_4$ are example of narrow bandgap semiconductors which are used for photodegradation of dyes [13–18]. Unfortunately, these catalysts cannot be used for commercial application due to the presence of toxic elements Cd and expensive elements like In. Recently, metal–organic frameworks (MOFs), Functional Mesoporous Silica Nanomaterials, MXene and MXene-based composites in nanoparticles forms are also used for photocatalysis application [19–21]. But they are used in nanoparticles form. So, it is necessary to explore a different photocatalyst semiconductor material which is originated by earth-abundant elements and it should be low cost. In this regard, CZTS and CTS are two semiconductors which fulfill these criteria though CZTS and CTS are mostly used in solar cell applications [22–24].

Recently, CZTS and CTS have gained researcher attention for the photocatalysis process. These materials are formed from inexpensive earth-abundant material and they have lower bandgap. Between these two semiconductors, it is easy to deposit/synthesis CTS than CZTS due to one less number of elements. Q. S. Ren et.al. have synthesized Cu_2ZnSnS_4 photocatalyst using facile solvothermal techniques. Prepared sample exhibits sphere and rose-shaped structure. Rose-shaped structure showed higher percentage of methylene blue (MB) dye degradation compare to sphere shape structure [25]. J. Zia et.al have prepared $Bi_2O_2CO_3$ microstructure using the hydrothermal method and they have reported that photocatalyst activity has been improved with iodine (I) doping [26]. Jiasong Zhong et.al. have reported the synthesis of 3D hierarchical Cu_2FeSnS_4 microstructure using solvothermal method and they have shown 73% of degradation of RhB dye under visible light irradiation [27]. The above works have encouraged us to prepare non-toxic semiconductor material Cu_2SnS_3 (CTS) thin films for photocatalytic activity under visible light irradiation. CTS, a p-type direct bandgap semiconductor (0.9 -1.7 eV), is mostly used for photovoltaic device and it has a very large absorption coefficient ($\geq 10^4$ cm^{-1})[28]. Yu. Tan et.al. have prepared Cu_2SnS_3 powder sample using solvothermal techniques to obtain the band gap of 2.42 eV [12]. They have reported the complete decomposition of MB dye under four hours of visible light irradiation, but the effect of different pH medium and repeatability test is not studied. S.Vadivel et.al reported the study the photocatalysis activity of Cu_2SnS_3/RGO nanocomposites which is synthesized using solvothermal method [29]. Most of these photocatalytic experiments were performed when the photocatalysts were in powder form. But a photocatalyst in powder forms has demerits in practical applications. Because after experiment these powders have to filter out from the solution [10, 11]. So, a filtration unit needs to install there which will increase the total process cost. In this regard, presently researchers are looking for thin film based photocatalysis application. There are various methods available to deposit thin film. Among all the available methods to deposit thin films, spray pyrolysis is a single step deposition method to deposit thin films.

There are no reports of CTS thin films deposited by ultrasonic spray pyrolysis (USP) for photocatalytic applications. In this paper, we have discussed the deposition of CTS thin films using USP. Then, these films were used for photocatalytic application. Here, photocatalysis activity of CTS thin film is explored for degradation of test dye (methylene blue) solution under 300 W halogen lamp. Repeatability studies and the effect of different pH medium in CTS thin film as a photocatalyst is also studied here.

2 Experimental details

2.1 Synthesis of Cu_2SnS_3 thin films by ultrasonic spray pyrolysis

CTS thin film was deposited on the soda-lime glass substrate using an automated ultrasonic spray pyrolysis system (frequency 1.63 MHz) as discussed in our earlier report [30]. In brief, at first, glass substrates were washed thoroughly in soap water. Then, it was cleaned in deionized water, acetone and IPA ultrasonically for 30 min, respectively; followed by drying in hot air oven for 3 h at 80 °C. Before deposition of the films, precursor solutions were prepared from copper chloride ($CuCl_2$, H_2O), tin chloride ($SnCl_2$, $2H_2O$) and thiourea [$CS(NH_2)_2$] in 60 ml DI water. The concentration of tin chloride, copper chloride and thiourea were 0.01 M, 0.02 M and 0.2 M, respectively. Finally, prepared precursor solution was sprayed on the heated glass substrate (500 °C) for 30 min reported elsewhere [30].

2.2 Characterization

X-ray diffraction (XRD) was used to determine the crystal structure of prepared films using Bruker D8. Cu Kα is used as a radiation source which has a wavelength of 1.5405Å and it is operated at 40 kV and 40 mA, respectively. Horiba LABRAM HR is used to characterize Raman spectroscopy of the CTS sample with a 532 nm wavelength of Argon laser as an excitation source. Surface morphology and energy dispersive X-ray analysis (EDX) of the CTS sample is determined using FESEM Quanta-200 for scanning electron microscopy (SEM). Absorption spectra of CTS sample are measured using the UV–Vis spectrometer (SPECORD S600), and the bandgap is measured from the absorbance spectra through Tauc plot. The resistivity of CTS thin film is calculated using the four-point probe method (Four Dimensions probe metal model 280).

2.3 Photocatalytic studies

Photocatalytic test was performed in a double jacket photoreactor under the irradiation of visible light from 300 W halogen lamp. 60 ml methylene blue (MB) solution (1×10^{-5} M) with three pieces of CTS thin films (1.5×2.5 cm^2) was used for the photocatalysis studies. The films under MB solution are kept on magnetic stirrer for 30 min in the dark environment before light irradiation to achieve absorption desorption equilibrium between MB and CTS photocatalyst. The solution is then exposed under visible light obtained from a 300 W halogen lamp. The absorbance spectra of the solution are measured after a certain interval using a UV–vis spectrometer. Coldwater was circulated in the double jacket reactor during the experiment process to prevent the evaporation of MB solution.

3 Results and discussion

3.1 XRD

Figure 1 shows the XRD pattern of the CTS films deposited on glass substrate at 500 °C. The sample shows the diffraction peak found at 28.47°, 33.1° 47.14° and 56.36°, respectively, which corresponds to (112), (200), (204) and (312) planes. Four peaks are observed in the XRD pattern which clearly indicates that this film is polycrystalline in nature.

The diffraction peak at (112) has sharp and high-intensity peaks compared to other orientation, so it indicates that the preferred orientation of CTS thin films is (112). High-intensity peaks indicated the high crystallinity of the film. The XRD pattern is well-matched with the standard tetragonal structure of the CTS films (JCPDS 89–4715). The average crystallite size of the CTS film is found to be 43 nm from Debye Scherrer's formula.

3.2 Raman analysis

Raman analysis is a nondestructive tool to reconfirm the structure of CTS thin film and it also reveals the presence of secondary phases, if presents, in the multicompound film. Figure 2 depicts the Raman spectrum of CTS thin film. The sample shows peak at 334 cm^{-1} which corresponds to A1 vibration mode of tetragonal Cu_2SnS_3 and other peaks at 287 cm^{-1} is also matching with tetragonal Cu_2SnS_3 [31, 32].

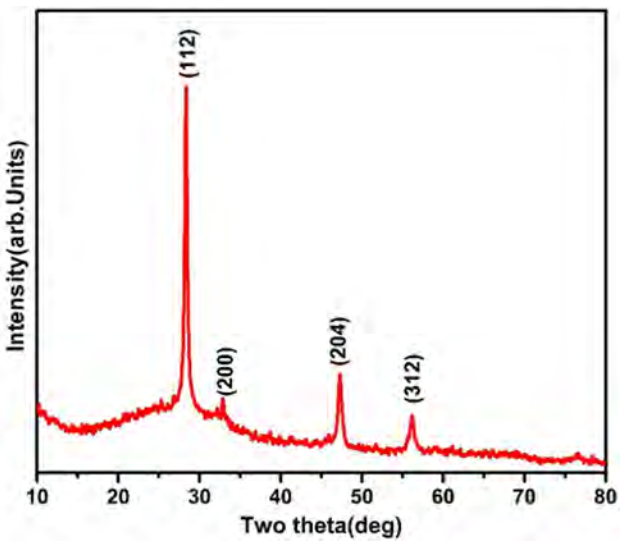


Fig. 1 X-ray diffraction pattern of CTS film prepared at 500 °C

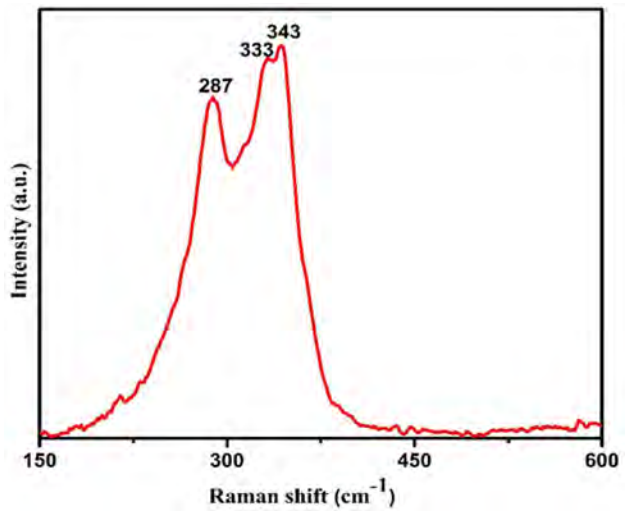


Fig. 2 Raman spectra of CTS film prepared at Ts = 500℃

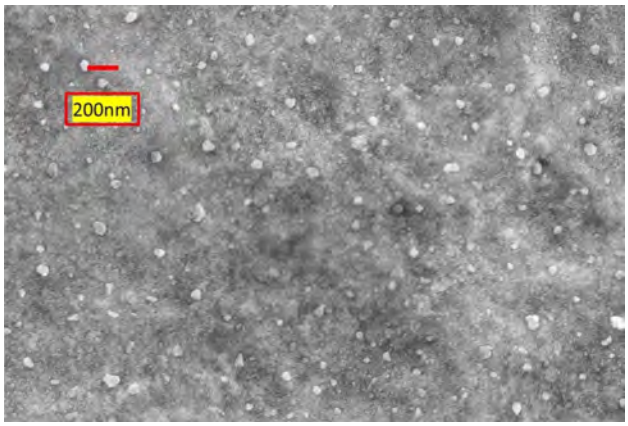


Fig. 3 The SEM image of the CTS sample prepared at Ts = 500℃

Similar kind of results also finds in the literature [31–34]. There is no binary phase observed in Raman spectra.

3.3 Morphological studies

Figure 3 depicts the SEM micrograph of CTS thin film deposited at temperature 500 °C by ultrasonic spray pyrolysis method. It is observed that the surface is quite smooth and uniform. There are no cracks and pinholes visible in the film.

EDX analysis is carried out to find the elemental composition in the films. Table 1 gives the atomic percentage of elements (Cu, Sn, and S) for the deposited sample. $Cu_{2.00}SnS_{2.71}$ is the stoichiometric ratio of as-deposited CTS thin film. It is clearly visible from the table that prepared film is found to be sulfur deficient due to the volatile nature of sulfur but it is almost near to Cu_2SnS_3 film.

Table 1 Elemental composition of CTS thin film deposited at Ts = 500 °C

Elements present in the CTS film	Atomic percentage (%)
Cu	35.02
Sn	17.50
S	47.48

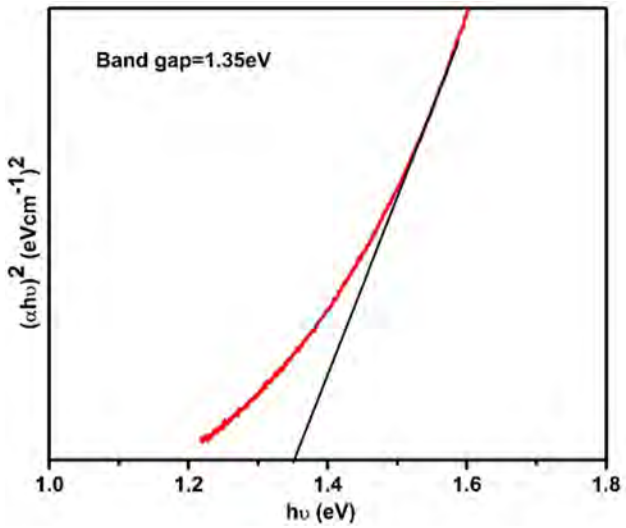


Fig. 4 Bandgap of the CTS sample prepared at 500 °C

3.4 Optical analysis

The optical property of semiconductor material plays a very important role to evaluate the efficiency of photodegradation of organic pollutants. Tauc formula is used to find the bandgap of CTS thin film using absorption spectra [35–37]. A graph is plotted between photon energy (hν) and $(\alpha h\nu)^2$. Then, a slope, plotted at the linear portion of $(\alpha h\nu)^2$, is extended to X-axis and the intersected point at X axis gives the band gap value of CTS thin film. The bandgap of CTS thin film is found to be 1.35 eV, which is matching with reported value[38]. Figure 4 shows the bandgap of the film.

3.5 Photocatalysis activity

CTS thin film as a photocatalyst for degradation of MB is presented here. The absorbance spectra of MB solution are measured at a different time(t) when it is illuminated with visible light. Under the visible light illumination with the presence of CTS thin film photocatalyst, MB has started to degrade. At different time interval, the absorption spectra of the solution are recroded. These absorbance spectra are proportionl to the concentration of the solution which helps to find the photodegradation efficiency. The photodegradation

efficiency of the deposited CTS sample for MB degradation is given by[39, 40]:

$$\text{Photocatalytic degradation efficiency} = \frac{C0 - Ct}{C0} \times 100\% \quad (1)$$

where C_0 is the concentration of the solution at time $t = 0$ and C_t is the concentration of the MB solution at time t in the presence of photocatalyst under visible light irradiation. Here, we have chosen the characteristic peak of the MB solution at $\lambda = 664$ nm. Figure 5a shows the absorbance spectra of MB at a different time under the illumination of visible light. It is clearly visible from the figure that the intensity of absorbance spectra is decreasing with increasing illuminated time. Figure 5b shows the color of the MB dye solution at different intervals of t. It shows that the color of MB dye changed from blue to transparent solution. Photocatalytic degradation efficiency is shown in Fig. 5c. Degradation efficiency is found to be ~ 90% under 3 h visible light irradiation. Degradation rate for a catalyst is calculated from the following equation

$$\ln\left(\frac{Ct}{C0}\right) = -Kt \quad (2)$$

where C_0, C_t and K are initial concentration, final concentration and pseudo-first-order kinetic constant, respectively.

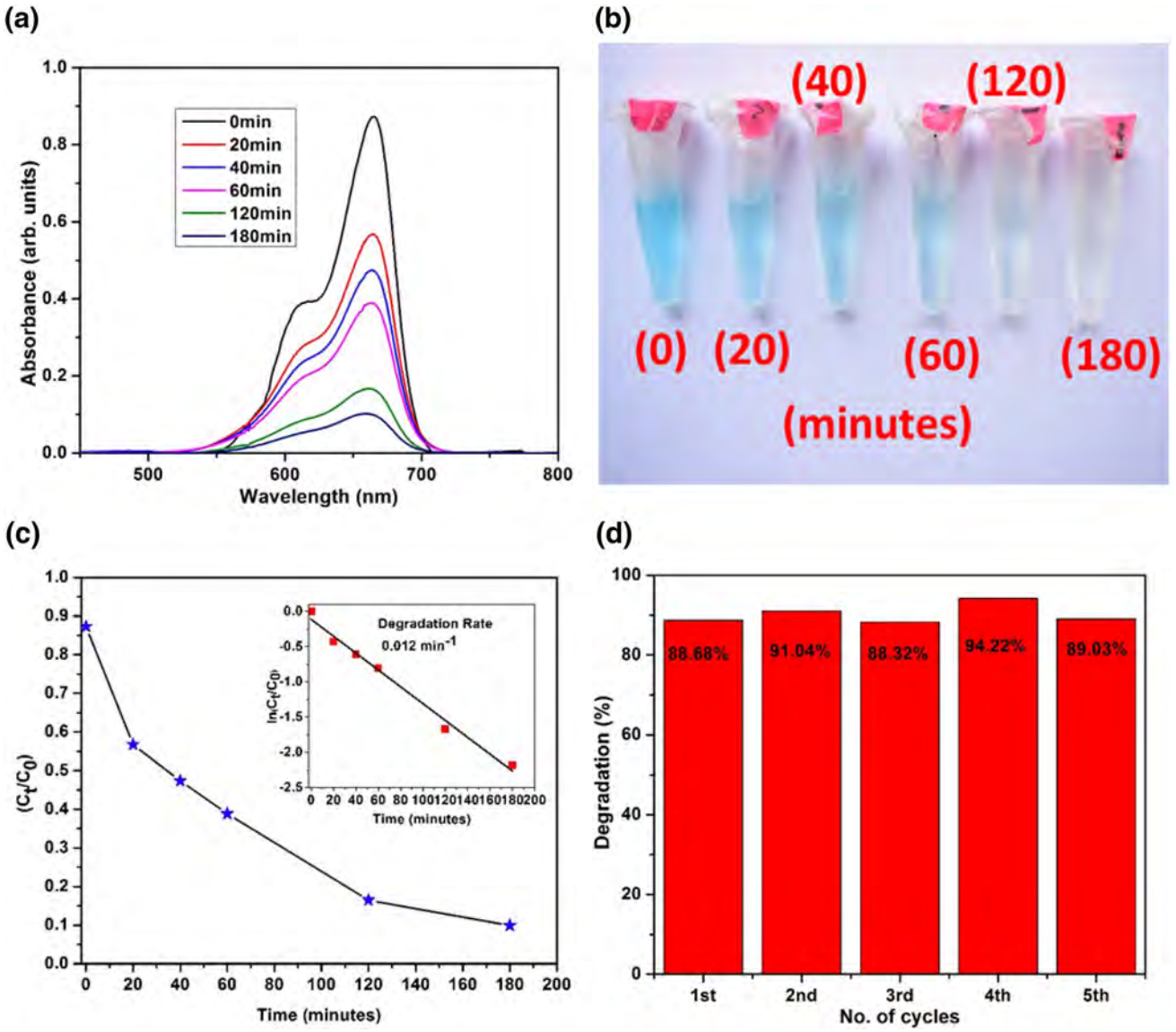


Fig. 5 **a** Absorbance spectra of MB under visible light illumination at different time, **b** discoloration of MB at different time using CTS photocatalyst, **c** Degradation of MB and (in inset) degradation of MB using CTS photocatalyst under visible light irradiation, and **d** Repeatability studies of CTS photocatalyst

Figure 5c (inset) shows the degradation rate of the photocatalyst and it is found to be 0.012/min. We have also studied the repeatability performance of CTS thin films for 5 times under the same condition to check the stability and reusability of the photocatalyst. Figure 5d shows the repeatability performance of CTS thin film as a photocatalyst. There is a negligible difference between the first and last degradation performance of the films and it shows the ~ 90% degradation of the MB solution.

3.6 Mechanism

Figure 6 shows the schematic visualization of the photocatalytic mechanism of CTS thin film under visible light irradiation. Bandgap of CTS is found to be 1.35 eV from the Tauc plot. As the band gap value is 1.35 eV, the CTS material starts to produce electron and hole pair when visible light is illuminated on its surface. These excited electrons at the conduction band react with oxygen and produce oxygen radicals. Similarly, holes at valence band produce the hydroxyl radicals. These radicals are highly reactive in nature and they react with the organic dyes. After the reactions, dyes are decomposed into less harmful products. The dye degradation mechanism is shown below equations.

3.7 pH study of CTS photocatalyst

Initially our experiments with CTS thin films for the degradation of MB are performed only in neutral water. But in practice, the solution containing dyes have different pH. To test the CTS thin film as a photocatalyst in different medium, we have also performed it photocatalytic activity in three different pH 9, 7 and 4 as basic, neutral and acidic medium, respectively. Figure 7 shows the degradation pattern and bar diagram of MB at different medium for 90 min of illumination using CTS photocatalyst. It is clearly visible that degradation rate is faster in the base medium. At base medium, extra OH ions are available

$$CTS + h\nu \longrightarrow CTS(e^- + h^+) \quad \text{.....(3)}$$

$$H_2O + h^+ \longrightarrow \bullet OH + H^+ \quad \text{.....(4)}$$

$$O_2 + e^- \longrightarrow \bullet O_2^- \quad \text{.....(5)}$$

$$O^-_2 + H^+ \longrightarrow \bullet HO_2 \quad \text{.....(6)}$$

$$\bullet HO_2 + H^+ \longrightarrow H_2O_2 \quad \text{.....(7)}$$

$$H_2O_2 + e^- \longrightarrow \bullet OH + OH^- \quad \text{.....(8)}$$

$$\bullet O_2^-, \bullet OH, \bullet HO_2 + \text{Dye} \longrightarrow \text{Degraded product} \quad \text{.....(9)}$$

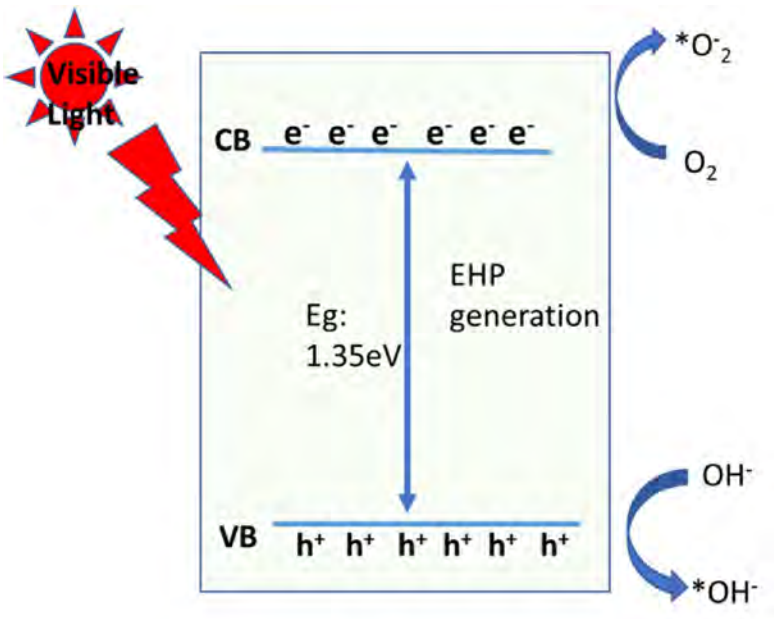


Fig. 6 Schematic diagram of photocatalytic mechanism

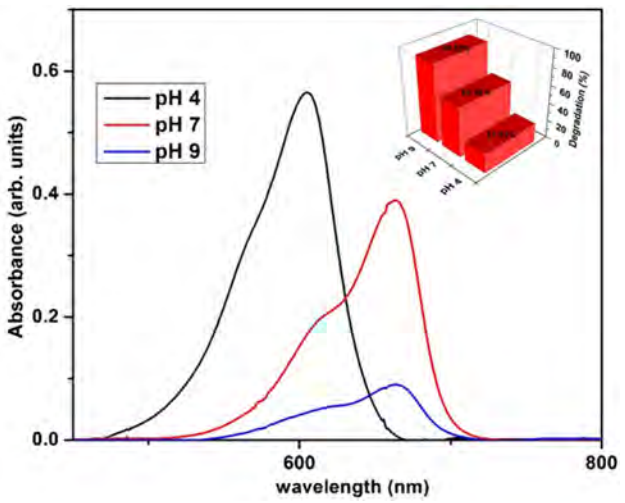


Fig. 7 Absorbance spectra of MB at different pH medium under 60 min visible light irradiation (inset: degradation % of MB bar diagram at different pH under 60 min light irradiation)

which later transforms into superhydroxide radicals. These excess radicals use to degrade the dye faster. Degradation performance of CTS film is lower in acidic medium than neutral pH solution and a blue shift of absorbance peak is found in the graph. Shifting of the peak is due to the promotion effect in acidic solution [41, 42]. As there are excess H^+ ions available in the acidic medium, it will then reacts with OH^- to form H_2O. As the number of OH radicals is reduced, the degradation process also reduces. So, the degradation rate is poor in the acidic medium. Samples are not corroded in the acidic medium as we have performed the next experiment in another pH 7 solution and found the same photocatalytic behavior of the film.

Different research groups have used thin films and nanoparticles as a photocatalyst, and Table 2 shows the small summary of their research in degrading different dyes. It is found from the table that thin films are rarely used for photocatalysis activity. There are huge advantages of thin film based photocatalysis compare to nanoparticles based photocatalyst as described in the introduction section. In this scenario, CTS thin films will be used not only as solar cell absorber material but as well as it will be used as photocatalysis process and CTS thin films are comparable with other thin film (metal oxide) based photocatalyst. Even our results show better performance for MB degradation under visible light environment compared to other earth abundant chalcogenide material. Our works also studied the photocatalytic activity in different pH medium.

Table 2 Comparative results of different thin film and nanoparticle photocatalyst

Photocatalyst material	Deposition technique	Dye solution	Light source	Degradation time and %	References
copper zinc tin sulfide (CZTS) thin film	SILAR	Basic Blue 41 (BB-41)	Four fluorescent lamps, of 4 W nominal power	180 min, 97.5%	[43]
copper zinc tin sulfide (CZTS) thin film	Spray pyrolysis	Sugarcane industry waste	UV lamps	After 120 min, 90%	[44]
Cu_2SnS_3/Ti_3+-TiO_2 thin film	Hydrothermal method	Tetracycline	1000 W halide lamp	98%	[45]
Cu_2FeSnS_4 thin film	Spray pyrolysis	Methylene blue (MB)	55 W Xenon cylindrical light	Four hours, 76%	[46]
ZnO thin film	Spray pyrolysis	MB	UV–Vis lamp	3.5 h, 90%	[47]
Cu_3SnS_4 nanoparticle	Ball milling (CTS-B) and solvothermal method (CTS-S)	MB	250 W halogen lamp	After 90 min, 42% degradation showed by CTS-B and 95% degradation for CTS-S	[48]
Titania nanoparticles	Hydrothermal method	MB	6 W Lamp (BoittonInstruments) as the light source	180 min. 97%	[49]
ZnO nanoparticles	Co-precipitation method	Rose Bengal (RB), (MB), Bromocresol green (BG)	15 W UV lamps	After 240 min, RB-83%, MB-81%, BG-56% for pH = 11	[50]
Copper zinc tin sulfide (CZTS) nanoparticles	Hydrothermal	MB	Visible light	After 45 min, 50%	[51]
Copper zinc tin sulfide (CZTS) nanoparticles	Solvothermal process	MB	Halogen lamp of 300 W	120 min, 91%	[41]
CTS Thin film	Ultrasonic Spray pyrolysis	MB	300 W Halogen lamp (visible light)	3 h, 90%, Different pH condition	This work

4 Conclusion

In a single step method, CTS thin films are successfully deposited on soda lime glass substrate at 500 °C by indigenously developed automated ultrasonic spray pyrolysis. It takes 2 h to complete the thin film deposition including precursor preparation time. Tetragonal single-phase crystal structure without any secondary impurity is found from the XRD. The crystal structure has been further reconfirmed by Raman analysis. A dense uniform surface morphology has been observed in CTS films. The photocatalytic activity showed 90% degradation of MB dye under 3 h visible light irradiation. Photocatalytic activity is repeated for five cycle and the good stability of CTS thin film is observed, i.e., CTS thin films are not photo corroded even after five repaetivive experiment. It consistently shows ~ 90% degradation of MB dye in each experiment. Not only these photocatalysis experiments performed in the neutral (pH 7) medium but also performed in acidic (pH 4) and base (pH 9) medium to mimic industrial effluent pH. It is found that the degradation rate is increased in base medium compare to acidic medium due to the presence of extra OH ions which are available in the base medium. These excess radicals are used to accelerate the degradation of the dye.

Acknowledgements We thank to MNCF, IISc to provide us to use characterization facilities for above work.

References

1. A.B. Djurišić, Y.H. Leung, A.M.C. Ng, Strategies for improving the efficiency of semiconductor metal oxide photocatalysis. Mater. Horizons. **1**, 400–410 (2014)
2. N. Shaham-Waldmann, Y. Paz, Away from TiO_2: a critical mini-review on the developing of new photocatalysts for degradation of contaminants in water. Mater. Sci. Semicond. Process. **42**, 72–80 (2016). https://doi.org/10.1016/j.mssp.2015.06.068
3. A. Mills, S. Le Hunte, An overview of semiconductor photocatalysis. J. Photochem. Photobiol. A Chem. **108**, 1–35 (1997)
4. I.A. Perales-Martínez, V. Rodríguez-González, S.W. Lee, S. Obregón, Facile synthesis of $InVO_4/TiO_2$ heterojunction photocatalysts with enhanced photocatalytic properties under UV-vis irradiation. J. Photochem. Photobiol. A Chem. **299**, 152–158 (2015). https://doi.org/10.1016/j.jphotochem.2014.11.021
5. H. Han, R. Bai, Highly effective buoyant photocatalyst prepared with a novel layered-TiO_2 configuration on polypropylene fabric and the degradation performance for methyl orange dye under UV–Vis and Vis lights. Sep. Purif. Technol. **73**, 142–150 (2010)
6. Y. Li, J. Yan, Q. Su, E. Xie, W. Lan, Preparation of graphene–TiO_2 nanotubes/nanofibers composites as an enhanced visible light photocatalyst using a hybrid synthetic strategy. Mater. Sci. Semicond. Process. **27**, 695–701 (2014)
7. 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**, 68–76 (2016)
8. A. Samad, M. Furukawa, H. Katsumata, T. Suzuki, S. Kaneco, Photocatalytic oxidation and simultaneous removal of arsenite with CuO/ZnO photocatalyst. J. Photochem. Photobiol. A Chem. **325**, 97–103 (2016)
9. M.K. Singha, A. Patra, V. Rojwal, K.G. Deepa, D. Kumar, Ultrasonic spray pyrolysis deposited ZnO thin film for photocatalytic activity. AIP Conf. Proc. **2082**, 030023 (2019). https://doi.org/10.1063/1.5093841
10. M.K. Singha, A. Patra, Highly efficient and reusable ZnO microflower photocatalyst on stainless steel mesh under UV-Vis and natural sunlight. Opt. Mater. **107**, 110000 (2020)
11. A. Samad, M. Furukawa, H. Katsumata, T. Suzuki, S. Kaneco, Single step fabrication of ZnO microflower thin films for highly efficient and reusable photocatalytic activity. J. Mater. Sci.: Mater. Electron. **325**, 13578–13587 (2020)
12. 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**, 240–242 (2012). https://doi.org/10.1016/j.matlet.2012.08.117
13. F. Vaquero, R.M. Navarro, J.L.G. Fierro, Influence of the solvent on the structure, morphology and performance for H_2 evolution of CdS photocatalysts prepared by solvothermal method. Appl. Catal. B Environ. **203**, 753–767 (2017)
14. S.K. Batabyal, S.E. Lu, J.J. Vittal, Synthesis, characterization, and photocatalytic properties of In_2S_3, $ZnIn_2S_4$, and $CdIn_2S_4$ nanocrystals. Cryst. Growth Des. **16**, 2231–2238 (2016)
15. R. Sumi, A.R. Warrier, C. Vijayan, Visible-light driven photocatalytic activity of $β$-indium sulfide (In_2S_3) quantum dots embedded in Nafion matrix. J. Phys. D. Appl. Phys. **47**, 105103 (2014)
16. N. Ding, Y. Fan, Y. Luo, D. Li, Q. Meng, Enhancement of H_2 evolution over new $ZnIn_2S_4$/RGO/MoS_2 photocatalysts under visible light. APL Mater. **3**, 104417 (2015)
17. M.A. Zhukovskiy, A.L. Stroyuk, V.V. Shvalagin, N.P. Smirnova, O.S. Lytvyn, A.M. Eremenko, Photocatalytic growth of CdS, PbS, and CuxS nanoparticles on the nanocrystalline TiO_2 films. J. Photochem. Photobiol. A Chem. **203**, 137–144 (2009)
18. X. Yang, Z. Wang, X. Lv, Y. Wang, H. Jia, Enhanced photocatalytic activity of Zn-doped dendritic-like CdS structures synthesized by hydrothermal synthesis. J. Photochem. Photobiol. A Chem. **329**, 175–181 (2016)
19. Q. Wang, Q. Gao, A.M. Al-Enizi, A. Nafady, S. Ma, Recent advances in MOF-based photocatalysis: environmental remediation under visible light. Inorg. Chem. Front. **7**, 300–339 (2020)
20. B. Singh, J. Na, M. Konarova, T. Wakihara, Y. Yamauchi, C. Salomon, M.B. Gawande, Functional mesoporous silica nanomaterials for catalysis and environmental applications. Bull. Chem. Soc. Jpn. **93**, 1459–1496 (2020)
21. X. Zhan, C. Si, J. Zhou, Z. Sun, MXene and MXene-based composites: synthesis, properties and environment-related applications. Nanoscale Horiz. **5**, 235–258 (2020)
22. S. Rahaman, M.K. Singha, M.A. Sunil, K. Ghosh, Effect of Copper concentration on CTS thin films for solar cell absorber and photocatalysis application. Superlatt. Microstruct. **147**, 106589 (2020)
23. S. Rahaman, M.A. Sunil, M.K. Singha, K. Ghosh, Studies of ultrasonically sprayed Cu_2SnS_3 thin films by varying Sn concentration. Mater. Today Proceed. (2021). https://doi.org/10.1016/j.matpr.2021.02.657
24. S. Rahaman, M.A. Sunil, M.K. Singha, K. Ghosh, Ultrasonic spray pyrolysis deposited CTS thin film: variation of thiourea concentration in the film. Mater. Today Proceed. (2021). https://doi.org/10.1016/j.matpr.2021.04.513
25. Q. Ren, W. Wang, H. Shi, Y. Liang, Synthesis and shape-dependent visible-light-driven photocatalytic activities of Cu_2ZnSnS_4 nanostructures, Micro. Nano Lett. **9**, 505–508 (2014)

26. J. Zai, F. Cao, N. Liang, K. Yu, Y. Tian, H. Sun, X. Qian, Rose-like I-doped $Bi_2O_2CO_3$ microspheres with enhanced visible light response: DFT calculation, synthesis and photocatalytic performance. J. Hazard. Mater. **321**, 464–472 (2017). https://doi.org/10.1016/j.jhazmat.2016.09.034
27. J. Zhong, Q. Wang, D. Chen, L. Chen, H. Yu, H. Lu, Z. Ji, Biomolecule-assisted solvothermal synthesis of 3D hierarchical Cu_2FeSnS_4 microspheres with enhanced photocatalytic activity. Appl. Surf. Sci. **343**, 28–32 (2015). https://doi.org/10.1016/j.apsusc.2015.03.066
28. M. Nakashima, T. Yamaguchi, H. Itani, J. Sasano, M. Izaki, 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**, 761–764 (2015)
29. 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 Cu_2SnS_3 flowers/RGO nanocomposites and their visible-light-driven photocatalytic activities. RSC Adv. **6**, 74177–74185 (2016). https://doi.org/10.1039/c6ra12068g
30. 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. Exp. **6**, 106417 (2019)
31. B. Patel, M. Waldiya, A. Ray, Highly phase-pure spray-pyrolysed Cu_2SnS_3 thin films prepared by hybrid thermal treatment for photovoltaic applications. J. Alloys Compd. **745**, 347–354 (2018). https://doi.org/10.1016/j.jallcom.2018.02.220
32. 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. Opt. **55**, B158–B162 (2016). https://doi.org/10.1364/AO.55.00B158
33. B. Patel, R.K. Pati, I. Mukhopadhyay, A. Ray, Effect of vacuum and sulphur annealing on the structural properties of spray deposited Cu_2SnS_3 thin films. Vacuum **158**, 263–270 (2018)
34. S. Thiruvenkadam, D. Jovina, A.L. Rajesh, The influence of deposition temperature in the photovoltaic properties of spray deposited CZTS thin films. Sol. Energy. **106**, 166–170 (2014)
35. D. Patidar, K.S. Rathore, N.S. Saxena, K. Sharma, T.P. Sharma, Energy band gap and conductivity measurement of CdSe thin films. Chalcogenide Lett. **5**, 21–25 (2008)
36. T.S. Moss, *Optical properties of semicoundutors* (Academic Press, NY, 1959)
37. K.L. Chopra, S.R. Das, Thin Film Solar Cells (Plenum Press, New York, 1983)., (n.d.)
38. 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 Cu_2SnS_3 film deposited by ultrasonic spray pyrolysis. J. Mater. Sci. Mater. Electron. **27**, 4636–4646 (2016)
39. 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**, 237–243 (2015)
40. 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**, 332–341 (2013)
41. K.G. Deepa, P.C. Ramamurthy, M.K. Singha, Mesoporous Cu_2ZnSnS_4 nanoparticle film as a flexible and reusable visible light photocatalyst. Opt. Mater. (2019). https://doi.org/10.1016/j.optmat.2019.109492
42. N.M. Diantoro, A. Kusumaatmaja, K. Triyana, Study on photocatalytic properties of TiO_2 nanoparticle in various pH condition IOP Conf Series. J Phys Conf Ser **1011**, 012069 (2018)
43. A. Apostolopoulou, S. Mahajan, R. Sharma, E. Stathatos, Novel development of nanocrystalline kesterite Cu_2ZnSnS_4 thin film with high photocatalytic activity under visible light illumination. J. Phys. Chem. Solids. **112**, 37–42 (2018)
44. Y.M. Hunge, M.A. Mahadik, V.L. Patil, A.R. Pawar, S.R. Gadakh, A.V. Moholkar, P.S. Patil, C.H. Bhosale, Visible light assisted photoelectrocatalytic degradation of sugarcane factory wastewater by sprayed CZTS thin films. J. Phys. Chem. Solids. **111**, 176–181 (2017)
45. K. Shahzad, M.B. Tahir, M. Sagir, N.A. Alhakamy, M.R. Kabli, Synthesis of novel p-n heterojunction Cu_2SnS_3/Ti_3+-TiO_2 for the complete tetracycline degradation in few minutes and photocatalytic activity under simulated solar irradiation. Ceramics Int **47**(22), 31337–31348 (2017)
46. C. Nefzi, M. Souli, Y. Cuminal, N. Kamoun-Turki, Effect of sulfur concentration on structural, optical and electrical properties of Cu_2FeSnS_4 thin films for solar cells and photocatalysis applications. Superlatt. Microstruct. **124**, 17–29 (2018)
47. M.K. Singha, A. Patra, V. Rojwal, K.G. Deepa, Single-step fabrication of ZnO microflower thin films for highly efficient and reusable photocatalytic activity. J. Mater. Sci. Mater. Electron. **31**, 13578–13587 (2020)
48. V. Maheskumar, B. Vidhya, Investigation on the morphology and photocatalytic activity of Cu_3SnS_4 synthesized by ball milling and solvothermal method. J. Photochem. Photobiol. A Chem. **356**, 521–529 (2018). https://doi.org/10.1016/j.jphotochem.2017.12.026
49. F. Azeez, E. Al-Hetlani, M. Arafa, Y. Abdelmonem, A.A. Nazeer, M.O. Amin, M. Madkour, The effect of surface charge on photocatalytic degradation of methylene blue dye using chargeable titania nanoparticles. Sci. Rep. **8**, 1–9 (2018). https://doi.org/10.1038/s41598-018-25673-5
50. I. Kazeminezhad, A. Sadollahkhani, Influence of pH on the photocatalytic activity of ZnO nanoparticles. J. Mater. Sci. Mater. Electron. **27**, 4206–4215 (2016). https://doi.org/10.1007/s10854-016-4284-0
51. S.A. Phaltane, S.A.V.T.S. Bhat, P.S.P.S.D. Sartale, Photocatalytic degradation of methylene blue by hydrothermally synthesized CZTS nanoparticles. J Mater Sci Mater Electron (2017). https://doi.org/10.1007/s10854-017-6527-0

Authors and Affiliations

Sabina Rahaman[1] · Monoj Kumar Singha[2,3]

✉ Monoj Kumar Singha
monojsingha@yahoo.co.in

[1] Department of Electronics & Communication Engineering, BMS Institute of Technology and Managment, Bangalore 560064, India

[2] Department of Instrumentation and Applied Physics, Indian Institute of Science, Bangalore 560012, India

[3] Department of Electronics and Communication Engineering, Punjab Engineering College (Deemed To Be University), Chandigarh 160012, India

Optical Materials 123 (2022) 111838

Contents lists available at ScienceDirect

Optical Materials

journal homepage: www.elsevier.com/locate/optmat

Optimization and fabrication of low cost Cu_2SnS_3/ZnS thin film heterojunction solar cell using ultrasonic spray pyrolysis

Sabina Rahaman [a], M. Anantha Sunil [b,d], Monoj Kumar Singha [c,d,*], Kaustab Ghosh [e]

[a] Department of Electronics & Communication Engineering, BMS Institute of Technology and Management, Bangalore, 560064, India
[b] Department of Electronics & Communication Engineering, BMS College of Engineering, Bangalore, 560019, India
[c] Department of Electronics & Communication Engineering, Punjab Engineering College (Deemed to Be University), Chandigarh, 160012, India
[d] Department of Instrumentation and Applied Physics, Indian Institute of Science, Bangalore, 560012, India
[e] Centre for Nanoelectronics and VLSI Design, Vellore Institute of Technology, Vandalur Kelambakkam Road, Chennai, 600127, India

ARTICLE INFO

Keywords:
ZnS
Ultrasonic spray pyrolysis
Transmittance spectra
Bandgap
Solar cell window layer

ABSTRACT

In this work, we present the fabrication of heterojunction Cu_2SnS_3/ZnS photovoltaic cell using indigenously developed cost-effective ultrasonic spray pyrolysis method. Substrate temperature of as-prepared films have been varied from 350 °C to 500 °C. Temperature played a key role in enhancing the optical properties of ZnS films making it useable as window layer for solar cell. The p-type Cu_2SnS_3 (CTS) absorber layer of the cell is grown at optimized substrate temperature (500 °C) and with optimized Cu and Sn composition. Finally, solar cell is fabricated with these optimized p and n layers and have the structure of glass/FTO/Cu_2SnS_3/ZnS/Ag with V_{oc} = 685 mV, J_{SC} = 1.8 mA/cm^2, fill factor = 39% and solar efficiency = 0.5%. A detailed discussion is given in overcoming the intricacies involved in the cell and improving the efficiency.

1. Introduction

Zinc sulfide (ZnS) is a wide bandgap n type semiconductor material which is extensively used for optoelectronics applications such as window layer for solar cell applications, light emitting diode (LED), flat panel display, and lasers [1–4]. ZnS is found in two crystal form: α structure (wurtzite) and β structure (cubic) [5]. Cubic phase of the ZnS is steady at low temperature whereas Wurtzite structure is formed at high temperature (1296 K). Generally, cubic structure of ZnS is used as window layer of solar cells. CdS is also another window layer which is mostly used for thin film solar cell application in CdTe or $CuInSe_2$ heterojunction solar cells [6,7]. But there are several disadvantages using CdS as buffer or window layer. Firstly, cadmium is a toxic element. Due to its toxicity, it cannot be used as a buffer layer in large production for manufacture solar cell. Secondly, compare to ZnS, it has lower band gap. Due to these, the researchers are motivated to develop ZnS as window layer which is nontoxic and earth abundant. The wide bandgap also improve light transmission in the blue wavelength region of the electromagnetic spectrum which results in the improvement of short circuit current in solar cell [8]. ZnS can be used as a anti reflection coating of solar cell as it has s large refractive index of 2.35 [9]. Lattice structure of ZnS matches well with I_2-IV-VI_3 group of semiconductor materials which forms the absorber layer of solar cell [10]. Performance of solar cell depends on the properties of the absorber layer. The absorber material needs to have high carrier concentration, low resistivity, good optical property such as high absorption coefficient and direct bandgap. In this work, ternary semiconductor Cu_2SnS_3 (CTS) is selected as a p-type absorber layer for fabrication of cell. The p-type layer is prepared by optimizing the substrate temperature, Cu and Sn concentration as described in our earlier publications [11–13]. This was carried out to obtain optimal band gap value of 1.35 eV along with high carrier concentration and absorption coefficient which is suitable for solar cell fabrication.

Various methods are used to grow ZnS films such as sputtering, thermal evaporation, pulsed laser deposition, chemical spray pyrolysis, electrodeposition and chemical bath deposition [6,7,14–19]. Amongst these, one of the popular methods is chemical spray pyrolysis, used extensively for synthesis of thin film. The merits of this method are lesser time requirement for preparation of one sample, cost effective for large scale deposition which is important criteria for mass scale production. The spray method is not only useful for thin film solar cell, even it is used for fabrication of die synthesized, polymer and quantum dot solar cells

* Corresponding author. Department of Electronics & Communication Engineering, Punjab Engineering College (Deemed to be University), Chandigarh, 160012, India.
E-mail address: monojsingha@yahoo.co.in (M.K. Singha).

https://doi.org/10.1016/j.optmat.2021.111838
Received 30 September 2021; Received in revised form 23 November 2021; Accepted 24 November 2021

[20].

Many researchers have used chemical spray pyrolysis (pneumatic) techniques for synthesis of ZnS films [21–24]. Zeng et al. deposited the thin film at very low substrate temperature (~310 °C) using thioacetamide precursor as a sulphur source and formed the wurtzite structure of ZnS film. S. Serkis Yesilkaya et al. found the decrease in resistivity with the addition of Na doping. Performance of solar cell also improved with Na doping in the film. K. Ben Bacha et al. carried out annealing under N_2/S ambience and studied the crystal structural, surface and optical properties of the film. Lopez et al. reported the high grain size, uniform surface and high transmittance when films are prepared using $Zn(O_2CCH_3)_2(H_2O)_2$ precursor. Heterojunction solar cell has also been fabricated using $Cu(In,Ga)Se_2$ (p-type) and ZnS (n-type) layer and solar efficiencies of 18.6%, 18% and 14.18% could be achieved [5,25,26]. $CuInS_2$/ZnS/ZnO structures had 7.8% efficiency, while Cu_2ZnSnS_4/ZnS structures had obtained 8.4% efficiency [1,27,28].

Here, indigenously developed ultrasonic spray pyrolysis (USP) set up is employed for the synthesis of ZnS thin films [29]. The ultrasonic frequency generates aerosols which are used to be deposited on the preheated substrate and no carrier gas is used here when compared with conventional spray technique. Narrow size distribution, homogeneity of films, high deposition rate are advantages of USP method and it also has gas flow rate independent of aerosol flow rate [30–32]. ZnS films prepared by this technique with different substrate temperatures are studied for improving the optical properties. After ensuring the quality of the optimized film through various characterization techniques, it is finally used to fabricate the heterojunction solar cell with CTS as absorber layer.

2. Experimental details

2.1. Deposition of ZnS

USP method is used for preparation of zinc sulphide films on glass substrates and substrate temperature is varied from 350 °C to 500 °C. $ZnCl_2$ and $SC(NH_2)_2$ are used as precursor materials which are purchased from sigma-Aldrich. In order to prepare ZnS solution, measured quantity of 0.1 M $ZnCl_2$ and 0.1 M $SC(NH_2)_2$ are mixed in deionized water. 30 ml of the solution is placed in ultrasonic nebulizer which is vibrated at a frequency of 1.7 MHz. At this high frequency, the solution is turned in the form of a mist which is sprayed at 2 ml/min through a nozzle and decomposed on the preheated glass substrate to form the ZnS thin film. Gap between substrate and spray nozzle is maintained at 3 cm. Once deposition is complete, films are kept at deposition temperature for another 30 min to complete the annealing process and after that they are eventually made to return to room temperature. Throughout the experiment, all spray parameters are kept constant. ZT350, ZT400, ZT450, and ZT500 are the samples prepared with different substrate temperatures (Ts) ranging from 350 °C to 500 °C in 50 °C increment.

2.2. Cu_2SnS_3/ZnS thin film solar cell fabrication

CTS films has been prepared using the same USP method and detailed optimization of this layer is given elsewhere [11–13]. The structure of Cu_2SnS_3/ZnS cell is shown in the schematic diagram of Fig. 1. It consists of conducting material which acts as a bottom electrode, n type semiconductor ZnS (buffer layer), p type semiconductor CTS (absorber layer) along with top electrodes. As a bottom electrode, FTO coated glass is used with a surface resistance of 8–12 Ω/sq and having transmission greater than 85%. Ultrasonic spray pyrolysis is used to deposit the CTS thin film on top of the FTO substrate. The deposition of optimized CTS film is carried out at a deposition temperature of 500 °C with $CuCl_2$. H_2O of 0.025 M, $SnCl_2$. $2H_2O$ of 0.01 M and $SC(NH_2)_2$ of 0.25 M in 60 ml solvent. Thickness of the CTS layer is approximately ~650 nm. ZnS thin film is deposited on top of CTS film.

The incident solar radiation should travel shorter path and enter the p-n junction, creating more electron and hole pairs. Silver is used as top electrode for the cell. The electrodes are deposited on FTO/CTS/ZnS structure using the thermal evaporation technique. Fig. 2 shows the schematic diagram of the fabrication process of CTS/ZnS heterojunction solar cell.

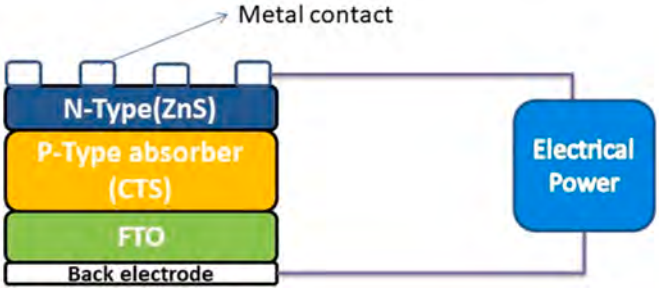


Fig. 1. Structure of Cu_2SnS_3/ZnS heterojunction solar cell.

2.3. Instrument used for characterization

High resolution XRD instrument (Bruker D8) is used for measurement of X-ray diffraction pattern. It is operated at a tube voltage in the range of 20–40 kV and tube current in the range of 2–40 mA. Wavelength of radiation source used is 1.5405 Å and a moving scintillation counter is used as detector. Scanning Electron Microscopy (SEM) and Energy Dispersive X-ray Analysis (EDX) of the films are measured using ZEISS Ultra 55 FE-SEM instrument. UV-VIS spectrophotometer (SPECORD S600 UV-VIS) records the absorption and transmittance spectra from 300 to 1100 nm. The HMS-3000 Hall measurement system having current range of 1nA-20mA and magnetic flux density of 0.5T is used for measuring the electrical properties of ZnS thin films. Fig. 3 shows the USP system setup which is utilized for preparation of thin film layer (both n-type and p-type).

3. Results and discussion

The crystalline structures of the ZnS films grown at different substrate temperatures is analyzed by X-ray diffraction method. The diffraction spectrum is exhibited in Fig. 4. All films are found to be of zinc blende cubic structure and prominent peak at 28.45° corresponds to the (111) plane. Other two weak peaks are observed at 46.72° and 56.43° which corresponds to (220) and (311) plane of ZnS cubic structure (JCPDS card number 05–0566) respectively. There are no additional peaks observed in XRD data. The presence of broad peaks from the image can be attributed to the smaller grain size of the ZnS films. With the increase of substrate temperature, crystallinity of as-deposited films improves slightly. Based on XRD data, lattice parameter value is measured to be 5.38 Å. The lattice constant measured is close to standard data of JCPDS card No. 05–0566.

Bandgap and elemental composition value of ZnS film are summarized in Table 1. The overall composition of the films is measured by using EDX analysis. For lower substrate temperature, films were slightly zinc rich resulting in higher Zn/S ratio. With the increase in the substrate temperature, Zn/S ratio reached to nearly one resulting in stochiometric films. Sample ZT450 is found to be stoichiometric. It can be concluded that composition of films depends on substrate temperature. Sulphur concentration drastically decreases at a substrate temperature 500 °C which is due to the volatile nature of the sulphur.

Fig. 5 displays the morphology of ZnS film using SEM. Films deposited at 350 °C had irregular shape with non-uniform distribution of grains. When temperature is more than 350 °C, morphology of films changed from irregular to spherical structure. The surface morphology of the ZnS film prepared at T_s = 450 °C is relatively more uniform, homogenous, and denser. The sample does not contain cracks or pin

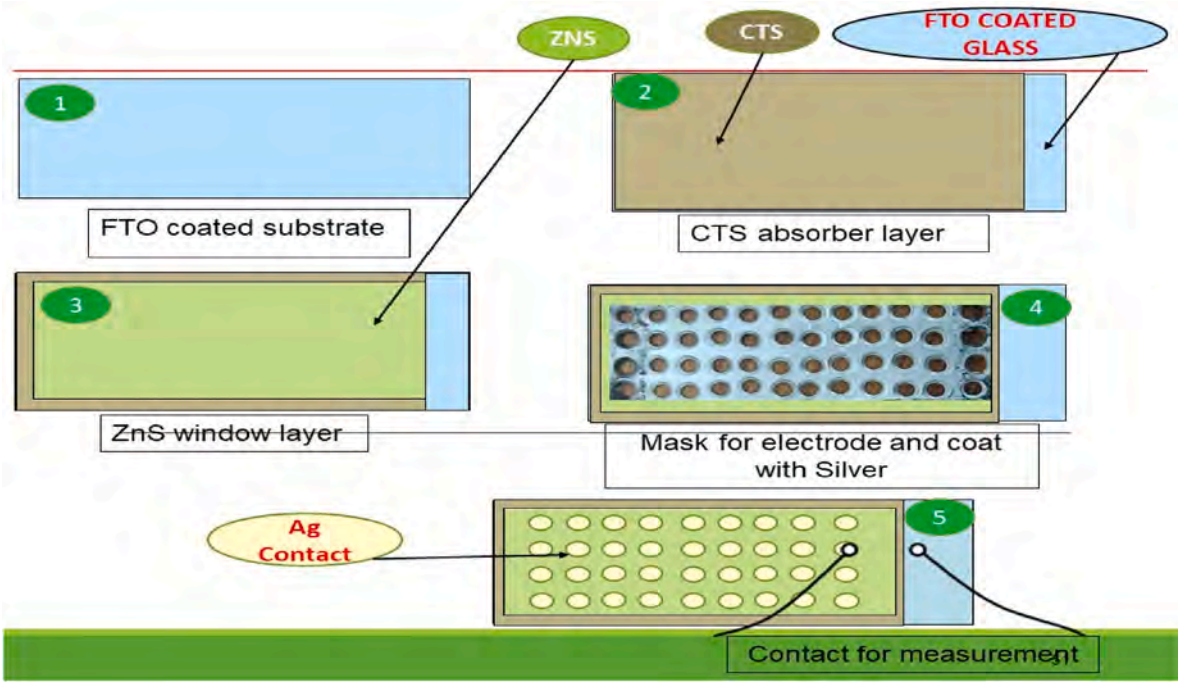


Fig. 2. Fabrication process of CTS/ZnS heterojunction solar cell.

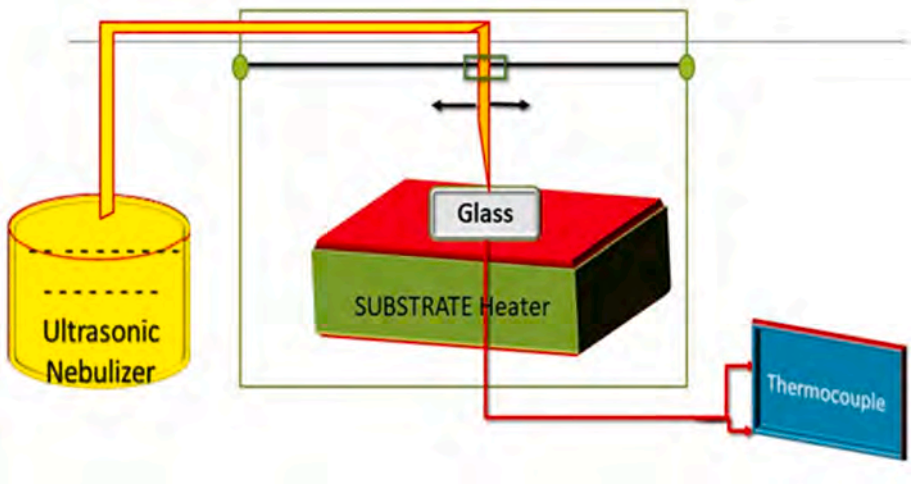


Fig. 3. Schematic diagram of automated ultrasonic spray setup.

holes.

Properties of the material like transmittance and bandgaps are important parameters for solar cell application. A good solar cell window must transmit most of the solar radiation. Hence, the transmittance spectra is measured for the films in the wavelength range from 300 to 1100 nm, which is shown in Fig. 6. The films prepared at higher substrate temperature are found to have more transmittance. For all the films, the transmittance is also seen to increase with increase in wavelength and the transmittance of ZT450 film is found to reach 88% at near infrared wavelength of 1000 nm. However, at such higher wavelength, the transmittance of the film prepared at the substrate temperature of 500 °C is found to be much lower than ZT450. The drop in transmittance is due to higher Zn concentration as zinc rich films exhibits higher reflectance. Similar results are reported for growth of ZnS films using chemical bath deposition techniques [10].

Fig. 7 shows the plot of band gap of the ZnS films at different temperatures. Here, X axis represents the photon energy (hν) and Y axis denotes the value of $(\alpha h\nu)^2$, where α value indicates the absorbance coefficient of material. From the Tauc equation, bandgap of ZnS films is measured [33].

$$\alpha h\upsilon = K(h\upsilon - E_g)^n \quad (1)$$

where K is a proportionality constant, n = 1/2 for direct bandgap semiconductor and Eg is the energy band gap calculated by extrapolating the linear region of the graph on to X-axis. For all the films, the bandgap is found between 3.56 eV and 3.42 eV. From Fig. 7, it can be seen that band gap tend to decrease with the increase in substrate temperature. There are many reasons associated with band gap variation. Many researchers suggest that morphology of the films (nanostructure to microstructure), defects, strain and doping into the films and particle confinement are the main reasons for the change in band gaps [34]. It may be possible that some defects are present in our films which may also have affected the bandgap. Chemical bath deposited films also showed the same behavior of decreasing band gap [35].

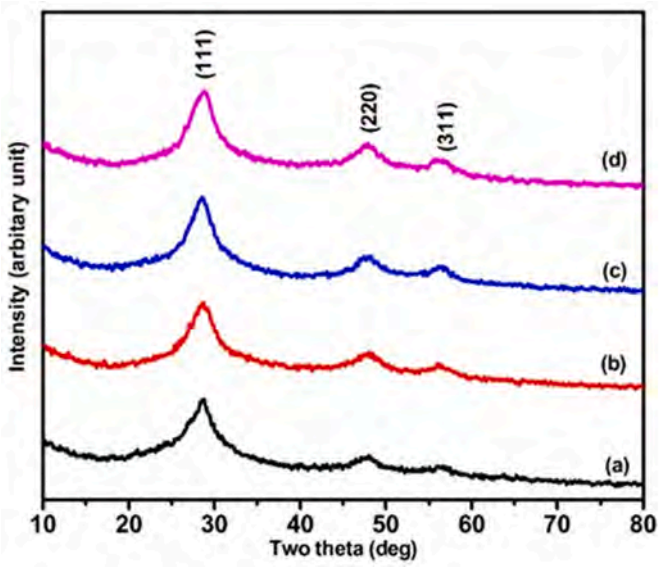


Fig. 4. X-ray diffraction spectra of sprayed ZnS films at different temperature (a) ZT350 (b) ZT400 (c) ZT450 (d) ZT500.

Table 1
Elemental composition and bandgap of ZnS films.

Sample name	The atomic percentage of elements		Bandgap(eV)
	Zn (%)	S (%)	
ZT350	54.58	45.42	3.56
ZT400	52.43	47.57	3.50
ZT450	49.96	50.04	3.43
ZT500	55.94	44.06	3.42

Hall measurement is performed to measure the different electrical parameter values of ZnS films. Table 2 shows the electrical properties of ZnS thin films at different substrate temperature measured using Hall method. All films are found to be n-type semiconductor, an essential requirement for forming the p-n junction. It is observed that there is no significant change in the carrier concentration and mobility of prepared ZnS thin films. The ZnS film resistivity is found to be in the order of 10^6 to 10^5 Ω-cm. As substrate temperature is increased, there is a reduction in resistivity, which may be due to improvement of crystallinity of the film. Reduced resistivity can also be attributed to an increase in grain size, which contributes to a decrease in the scattering at the grain boundaries.

Based on the above study, it is found that ZnS film prepared at 450 °C exhibits the optimal features such as high crystallinity, suitable band gap and high optical transmittance, which can be utilized as n-type semiconductor in solar cell. Thus, the thin film solar cell is prepared using p-type Cu_2SnS_3 and n-type ZnS (ZT450) layer. The J-V characteristic of the Cu_2SnS_3/ZnS solar cell under the dark and AM1.5 illumination is showed in Fig. 8. The intensity of input light is set to 100 mW/cm^2.

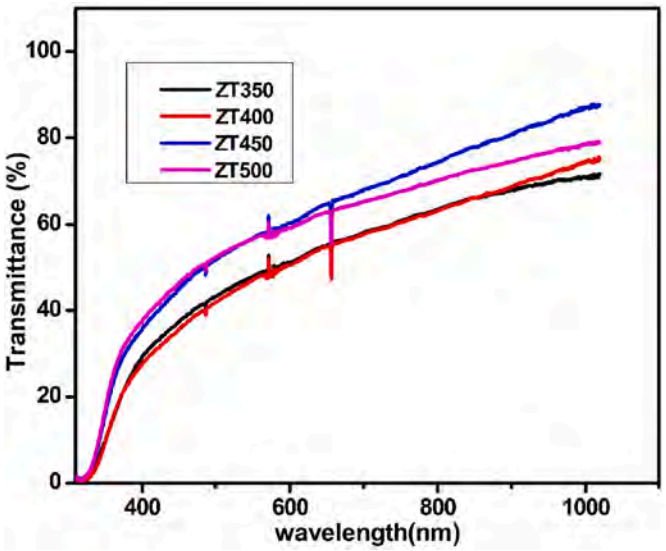


Fig. 6. Transmittance of ZnS films of different substrate temperature.

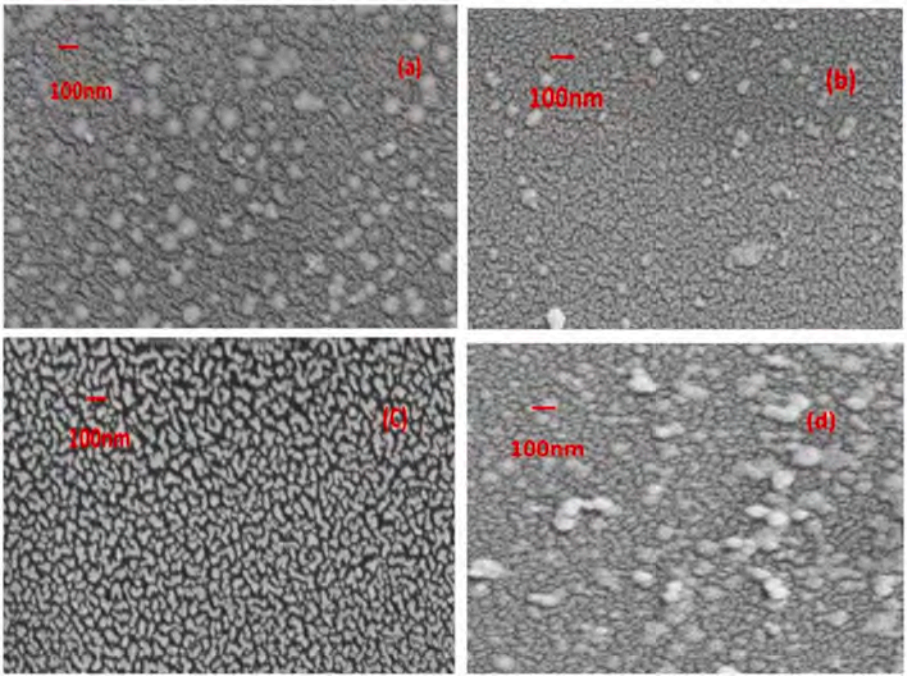


Fig. 5. SEM images of ZnS films (a) ZT350 (b) ZT400 (c) ZT450 (d) ZT500.

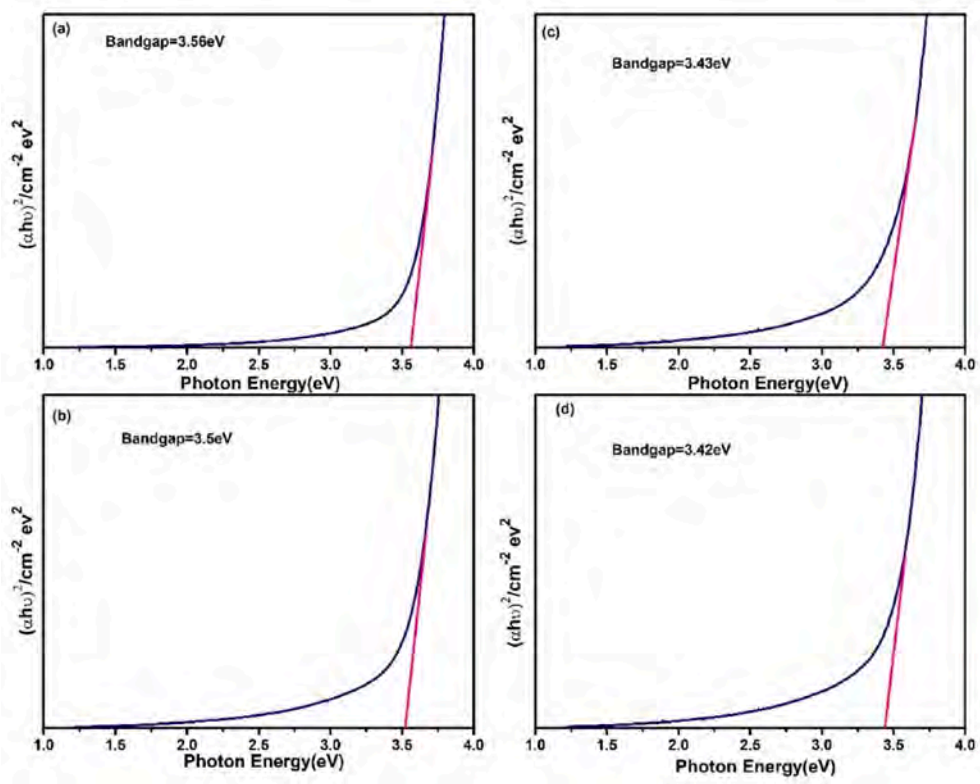


Fig. 7. Bandgap of ZnS films (a) ZT350 (b) ZT400 (c) ZT450 (d) ZT500.

Table 2
Electrical properties of ZnS thin films at different substrate temperature.

Sample name	Type of conductivity	Resistivity (Ω cm)	Carrier Concentration (ions/cm^3)	Mobility (cm^2/V.s)
ZT350	n	1.45×10^6	2.37×10^{12}	1.82
ZT400	n	2.94×10^5	6.13×10^{12}	3.47
ZT450	n	1.85×10^5	7.93×10^{12}	4.21
ZT500	n	1.12×10^5	8.56×10^{12}	6.52

Under the illuminated condition, we obtain the short circuit current density (Jsc) of 1.8 mA/cm^2, open circuit voltage (Voc) of 0.685 V, fill factor of 39% and series and shunt resistances to be of 100 Ω/cm^2 and 230 Ω/cm^2 respectively. However, the efficiency of the solar cell is found to be 0.5% although having 88% high transmittance of ZnS buffer layer at 1000 nm wavelength. To investigate this matter, we have measured the transmittance spectra of the CTS film as shown in Fig. 9. Since, binary phase Cu_2S is found in sample CTS4, so we have selected CTS3 (Cu concentration 0.025 M) as the absorber layer of this solar cell and detailed study is presented elsewhere [12]. Very low transmittance varying from 0 to 5% could be seen in the entire spectra, which justifies better absorbance in this layer. Thus, other factors are responsible for reduction in efficiency such as p-n junction interface which introduces surface states and impurity related traps or defects. Rough and

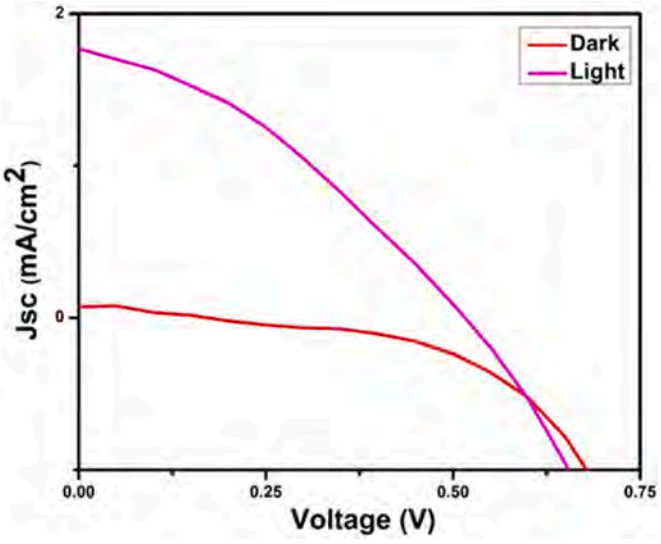


Fig. 8. J-V characteristics of Cu_2SnS_3/ZnS heterojunction solar cell.

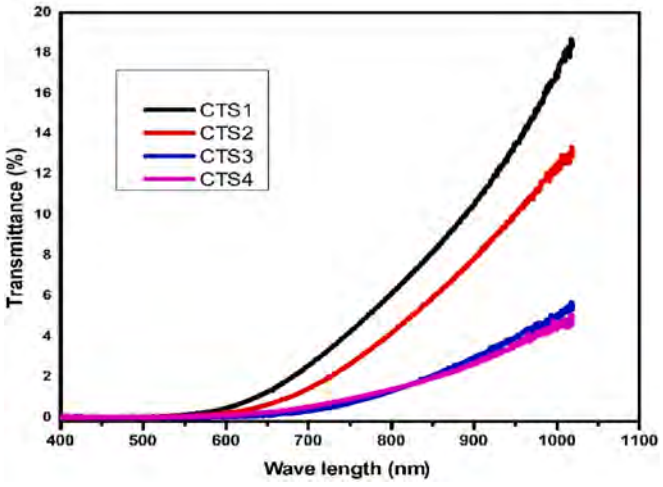


Fig. 9. Transmission spectra of Cu_2SnS_3.

non-uniform interface in the p-n junction and the impurities in the sample act as recombination centers which trap the generated electrons and lowers the efficiency. Post-growth thermal annealing is a well-known method to reduce these traps and improve the optical and electrical properties in nanostructure materials like quantum dots [36, 37]. Usage of ultrahigh purity starting materials has also been found to improve the quality of the compound semiconductors required for device fabrication [38,39]. Thus, research needs to be focused more in optimizing the annealing temperature and purity of the starting material of these thin films, which can pave the way in improving the performance of cost-effective solar cells.

4. Conclusions

N-type semiconductor (ZnS) thin film materials are prepared by USP method with temperature variation of the substrate from 350 to 500 °C in steps of 50 °C. The bandgap of the films also tend to decrease with increase in substrate temperature and the resistivity also reduced from 10^6 to 10^5 Ω-cm. XRD studies showed that the crystalline nature of the films exhibiting zinc blende cubic structure and orientation along (111) plane. From SEM images, the film ZT450 (prepared at substrate temperature of 450 °C) is found to be homogeneous, uniform, dense and void free. At 450 °C substrate temperature with 1:1 ratio of Zn/S, we obtained a very high transmittance reaching to 88% at 1000 nm wavelength. Thus, by analyzing the structural and optical properties, we found ZT450 to be a suitable material as window layer and used the same as a buffer layer for fabrication of Cu_2SnS_3/ZnS solar cell. Device parameters like Jsc, Voc, FF and efficiency are determined from the J-V curve. However, the efficiency of the solar cell is found to be 0.5%. It is presumed that post growth thermal annealing and use of ultrahigh purity starting materials could be effective in reducing the interface traps and improving the efficiency of this low cost and indigenously fabricated solar cell.

CRediT authorship contribution statement

Sabina Rahaman: Investigation, Formal analysis, Resources, Writing – original draft, Writing – review & editing. **M. Anantha Sunil:** Formal analysis, Writing – review & editing, Visualization. **Monoj Kumar Singha:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Kaustab Ghosh:** Writing – review & editing, Visualization.

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

One of the author would like to thank TEQIP-III for providing the grant with file no BMS/COE/TEQIP-III/2020–21/ECE/(03 -02). He also would like to thank Science and Engineering Board (SERB), Govt of India for providing the grant with file No: TAR/2020/000293 under Teachers Associateship for Research Excellence scheme. We thank Micro-nano Characterization facility, Centre for Nanoscience and Engineering at Indian Institute of Science (IISc) for providing the necessary facility to carry out characterization.

References

[1] B. Asenjo, A.M. Chaparro, M.T. Gutiérrez, J. Herrero, J. Klaer, Study of CuInS2/ZnS/ZnO solar cells, with chemically deposited ZnS buffer layers from acidic solutions, Sol. Energy Mater. Sol. Cells 92 (2008) 302–306.

[2] I.K. Sou, Z.H. Ma, Z.Q. Zhang, G.K.L. Wong, Temperature dependence of the responsivity of II–VI ultraviolet photodiodes, Appl. Phys. Lett. 76 (2000) 1098–1100.

[3] C. Jin, H. Kim, H.W. Kim, C. Lee, Enhancement in the photoluminescence of ZnS nanowires by TiO2 coating and thermal annealing, J. Lumin. 130 (2010) 516–519.

[4] Y. Shuai, C. Liu, J. Wang, X. Cui, L. Nie, Pseudo-bi-enzyme glucose sensor: ZnS hollow spheres and glucose oxidase concerted catalysis glucose, Analyst 138 (2013) 3259–3263.

[5] D. Hariskos, S. Spiering, M. Powalla, Buffer layers in Cu (In, Ga) Se2 solar cells and modules, Thin Solid Films 480 (2005) 99–109.

[6] S.M.A. Durrani, A.M. Al-Shukri, A. Iob, E.E. Khawaja, Optical constants of zinc sulfide films determined from transmittance measurements, Thin Solid Films 379 (2000) 199–202, https://doi.org/10.1016/S0040-6090(00)01539-X.

[7] D.H. Hwang, J.H. Ahn, K.N. Hui, K. San Hui, Y.G. Son, Structural and optical properties of ZnS thin films deposited by RF magnetron sputtering, Nanoscale Res. Lett. 7 (2012) 1–7.

[8] X. Fang, T. Zhai, U.K. Gautam, L. Li, L. Wu, Y. Bando, D. Golberg, ZnS nanostructures: from synthesis to applications, Prog. Mater. Sci. 56 (2011) 175–287.

[9] A. Nakaruk, D. Ragazzon, C.C. Sorrell, Anatase thin films by ultrasonic spray pyrolysis, J. Anal. Appl. Pyrolysis 88 (2010) 98–101, https://doi.org/10.1016/j.jaap.2010.03.001.

[10] U. Gangopadhyay, K. Kim, D. Mangalaraj, J. Yi, Low cost CBD ZnS antireflection coating on large area commercial mono-crystalline silicon solar cells, Appl. Surf. Sci. 230 (2004) 364–370.

[11] S. Rahaman, M.A. Sunil, M.K. Singha, K. Ghosh, Temperature dependent growth of Cu2SnS3 thin films using ultrasonic spray pyrolysis for solar cell absorber layer and photocatalytic application, Mater. Res. Express 6 (2019), 106417, https://doi.org/10.1088/2053-1591/ab3928.

[12] S. Rahaman, M.K. Singha, M.A. Sunil, K. Ghosh, Effect of copper concentration on CTS thin films for solar cell absorber layer and photocatalysis applications, Superlattices and Microstructures 145 (2020), 106589, https://doi.org/10.1016/j.spmi.2020.106589.

[13] S. Rahaman, M. Anantha Sunil, M. Kumar Singha, K. Ghosh, Studies of ultrasonically sprayed Cu2SnS3thin films by varying Sn concentration, Mater. Today Proc. 43 (2020) 3938–3941, https://doi.org/10.1016/j.matpr.2021.02.657.

[14] Y.-Z. Yoo, Y. Osaka, T. Fukumura, Z. Jin, M. Kawasaki, H. Koinuma, T. Chikyow, P. Ahmet, A. Setoguchi, S.F. Chichibu, High temperature growth of ZnS films on bare Si and transformation of ZnS to ZnO by thermal oxidation, Appl. Phys. Lett. 78 (2001) 616–618.

[15] I.O. Oladeji, L. Chow, A study of the effects of ammonium salts on chemical bath deposited zinc sulfide thin films, Thin Solid Films 339 (1999) 148–153.

[16] T. Ben Nasr, N. Kamoun, C. Guasch, Physical properties of ZnS thin films prepared by chemical bath deposition, Appl. Surf. Sci. 254 (2008) 5039–5043.

[17] N.K. Allouche, T. Ben Nasr, N.T. Kamoun, C. Guasch, Synthesis and properties of chemical bath deposited ZnS multilayer films, Mater. Chem. Phys. 123 (2010) 620–624.

[18] A. Goudarzi, G.M. Aval, R. Sahraei, H. Ahmadpoor, Ammonia-free chemical bath deposition of nanocrystalline ZnS thin film buffer layer for solar cells, Thin Solid Films 516 (2008) 4953–4957.

[19] H. Hennayaka, H.S. Lee, Structural and optical properties of ZnS thin film grown by pulsed electrodeposition, Thin Solid Films 548 (2013) 86–90.

[20] M. Eslamian, Spray-on thin film PV solar cells: advances, potentials and challenges, Coatings 4 (2014) 60–84, https://doi.org/10.3390/coatings4010060.

[21] X. Zeng, S.S. Pramana, S.K. Batabyal, S.G. Mhaisalkar, X. Chen, K.B. Jinesh, Low temperature synthesis of wurtzite zinc sulfide (ZnS) thin films by chemical spray pyrolysis, Phys. Chem. Chem. Phys. 15 (2013) 6763–6768.

[22] S.S. Yesilkaya, U. Ulutas, H.M.A. Alqader, Effect of Na doping on the properties of ZnS thin films and ZnS/Si heterojunction cells, Mater. Lett. 288 (2021) 1–11, https://doi.org/10.1016/j.matlet.2021.129347.

[23] K. Ben Bacha, N. Bitri, H. Bouzouita, Effect of annealing parameters on structural and morphological properties of sprayed ZnS thin films, Optik 127 (2016) 3100–3104, https://doi.org/10.1016/j.ijleo.2015.12.083.

[24] M.C. Lopez, J.P. Espinos, F. Martin, D. Leinen, J.R. Ramos-Barrado, Growth of ZnS thin films obtained by chemical spray pyrolysis: the influence of precursors, J. Cryst. Growth 285 (2005) 66–75.

[25] T. Nakada, M. Mizutani, Qficiency Cd-free Cu (In, Ga) Se2 thin-film solar cells fabricated using chemical bath deposition (CBD)-ZnS buffer layers, Jpn. J. Appl. Phys. 41 (2002) L165.

[26] D.H. Jeon, D.K. Hwang, D.H. Kim, J.K. Kang, C.S. Lee, Optimization of the ZnS buffer layer by chemical bath deposition for Cu(In,Ga)Se2 solar cells, J. Nanosci. Nanotechnol. 16 (2016) 5398–5402, https://doi.org/10.1166/jnn.2016.12217.

[27] B. Shin, O. Gunawan, Y. Zhu, N.A. Bojarczuk, S.J. Chey, S. Guha, Thin film solar cell with 8.4% power conversion efficiency using an earth-abundant Cu_2ZnSnS_4 absorber, Prog. Photovoltaics Res. Appl. 21 (2013) 72–76, https://doi.org/10.1002/pip.1174.

[28] K.G. Deepa, P.C. Ramamurthy, M.K. Singha, Mesoporous Cu2ZnSnS4 nanoparticle film as a flexible and reusable visible light photocatalyst, Opt. Mater. 98 (2019), 109492.

[29] M.A. Sunil, K.G. Deepa, J. Nagaraju, Growth of AgInS2thin films by ultrasonic spray pyrolysis technique, Thin Solid Films (2014), https://doi.org/10.1016/j.tsf.2013.10.053.

[30] P.S. Patil, Versatility of chemical spray pyrolysis technique, Mater. Chem. Phys. 59 (1999) 185–198.

[31] Monoj Kumar Singha, Aniket Patra, Highly efficient and Reusable ZnO microflower photocatalyst on stainless steel mesh under UV–Vis and natural sunlight, Optic. Mater. 107 (2020), https://doi.org/10.1016/j.optmat.2020.110000.
[32] Monoj Kumar Singha, Aniket Patra, Vineet Rojwal, K G Deepa, Single-step fabrication of ZnO microflower thin films for highly efficient and reusable photocatalytic activity, J. Mater. Sci. Mater. Electron. 31 (2020), https://doi.org/10.1007/s10854-020-03914-6.
[33] J. Tauc, Amorphous and Liquid Semiconductors, first ed., Springer, 1974.
[34] Y.G. Wang, S.P. Lau, H.W. Lee, S.F. Yu, B.K. Tay, X.H. Zhang, H.H. Hng, Photoluminescence study of ZnO films prepared by thermal oxidation of Zn metallic films in air, J. Appl. Phys. 94 (2003) 354–358.
[35] T. Yamaguchi, Y. Yamamoto, T. Tanaka, Y. Demizu, A. Yoshida, (Cd, Zn) S thin films prepared by chemical bath deposition for photovoltaic devices, Thin Solid Films 281 (1996) 375–378.
[36] K. Ghosh, Y. Naresh, N.S. Reddy, Theoretical Optimization of Multi-Layer InAs/GaAs Quantum Dots Subject to Post-growth Thermal Annealing for Tailoring the Photoluminescence Emission beyond 1 . 3μm Theoretical Optimization of Multi-Layer InAs/GaAs Quantum Dots Subject to Post-growth Thermal Annealing for Tailoring the Photoluminescence Emission beyond 1, vol. 3, 2012, https://doi.org/10.1063/1.4739457 l m, 024315.
[37] K. Ghosh, S. Kundu, N. Halder, M. Srujan, S. Sengupta, S. Chakrabarti, Annealing of $In_{0.45}Ga_{0.55}As$/GaAs quantum dots overgrown with large monolayer (11 ML) coverage for applications in thermally stable optoelectronic devices, Solid State Commun. 151 (2011) 1394–1399, https://doi.org/10.1016/j.ssc.2011.06.007.
[38] K. Ghosh, V.N. Mani, S. Dhar, A Modeling Approach for the Purification of Group III Metals „ Ga and in … by Zone Refining, 2008, pp. 1–8, https://doi.org/10.1063/1.2959832.
[39] K. Ghosh, V.N. Mani, S. Dhar, Numerical Study and Experimental Investigation of Zone Refining in Ultra-high Purification of Gallium and its Use in the Growth of GaAs Epitaxial Layers, vol. 311, 2009, pp. 1521–1528, https://doi.org/10.1016/j.jcrysgro.2009.01.102.

Optical Materials 123 (2022) 111838

Contents lists available at ScienceDirect

Optical Materials

journal homepage: www.elsevier.com/locate/optmat

Optimization and fabrication of low cost Cu_2SnS_3/ZnS thin film heterojunction solar cell using ultrasonic spray pyrolysis

Sabina Rahaman [a], M. Anantha Sunil [b,d], Monoj Kumar Singha [c,d,*], Kaustab Ghosh [e]

[a] Department of Electronics & Communication Engineering, BMS Institute of Technology and Management, Bangalore, 560064, India
[b] Department of Electronics & Communication Engineering, BMS College of Engineering, Bangalore, 560019, India
[c] Department of Electronics & Communication Engineering, Punjab Engineering College (Deemed to Be University), Chandigarh, 160012, India
[d] Department of Instrumentation and Applied Physics, Indian Institute of Science, Bangalore, 560012, India
[e] Centre for Nanoelectronics and VLSI Design, Vellore Institute of Technology, Vandalur Kelambakkam Road, Chennai, 600127, India

ARTICLE INFO

Keywords:
ZnS
Ultrasonic spray pyrolysis
Transmittance spectra
Bandgap
Solar cell window layer

ABSTRACT

In this work, we present the fabrication of heterojunction Cu_2SnS_3/ZnS photovoltaic cell using indigenously developed cost-effective ultrasonic spray pyrolysis method. Substrate temperature of as-prepared films have been varied from 350 °C to 500 °C. Temperature played a key role in enhancing the optical properties of ZnS films making it useable as window layer for solar cell. The p-type Cu_2SnS_3 (CTS) absorber layer of the cell is grown at optimized substrate temperature (500 °C) and with optimized Cu and Sn composition. Finally, solar cell is fabricated with these optimized p and n layers and have the structure of glass/FTO/Cu_2SnS_3/ZnS/Ag with V_{oc} = 685 mV, J_{SC} = 1.8 mA/cm^2, fill factor = 39% and solar efficiency = 0.5%. A detailed discussion is given in overcoming the intricacies involved in the cell and improving the efficiency.

1. Introduction

Zinc sulfide (ZnS) is a wide bandgap n type semiconductor material which is extensively used for optoelectronics applications such as window layer for solar cell applications, light emitting diode (LED), flat panel display, and lasers [1–4]. ZnS is found in two crystal form: α structure (wurtzite) and β structure (cubic) [5]. Cubic phase of the ZnS is steady at low temperature whereas Wurtzite structure is formed at high temperature (1296 K). Generally, cubic structure of ZnS is used as window layer of solar cells. CdS is also another window layer which is mostly used for thin film solar cell application in CdTe or $CuInSe_2$ heterojunction solar cells [6,7]. But there are several disadvantages using CdS as buffer or window layer. Firstly, cadmium is a toxic element. Due to its toxicity, it cannot be used as a buffer layer in large production for manufacture solar cell. Secondly, compare to ZnS, it has lower band gap. Due to these, the researchers are motivated to develop ZnS as window layer which is nontoxic and earth abundant. The wide bandgap also improve light transmission in the blue wavelength region of the electromagnetic spectrum which results in the improvement of short circuit current in solar cell [8]. ZnS can be used as a anti reflection coating of solar cell as it has s large refractive index of 2.35 [9]. Lattice structure of ZnS matches well with I_2-IV-VI_3 group of semiconductor materials which forms the absorber layer of solar cell [10]. Performance of solar cell depends on the properties of the absorber layer. The absorber material needs to have high carrier concentration, low resistivity, good optical property such as high absorption coefficient and direct bandgap. In this work, ternary semiconductor Cu_2SnS_3 (CTS) is selected as a p-type absorber layer for fabrication of cell. The p-type layer is prepared by optimizing the substrate temperature, Cu and Sn concentration as described in our earlier publications [11–13]. This was carried out to obtain optimal band gap value of 1.35 eV along with high carrier concentration and absorption coefficient which is suitable for solar cell fabrication.

Various methods are used to grow ZnS films such as sputtering, thermal evaporation, pulsed laser deposition, chemical spray pyrolysis, electrodeposition and chemical bath deposition [6,7,14–19]. Amongst these, one of the popular methods is chemical spray pyrolysis, used extensively for synthesis of thin film. The merits of this method are lesser time requirement for preparation of one sample, cost effective for large scale deposition which is important criteria for mass scale production. The spray method is not only useful for thin film solar cell, even it is used for fabrication of die synthesized, polymer and quantum dot solar cells

* Corresponding author. Department of Electronics & Communication Engineering, Punjab Engineering College (Deemed to be University), Chandigarh, 160012, India.
E-mail address: monojsingha@yahoo.co.in (M.K. Singha).

https://doi.org/10.1016/j.optmat.2021.111838
Received 30 September 2021; Received in revised form 23 November 2021; Accepted 24 November 2021

[20].

Many researchers have used chemical spray pyrolysis (pneumatic) techniques for synthesis of ZnS films [21–24]. Zeng et al. deposited the thin film at very low substrate temperature (~310 °C) using thioacetamide precursor as a sulphur source and formed the wurtzite structure of ZnS film. S. Serkis Yesilkaya et al. found the decrease in resistivity with the addition of Na doping. Performance of solar cell also improved with Na doping in the film. K. Ben Bacha et al. carried out annealing under N_2/S ambience and studied the crystal structural, surface and optical properties of the film. Lopez et al. reported the high grain size, uniform surface and high transmittance when films are prepared using $Zn(O_2CCH_3)_2(H_2O)_2$ precursor. Heterojunction solar cell has also been fabricated using $Cu(In,Ga)Se_2$ (p-type) and ZnS (n-type) layer and solar efficiencies of 18.6%, 18% and 14.18% could be achieved [5,25,26]. $CuInS_2$/ZnS/ZnO structures had 7.8% efficiency, while Cu_2ZnSnS_4/ZnS structures had obtained 8.4% efficiency [1,27,28].

Here, indigenously developed ultrasonic spray pyrolysis (USP) set up is employed for the synthesis of ZnS thin films [29]. The ultrasonic frequency generates aerosols which are used to be deposited on the preheated substrate and no carrier gas is used here when compared with conventional spray technique. Narrow size distribution, homogeneity of films, high deposition rate are advantages of USP method and it also has gas flow rate independent of aerosol flow rate [30–32]. ZnS films prepared by this technique with different substrate temperatures are studied for improving the optical properties. After ensuring the quality of the optimized film through various characterization techniques, it is finally used to fabricate the heterojunction solar cell with CTS as absorber layer.

2. Experimental details

2.1. Deposition of ZnS

USP method is used for preparation of zinc sulphide films on glass substrates and substrate temperature is varied from 350 °C to 500 °C. $ZnCl_2$ and $SC(NH_2)_2$ are used as precursor materials which are purchased from sigma-Aldrich. In order to prepare ZnS solution, measured quantity of 0.1 M $ZnCl_2$ and 0.1 M $SC(NH_2)_2$ are mixed in deionized water. 30 ml of the solution is placed in ultrasonic nebulizer which is vibrated at a frequency of 1.7 MHz. At this high frequency, the solution is turned in the form of a mist which is sprayed at 2 ml/min through a nozzle and decomposed on the preheated glass substrate to form the ZnS thin film. Gap between substrate and spray nozzle is maintained at 3 cm. Once deposition is complete, films are kept at deposition temperature for another 30 min to complete the annealing process and after that they are eventually made to return to room temperature. Throughout the experiment, all spray parameters are kept constant. ZT350, ZT400, ZT450, and ZT500 are the samples prepared with different substrate temperatures (Ts) ranging from 350 °C to 500 °C in 50 °C increment.

2.2. Cu_2SnS_3/ZnS thin film solar cell fabrication

CTS films has been prepared using the same USP method and detailed optimization of this layer is given elsewhere [11–13]. The structure of Cu_2SnS_3/ZnS cell is shown in the schematic diagram of Fig. 1. It consists of conducting material which acts as a bottom electrode, n type semiconductor ZnS (buffer layer), p type semiconductor CTS (absorber layer) along with top electrodes. As a bottom electrode, FTO coated glass is used with a surface resistance of 8–12 Ω/sq and having transmission greater than 85%. Ultrasonic spray pyrolysis is used to deposit the CTS thin film on top of the FTO substrate. The deposition of optimized CTS film is carried out at a deposition temperature of 500 °C with $CuCl_2$. H_2O of 0.025 M, $SnCl_2$. $2H_2O$ of 0.01 M and $SC(NH_2)_2$ of 0.25 M in 60 ml solvent. Thickness of the CTS layer is approximately ~650 nm. ZnS thin film is deposited on top of CTS film.

The incident solar radiation should travel shorter path and enter the p-n junction, creating more electron and hole pairs. Silver is used as top electrode for the cell. The electrodes are deposited on FTO/CTS/ZnS structure using the thermal evaporation technique. Fig. 2 shows the schematic diagram of the fabrication process of CTS/ZnS heterojunction solar cell.

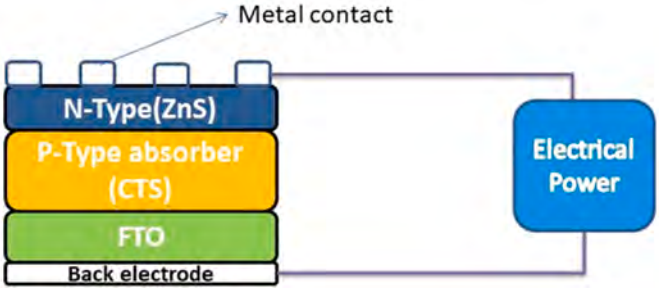


Fig. 1. Structure of Cu_2SnS_3/ZnS heterojunction solar cell.

2.3. Instrument used for characterization

High resolution XRD instrument (Bruker D8) is used for measurement of X-ray diffraction pattern. It is operated at a tube voltage in the range of 20–40 kV and tube current in the range of 2–40 mA. Wavelength of radiation source used is 1.5405 Å and a moving scintillation counter is used as detector. Scanning Electron Microscopy (SEM) and Energy Dispersive X-ray Analysis (EDX) of the films are measured using ZEISS Ultra 55 FE-SEM instrument. UV-VIS spectrophotometer (SPECORD S600 UV-VIS) records the absorption and transmittance spectra from 300 to 1100 nm. The HMS-3000 Hall measurement system having current range of 1nA-20mA and magnetic flux density of 0.5T is used for measuring the electrical properties of ZnS thin films. Fig. 3 shows the USP system setup which is utilized for preparation of thin film layer (both n-type and p-type).

3. Results and discussion

The crystalline structures of the ZnS films grown at different substrate temperatures is analyzed by X-ray diffraction method. The diffraction spectrum is exhibited in Fig. 4. All films are found to be of zinc blende cubic structure and prominent peak at 28.45° corresponds to the (111) plane. Other two weak peaks are observed at 46.72° and 56.43° which corresponds to (220) and (311) plane of ZnS cubic structure (JCPDS card number 05–0566) respectively. There are no additional peaks observed in XRD data. The presence of broad peaks from the image can be attributed to the smaller grain size of the ZnS films. With the increase of substrate temperature, crystallinity of as-deposited films improves slightly. Based on XRD data, lattice parameter value is measured to be 5.38 Å. The lattice constant measured is close to standard data of JCPDS card No. 05–0566.

Bandgap and elemental composition value of ZnS film are summarized in Table 1. The overall composition of the films is measured by using EDX analysis. For lower substrate temperature, films were slightly zinc rich resulting in higher Zn/S ratio. With the increase in the substrate temperature, Zn/S ratio reached to nearly one resulting in stochiometric films. Sample ZT450 is found to be stoichiometric. It can be concluded that composition of films depends on substrate temperature. Sulphur concentration drastically decreases at a substrate temperature 500 °C which is due to the volatile nature of the sulphur.

Fig. 5 displays the morphology of ZnS film using SEM. Films deposited at 350 °C had irregular shape with non-uniform distribution of grains. When temperature is more than 350 °C, morphology of films changed from irregular to spherical structure. The surface morphology of the ZnS film prepared at T_s = 450 °C is relatively more uniform, homogenous, and denser. The sample does not contain cracks or pin

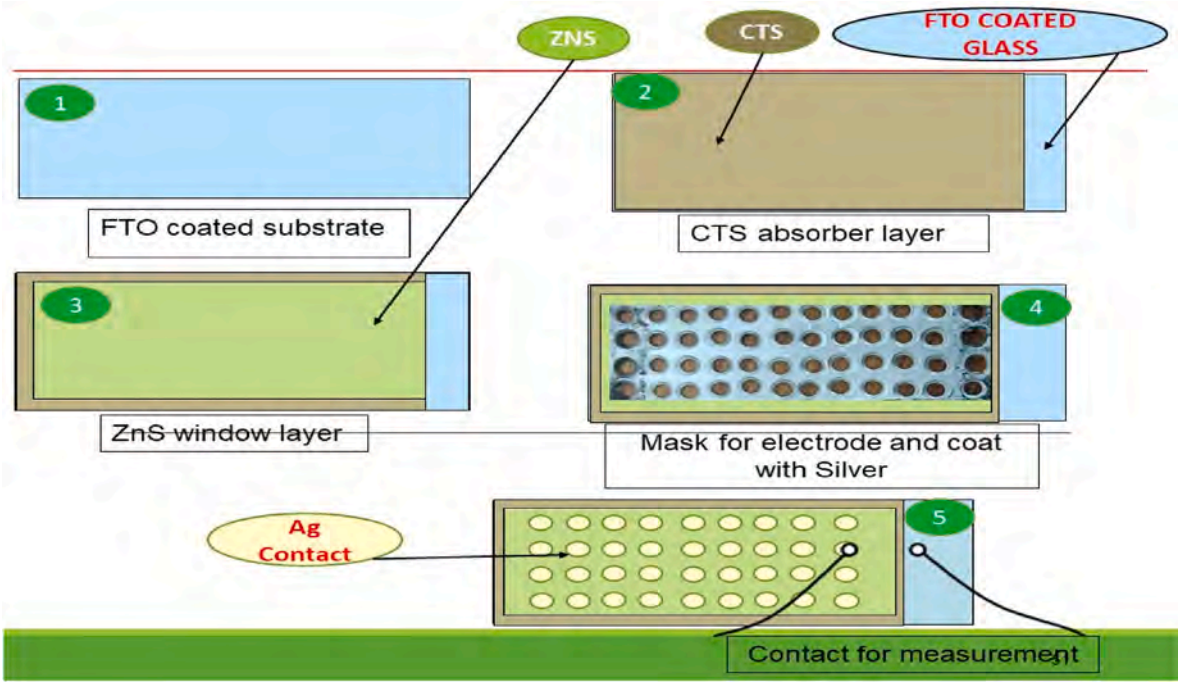


Fig. 2. Fabrication process of CTS/ZnS heterojunction solar cell.

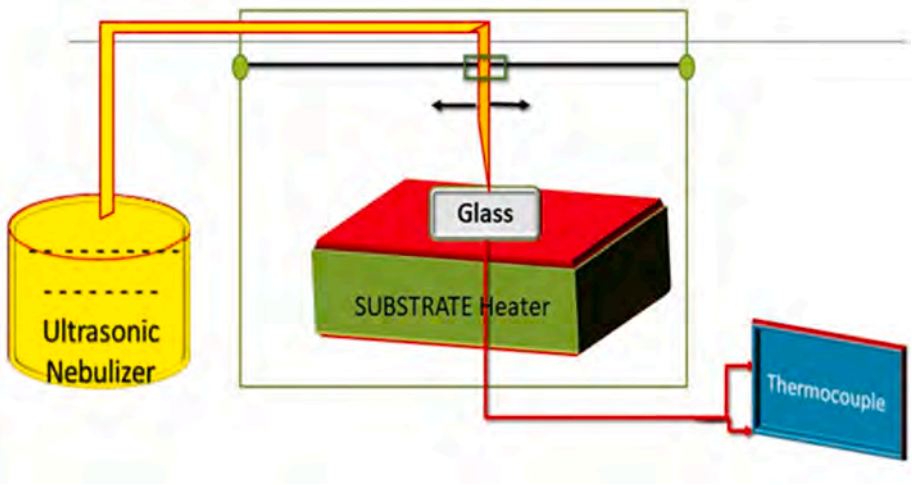


Fig. 3. Schematic diagram of automated ultrasonic spray setup.

holes.

Properties of the material like transmittance and bandgaps are important parameters for solar cell application. A good solar cell window must transmit most of the solar radiation. Hence, the transmittance spectra is measured for the films in the wavelength range from 300 to 1100 nm, which is shown in Fig. 6. The films prepared at higher substrate temperature are found to have more transmittance. For all the films, the transmittance is also seen to increase with increase in wavelength and the transmittance of ZT450 film is found to reach 88% at near infrared wavelength of 1000 nm. However, at such higher wavelength, the transmittance of the film prepared at the substrate temperature of 500 °C is found to be much lower than ZT450. The drop in transmittance is due to higher Zn concentration as zinc rich films exhibits higher reflectance. Similar results are reported for growth of ZnS films using chemical bath deposition techniques [10].

Fig. 7 shows the plot of band gap of the ZnS films at different temperatures. Here, X axis represents the photon energy (hν) and Y axis denotes the value of $(\alpha h\nu)^2$, where α value indicates the absorbance coefficient of material. From the Tauc equation, bandgap of ZnS films is measured [33].

$$\alpha h\upsilon = K(h\upsilon - E_g)^n \quad (1)$$

where K is a proportionality constant, n = 1/2 for direct bandgap semiconductor and Eg is the energy band gap calculated by extrapolating the linear region of the graph on to X-axis. For all the films, the bandgap is found between 3.56 eV and 3.42 eV. From Fig. 7, it can be seen that band gap tend to decrease with the increase in substrate temperature. There are many reasons associated with band gap variation. Many researchers suggest that morphology of the films (nanostructure to microstructure), defects, strain and doping into the films and particle confinement are the main reasons for the change in band gaps [34]. It may be possible that some defects are present in our films which may also have affected the bandgap. Chemical bath deposited films also showed the same behavior of decreasing band gap [35].

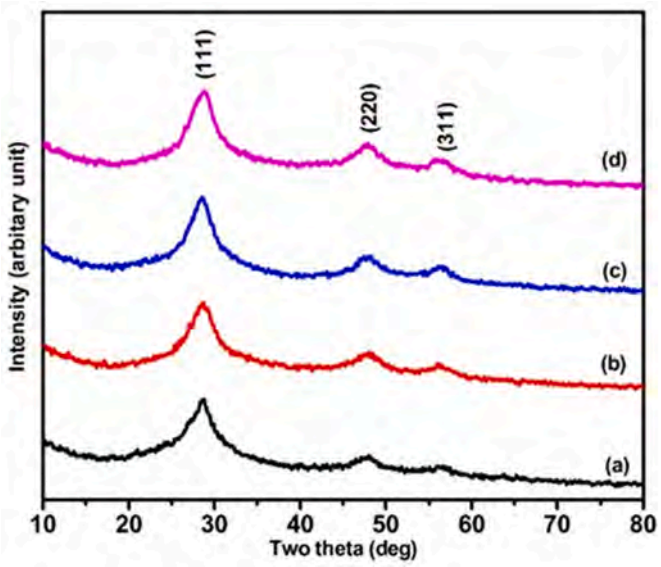


Fig. 4. X-ray diffraction spectra of sprayed ZnS films at different temperature (a) ZT350 (b) ZT400 (c) ZT450 (d) ZT500.

Table 1
Elemental composition and bandgap of ZnS films.

Sample name	The atomic percentage of elements		Bandgap(eV)
	Zn (%)	S (%)	
ZT350	54.58	45.42	3.56
ZT400	52.43	47.57	3.50
ZT450	49.96	50.04	3.43
ZT500	55.94	44.06	3.42

Hall measurement is performed to measure the different electrical parameter values of ZnS films. Table 2 shows the electrical properties of ZnS thin films at different substrate temperature measured using Hall method. All films are found to be n-type semiconductor, an essential requirement for forming the p-n junction. It is observed that there is no significant change in the carrier concentration and mobility of prepared ZnS thin films. The ZnS film resistivity is found to be in the order of 10^6 to 10^5 Ω-cm. As substrate temperature is increased, there is a reduction in resistivity, which may be due to improvement of crystallinity of the film. Reduced resistivity can also be attributed to an increase in grain size, which contributes to a decrease in the scattering at the grain boundaries.

Based on the above study, it is found that ZnS film prepared at 450 °C exhibits the optimal features such as high crystallinity, suitable band gap and high optical transmittance, which can be utilized as n-type semiconductor in solar cell. Thus, the thin film solar cell is prepared using p-type Cu_2SnS_3 and n-type ZnS (ZT450) layer. The J-V characteristic of the Cu_2SnS_3/ZnS solar cell under the dark and AM1.5 illumination is showed in Fig. 8. The intensity of input light is set to 100 mW/cm^2.

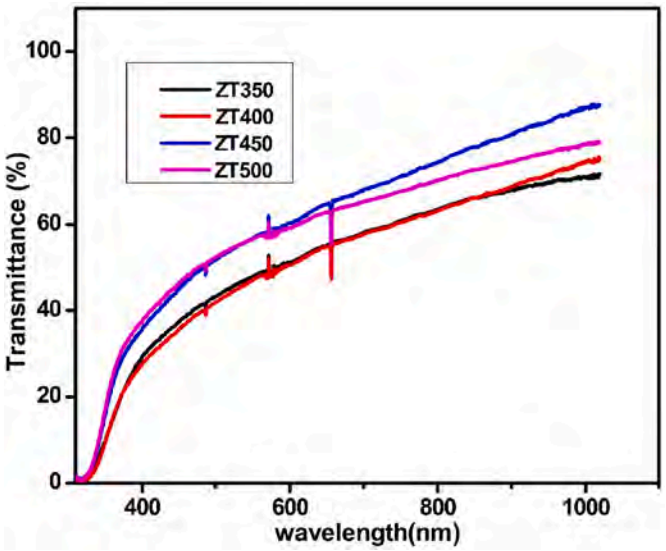


Fig. 6. Transmittance of ZnS films of different substrate temperature.

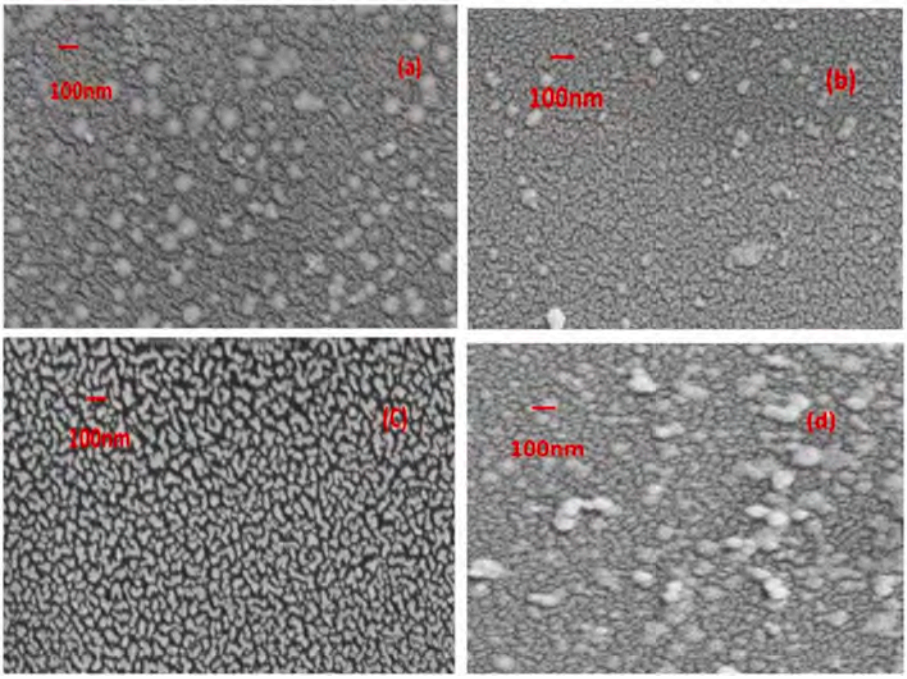


Fig. 5. SEM images of ZnS films (a) ZT350 (b) ZT400 (c) ZT450 (d) ZT500.

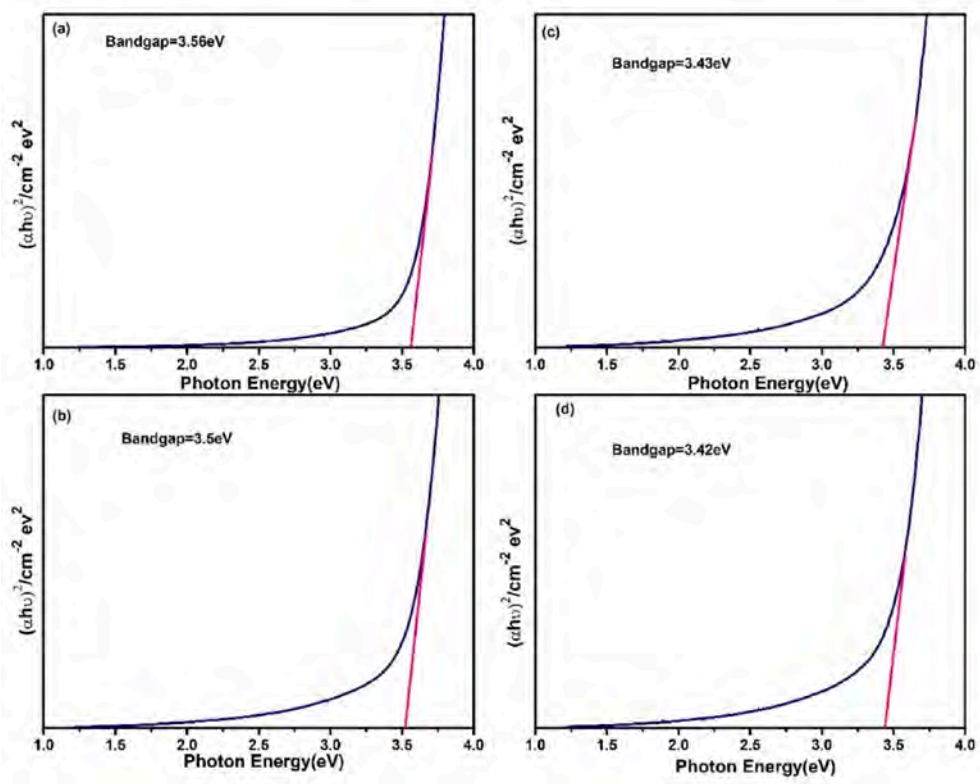


Fig. 7. Bandgap of ZnS films (a) ZT350 (b) ZT400 (c) ZT450 (d) ZT500.

Table 2
Electrical properties of ZnS thin films at different substrate temperature.

Sample name	Type of conductivity	Resistivity (Ω cm)	Carrier Concentration (ions/cm^3)	Mobility (cm^2/V.s)
ZT350	n	1.45×10^6	2.37×10^{12}	1.82
ZT400	n	2.94×10^5	6.13×10^{12}	3.47
ZT450	n	1.85×10^5	7.93×10^{12}	4.21
ZT500	n	1.12×10^5	8.56×10^{12}	6.52

Under the illuminated condition, we obtain the short circuit current density (Jsc) of 1.8 mA/cm^2, open circuit voltage (Voc) of 0.685 V, fill factor of 39% and series and shunt resistances to be of 100 Ω/cm^2 and 230 Ω/cm^2 respectively. However, the efficiency of the solar cell is found to be 0.5% although having 88% high transmittance of ZnS buffer layer at 1000 nm wavelength. To investigate this matter, we have measured the transmittance spectra of the CTS film as shown in Fig. 9. Since, binary phase Cu_2S is found in sample CTS4, so we have selected CTS3 (Cu concentration 0.025 M) as the absorber layer of this solar cell and detailed study is presented elsewhere [12]. Very low transmittance varying from 0 to 5% could be seen in the entire spectra, which justifies better absorbance in this layer. Thus, other factors are responsible for reduction in efficiency such as p-n junction interface which introduces surface states and impurity related traps or defects. Rough and

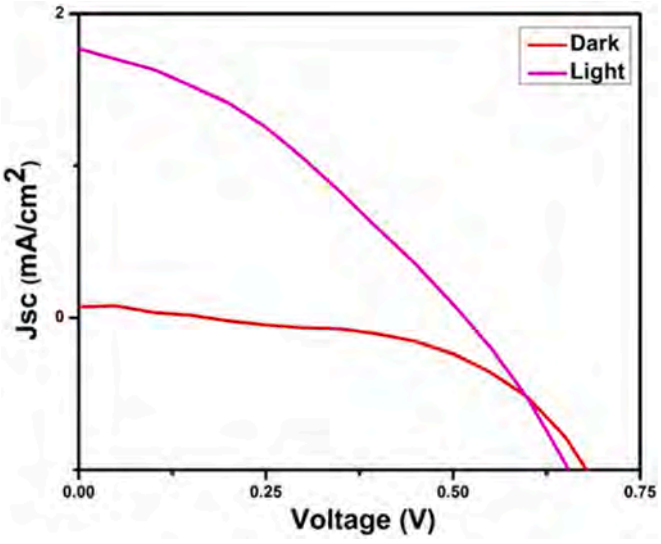


Fig. 8. J-V characteristics of Cu_2SnS_3/ZnS heterojunction solar cell.

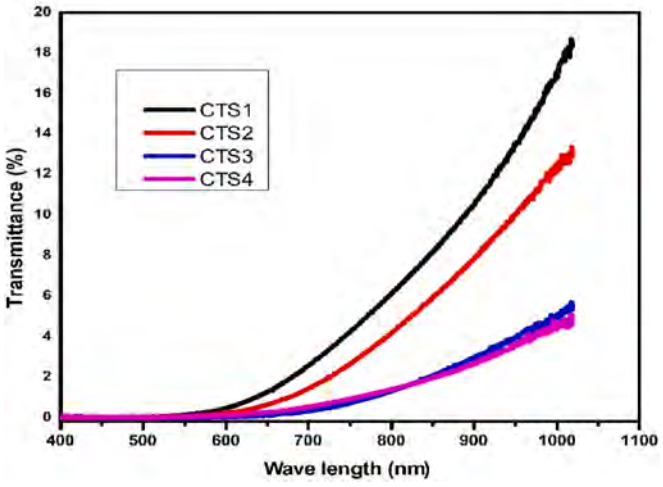


Fig. 9. Transmission spectra of Cu_2SnS_3.

non-uniform interface in the p-n junction and the impurities in the sample act as recombination centers which trap the generated electrons and lowers the efficiency. Post-growth thermal annealing is a well-known method to reduce these traps and improve the optical and electrical properties in nanostructure materials like quantum dots [36, 37]. Usage of ultrahigh purity starting materials has also been found to improve the quality of the compound semiconductors required for device fabrication [38,39]. Thus, research needs to be focused more in optimizing the annealing temperature and purity of the starting material of these thin films, which can pave the way in improving the performance of cost-effective solar cells.

4. Conclusions

N-type semiconductor (ZnS) thin film materials are prepared by USP method with temperature variation of the substrate from 350 to 500 °C in steps of 50 °C. The bandgap of the films also tend to decrease with increase in substrate temperature and the resistivity also reduced from 10^6 to 10^5 Ω-cm. XRD studies showed that the crystalline nature of the films exhibiting zinc blende cubic structure and orientation along (111) plane. From SEM images, the film ZT450 (prepared at substrate temperature of 450 °C) is found to be homogeneous, uniform, dense and void free. At 450 °C substrate temperature with 1:1 ratio of Zn/S, we obtained a very high transmittance reaching to 88% at 1000 nm wavelength. Thus, by analyzing the structural and optical properties, we found ZT450 to be a suitable material as window layer and used the same as a buffer layer for fabrication of Cu_2SnS_3/ZnS solar cell. Device parameters like Jsc, Voc, FF and efficiency are determined from the J-V curve. However, the efficiency of the solar cell is found to be 0.5%. It is presumed that post growth thermal annealing and use of ultrahigh purity starting materials could be effective in reducing the interface traps and improving the efficiency of this low cost and indigenously fabricated solar cell.

CRediT authorship contribution statement

Sabina Rahaman: Investigation, Formal analysis, Resources, Writing – original draft, Writing – review & editing. **M. Anantha Sunil:** Formal analysis, Writing – review & editing, Visualization. **Monoj Kumar Singha:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Kaustab Ghosh:** Writing – review & editing, Visualization.

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

One of the author would like to thank TEQIP-III for providing the grant with file no BMS/COE/TEQIP-III/2020–21/ECE/(03 -02). He also would like to thank Science and Engineering Board (SERB), Govt of India for providing the grant with file No: TAR/2020/000293 under Teachers Associateship for Research Excellence scheme. We thank Micro-nano Characterization facility, Centre for Nanoscience and Engineering at Indian Institute of Science (IISc) for providing the necessary facility to carry out characterization.

References

[1] B. Asenjo, A.M. Chaparro, M.T. Gutiérrez, J. Herrero, J. Klaer, Study of CuInS2/ZnS/ZnO solar cells, with chemically deposited ZnS buffer layers from acidic solutions, Sol. Energy Mater. Sol. Cells 92 (2008) 302–306.

[2] I.K. Sou, Z.H. Ma, Z.Q. Zhang, G.K.L. Wong, Temperature dependence of the responsivity of II–VI ultraviolet photodiodes, Appl. Phys. Lett. 76 (2000) 1098–1100.

[3] C. Jin, H. Kim, H.W. Kim, C. Lee, Enhancement in the photoluminescence of ZnS nanowires by TiO2 coating and thermal annealing, J. Lumin. 130 (2010) 516–519.

[4] Y. Shuai, C. Liu, J. Wang, X. Cui, L. Nie, Pseudo-bi-enzyme glucose sensor: ZnS hollow spheres and glucose oxidase concerted catalysis glucose, Analyst 138 (2013) 3259–3263.

[5] D. Hariskos, S. Spiering, M. Powalla, Buffer layers in Cu (In, Ga) Se2 solar cells and modules, Thin Solid Films 480 (2005) 99–109.

[6] S.M.A. Durrani, A.M. Al-Shukri, A. Iob, E.E. Khawaja, Optical constants of zinc sulfide films determined from transmittance measurements, Thin Solid Films 379 (2000) 199–202, https://doi.org/10.1016/S0040-6090(00)01539-X.

[7] D.H. Hwang, J.H. Ahn, K.N. Hui, K. San Hui, Y.G. Son, Structural and optical properties of ZnS thin films deposited by RF magnetron sputtering, Nanoscale Res. Lett. 7 (2012) 1–7.

[8] X. Fang, T. Zhai, U.K. Gautam, L. Li, L. Wu, Y. Bando, D. Golberg, ZnS nanostructures: from synthesis to applications, Prog. Mater. Sci. 56 (2011) 175–287.

[9] A. Nakaruk, D. Ragazzon, C.C. Sorrell, Anatase thin films by ultrasonic spray pyrolysis, J. Anal. Appl. Pyrolysis 88 (2010) 98–101, https://doi.org/10.1016/j.jaap.2010.03.001.

[10] U. Gangopadhyay, K. Kim, D. Mangalaraj, J. Yi, Low cost CBD ZnS antireflection coating on large area commercial mono-crystalline silicon solar cells, Appl. Surf. Sci. 230 (2004) 364–370.

[11] S. Rahaman, M.A. Sunil, M.K. Singha, K. Ghosh, Temperature dependent growth of Cu2SnS3 thin films using ultrasonic spray pyrolysis for solar cell absorber layer and photocatalytic application, Mater. Res. Express 6 (2019), 106417, https://doi.org/10.1088/2053-1591/ab3928.

[12] S. Rahaman, M.K. Singha, M.A. Sunil, K. Ghosh, Effect of copper concentration on CTS thin films for solar cell absorber layer and photocatalysis applications, Superlattices and Microstructures 145 (2020), 106589, https://doi.org/10.1016/j.spmi.2020.106589.

[13] S. Rahaman, M. Anantha Sunil, M. Kumar Singha, K. Ghosh, Studies of ultrasonically sprayed Cu2SnS3thin films by varying Sn concentration, Mater. Today Proc. 43 (2020) 3938–3941, https://doi.org/10.1016/j.matpr.2021.02.657.

[14] Y.-Z. Yoo, Y. Osaka, T. Fukumura, Z. Jin, M. Kawasaki, H. Koinuma, T. Chikyow, P. Ahmet, A. Setoguchi, S.F. Chichibu, High temperature growth of ZnS films on bare Si and transformation of ZnS to ZnO by thermal oxidation, Appl. Phys. Lett. 78 (2001) 616–618.

[15] I.O. Oladeji, L. Chow, A study of the effects of ammonium salts on chemical bath deposited zinc sulfide thin films, Thin Solid Films 339 (1999) 148–153.

[16] T. Ben Nasr, N. Kamoun, C. Guasch, Physical properties of ZnS thin films prepared by chemical bath deposition, Appl. Surf. Sci. 254 (2008) 5039–5043.

[17] N.K. Allouche, T. Ben Nasr, N.T. Kamoun, C. Guasch, Synthesis and properties of chemical bath deposited ZnS multilayer films, Mater. Chem. Phys. 123 (2010) 620–624.

[18] A. Goudarzi, G.M. Aval, R. Sahraei, H. Ahmadpoor, Ammonia-free chemical bath deposition of nanocrystalline ZnS thin film buffer layer for solar cells, Thin Solid Films 516 (2008) 4953–4957.

[19] H. Hennayaka, H.S. Lee, Structural and optical properties of ZnS thin film grown by pulsed electrodeposition, Thin Solid Films 548 (2013) 86–90.

[20] M. Eslamian, Spray-on thin film PV solar cells: advances, potentials and challenges, Coatings 4 (2014) 60–84, https://doi.org/10.3390/coatings4010060.

[21] X. Zeng, S.S. Pramana, S.K. Batabyal, S.G. Mhaisalkar, X. Chen, K.B. Jinesh, Low temperature synthesis of wurtzite zinc sulfide (ZnS) thin films by chemical spray pyrolysis, Phys. Chem. Chem. Phys. 15 (2013) 6763–6768.

[22] S.S. Yesilkaya, U. Ulutas, H.M.A. Alqader, Effect of Na doping on the properties of ZnS thin films and ZnS/Si heterojunction cells, Mater. Lett. 288 (2021) 1–11, https://doi.org/10.1016/j.matlet.2021.129347.

[23] K. Ben Bacha, N. Bitri, H. Bouzouita, Effect of annealing parameters on structural and morphological properties of sprayed ZnS thin films, Optik 127 (2016) 3100–3104, https://doi.org/10.1016/j.ijleo.2015.12.083.

[24] M.C. Lopez, J.P. Espinos, F. Martin, D. Leinen, J.R. Ramos-Barrado, Growth of ZnS thin films obtained by chemical spray pyrolysis: the influence of precursors, J. Cryst. Growth 285 (2005) 66–75.

[25] T. Nakada, M. Mizutani, Qficiency Cd-free Cu (In, Ga) Se2 thin-film solar cells fabricated using chemical bath deposition (CBD)-ZnS buffer layers, Jpn. J. Appl. Phys. 41 (2002) L165.

[26] D.H. Jeon, D.K. Hwang, D.H. Kim, J.K. Kang, C.S. Lee, Optimization of the ZnS buffer layer by chemical bath deposition for Cu(In,Ga)Se2 solar cells, J. Nanosci. Nanotechnol. 16 (2016) 5398–5402, https://doi.org/10.1166/jnn.2016.12217.

[27] B. Shin, O. Gunawan, Y. Zhu, N.A. Bojarczuk, S.J. Chey, S. Guha, Thin film solar cell with 8.4% power conversion efficiency using an earth-abundant Cu_2ZnSnS_4 absorber, Prog. Photovoltaics Res. Appl. 21 (2013) 72–76, https://doi.org/10.1002/pip.1174.

[28] K.G. Deepa, P.C. Ramamurthy, M.K. Singha, Mesoporous Cu2ZnSnS4 nanoparticle film as a flexible and reusable visible light photocatalyst, Opt. Mater. 98 (2019), 109492.

[29] M.A. Sunil, K.G. Deepa, J. Nagaraju, Growth of AgInS2thin films by ultrasonic spray pyrolysis technique, Thin Solid Films (2014), https://doi.org/10.1016/j.tsf.2013.10.053.

[30] P.S. Patil, Versatility of chemical spray pyrolysis technique, Mater. Chem. Phys. 59 (1999) 185–198.

[31] Monoj Kumar Singha, Aniket Patra, Highly efficient and Reusable ZnO microflower photocatalyst on stainless steel mesh under UV–Vis and natural sunlight, Optic. Mater. 107 (2020), https://doi.org/10.1016/j.optmat.2020.110000.
[32] Monoj Kumar Singha, Aniket Patra, Vineet Rojwal, K G Deepa, Single-step fabrication of ZnO microflower thin films for highly efficient and reusable photocatalytic activity, J. Mater. Sci. Mater. Electron. 31 (2020), https://doi.org/10.1007/s10854-020-03914-6.
[33] J. Tauc, Amorphous and Liquid Semiconductors, first ed., Springer, 1974.
[34] Y.G. Wang, S.P. Lau, H.W. Lee, S.F. Yu, B.K. Tay, X.H. Zhang, H.H. Hng, Photoluminescence study of ZnO films prepared by thermal oxidation of Zn metallic films in air, J. Appl. Phys. 94 (2003) 354–358.
[35] T. Yamaguchi, Y. Yamamoto, T. Tanaka, Y. Demizu, A. Yoshida, (Cd, Zn) S thin films prepared by chemical bath deposition for photovoltaic devices, Thin Solid Films 281 (1996) 375–378.
[36] K. Ghosh, Y. Naresh, N.S. Reddy, Theoretical Optimization of Multi-Layer InAs/GaAs Quantum Dots Subject to Post-growth Thermal Annealing for Tailoring the Photoluminescence Emission beyond 1 . 3μm Theoretical Optimization of Multi-Layer InAs/GaAs Quantum Dots Subject to Post-growth Thermal Annealing for Tailoring the Photoluminescence Emission beyond 1, vol. 3, 2012, https://doi.org/10.1063/1.4739457 l m, 024315.
[37] K. Ghosh, S. Kundu, N. Halder, M. Srujan, S. Sengupta, S. Chakrabarti, Annealing of $In_{0.45}Ga_{0.55}As$/GaAs quantum dots overgrown with large monolayer (11 ML) coverage for applications in thermally stable optoelectronic devices, Solid State Commun. 151 (2011) 1394–1399, https://doi.org/10.1016/j.ssc.2011.06.007.
[38] K. Ghosh, V.N. Mani, S. Dhar, A Modeling Approach for the Purification of Group III Metals „ Ga and in … by Zone Refining, 2008, pp. 1–8, https://doi.org/10.1063/1.2959832.
[39] K. Ghosh, V.N. Mani, S. Dhar, Numerical Study and Experimental Investigation of Zone Refining in Ultra-high Purification of Gallium and its Use in the Growth of GaAs Epitaxial Layers, vol. 311, 2009, pp. 1521–1528, https://doi.org/10.1016/j.jcrysgro.2009.01.102.

Detection of Breast Cancer with Lightweight Deep Neural Networks for Histology Image Classification

H.S. Laxmisagar* & M.C. Hanumantharaju

Department of Electronics and Communication Engineering, BMS Institute of Technology Management, Bengaluru 560064, India

*Address all correspondence to: H.S. Laxmisagar, Department of Electronics and Communication Engineering, BMS Institute of Technology Management, Bengaluru 560064, India; Tel.: +08123008252, E-mail: sagar8.hs@bmsit.in

ABSTRACT: Many researchers have developed computer-assisted diagnostic (CAD) methods to diagnose breast cancer using histopathology microscopic images. These techniques help to improve the accuracy of biopsy diagnosis with hematoxylin and eosin-stained images. On the other hand, most CAD systems usually rely on inefficient and time-consuming manual feature extraction methods. Using a deep learning (DL) model with convolutional layers, we present a method to extract the most useful pictorial information for breast cancer classification. Breast biopsy images stained with hematoxylin and eosin can be categorized into four groups namely benign lesions, normal tissue, carcinoma *in situ*, and invasive carcinoma. To correctly classify different types of breast cancer, it is important to classify histopathological images accurately. The MobileNet architecture model is used to obtain high accuracy with less resource utilization. The proposed model is fast, inexpensive, and safe due to which it is suitable for the detection of breast cancer at an early stage. This lightweight deep neural network can be accelerated using field-programmable gate arrays for the detection of breast cancer. DL has been implemented to successfully classify breast cancer. The model uses categorical cross-entropy to learn to give the correct class a high probability and other classes a low probability. It is used in the classification stage of the convolutional neural network (CNN) after the clustering stage, thereby improving the performance of the proposed system. To measure training and validation accuracy, the model was trained on Google Colab for 280 epochs with a powerful GPU with 2496 CUDA cores, 12 GB GDDR5 VRAM, and 12.6 GB RAM. Our results demonstrate that deep CNN with a chi-square test has improved the accuracy of histopathological image classification of breast cancer by greater than 11% compared with other state-of-the-art methods.

KEY WORDS: CAD, breast cancer, CNN, hematoxylin and eosin, MobileNet, GPU

I. INTRODUCTION

According to the American Cancer Society, cancer is the second most common cause of death for women. In 2018, an estimated 627,000 women died from cancer, accounting for roughly 15% of total cancer deaths in women.[1] The most essential techniques for preventing breast cancer fatalities are early identification[2] and receiving cutting-edge cancer therapy. X-ray radiography, three-dimensional ultrasonography, positron emission tomography, magnetic resonance imaging, tomography, and breast temperature monitoring are all approaches for screening for breast cancer.[3–7] Pathological diagnosis, on the other hand, is frequently viewed as a "golden standard." The extracted tissues are frequently stained for better observation and analysis, including hematoxylin and eosin (H&E) staining is the most common procedure. Hematoxylin paints the nuclei a dark purple color, while eosin dyes the cytoplasm, stroma, and other structures a pink color. However, even experienced pathologists find it difficult to analyze such images, resulting in inter and intra observer differences. Additionally, such analysis is time and resource consuming.[8] The different kinds of breast tissue scan stained with H&E[9] are shown in Fig. 1.

Normal, benign, carcinoma *in situ* (CIS), and invasive ductal carcinoma (IDC) are the four forms of breast tissue.[10] Benign tissue is a slight variation in the structure of the breast that is not cancerous and, in most situations, does not pose a health risk. CIS cancer is limited to the mammary duct lobule system and does not affect the rest of the body. *In situ* carcinoma can be treated if detected early enough. Invasive carcinoma, on the other hand, is a cancerous tumor that spreads to other organs. The most common[11] method for detecting many cancer kinds is histopathological imaging. A histopathologist examines

0278-940X/22/$35.00

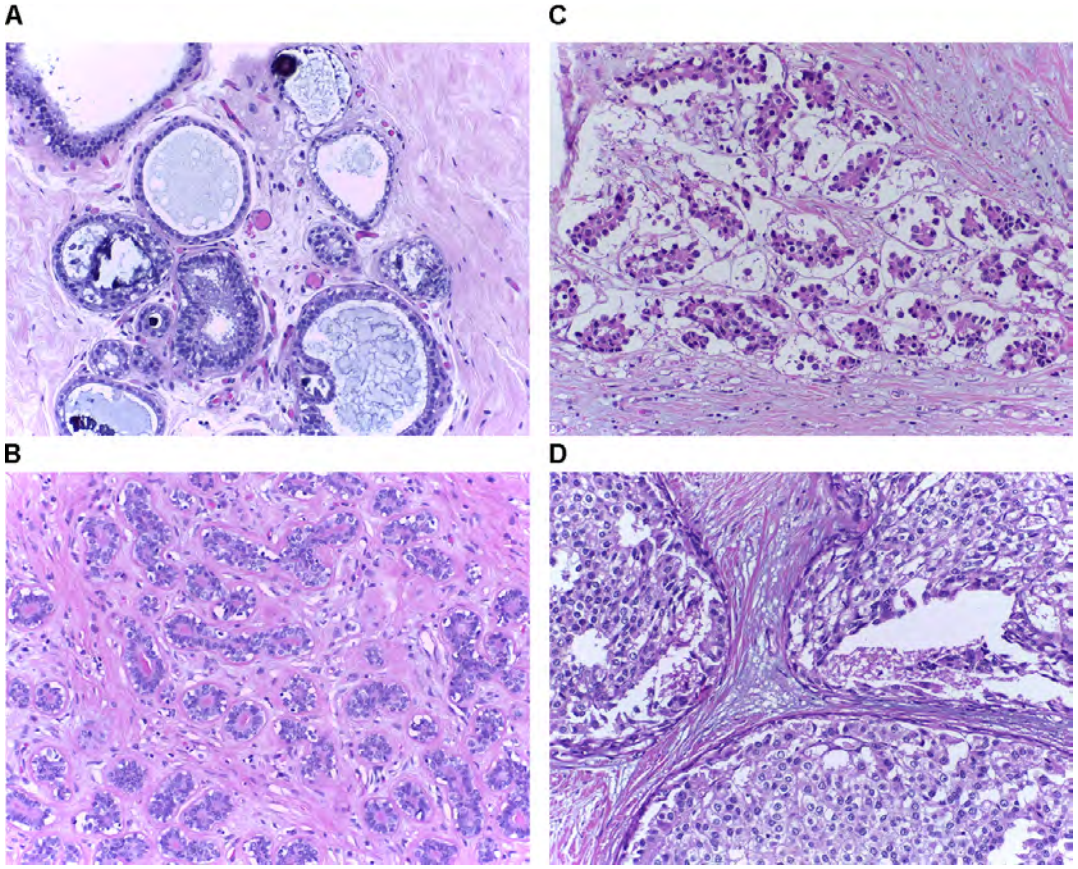


FIG. 1: Sample histopathology images obtained from ICIAR 2018 grand challenge (BACH). (A) Normal. (B) Benign. (C) Invasive. (D) *In situ*.

the sections under an optical microscope to investigate the tissues' features and properties in histopathological research.[12] Traditionally, the histopathologist observes tissue sections with his or her naked eyes, and the visual data is manually examined based on existing medical knowledge. This manual analysis can take a long time due to the intricacy and diversity of histopathological pictures. The accuracy rate of diagnostician's manual classification of cancer kinds appears to be over 75%.[13] The CAD has been developed to distinguish and detect cancer kinds to tackle this challenge. Furthermore, CAD relies on machine learning (ML) algorithms to categorize and identify breast cancer, which is based on histopathological image discriminatory features.[14]

Artificial intelligence (AI) technology is advancing fast in recent years. It has made significant advances in computer vision, image processing, and analysis. In addition, AI has the potential to improve histopathology analysis. Fortunately, advances in computer vision and ML may provide more reliable classification approaches for hematoxylin and eosin-stained sections histological assessment. These approaches have a good classification rate for automatically classifying breast tissues into distinct categories. As a result, numerous researchers have created image examination procedures for breast cancer screening that are quick and exact. Their findings, however, fall far short of what is expected in clinical practice. As a result, the majority of investigators' efforts have gone into developing novel algorithms for histopathological image processing.[15–17] Pathologists' errors can be decreased to some extent, and their productivity can be increased. Deep learning (DL) with convolutional neural network (CNN) has appeared as one of the most powerful ML technologies in image

classification in recent years.[18] According to new research, CNN-based algorithms outperform the best classical ML approaches in terms of performance. In the field of histopathology, DL algorithms have been developed to improve diagnosis accuracy which is more accurate compared with trained pathologists for detection of breast cancer.[19]

Section II develops a comparative table of performances of neural network algorithms on image datasets such as BreakHis, ICIAR BACH, and IMEDIATREAT. Based on the inferences from the literature survey, section III proposes a model to improve the classification performance parameter. The architectural framework and other specifications of the model proposed are discussed in section IV. In section V, the summary of various CAD system development with percentage accuracy is presented in more detail. Finally, a summary and future implementation are presented in section VI.

II. RELATED WORK

The state-of-the-art algorithms for breast cancer identification and related disciplines, such as feature extraction utilizing CAD systems and DL for breast cancer detection are briefly reviewed and discussed in this part. In many hospitals, CAD has quickly grown in recent years to support clinicians in detecting cancer patients faster and more accurately. CAD is utilized to produce objective results and has aided in the diagnosis of medical images in general. The ability to discern between benign, normal, CIS, and malignant carcinoma is one of the most important CAD applications. ResNet50 and DenseNet121 deep CNN approaches for CAD tool development for breast cancer diagnosis were proposed by Yasin et al.[20] The total number of parameters in ResNet50 and DenseNet121 is insufficient to adequately cover the fine detail of 400 images. Only 82 patients were used to create the BreakHis dataset, resulting in a high level of data variability. Erfankhah et al.[22] proposed a local binary pattern (LBP) model with SVM classifier. IDC dataset's main limitations may be related to its small patch size (50 × 50) and binary classification.

Gour et al.[24] proposed a 152-layered CNN based on residual learning. In spite of the high performance of ResHist in detecting breast cancer, it must be validated on a larger dataset before it can be used in healthcare. Hu et al.[25] proposed the myResNet-34, a residual learning-based convolutional neural network. It is required to analyse the myResNet-34 with pre-training on ImageNet and investigate the outcomes of the proposed framework on various datasets, based on the effectiveness of transfer learning reported in recent papers. Yang et al.[29] suggested a multiscale network ensemble (EMS-Net) which is used to find and use the most discriminative areas in each microscope image. Table 1 describes the summary of different techniques proposed by various authors for the detection of breast cancer using BreakHis and IMEDIATREAT dataset.

Table 2 describes the summary of different techniques proposed by various authors for the identification of breast cancer using the BACH dataset. The uniqueness of this study consists in the development of a cost-effective CAD system based on feature extraction and classification utilizing DL approaches to differentiate between normal and abnormal breast cancer lesions.

The BACH challenge dataset was used to examine this CAD system. Multiple articles in the literature have used DCNNs to categorize breast cancer as part of their CAD systems,[20–30] but the accuracy of 83–96% was insufficient for a dependable and powerful CAD system. Furthermore, they did not look at ways to lower the CAD system's computing costs. To address these flaws, this research proposes a revolutionary CAD method for developing low-cost deep neural networks, which enhances the CAD system's accuracy. In medicine, images are vital[31] to improve the processes to treat patients in addition to helping doctors perform better diagnostics.

As per our knowledge, no paper is found related to the implementation of an efficient lightweight DCNN algorithm for the detection of breast cancer, and most existing development is related to other DCNNs such as ResNet, AlexNet, VggNet, and EMS-Net as discussed in the literature survey of state-of-the-art methods.

III. RESEARCH METHODOLOGY

The outline of the proposed work is shown in Fig. 2. In the data preprocessing steps such as data

TABLE 1: Summary of different techniques proposed by various authors for the detection of breast cancer using the BreakHis[34] and IMEDIATREAT[35] datasets

Author	Publisher, year	Dataset	Accuracy rate	Limitations
Yari et al.[20]	IEEE Transaction, 2020	Histopathology Image: BreakHis Dataset Method: DCNN: ResNet50 and DenseNet121	CAD tool development for breast cancer detection with 98 (100×) % accuracy	Another future scope could be to investigate the lower precision in the 400× image. The investigation of lower accuracy in the 400× image can be another future work.
Mansour et al.[21]	World Scientific Journal, 2020	Histopathology Image: BreakHis Dataset Method: CNN: AlexNet-DNN	CAD tool development for breast cancer detection with 96.20%	The proposed method can be examined with additional cutting-edge GPUs such as Nvidia and others. To achieve a more time-efficient method that produces the highest quality BC-CAD solution.
Erfankhah et al.[22]	IEEE Journal, 2019	Histopathology Image: BreakHis Dataset 400× Method: Local binary pattern (LBP) model with SVM classifier	CAD tool development for breast cancer detection with 83.40%	The fundamental restriction of this dataset is the nature of its classes. Two test portions of the same tissue type (for example, fat tissue) may come from two distinct scans.
Lichtblau, et al.[23]	IEEE, 2019	Histopathology Image: IMEDIATREAT data set Method: AlexNet model with various Classifier	CAD tool development for breast cancer detection with 87%	FTT+PCA makes a mistake by relying on the cumulative weight of the correct classifiers and assigning the proper FTT+PCA label where other classifiers may have performed better.
Gour et al.[24]	Wiley Publication, 2020	Histopathology Image: BreakHis Dataset Method: Residual learning-based CNN	CAD tool development for breast cancer detection with 84.34% accuracy	The suggested approach has numerous advantages over hand-crafted local feature descriptors; nonetheless, training the ResHist model requires enormous computational power and time when compared with feature descriptors.
Hu et al.[25]	Wiley Publication, 2021	BreakHis dataset: Method: Convolutional neural network named myResNet-34	CAD tool development for breast cancer detection with 94.03% accuracy	The classification accuracy may be improved by employing other DL-based classifier models and better stain normalization approaches.

augmentation and normalization techniques were used to increase the number of samples and model accuracy by removing image staining. The term "data augmentation" denotes the process of adding comparable data to an existing dataset to strengthen the model. Normalization is a pre-processing procedure for preparing biopsy imaging slides that employ a staining approach to eliminate staining**.** Finally, we used different MobileNet models to find out model accuracy.

TABLE 2: Summary of different techniques proposed by various authors for the identification of breast cancer using the BACH dataset

Author	Publisher, year	Dataset	Accuracy rate	Limitations
Aresta et al.[26]	Elsevier, 2019	Histopathology Image: ICIAR, 2018, BACH challenge dataset Method: Convolutional neuronal networks	CAD tool development for breast cancer detection with 87% accuracy	The BACH dataset remains freely accessible to support additional developments in digital pathology for CAD system development. The accuracy obtained is less due to the CNN model trained without a transfer learning approach.
Kassani et al.[27]	IEEE ICTC, 2019	Histopathology Image: ICIAR, 2018, BACH challenge dataset Method: Deep Convolutional Neural Network (DCNN)	CAD tool development for breast cancer detection with 92.50% accuracy	The classification accuracy may be improved by employing a DL-based classifier model and better stain normalization approaches.
Kausar et al.[28]	Elsevier, 2019	Histopathology Image: ICIAR, 2018 and BreakHis dataset Method: Deep convolution neural network (CNN)	ICIAR: CAD tool development for breast cancer detection with 91% accuracy BreakHis: 96.85% accuracy	The insufficient amount of accessible standard datasets lowers the refinement capacity of the deep CNN model.
Yang et al.[29]	Elsevier, 2019	Histopathology Image: ICIAR, 2018, BACH challenge dataset Method: Ensemble of MultiScale Networks (EMS-Net)	CAD tool development for detection of breast cancer with 90% accuracy	The network is time-consuming during training, it may be done offline. Introduce the new approach to locate and use the furthermost exclusionary areas in each microscope image.
Kumar et al.[30]	Cancer Management and Research, 2020	Histopathology Whole slide Images: ICIAR, 2018, BACH challenge dataset Method: k-means clustering algorithm	Normal Class: 85.5% accuracy	The outcome can be enhanced, if there is a significant number and variety of labeled whole slide image sample datasets.

FIG. 2: Overview of the proposed model

A. Dataset Description

A collection of 400 high-resolution histology images[33] is included in this archive. The resolution of each image ranges from 2040 × 1536 pixels. This image is made of histology microscopic images stained with H&E that were used in the bio-imaging breast histology categorization challenge 2018. There are four categories of image classes: benign, normal, *in situ*, and invasive. A diagnosis was produced from the contents of the image classes by two experienced pathologists. The bio-imaging challenge's purpose is to conduct data categorization amongst distinct classes. The collection contains 400 microscopic pictures, with 100 images each for benign, normal, invasive carcinoma, and carcinoma *in situ* conditions. The tiny pictures are 2048 × 1536 pixels and have an RGB color model with a .tiff format.

B. Data Pre-Processing

We pre-process the dataset which is collected in two steps such as normalization and data augmentation. Normalization is a pre-processing procedure for preparing biopsy imaging slides that employ a staining approach to eliminate staining as shown in Fig. 3. A logarithmic[13,14] transformation is used to convert the high-resolution pictures to an optical density. The optical density tuples are then subjected to the singular value decomposition, and the result in color space transform is utilized to produce the odd biopsy picture. The histogram picture[15] is then expanded to encompass the lower 90% of the dataset. The picture samples before and after normalizing are shown in the diagram below. Computers don't need to be stained; only pathologists need to dye histology pictures to distinguish between different regions of interest.

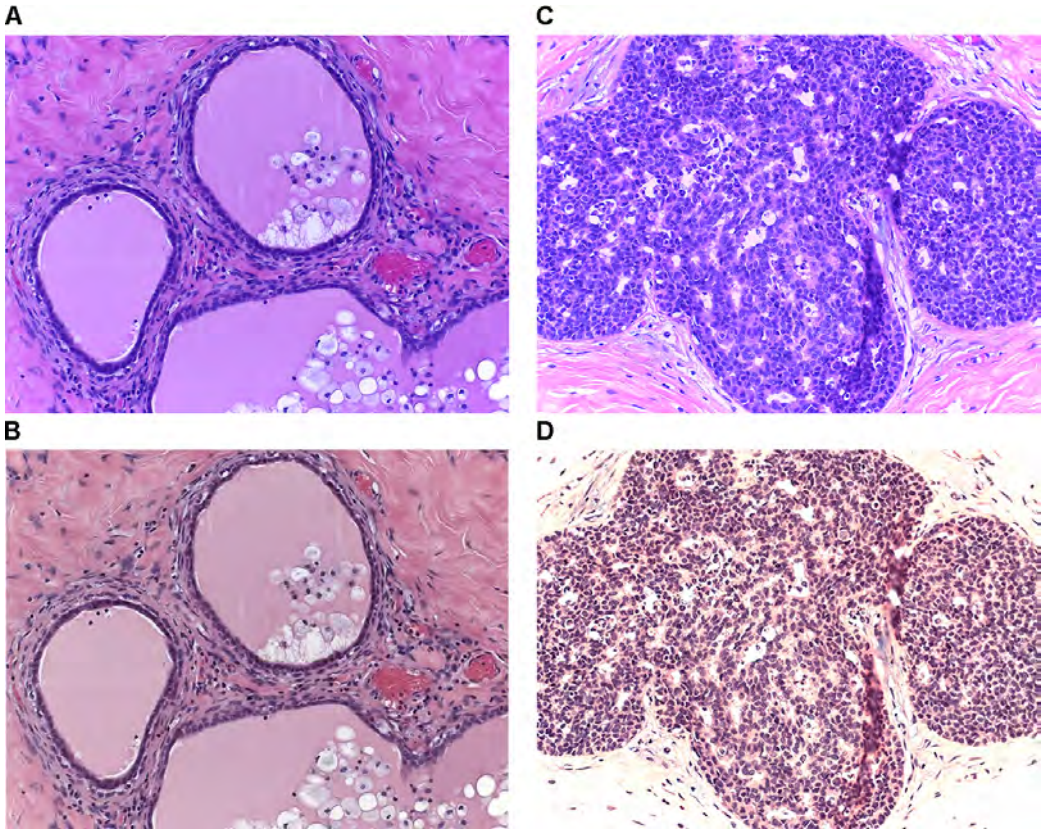


FIG. 3: Normalization of images (before vs. after): (A,C) before normalization of images; (B,D) after normalization of images

The term "data augmentation" refers to the process of adding comparable data to an existing dataset to strengthen the model. In addition, the normalized pictures are used to build a dataset that is used to rotate the image in the process. A data augmentation strategy based on picture rotation and flipping is used to enhance[15] the dataset without impacting diagnostic or image quality to avoid underfitting owing to a small number of samples. The data is supplied into the CNN model once it has been normalized and enhanced. Once the model has learned the underlying attributes of these four classes, it will be able to predict images without labels. Image Data Generator is a class in Keras that allows handlers to perform image augmentation in a very basic method.

For our deep neural network to train the model, the dataset, whatever it is utilized for, is extremely little. To avoid this, researchers employed several data augmentations on pre-processed histology pictures, such as rotating the x- and y-axis, flipping and cropping the image, to construct a robust model and avoid overfitting issues. As illustrated in Fig. 4, data augmentation of the original image with respect to the axis has been performed.

To perform high-precision diagnostics, convolutional neural networks require a large dataset size. One of the challenges in CNN is insufficient medical

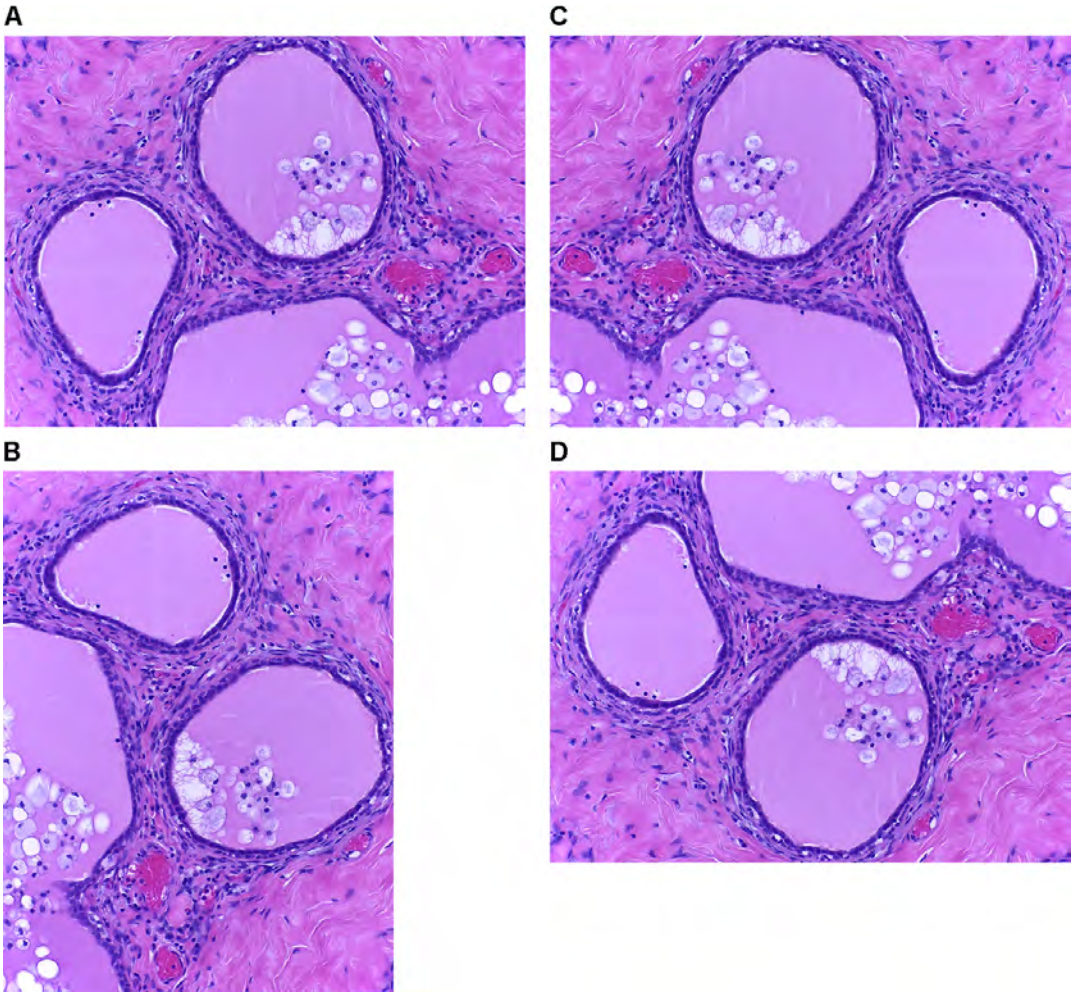


FIG. 4: Image augmentation (rotation about the axis). (A) Original image. (B) Data augmentation 1. (C) Data augmentation 2. (D) Data augmentation 3.

data set size, and the over-fitting problem which means CNN performed in terms of accuracy is high for the training dataset, but it performs less with the test dataset. Therefore, in the proposed system, the present study has used the data augmentation technique to solve the problems of data set size and over-fitting. In the data augmentation technique, the number of each microscopic image was increased by geometric transformations such as translation, scaling, rotation, and flipping.

IV. MODEL ARCHITECTURE

Convolutional Neural Networks are a family of deep, feed-forward artificial neural networks, most typically employed to evaluate visual data. CNNs employ a form of multilayer perceptron's designed to need little preprocessing. In a convolutional neural network, each neuron is only linked with a few local neurons in the previous layer, and the weight is distributed for every neuron in that layer. Convolutional neural networks are useful for image classification tasks because the convolution procedure generates information on spatially associated elements of the image.

With MobileNet, you are primarily going to use depth-wise discrete convolutions, which multiply a regular convolution by a depth-wise convolution as well as a 1 × 1 convolution can be called a pointwise convolution.[8] As shown in Fig. 5, the convolutional neural network technique MobileNet 2.10ex is employed to implement the operation, which includes two layers, one of which performs depth-wise convolutions and another of which performs pointwise convolutions. The ReLu activation function is employed at the output of each neuron to prevent neuron loss. We can use depth-wise convolutional layers and a convolutional layer of 1 × 1 for combining.[19] After the first phase is completed, the next phase involves the extraction of a feature matrix that contains labeled data. Four hidden layers of the deep neural network analyze the feature data and classify the data into four different categories. Models with low latency such as MobileNet are ideal for systems and applications that require models to be small. The underlying architecture of MobileNet is tiny and has low latency, making it ideal for the needs of embedded vision and mobile applications. In order to improve the classification performance on a small dataset, it is required to incorporate a trained weight transfer model using ImageNet database as shown in Fig. 6.

The proposed convolutional neural network, MobileNet model, was used to classify four different classes of data for the detection of breast cancer. This method has included four steps such

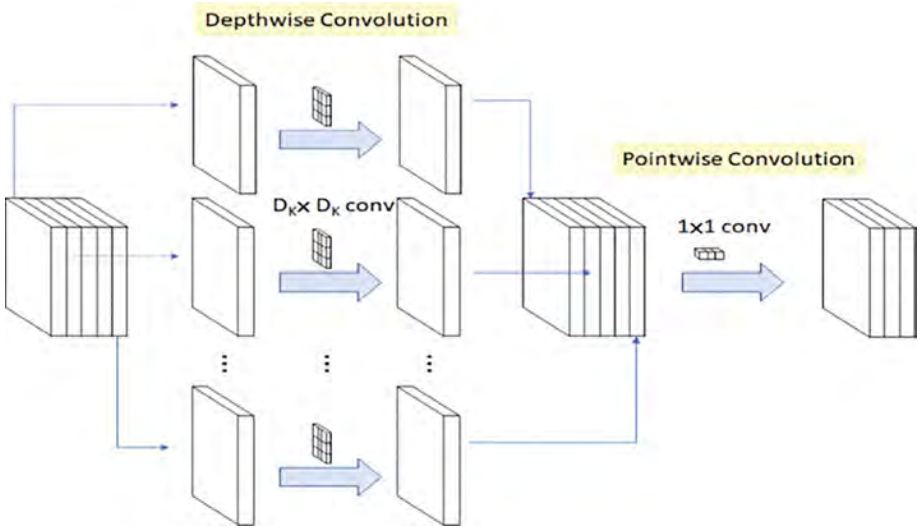


FIG. 5: MobileNet architecture[8]

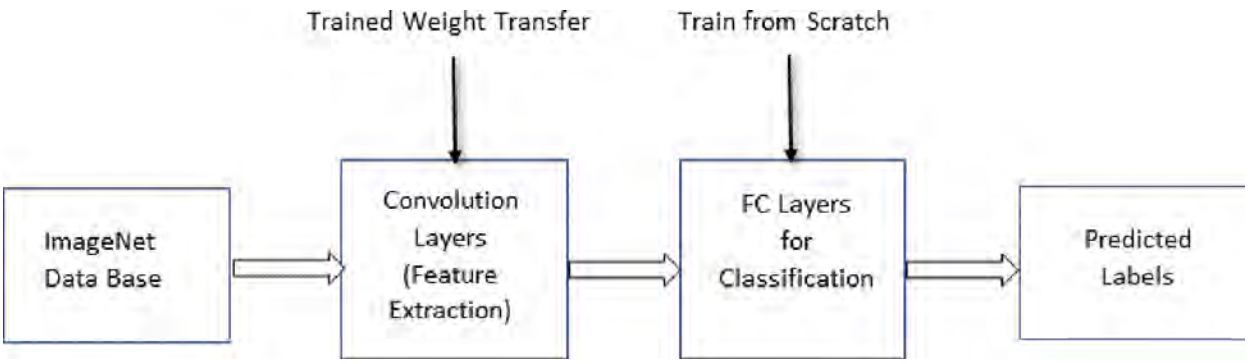


FIG. 6: Trained weight transfer model using the ImageNet database

as data normalization, data augmentation, feature extraction, and classification as shown in Fig. 7. First, pre-treatment is necessary to remove noise from histology images, and normalization is done as a pre-processing procedure to eliminate staining from microscopic images. This step improves diagnostic performance. Second, the data augmentation technique is used to increase the amount of training data effectively. A data augmentation strategy based on picture rotation and flipping is used to enhance the dataset without impacting diagnostic or image quality to avoid underfitting owing to a small number of samples. Third, the augmented dataset is used to train the network on images of various multi-scales,

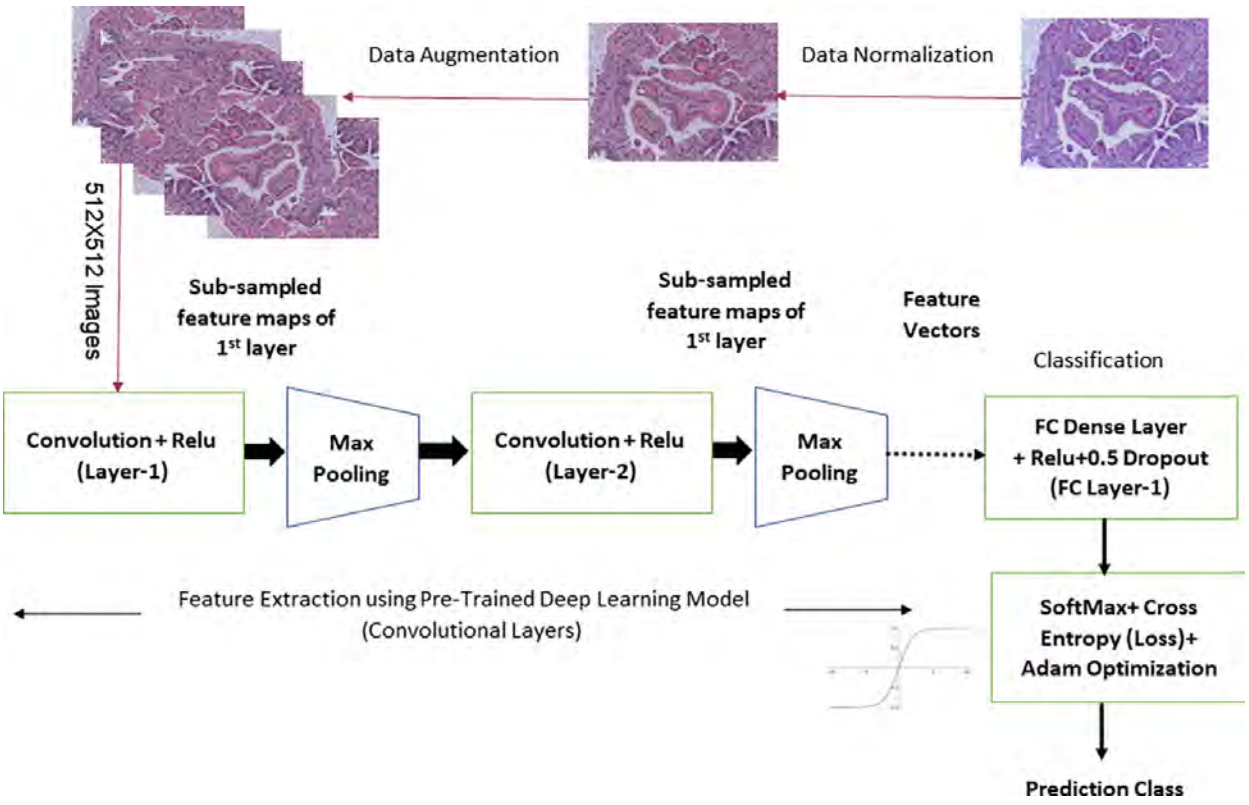


FIG. 7: Framework for histology image classification

and then the deep features of the breast histology are extracted. In the feature extraction downsampling is performed using max-pooling at various stages to get feature vectors. Once the model has learned the underlying attributes of these four classes, it will be able to predict images without labels. In the proposed system, the MobileNet model was used to extract deep features. Fourth, the deep features are fed into the fully connected layer, which classifies the microscopic images into four different classes.

All layers of MobileNet follow a batch-normalization layer and include rectified linear unit (ReLu) nonlinearity except for the last fully connected layer, on which there is no nonlinearity. Additionally, there is the output dense layer for prediction with four different classes of data. The equation of rectified linear unit is given below,

$$R(Z) = \max(0, Z) \tag{1}$$

where R(Z) is the activation function and Z denotes the input to a neuron.

This function value is set to 0 if it receives a negative input, and if it receives a positive value, the function will retain the same positive value. The softmax function is very essential for multiple classes which are more than two.

It converts real numbers to normalized probability distribution by taking the exponents of each output. Once that has been done, regularize each number by the sum of each exponent so that the entire output vector adds up to one. At the last stage of an image classifier, a softmax function is used, such as those used in ImageNet competitions, such as CNN (VggNet).

$$\sum_{i=1}^{C} P_i = 1 \tag{2}$$

The sum of all probabilities at the output of softmax equals to 1 and the number of output probabilities varies from i = 1 to C, where C denotes total number of output probabilities of softmax function. Using the following equation, the softmax classification result may be predicted with more confidence as one of several outcomes,

$$\text{Confidence} = max_i P_i \tag{3}$$

where $max_i P_i$ is the maximum probability value at the output of softmax function.

The softmax function output is selected based on probability value which is higher compared with all others. The higher probability value is selected with more confidence to classify correct class of data in case of multiple classes to be classified.

$$P_i \geq 0,\ P_i \leq 1, \text{ where, } i = 1,2,3,\ldots\ldots,C$$

The output probability value of softmax function lies between the range 0 to 1 ($P_i \geq 0$, $P_i \leq 1$) as mentioned above. Softmax function converts all score values into normalized probability distribution,

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \tag{4}$$

where $S(y_i)$ is the softmax function for input real number y_i, $\sum_j e^{y_j}$ is the sum of all the exponents of input data y_i and it is a single number, e^{y_i} denotes exponent of input data y_i.

Categorical cross-entropy loss is a loss function used in multiclass classification problems. A probability density function model is designed to measure how different two probability densities differ when an example relates only to one of several possible categories. Its formal purpose is to identify the difference between two probability densities. The categorical cross-entropy is used by the model to learn to assign a high probability to the right class and a low probability to the other categories. The error is calculated using cross-entropy and the equation below

$$L = -\sum t_i \log y_i \quad (5)$$

where L denotes categorical cross-entropy loss, t_i is the truth label and y_i is the softmax probability for i*th* class. The aim is to minimize the loss, i.e., the smaller the loss the better the model. As an alternative to stochastic gradient descent, Adam can be used to iteratively update the weights of networks based on training data, instead of the traditional stochastic gradient descent process. The Adam technique is a stochastic gradient descent replacement method for training DL models. Table 3 shows pseudo-code for data normalization and augmentation, mobileNet model creation, training and validation of images, model evolution, and inferring the class of an image.

The notion of transfer learning is utilized throughout the model training process. Transfer learning in computer vision may be depicted using pre-trained models. Transfer learning models outperform typical ML models in terms of performance. This is because models that use information (features, weights, etc.) from previously trained models already comprehend the features. It is speedier than training neural networks from the ground truth images. Thus, it is best to use models such as VggNet, Inception, and MobileNet to reduce the need for multiplication when training such models.

In this approach, a transfer learning method is used for classifying breast cancer histology images. The network has learned the deep features of the MobileNet architecture, which is pre-trained on ImageNet.

The ICIAR BACH dataset has been classified into benign, normal, *in situ*, and invasive. ImageNet pre-trained models were used for transfer learning to get better accuracy. One of the main challenges of employing DL models in the field of medicine is a lack of training data due to difficulty in collecting and labeling data, which needs to be performed by experts. To overcome this drawback, transfer learning (TL) has been utilized to solve several medical imaging tasks using pre-trained state-of-the-art models from the ImageNet dataset as shown in Fig. 8.

Figure 9 depicts steps involved to perform histopathology image classification for detection of breast cancer. Breast cancer histopathology image datasets used in the present work have been obtained from https://iciar2018-challenge.grand-challenge.org/Dataset/. The resolution of each image ranges from 2040 × 1536 pixels. This image is made of histology microscopic images stained with H&E that were used in the bio-imaging breast histology categorization challenge 2018. There are four categories of image classes: benign, normal, *in situ*, and invasive. After collecting the dataset, it is required to pre-process the dataset in two steps, such as normalization and data augmentation, to remove staining from biopsy images and to increase the size of the dataset. To perform high-precision diagnostics, convolutional neural networks require a large dataset size. In the data augmentation technique, the number of each microscopic image was increased by geometric transformations such as translation, scaling, rotation, and flipping. Image Data Generator is a class in Keras that allows handlers to perform image augmentation in a very basic method. In the proposed system, the MobileNet model was used to extract deep features. In the feature extraction downsampling is done using max-pooling at various stages to get feature vectors. Finally, the deep features are fed into the fully connected layer, which classifies the microscopic images into four different classes. Four hidden layers of the deep

TABLE 3: Overview of the results obtained through the process of model validation

Models	Benign	*In Situ*	Invasive	Normal	Model accuracy
Model 2.00	10	10	4	12	75.0%
Model 2.10ex	10	11	7	10	79.2%
Model 1.40	10	11	10	11	87.5%
Model 1.40ex	11	11	7	6	72.9%

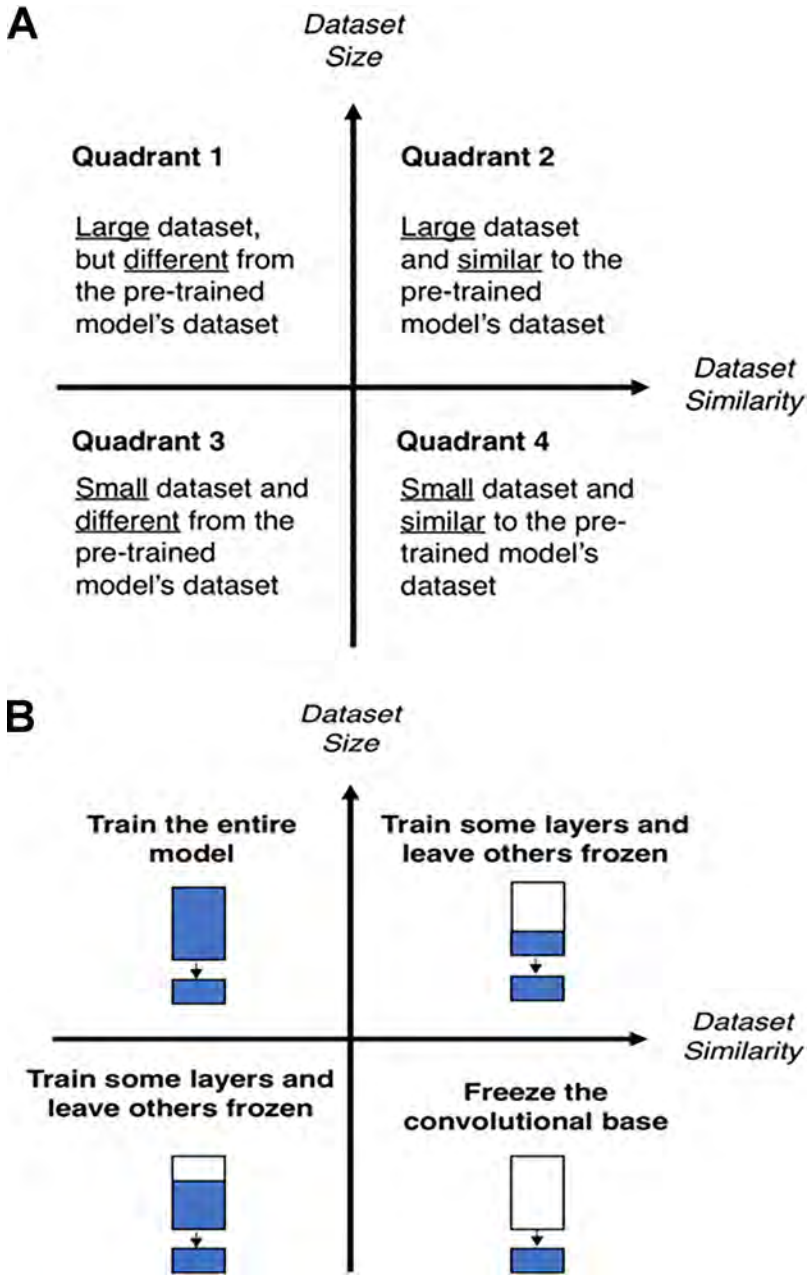


FIG. 8: Transfer learning strategies.[32] (a) Dataset size-similarity matrix; (b) Fine-tuning pre-trained models.

neural network analyze the feature data and classify the data into four categories.

V. RESULTS AND ANALYSIS

The model was trained up to 280 epochs in total to achieve the best results feasible given the technological restrictions. An overview of the results obtained through the process of model validation with respect to different MobileNet models is depicted in Table 3.

For our tests, we utilized the Tensorflow backend and the Keras frameworks for neural - network development as well as dataset preparation. We

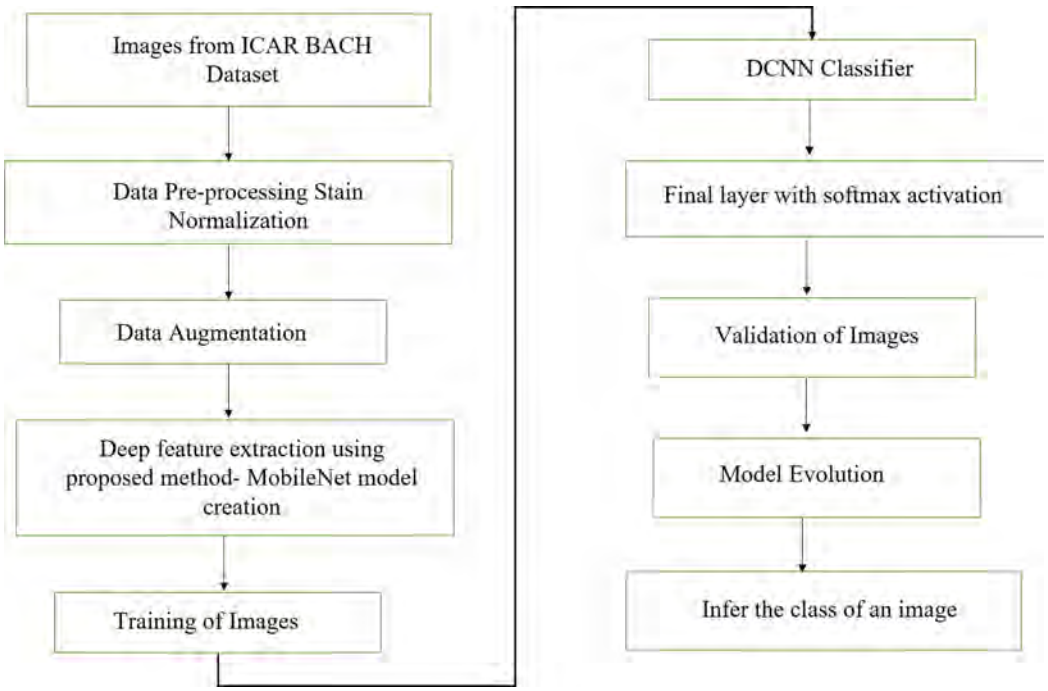


FIG. 9: Flowchart depicting steps involved to perform histopathology image classification for detection of breast cancer

divided the data into three groups, with 88% of the instances used for training and 12% utilized for testing. The training period was set at 280 epochs with 25 training steps each epoch. The overall graphical representation of model training and validation accuracy is shown in Fig. 10.

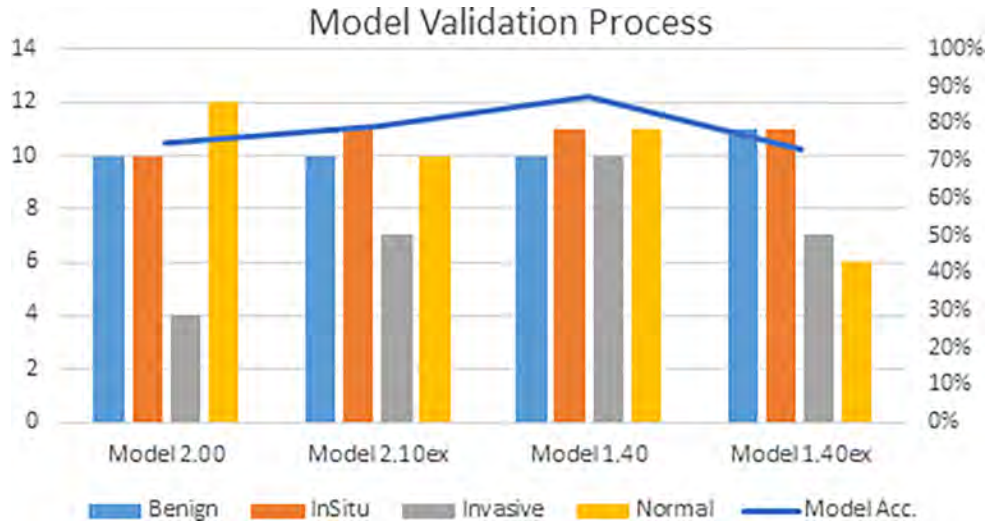


FIG. 10: Graphical overview of the results obtained over the process of the model validation. The x-axis denotes different MobileNet models and the y-axis represents model accuracy (RHS) and the number of correct classifications (LHS).

Figure 11 shows the learning approach for the four classifiers in the BACH dataset with 400 tissue pictures. The model's accuracy improves with time as the epochs continue. In the meanwhile, we can identify some unexpected declines inaccuracy. This is mostly due to the learning rate, which is set at 0.0001. According to our data and the model's properties, this learning rate seems to be rather high. Google Colab trained the model for 280 epochs using its powerful GPU with 2496 compute units, 12.6 GB RAM, and 12 GB GDDR5 memory. The training and validation accuracy of the model can be plotted by the number of DL epochs applied once it has been trained. This depicts that they are moving extremely near to each other, with the distance between the two being as small as possible.

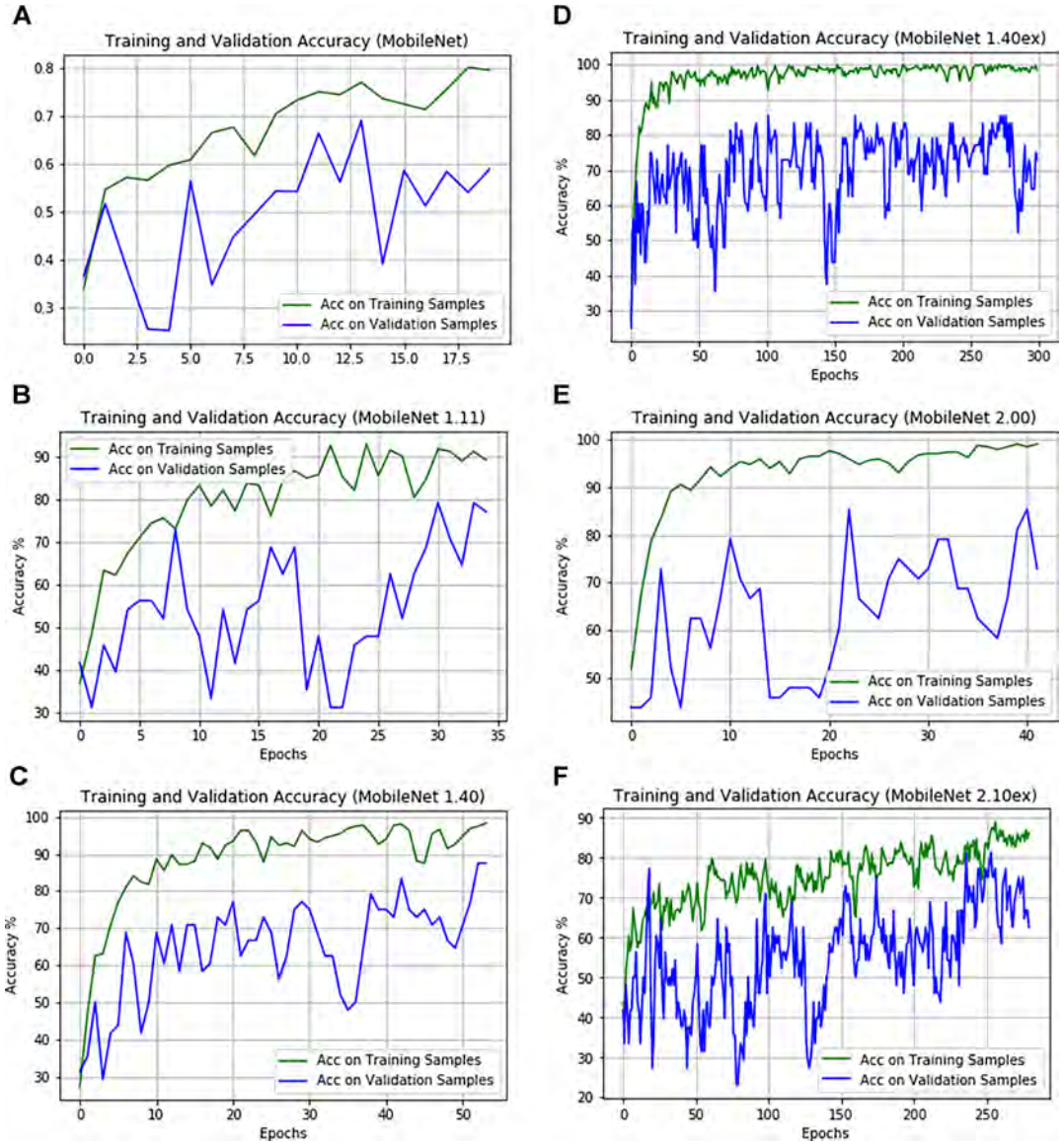


FIG. 11: Training and validation accuracy of different MobileNet models. The x-axis denotes the number of epochs corresponding to each model and the y-axis represents model accuracy in terms of percentage accuracy for MobleNet models. (A) Model 1.10; (B) model 1.11; (C) model 1.40; (D) model 1.40ex; (E) model 2.00; (F) model 2.10ex.

Figure 12 shows the confusion matrix for the classification of histopathology images into normal, benign, *in situ*, and invasive by the number of models trained. The feature score can be computed using the chi-square test with degrees of freedom, k = 10 for selecting the best 30 features out of 1024 features extracted using deep CNN as shown in Table 4. Figure 13 shows the learning approach for the four classifiers in the BACH dataset with 400 tissue pictures.

The scikit train, test and split technique is used to further separate features data (x data) and target data (y data) into train and test. The first 80% of the data was allocated to training, while the remaining 20% was allocated to testing. Setting the test size option to 0.2 controls the train, test, and split fraction. The data that is prepared can be fed to the ML model such as linear support vector classifier kernel to compute percentage accuracy. Similarly, the same steps are carried out to select the 39 best features out of 1024 features.

The graphical representation of the chi-square test score computed for the 30 best features to compute percentage accuracy is shown in Fig. 14. The summary of CAD system development with percentage accuracy is presented in Table 5.

VI. CONCLUSIONS

The main contribution of this article is the categorization of four distinct classes of data using microscopic images as benign, normal, invasive, and *in situ* carcinoma. The ICIAR-BACH competition dataset was enhanced to make it more adaptable for convolutional neural networks. MobileNet's deep CNN architecture is employed in the feature extraction step. This MobileNet has been trained to identify microscopic medical images, and its parameters have been modified to differentiate between four types of input. With multiple architectural versions developed and tested, we created a high-performance design that has an estimated accuracy of 87.5%.

As a result of its excellent accuracy, as well as its resource-efficient architecture, this version of MobileNet design has high potential. This low-cost model presents an intriguing prospect for industry-related studies including breast cancer screening. In the future, we can use modern FPGA boards to build this lightweight deep neural network model. Other deep neural network architectures, such as the residual network (ResNet) and visual geometry group (VggNet) will be proposed in the future to analyze performance parameters.

It is critical to correctly categorize histopathological pictures to correctly determine the kind of breast cancer. DL was used to effectively classify breast cancer. The categorical cross-entropy is used by the model to learn to assign a high probability to the right class and a low probability to the other classes. It is employed in the classification step of

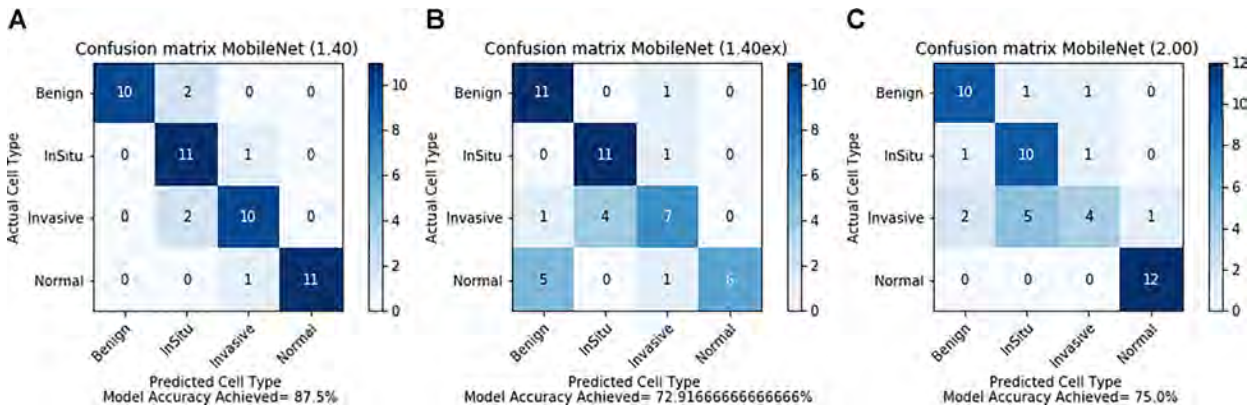


FIG. 12: Confusion matrix for the four classes normal, benign, invasive carcinoma, and *in situ* carcinoma. The x-axis denotes the predicted cell type corresponding to each class of data and the y-axis represents the actual cell type. (A) Model 1.40; (B) model 1.40 ex; (C) model 2.00.

TABLE 4: Summary of the chi-square test score for the 30 best features to compute accuracy

Best feature	Specs	Score
630	630	719.05
798	798	712.83
371	371	697.30
533	533	665.06
183	183	645.84
175	175	612.44
669	669	594.89
830	830	590.01
48	48	589.36
20	20	574.84
210	210	558.34
777	777	551.21
836	836	541.32
949	949	540.86
530	530	522.40
758	758	515.41
540	540	513.50
121	121	511.61
712	712	499.24
914	914	494.21
198	198	493.98
330	330	493.71
806	806	485.31
574	574	481.13
682	682	472.58
282	282	469.01
802	802	467.02
888	888	466.09
877	877	453.98
335	335	442.52

the CNN following the clustering stage, which improves the performance of the suggested system. The suggested method made use of the Python programming language, which enhanced classification accuracy while shortening the model's processing time. To measure training and validation accuracy, the model was trained on Google Colab for 280 epochs with a powerful GPU with 2496 CUDA cores, 12 GB GDDR5 VRAM, and 12.6 GB RAM. The feature score is computed using the chi-square test with degrees of freedom, k = 10 for selecting the best 30 features out of 1024 features extracted using deep CNN. The data that is prepared using the train_test_split technique is fed to the ML model such as linear support vector classifier kernel to compute percentage accuracy. Therefore, deep CNNs with chi-square tests have improved the histopathological image classification of breast cancer

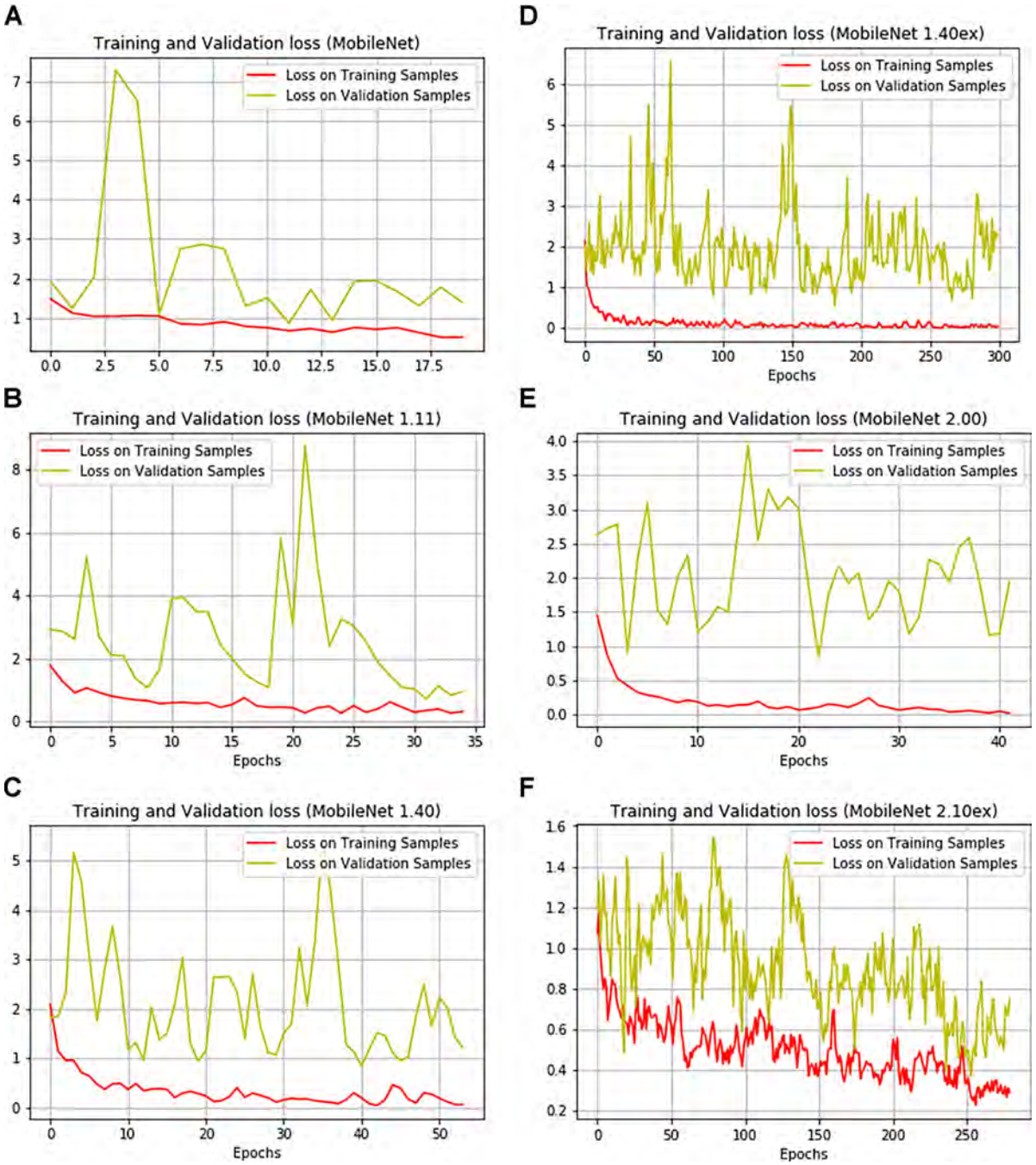


FIG. 13: Training and validation loss (red or bottom lines and yellow or top lines, respectively). The x-axis denotes the number of epochs and the y-axis represents model loss for MobileNet models. (A) Model 1.10; (B) model 1.11; (C) model 1.40; (D) model 1.40ex; (E) model 2.00; (F) model 2.10ex.

with an increase of 11.25% accuracy. The proposed system might be improved further by future studies utilizing a huge dataset of histopathology images in the CNN training process, and even by using huge feature extractions in CNN to build models that can further improve the proposed system. Furthermore,

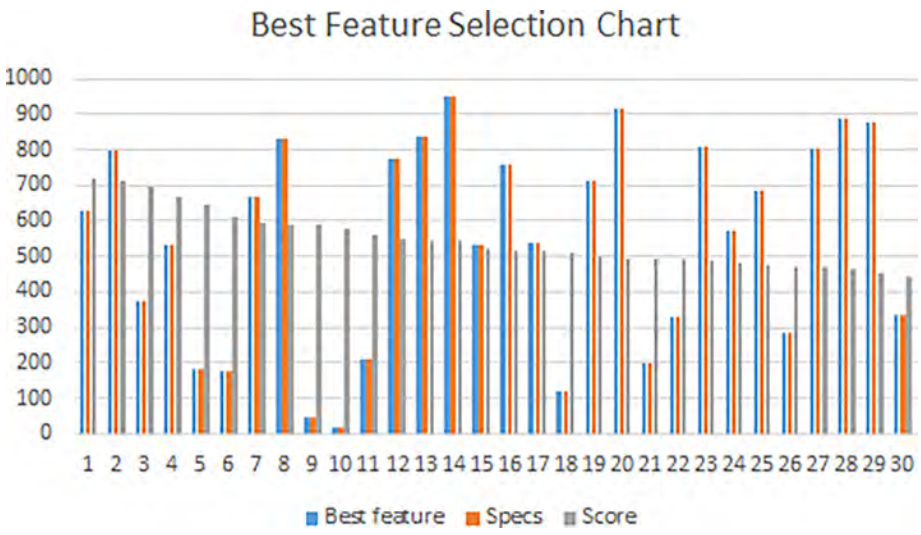


FIG. 14: Graphical chart of chi-square test score computed for 30 best features. The x-axis denotes the 30 best features and the y-axis represents the best score value corresponding to each feature.

TABLE 5: Summary of CAD system development with percentage accuracy

Method	Dataset	Features	Percentage accuracy	Number of classes	Model
Software	Histopathology	1024	87.50	Four	DCNN-Mobilenet 1.40
Software	WBCD	30	97.20	Two	SVM
Software	Histopathology	30	98.75	Four	SVM
Software	Histopathology	39	94.00	Four	SVM

more research is needed on the multi-level categorization of breast cancer.

REFERENCES

1. World Health Organization. Preventing cancer. Available from: http://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/.
2. American Cancer Society. Breast cancer early detection and diagnosis. Available from: https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection.html.
3. Moghbel M, Ooi CY, Ismail N, Hau YW, Memari N. A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography. Artif Intell Rev. 2020;53(3):1873–918.
4. Kozegar E, Soryani M, Behnam H, Salamati M, Tan T. Computer aided detection in automated 3-D breast ultrasound images: A survey. Artif Intell Rev. 2020;53(3):1919–41.
5. Domingues I, Pereira G, Martins P, Duarte H, Santos J, Abreu PH. Using deep learning techniques in medical imaging: A systematic review of applications on CT and PET. Artif Intell Rev. 2020;53(6):4093–160.
6. Murtaza G, Shuib L, Abdul Wahab AW, Mujtaba G, Nweke HF, Al-Garadi MA, Zulfiqar F, Raza G, Azmi NA. Deep learning-based breast cancer classification through medical imaging modalities: State of the art and research challenges. Artif Intell Rev. 2020;53(3):1655–720.
7. Moghbel M, Mashohor S. A review of computer assisted detection/diagnosis (CAD) in breast thermography for breast cancer detection. Artif Intell Rev. 2013;39(4):305–13.
8. De Matos J, Britto Jr AD, Oliveira LE, Koerich AL. Histopathologic image processing: A review. arXiv preprint arXiv:1904.07900. 2019 Apr 16.
9. ICIAR Grand Challenge 2018 on Breast Cancer Histology Images. Available from: https://zenodo.org/record/3632035#.YqgaL6FBxPY.
10. Zhou X, Li C, Rahaman MM, Yao Y, Ai S, Sun C, Wang Q, Zhang Y, Li M, Li X, Jiang T. A comprehensive review for breast histopathology image analysis using classical and deep neural networks. IEEE Access. 2020;8:90931–56.
11. Vo DM, Nguyen NQ, Lee SW. Classification of breast

cancer histology images using incremental boosting convolution networks. Inf Sci. 2019;482:123–38.
12. Li C, Chen H, Li X, Xu N, Hu Z, Xue D, Qi S, Ma H, Zhang L, Sun H. A review for cervical histopathology image analysis using machine vision approaches. Artif Intell Rev. 2020;53(7):4821–62.
13. Jadoon MM, Zhang Q, Haq IU, Butt S, Jadoon A. Three-class mammogram classification based on descriptive CNN features. BioMed Res Int. 2017;2017:3640901.
14. Loukas C, Kostopoulos S, Tanoglidi A, Glotsos D, Sfikas C, Cavouras D. Breast cancer characterization based on image classification of tissue sections visualized under low magnification. Comput Math Methods Med. 2013;2013:829461.
15. Gurcan MN, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B. Histopathological image analysis: A review. IEEE Rev Biomed Eng. 2009;2:147–71.
16. Kowal M, Filipczuk P, Obuchowicz A, Korbicz J, Monczak R. Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. Comput Biol Med. 2013;43(10):1563–72.
17. Tang J, Rangayyan RM, Xu J, El Naqa I, Yang Y. Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances. IEEE Trans Inf Technol Biomed. 2009;13(2):236–51.
18. Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. IEEE Trans Pattern Anal Mach Intell. 2013;35(8):1798–828.
19. Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I, Hulsbergen-Van De Kaa C, Bult P, Van Ginneken B, Van Der Laak J. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci Rep. 2016;6:26286.
20. Yari Y, Nguyen TV, Nguyen HT. Deep learning applied for histological diagnosis of breast cancer. IEEE Access. 2020;8:162432–48.
21. Mansour RF. A robust deep neural network based breast cancer detection and classification. Int J Comput Intell Appl. 2020;19(1):2050007.
22. Erfankhah H, Yazdi M, Babaie M, Tizhoosh HR. Heterogeneity-aware local binary patterns for retrieval of histopathology images. IEEE Access. 2019;7:18354–67.
23. Lichtblau D, Stoean C. Cancer diagnosis through a tandem of classifiers for digitized histopathological slides. PLoS One. 2019;14(1):e0209274.
24. Gour M, Jain S, Sunil Kumar T. Residual learning based CNN for breast cancer histopathological image classification. Int J Imaging Syst Technol. 2020;30(3):621–35.
25. Hu C, Sun X, Yuan Z, Wu Y. Classification of breast cancer histopathological image with deep residual learning. Int J Imaging Syst Technol. 2021;31(3):1583–94.
26. Aresta G, Araújo T, Kwok S, Chennamsetty SS, Safwan M, Alex V, Marami B, Prastawa M, Chan M, Donovan M, Fernandez G. Bach: Grand challenge on breast cancer histology images. Med Image Anal. 2019;56:122–39.
27. Kassani SH, Kassani PH, Wesolowski MJ, Schneider KA, Deters R. Classification of histopathological biopsy images using ensemble of deep learning networks. arXiv preprint arXiv:1909.11870. 2019 Sep 26. Available from: https://arxiv.org/abs/1909.11870
28. Kausar T, Wang M, Idrees M, Lu Y. HWDCNN: Multi-class recognition in breast histopathology with Haar wavelet decomposed image based convolution neural network. Biocybern Biomed Eng. 2019;39(4):967–82.
29. Yang Z, Ran L, Zhang S, Xia Y, Zhang Y. EMS-Net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images. Neurocomputing. 2019;366:46–53.
30. Kumar A, Prateek M. Localization of nuclei in breast cancer using whole slide imaging system supported by morphological features and shape formulas. Cancer Manag Res. 2020;12:4573.
31. Chorianopoulos AM, Daramouskas I, Perikos I, Grivokostopoulou F, Hatzilygeroudis I. Deep learning methods in medical imaging for the recognition of breast cancer. In 2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA). IEEE Computer Society; 2020. p. 1–8.
32. Marcelino P. Transfer learning from pre-trained models: How to solve any image classification problem quickly and easily. Available from: https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751.
33. Aresta G, Araújo T, Kwok S, Chennamsetty SS, Safwan M, Alex V, Marami B, Prastawa M, Chan M, Donovan M, Fernandez G. Bach: Grand challenge on breast cancer histology images. Med Image Anal. 2019;56:122–39.
34. Breast Cancer Histopathological Database (BreakHis). Available from: https://www.kaggle.com/datasets/ambarish/breakhis.
35. IMEDIATREAT Dataset. Available from: https://sites.google.com/site/imediatreat/data-sets.

SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348 – 8379 / https://doi.org/10.14445/23488379/IJEEE-V9I9P106

Volume 9 Issue 9, 47-53, September 2022

Original Article

Evaluating the Statistical Stability of POSIT Arithmetic and IEEE 754 Float to Accelerate Data for Detection of Breast Cancer

H. S. Laxmisagar[1], M. C. Hanumantharaju[2]

[1,2]*Department of Electronics and Communication Engineering, BMS Institute of Technology and Management, Bengaluru, India.*

[1]*Corresponding Author : sagar8.hs@bmsit.in*

Received: 25 July 2022 Revised: 12 September 2022 Accepted: 19 September 2022 Published: 30 September 2022

Abstract *- In recent years, the researcher has focused on the hardware implementation of Floating Point Units (FPUs), which have a huge area and energy footprint. Due to their greater accuracy, speed, and simpler hardware design, Posit Arithmetic Units (PAUs) are proposed to replace IEEE 754-2008 compliant FPUs. It is important to improve existing floating-point IP cores for field-programmable gate array (FPGA) based applications. In comparison with other floating-point formats, posits number representation offers greater dynamic range and numerical accuracy. Various researchers attempted to implement a support vector machine (SVM) using hardware implemented on FPGA platforms to achieve high performance at lower power consumption and cost. As a result, the algorithm is unsuitable for embedded real-time applications. Therefore, SVM linear classifier is implemented on hardware, decreasing the latency and executing the task in real-time. This paper proposes an SVM linear classifier with pipeline architecture for fast processing in Verilog HDL using a single-precision IEEE standard 754 number format with 32-bit representation. The POSIT algorithm design is done with 24-bit representation using a software approach to determine the accuracy of prediction for the detection of Breast Cancer. The accuracy rate is computed both using software and hardware for performance evaluation. The pipelined SVM architecture is designed using Verilog HDL and synthesized using the Vivado simulation tool. The design is configured to the Xilinx KC705 Kintex-7 evaluation board for implementation.*

Keywords *- Breast Cancer, FPGA, IEEE 754 format, SVM, Vivado tool, Verilog HDL.*

1. Introduction

Breast cancer seriously risks women's lives and health [1]. The illness is treatable in the early stages, but it is not recognized until later, which is the primary reason for the death of so many women worldwide. Breast cancer detection at an early stage is critical for successful treatment and lowering the mortality rate. In the United States, 12.15 percent of women [2] will develop this cancer during their lifetime. For the last two decades, machine-learning researchers have been improving classifier effectiveness. As a result, machine-learning research has resulted in a new generation of cutting-edge supervised machine learning [29] classification algorithms, such as SVMs.

SVMs are considered a very good supervised learning algorithm that offers cutting-edge accuracy at the cost of high computational complications [4]. On the other hand, the standard versions of the SVM algorithm are very time-consuming and computationally intensive, presenting a challenge to investigate other hardware architectures than the CPU. It motivates the SVM to be realized on hardware platforms to attain high-performance computation at low-power consumption and cost.

Posit has various benefits over floating point numbers, including greater accuracy, a wide dynamic range, the capacity to provide results that are bitwise equal across platforms, simpler hardware, and the capacity to manage exceptions more effectively [5]. Compared to IEEE float floating point units, posit processing units require less circuitry. It is more efficient to employ low-precision posits than inexact computing methods that attempt to compensate for decreased quality of answers [6]. A wide range of applications is accelerated and performed with a substantial performance by FPGAs, proving that they are superior to other comparable platforms, such as graphics processing units (GPUs) and general-purpose processors [25]. As a result, FPGAs have become increasingly popular for realizing and accelerating SVM on hardware for low-power embedded system applications.

Section 2 describes the literature review related to the identification of Breast cancer, SVM, IEEE 754 float, POSIT arithmetic, and FPGAs, which are necessary for understanding the suggested method for the identification of cervical cancer. Section 3 discusses the system architecture implementation with floating point multipliers and adder trees in more detail. Section 4 presents the proposed design simulation, synthesis, and implementation

in more detail. Finally, a brief conclusion and future implementation are presented in section 5.

2. Related Work

In real-time scenarios, Deep Neural Networks (DNNs) are increasingly constrained by real-time constraints, so information representation must be re-evaluated. Encoding information in a way that can be processed rapidly and represented in a hardware-friendly way is a very challenging endeavor. In recent studies, posit formats have been shown to satisfy real-time constraints better than IEEE 754 standards for floating point representations of real numbers. According to Cococcioni et al. [8], Posit-based DNNs can be activated in the same way as 32-bit floats with 16-bit down to 10 bits. Despite their lower performance, 8-bit Posits could be a good substitute for 32-bit floats due to their high speed and low storage properties.

According to Van Dam et al. [9], there are approximately 160 MPOPS for multiplication, 250 MPOPS for addition, and 180 MPOPS for accumulations. As a result, intermediate results can be rounded off without affecting decimal accuracy. For the same number of bits, posit arithmetic gives higher decimal accuracy than the IEEE floating point format. Selvathi et al. [10] proposed integrating a computerized diagnostic system for breast cancer identification onto FPGA using Artificial Neural Network (ANN). The features of the WBCD are used to train and test the Multilayer Perceptron Neural Network (MLPNN). Baez et al. [11] proposed using Xilinx SDSoCTM to optimize the HLS methodology (Software-Defined SoC (System on chip)). The multiclass SVM classifier algorithm is designed using C programming and implemented on ZC7020 (ZedBoard) and ZC7045 (ZC706) devices.

Mohammadi et al. [12] proposed hardware implementation of SVM with parallel processing using Stochastic Gradient Descent for implementation onto FPGA. Fiolhais et al. [26] proposed a fused dot product processor with exact operand arithmetic that can output an incomplete result per cycle. The processor employs a long interval accumulator with full fixed-point precision to achieve full accuracy. Bassoli et al. [14] investigated FPGA implementations of IEEE 754 floating-point adder and multiplier for fast data processing using a conventional approach and customizations to save the chip area. The binary parallel multiplier and adder, which is a digit-serial multiplier, were proposed by Louca et al. [15]. One of the most significant limitations of existing embedded FPGAs is their major hardware development effort and time-to-market.

Table 1. Summary of different methods for breast cancer detection with limitations for the WBCD dataset

Author	Method	Application	Hardware	Acc. Rate	Area, Timing, and Power	Limitations
T Chandrasekaran et al., 2021 [16]	ANN based machine learning classifier	Breast Cancer Detection	Hardware-Software Co-design (CMOS-IC)	ANN=96.9%	-------	The limited dataset has been used for training. The ANN requires a huge volume of data for proper training.
Selvathi, D et al. 2018 [17]	ANN implementation using FPGA	Breast Cancer Classification	FPGA Virtex 5 board using XILINX ISE tool	90.83%	-------	limitation is optimization is not done, and no on-board Implementation
Batista et al. 2020 [18]	Pseudo-logarithmic number representation	Data classification	FPGA	-------	--------	Accuracy rate not computed Timing optimization not done Expensive multipliers with simpler adders
González-Díaz_Conti et al. 2022 [19]	ANN	Breast Cancer Detection	Xilinx Vivado 2016.1 FPGA Device: Zynq-7 ZC702	-------	LUTs:2267 Power:1.752 W Timing: ------	Accuracy rate not computed Timing optimization is not done

Numerous researchers [20]–[23] have contributed their work to implement the SVM machine learning algorithm, which is becoming increasingly popular in research due to the promising opportunities it provides in this field. Kim et al. [28] presented an IEEE 754 single precision multiplication method to cut down on the number of logic blocks. The suggested work estimates only one of the operands and iteratively compensates for the calculation error to attain limited errors in respective applications.

Most authors proposed CAD system development for the detection of Breast cancer, and very few papers report FPGA implementation for the detection of various types of cancer. As per the literature review, no paper is found on implementing an efficient SVM classification system for the identification of Breast cancer onto FPGA. Most existing implementations other than Breast cancers are not realized on modern FPGAs board, as evident from the literature review.

3. Proposed Architecture for Faster Data Classification

3.1. System Architecture

The SVM linear classifier's proposed FPGA design is shown in Figure 3. The architecture's adder tree constructs the classification function f(X) with the bias 'b' and polynomials such as patient data and weight value. Once testing data is passed into the FPGA board, the trained model uses it to perform a classifying procedure. The results are compared using the decision function to obtain either benign or malignant.

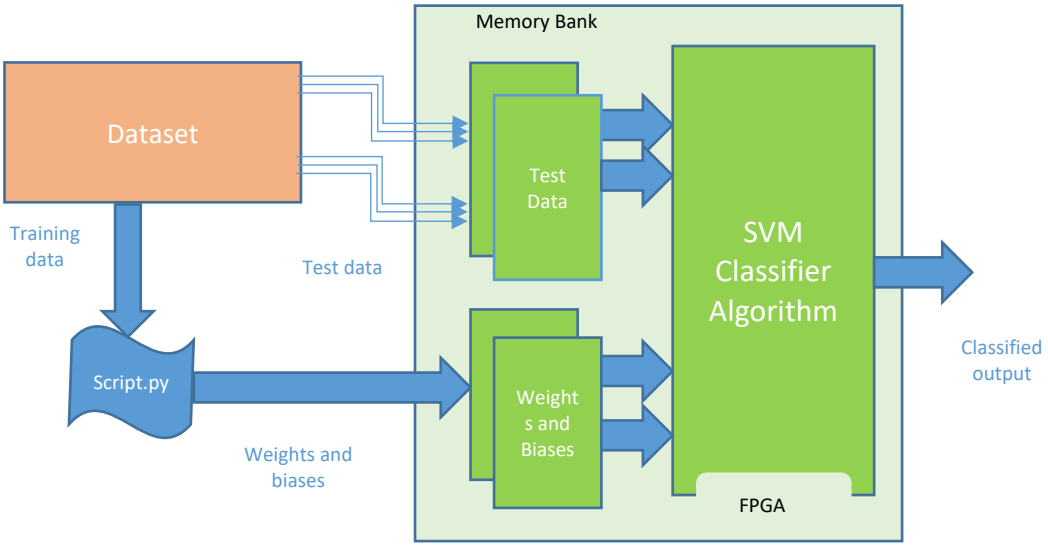


Fig. 1 System Architecture for Classification

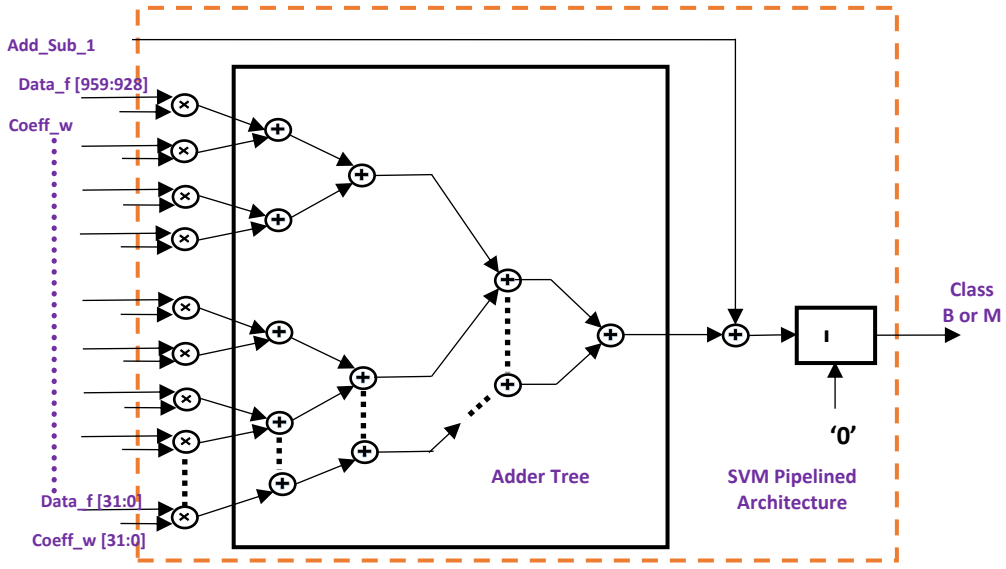


Fig. 2 SVM Linear Classifier Architecture

3.2. Design of BRAM Controller

The block RAM controller's design is essential to load breast cancer patient data and computed weight values using python script corresponding to each feature in Wisconsin Breast Cancer Dataset. The loaded data can be used further as input by designing a linear SVM classifier to create a training model to classify two data classes.

3.3. Computing for Linear SVM Classifier

32-bit handles the system's computations signed floating-point multipliers and adders with a single sign bit, a 23-bit mantissa, and an 8-bit exponent. The floating-point math packages are based on the VHDL 1164 numeric_std package and use the package's signed and unsigned arithmetic. Because the numeric_std package is well supported by simulation and synthesis tools, they are extremely efficient.

4. Experimental Results and Discussions

This study seeks to determine the most useful features in identifying malignant or normal cancer. The main objective is to discover whether breast cancer belongs to malignant or benign. Here, the standard WBCD dataset is analyzed for 569 patients with 30 best features to categorize between two data classes for the detection of Breast Cancer.

4.1. Simulation Output Waveform Analysis of Top Module of SVM Linear Classifier

The pipelined SVM architecture is designed using Verilog HDL, then synthesized by Vivado (ver. 2017.4) simulation tool. The design is implemented through translating, mapping, placing and routing, and generating a bit stream. Ultimately, it is configured to the Xilinx KC705 Kintex-7 evaluation board. Below are the simulation results for the single precision IEEE standard 754-based SVM linear classifier algorithm.

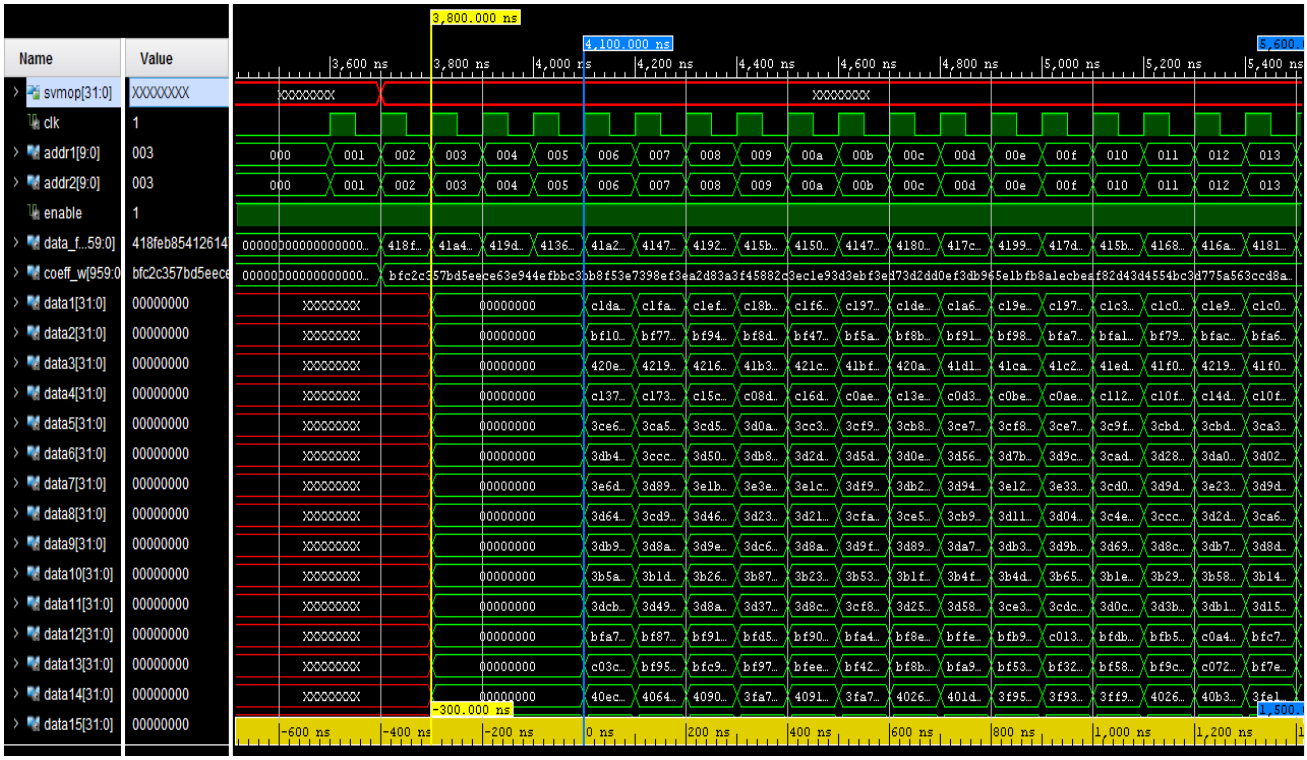


Fig. 3 The Output Waveform for Top Module of SVM Classifier: Depicts Data Loaded into BRAM at 3700 ns

The Simulation result of the SVM linear classifier depicts data loaded into BRAM for training to the positive edge of the clock, as shown in Figure 3. There are 30 features in the WBCD dataset, and each feature is represented with a 32-bit floating point number. So total bandwidth required to accommodate each patient's data is 960 bits wide. The simulation window shown in Figure 4 depicts pipelined adder tree operation with different stages of operation.

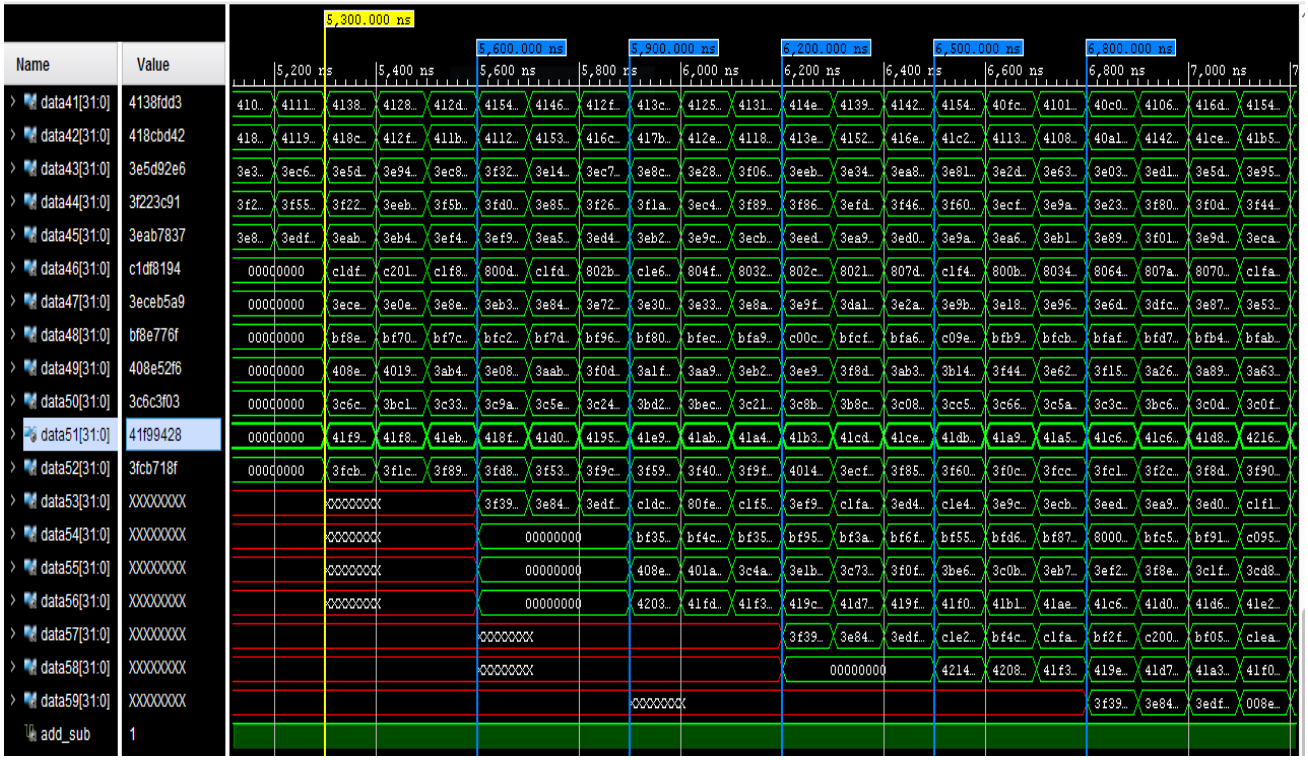


Fig. 4 Simulation Waveform of Adder Tree Pipelined Operation in SVM Linear Classifier at Different Stages

4.2. Logic Synthesis of top module SVM classifier system

After synthesis and implementation, it becomes very easy to check and analyse the different metrics such as area, power, and timing. The implementation result is more accurate in evaluating the performance analysis of the system. The schematic view of the top module SVM classifier after synthesis is shown in Fig. 5.

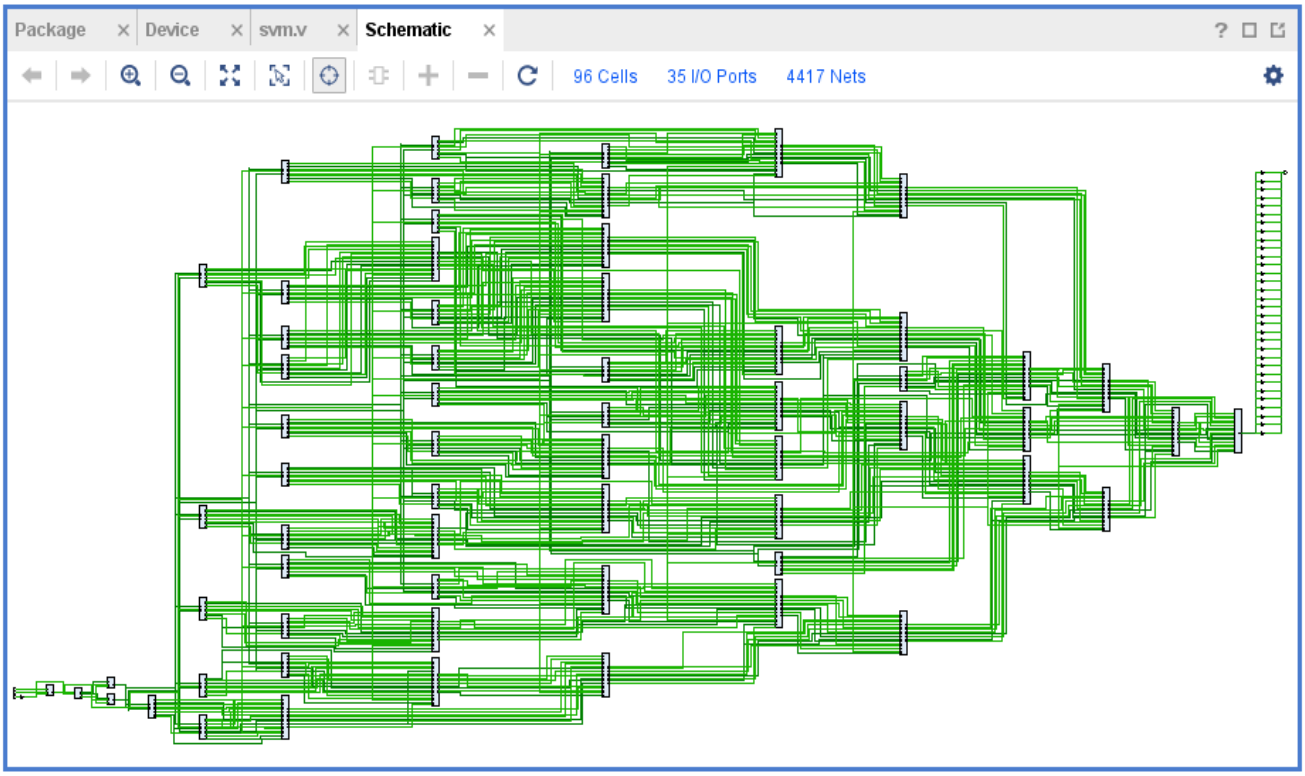


Fig. 5 Schematic View of Top Module SVM Classifier after Synthesis

4.3. Analysis of SVM output

The obtained hexadecimal number needs to be converted into a real number for comparison to find out the classified output, whether it belongs to 'Benign' or 'Malignant.' If the SVM output is positive or equal to '0', then the decision is taken w.r.t 'Malignant.' Similarly, if the SVM output is negative or less than '0', then a decision is taken w.r.t 'Benign.' It results in more inaccuracy than only positive, negative, and '0' for classification. To overcome this problem, a confusion matrix needs to be incorporated to increase the percentage accuracy by considering true positive or negative and false positive or negative. The accurate classification rate of the proposed system is 91%. The classification accuracy can further be increased by using an exact dot product accumulator with additional hardware cost.

There is a slight improvement in accuracy percentage obtained using Posit 24-bit format to IEEE 754 floating point algorithm implementation for detecting breast cancer using the WBCD dataset. The accuracy of precision for detection of breast cancer is obtained after running software code in Java for 24-bit posit format. The accuracy obtained is about 0.9122 using the Posit 24-bit format for IEEE 754 single precision 32-bit representation. The obtained result indicates that using the Posit 24-bit format, it is possible to obtain a better accuracy of IEEE 754 single precision 32-bit representation.

Table 2. Summary of POSIT arithmetic and IEEE 754 float

Format	Representation	Accuracy
IEEE 754 float	32 bits	91%
POSIT arithmetic	24 bits	91.22%

The summary of performance evaluation of POSIT arithmetic and IEEE 754 float is presented in table 2.

5. Conclusion

The proposed parallel implementation of the SVM algorithm for detecting breast cancer using the Vivado simulation tool as a training system is presented. The major objective of this solution is to reach a high data processing rate to satisfy the needs of computationally demanding applications. As a result, all possible multiplications and additions were parallelized. Finally, based on interpretations using the synthesis results, it can be concluded that implementing this technique in hardware would result in significant performance enhancements over the existing methods. The simulation and synthesis results exhibit that the SVM linear classification system is more effective in fast data classification. According to experimental results with a maximum frequency of 100 MHz in linear classification, the classification accuracy of about 91% in linear one has been achieved. The accuracy obtained is about 0.9122 using the Posit 24-bit format with respect to IEEE 754 single precision 32-bit representation. The obtained result indicates that using the Posit 24-bit format, it is possible to obtain a better accuracy of IEEE 754 single precision 32-bit representation.

The FPGA implementation of the SVM linear classifier with an exact dot product accumulator improves the percentage of accuracy related to state-of-the-art. It provides more flexibility to alter the structure in the future.

Conflicts of Interest

"Laxmisagar H.S and Hanumantharaju M.C declare that there is no conflict of interest regarding the publication of this paper."

References

[1] J. Ferlay et al., "Cancer Statistics for the Year 2020: An Overview," *Int. J. Cancer,* vol. 149, no. 4, pp. 778–789, 2021. Doi: https://doi.org/10.1002/ijc.33588.

[2] P. Mathur et al., "Cancer Statistics, 2020: Report from National Cancer Registry Programme, India," *JCO Glob. Oncol.,* no. 6, pp. 1063–1075, 2020. Doi: 10.1200/GO.20.00122.

[3] Dr.A.Nitesh and J.Deveesh, "Analysis and Optimization of a Floating Point Representation of a Complex Numbers using FPGA," *SSRG International Journal of VLSI & Signal Processing,* vol. 3, no. 2, pp. 17-21, 2016. *Crossref,* https://doi.org/10.14445/23942584/IJVSP-V3I2P104

[4] M. Papadonikolakis and C.S. Bouganis, "A Novel FPGA-Based SVM Classifier," in *International Conference on Field-Programmable Technology,* pp. 283–286, 2010. Doi: 10.1109/FPT.2010.5681485.

[5] J. Gustafson, "Posit Arithmetic," *Math. Noteb. Descr. Posit Number Syst.*, vol. 30, 2017.

[6] J. L. Gustafson and I. T. Yonemoto, "Beating Floating Point at its Own Game: Posit Arithmetic," *Supercomput. Front. Innov.,* vol. 4, no. 2, pp. 71–86, 2017.

[7] Chiranjeevi G N, Dr Subhash Kulkarni, "Enhanced MAC Controller Design for 2D Convolution Image Processing on FPGA," *SSRG International Journal of Engineering Trends and Technology,* vol. 69, no. 9, pp. 51-55, 2021. *Crossref,* https://doi.org/10.14445/22315381/IJETT-V69I9P207

[8] M. Cococcioni, F. Rossi, E. Ruffaldi, and S. Saponara, "Fast Approximations of Activation Functions in Deep Neural Networks when using Posit Arithmetic," *Sensors,* vol. 20, no. 5, 2020. Doi: 10.3390/s20051515.

[9] L. Van Dam, J. Peltenburg, Z. Al-Ars, and H. P. Hofstee, "An Accelerator for Posit Arithmetic Targeting Posit Level 1 Blas Routines and Pair-Hmm," in *Proceedings of the Conference for Next Generation Arithmetic,* vol. 2019, pp. 1–10, 2019.

[10] D. Selvathi and R. D. Nayagam, "FPGA Implementation of on-Chip ANN for Breast Cancer Diagnosis," *Intell. Decis. Technol.* vol. 10, no. 4, pp. 341–352, 2016.

[11] A. Baez et al., "High-Level Synthesis of Multiclass SVM using Code Refactoring to Classify Brain Cancer from Hyperspectral Images," *Electronics,* vol. 8, no. 12, 2019. Doi: 10.3390/electronics8121494.

[12] M. Mohammadi, A. Krishna, N. S., and S. K. Nandy, "A Hardware Architecture for Radial Basis Function Neural Network Classifier," *IEEE Trans. Parallel Distrib. Syst.,* vol. 29, no. 3, pp. 481–495, 2018. Doi: 10.1109/TPDS.2017.2768366.

[13] Ramya R, Madhura R, "FPGA Implementation of Optimized BIST Architecture for Testing of Logic Circuits," *SSRG International Journal of VLSI & Signal Processing,* vol. 7, no. 2, pp. 36-42, 2020. *Crossref,* https://doi.org/10.14445/23942584/IJVSP-V7I2P106

[14] M. Bassoli, V. Bianchi, and I. De Munari, "A Simulink Model-Based Design of a Floating-Point Pipelined Accumulator with HDL Coder Compatibility for FPGA Implementation," in *Applications in Electronics Pervading Industry, Environment and Society,* pp. 163–171, 2020.

[15] Louca, Cook, and Johnson, "Implementation of IEEE Single Precision Floating Point Addition and Multiplication on FPGAs," in *Proceedings IEEE Symposium on FPGAs for Custom Computing Machines,* pp. 107–116, 1996. Doi: 10.1109/FPGA.1996.564761.

[16] S. T. Chandrasekaran, R. Hua, I. Banerjee, and A. Sanyal, "A Fully-Integrated Analog Machine Learning Classifier for Breast Cancer Classification," *Electronics,* vol. 9, no. 3, 2020. Doi: 10.3390/electronics9030515.

[17] D. Selvathi and A. Aarthy Poornila, "Deep Learning Techniques for Breast Cancer Detection Using Medical Image Analysis," in *Biologically Rationalized Computing Techniques for Image Processing Applications*, J. Hemanth and B. V. Emilia, Eds. Cham: Springer International Publishing, pp. 159–186, 2018. Doi: 10.1007/978-3-319-61316-1_8.

[18] G. C. Batista, D. L. Oliveira, O. Saotome, and W. L. S. Silva, "A Low-Power Asynchronous Hardware Implementation of a Novel SVM Classifier, with an Application in a Speech Recognition System," *Microelectronics J.,* vol. 105, p. 104907, 2020. Doi: https://doi.org/10.1016/j.mejo.2020.104907.

[19] G. González-Díaz_Conti et al., "Hardware-Based Activation Function-Core for Neural Network Implementations," *Electronics,* vol. 11, no. 1, 2022. Doi: 10.3390/electronics11010014.

[20] H. M. Hussain, K. Benkrid, and H. Seker, "The Role of FPGAs as High Performance Computing Solution to Bioinformatics and Computational Biology Data," *AIHLS*, vol. 102, 2013.

[21] A. Eklund, P. Dufort, D. Forsberg, and S. M. LaConte, "Medical Image Processing on the GPU – Past, Present and Future," *Med. Image Anal.,* vol. 17, no. 8, pp. 1073–1094, 2013. Doi: https://doi.org/10.1016/j.media.2013.05.008.

[22] A. HajiRassouliha, A. J. Taberner, M. P. Nash, and P. M. F. Nielsen, "Suitability of Recent Hardware Accelerators DSPS, FPGAS, and GPUS for Computer Vision and Image Processing Algorithms," *Signal Process. Image Commun.,* vol. 68, pp. 101–119, 2018. Doi: https://doi.org/10.1016/j.image.2018.07.007.

[23] S. Asano, T. Maruyama, and Y. Yamaguchi, "Performance comparison of FPGA, GPU and CPU in Image Processing," in *International Conference on Field Programmable Logic and Applications*, pp. 126–131, 2009. Doi: 10.1109/FPL.2009.5272532.

[24] Mehboob Hasan Ahmed, Rutuja Jagtap, Roopal Pantode and Prof. S. S. Phule, "An FPGA Chip Identification Generator using Configurable Ring Oscillator," *SSRG International Journal of Electronics and Communication Engineering,* vol. 3, no. 4, pp. 10-14, 2016. *Crossref,* https://doi.org/10.14445/23488549/IJECE-V3I4P103

[25] S. Afifi, H. GholamHosseini, and R. Sinha, "A System on Chip for Melanoma Detection using FPGA-Based SVM Classifier," *Microprocess. Microsyst.,* vol. 65, pp. 57–68, 2019. Doi: https://doi.org/10.1016/j.micpro.2018.12.005.

[26] L. Fiolhais and H. Neto, "An Efficient Exact Fused Dot Product Processor in FPGA," in *28th International Conference on Field Programmable Logic and Applications FPL,* pp. 327–3273, 2018. Doi: 10.1109/FPL.2018.00062.

[27] E.Ganesan and V.Sakthivel, "A Novel FPGA Design with Hybrid LUT / MUX Architecture," *SSRG International Journal of Electronics and Communication Engineering,* vol. 3, no. 11, pp. 6-8, 2016. *Crossref,* https://doi.org/10.14445/23488549/IJECE-V3I11P112

[28] S. Kim and R. A. Rutenbar, "An Area-Efficient Iterative Single-Precision Floating-Point Multiplier Architecture for FPGA," in *Proceedings of the on Great Lakes Symposium on VLSI,* pp. 87–92, 2019. Doi: 10.1145/3299874.3318002.

[29] W. Chen, H. R. Pourghasemi, A. Kornejady, and N. Zhang, "Landslide Spatial Modeling: Introducing New Ensembles of ANN, Maxent, and SVM Machine Learning Techniques," *Geoderma,* vol. 305, pp. 314–327, 2017. Doi: https://doi.org/10.1016/j.geoderma.2017.06.020.

Catalysis Communications 162 (2022) 106381

Contents lists available at ScienceDirect

Catalysis Communications

journal homepage: www.elsevier.com/locate/catcom

Short communication

Synthesis of highly sensitive nanomaterial for ultra-fast photocatalytic activity: A detailed study on photocatalytic capabilities of rod-shaped TiS_3 nanostructures

Jaspal Singh [a], Prachi Sharma [b,e], Nishant Tripathi [b,*], Daria Shishkina [b], Anastasiia Rymzhina [b], Elisey A. Boltov [b], Vladimir Platonov [b], Vladimir Pavelyev [b], Valentyn S. Volkov [c], Aleksey V. Arsenin [c], Rishikesh Singh [a], R.K. Soni [a], Mohammad Talib [d], Samrah Manzoor [d], Debosmita Banerjee [f], P.M.Z. Hasan [g], Ahmed Alshahrie [g], Reem Darwesh [h], Maria A. Anikina [i], Prabhash Mishra [b,c,d,*]

[a] Laser spectroscopy Lab, Department of Physics, Indian Institute of Technology Delhi, New Delhi 110016, India
[b] Samara National Research University, 34, Moskovskoye Shosse, Samara 443086, Russia
[c] Center for Photonics & 2D Materials, Moscow Institute of Physics and Technology (MIPT), Dolgoprudny 141700, Russia
[d] Centre for Nanoscience and Nanotechnology, Jamia Millia Islamia (A Central University), New Delhi 110025, India
[e] Electronics and Communication Department, BMS Institute of Technology and Management, Bangalore 560064, India
[f] Department of Physics, School of Natural Sciences, Shiv Nadar University, NH-91, Gautam Buddha Nagar, Uttar Pradesh 201314, India
[g] Center of Nanotechnology, King Abdulaziz University, Jeddah, Saudi Arabia
[h] Department of Physics, King Abdulaziz University, Saudi Arabia
[i] Scientific Research Institute of the Department of Gas Processing, Hydrogen and Special Technologies, Samara State Technical University, Russia

ARTICLE INFO

Keywords:
TiS_3
Photocatalysis
Nanoribbons
Industrial pollutants
Near infrared

ABSTRACT

Two-dimensional TiS_3-nanostructures have been used to successfully remove toxic-industrial pollutants such as methyl-orange(MO) and methyl-blue(MB) from water bodies. Two-dimensional TiS_3-nanostructures with exceptional mobility properties have been produced using the chemical vapor transport(CVT) technique and its photocatalytic capabilities are examined for potential environmental applications.

Under infrared(IR) irradiation, the produced TiS_3-nanoribbons were evaluated for their ability to degrade MO and MB solutions. TiS_3-nanoribbons reveal significant high disintegration percentage; it is capable to decompose 90.3% of 50 μM MO pollutant solution in 5 min. Moreover it decomposes 94.4% of 50 μM MO solution as well as 10 μM MB solution in 20 min and 12 min, respectively.

1. Introduction

The continual desire of humans to improve their living standards has resulted in some of the most remarkable discoveries and inventions. Though these improvements appear to be very advantageous, they have come at a high cost in terms of environmental deterioration [1]. Our environment has been fatally polluted, and we have reached a critical juncture that requires immediate action to safeguard it. The two primary components of nature that have suffered the brunt of these industrial and technological advances have been air and water [2]. Water contamination has been continuously increasing due to effluent discharges from industrial waste. They are often released into bodies of water without being appropriately treated, posing a threat to humans and aquatic life [3]. Consequently, treatment of these effluents is urgently required.

Researchers have investigated many techniques of wastewater treatment, with photocatalysis being one of the most significant among them [4]. Photocatalysis is premised on the theory that when pollutants are exposed to sunshine, they are more easily degraded by catalysts. The choice of photocatalyst depends on the rate at which pollutants are degraded by the photocatalyst. A photocatalyst with a greater rate of pollutant degradation is most desirable. The photocatalyst must be very sensitive to specific exposure of light to ensure its maximum absorption [5]. Zinc oxide and titanium oxide have been the most extensively

* Corresponding authors.
E-mail addresses: nishant.tripathi.11@gmail.com, tripati.n@ssau.ru (N. Tripathi), pmishra@jmi.ac.in (P. Mishra).

https://doi.org/10.1016/j.catcom.2021.106381
Received 16 October 2021; Received in revised form 20 November 2021; Accepted 3 December 2021
Available online 6 December 2021

investigated and utilized photocatalysts, but their absorption is significantly restricted in the ultra-violet range [6]. The UV region makes up only 4% of the solar spectrum. Thus, the researchers are focusing on materials that are sensitive to visible and infrared wavelengths, which comprise most of the solar spectrum [7]. Due to their advantages over its bulk equivalents, two-dimensional (2D) materials have been at the forefront of research in recent years [8–12].

Graphene was the first 2D material to be studied, and it exhibited amazing electrical and optical characteristics. However, its zero bandgap was a major drawback [13]. Due to graphene's limitations, the research shifted to two-dimensional layered materials. The distinctive characteristic of 2D layered materials is that their electrical and optical properties can be easily adjusted by changing the number of layers [14]. Numerous techniques, including mechanical and chemical exfoliation, can be used to change the number of layers. The Transition Metal Dichalcogenides (TMDs) are extensively investigated layered materials that have emerged as a potential alternative to graphene due to their unique characteristic of adjustable bandgap structure [15]. Their bandgap shifts from indirect in bulk to direct in monolayers, allowing them to be used for a wide range of electrical, optical, and catalytic applications [16]. The direct bandgap is achieved when the TMDs are in their monolayer state. But the probability of producing monolayers is minimal, and this is a disadvantage of TMD materials [16,17].

This disadvantage is mitigated when we consider another member of chalcogenide family known as transition metal trichalcogenides (TMTC) [18]. The standard formula for transition metal trichalcogenides (TMTCs) is MX_3, where M stands for Ti, Zr, Hf, and other transition metals, and X stands for chalcogens such as S, Se, and Te. TMTCs can be found in both layered and bulk forms [18]. The interlayer bonds are covalent and are consequently firmly bounded, whereas the intralayer bonds are linked by the van der Waal forces and hence extremely weak making it easier to achieve layered structure from bulk [19]. The distinctive advantage of TMTCs over TMDs is that direct bandgap structures are achievable in monolayers and up to a few layers in TMTCs. On the other hand, a direct bandgap is only conceivable in the monolayer structure of TMDs [20]. Because TMTCs have greater electron mobility than TMDs, they are preferable for a variety of electrical applications [20– 21].

This paper provides an interesting investigation into the photocatalytic performance of TiS_3, the newest member of the TMTC family. TiS_3 monolayers are composed of a 1D quasi rigid chain. Each of these chains is made up of constituent atoms arranged in a trigonal pattern. The high anisotropy in the electronic and thermal properties of the TiS_3 has sparked a new wave of study into the material, as these qualities promise an improvement in the performance of various electronic and optoelectronic devices. TiS_3 can be synthesized using various methods such as mechanical and chemical exfoliation, chemical vapor technique, vapor-liquid-solid (VLS), and chemical vapor transport (CVT). In this work, we employed a physical method to manufacture TiS_3 and tested its dye degrading capabilities under NIR light. Additionally, this research illustrates the appreciable photodegradation performance of one-dimensional TiS_3 nanoribbons, owing to their high mobility and optical absorption.

2. Experimental

2.1. Materials

For the formation of TiS_3 nanoribbons, Titanium powder and sulfur powder were procured from Good fellow and Fisher scientific respectively. Methyl orange and Methylene blue pollutants molecules were procured from SRL, India. These chemicals were employed as purchased.

2.2. Fabrication of TiS_3 nanoribbons

Nanoribbons of TiS_3 were fabricated through the unique chemical vapor transport process. In this method, firstly the sulfur and titanium powder were taken in a ratio of 3:1 respectively then the mixture was thoroughly grinded in a mortar- pestle to obtain ultrafine nanoparticles. In the next step, the obtained mixed powder was placed in a silica ampoule tube with a diameter of 12 mm. After that silica ampoule tube was sealed in a vacuum (pressure $\sim 2 \times 10^{-5}$ mbar). Finally, the vacuum-sealed silica ampoule was put in a furnace at 500 °C temperature for 24 h and, finally TiS_3 is obtained in nanoribbons like nanostructures [22].

2.3. Characterization of TiS_3 nanoribbons

The nanostructures' shape, crystallinity and size distribution were determined using a FESEM (ZESIS, Sigma 5.05) equipped with chemical mapping capability and TEM (FEI, Eindhoven, Netherland) microscopy. The optical behavior of the fabricated nanoribbons sample was investigated by UV–Visible absorption spectroscopy (Shimadzu Corp., Japan). Raman spectroscopy (Horiba LabRAM HR 800, $\lambda = 532$ nm) and X-ray diffraction (Smart lab Rigaku, source Cu Kα, $\lambda = 1.54$ A) were used to investigate the crystal structures, vibrations, and composition of the prepared TiS_3 nanoribbons sample. To investigate the surface chemical composition of the prepared sample, X-ray photoelectron spectroscopy study (Multiprobe surface analysis system, Omicron) was performed.

2.4. Photodegradation study of TiS_3 nanoribbons

The photodegradation capability of the fabricated TiS_3 nanoribbons was investigated by the decomposition of methyl orange (MO) and methylene blue (MB) solution under IR light illumination. In photodegradation studies, 1 mg/10 mL of TiS_3 nanoribbons was dispersed into a 20 mL solution of MO while the effective concentration of MO is 50 μM. After that, the reaction vessel containing MO solution and nanoribbons sample is exposed with an IR lamp (1050 W/m^2). The distance between the glass vessel and the lamp was 25 cm. The reaction suspension was placed under lamp exposure for 5, 10, 15, and 20 min. The variations in the concentration of MO molecule solution were observed through UV–visible absorption spectroscopy. The capability of TiS_3 nanoribbons sample was computed through the following relation:

$$\eta = \frac{C_o - C_t}{C_o} \text{ X } 100 \quad (1)$$

Where C_o is the starting concentration of the MO solution while Ct is the concentration of MO solution at a particular time 't'. Similar process was repeated for performing the degradation of the MB solution. In order to explore the particular radicals which were majorly responsible for the photocatalytic activity of TiS_3 nanostructure, charge trapping studies were performed. To capture the electrons, holes, hydroxyl radicals and superoxide radicals in the photodegradation process copper nitrate (CN), formic acid (FA), isopropyl alcohol(IPA) and benzoquinone (BQ) were used respectively.

3. Results and discussion

The FESEM imaging was used to investigate the surface morphology of a nanostructured material, as seen in Fig. 1. (a). FESEM analysis revealed that two-dimensional nanoribbon-like structures existed uniformly with a high density, with an average length and breadth of 5436 nm and 447 nm for ribbon-like nanostructures, respectively. TEM examination of produced samples confirmed the development of nanoribbon-like nanostructures. To verify the homogeneity of the Ti and S atoms in the synthesized sample, elemental mapping was performed, as seen in inset of Fig. 1 (a). Elemental pictures of Ti and S atoms provide insight into their spatial distribution in the TiS_3 sample. Additionally,

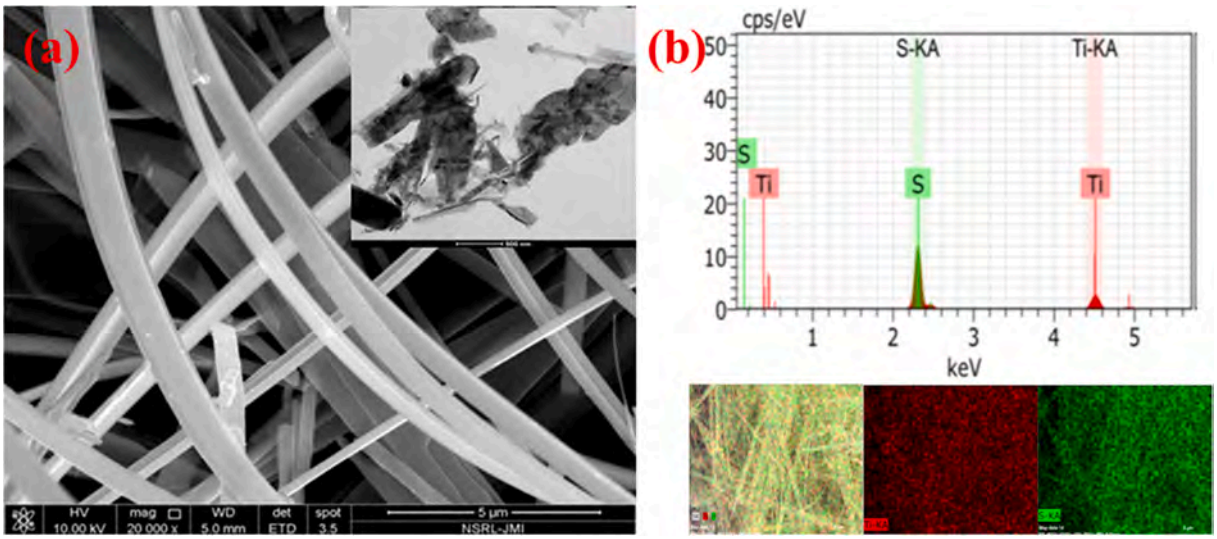


Fig. 1. (a) Shows the FE-SEM image of TiS_3 nanoribbons; inset image shows the respective TEM monograph, (b) EDX spectrum of TiS_3 nanoribbons showing the presence of Ti and S atoms. Elemental mapping assures the uniform distribution of Ti and S atoms in TiS_3 nanoribbons.

mapping for the composite (Fig. 1(b)) confirmed the presence of a Ti and S-based composite. The EDX spectrum of the produced sample shows clear peaks for Ti and S elements, indicating that no elemental impurity occurs in it. EDX studies highlights that the atomic percentage of Ti and S atoms are 12% and 35% respectively (see Table 1, Supplementary Information (SI)).

The crystallographic characteristic of the synthesized sample was investigated using the XRD technique and is depicted in Fig. 2. (a). The observed X-ray diffraction patterns (002, 003, 201, 210, 006, 202) were almost identical to the TiS_3 JCPDS card number (15–0783). These diffraction patterns show that TiS_3 exists in a monoclinic phase with a high degree of crystallinity. The small diffraction patterns among (2θ = 20-30^o) may be caused by the presence of sulfur in free form due to the use of sulfur powder as a precursor in the experiment. Thus, XRD investigations established that TiS_3 exists in a monoclinic phase.

Raman spectroscopy was also used to verify the production of TiS_3 and its phase confirmation by observing the fingerprint vibrations of TiS_3. Fig. 2 (b) shows the Raman profile of the prepared samples. (b). The Raman spectra of TiS_3 nanoribbon displays four distinct peaks at 177 cm^{-1}, 301 cm^{-1}, 371 cm^{-1}, and 558 cm^{-1}, confirming Ag Raman mode of TiS_3 [18]. A prominent peak at 177 cm^{-1}, is attributed to I-A_g^{rigid} mode, which is formed by the out-of-plane rigid vibration of TiS_3 chains extending along the b-axis [18,23]. The peak at 177 cm^{-1} represents the unique vibration associated with the production of TiS_3. The peaks at 301 cm^{-1}, 371 cm^{-1} can be marked as II-$A_g^{internal}$ and, III-$A_g^{internal}$ respectively. These two peaks appear due to the internal vibration of the single TiS_3 layer [24]. The peak at 558 cm^{-1} can be ascribed to the IV-$A_g^{S–S}$. This peak originated because of the in-plane out-of-phase vibration of S—S. A peak at 558 cm^{-1} provides the vibration information regarding out-of-plane respective motion among Ti atoms with bridge S atoms [25,26]. Optical absorption spectroscopy with a wavelength range of 400–1200 nm was used to examine the optical features of the produced TiS_3 nanoribbon. The absorption spectra of TiS_3 nanoribbons in the visible-NIR range is shown in Fig. 2 (c). The optical absorption spectrum exhibits two broad peaks in the visible and NIR regions, indicating that the prepared sample has exceptional light absorption capabilities in both the visible and NIR regions. Fig. 2(d) shows the direct bandgap value of TiS_3 nanoribbons as calculated by the Tauc plot method. The TiS_3 nanoribbons sample has a calculated bandgap of 1.18 eV. According to Tauc plot studies, the photodecomposition performance of TiS_3 nanoribbons is predicted to increase in the NIR light exposure due to the narrow bandgap. X-ray photoelectron spectroscopy (XPS) was carried out to understand the chemical bonding characteristics of the material. Fig. 2(e) represents the Gaussian fitted Ti 2p spectrum which consists of four prominent peaks arising due to the spin-orbit doublet splitting. Among them, the first doublet observed on the lower binding energy (B.E.) side at 456.2 eV and 462.1 eV corresponds to Ti $2p_{3/2}$ and Ti $2p_{1/2}$ of TiS_3 [27]. Whereas the other pair of peaks obtained at 459.2 eV and 464.9 eV are assigned to Ti $2p_{3/2}$ and Ti $2p_{1/2}$ doublets of TiO_2 [28]. All the peaks are in a stoichiometric Ti^{4+} state either in TiS_3 or TiO_2. This reveals the presence of a significant quantity of oxygenated Ti, as well as Ti—S bonding, which indicates the presence of sulfur vacancies [18]. Due to the sulfur vacancies, oxygen atoms can interact with Ti^{+4} and produce TiO_2 during the production of TiS_3 [18]. The fitted S2p spectrum is shown in Fig. 2 (f). Three significant core level peaks at 161.3 eV, 162.5 eV, and 163.6 eV are suggested by the contributions from two sulfur species known as sulfide (S^{2-}) and disulfide (S_2^{2-}) [27,29]. The first peak at lower B.E. is assigned to S $2p_{3/2}$ of sulfide unit (S), the second peak is an overlap of S $2p_{1/2}$ of sulfide (S^{2-}) and S $2p_{3/2}$ of disulfide (S_2^{2-}) unit, and the last one comes from S $2p_{1/2}$ of disulfide (S_2^{2-}) [30]. The corresponding peaks from two species are separated by 1.2 eV and for both cases, S($2p_{3/2}$):S ($2p_{1/2}$) is 2:1. A small satellite peak is also observed at 164.3 eV which could not be identified. This may be attributed to the S—O bond coming from the atmospheric oxidation of the surface. In order to check the specific surface area of the prepared TiS_3 nanostructures, BET measurements were performed and illustrated in Fig. 1(SI). The BET studies reveals that the specific surface of TiS_3 nanostructures is 13.93 mL/g.

The photodegradation efficiency of TiS_3 nanoribbons under NIR light is demonstrated by the breakdown of a 50 μM of MO and 10 μM MB contaminants solution. Fig. 3(a) depicts the breakdown of MB pollutant solution over time using a TiS_3 photocatalyst sample. Optical absorption tests indicated that the MB pollutant solution decomposed after 12 min. Fig. 3(b) shows that a TiS_3 nanoribbons sample decomposes 84.2% of a 10 μM MB pollutan solution in a short duration of 12 min.

The calculated rate constant for the decomposition rate kinetics of MB pollutant solution is 0.1468/min (Fig. 3(c)). Fig. 3(d) shows the UV–Vis analysis of a degraded MO pollutant solution using TiS_3 nanoribbons as a photocatalyst under NIR light exposure. The use of TiS_3 nanoribbons as photocatalysts resulted in a substantial decrease in the intensity of main peaks of MO pollutant solution.

Fig. 3(e) shows the rate of photodecomposition of the MO pollutant solution with exposure time. According to photodegradation rate kinetics, 94.4% of the MO pollutant solution is decomposed in 20 min under NIR light. The calculated rate constant value for the photodecomposition process in MO pollutant solution is 0.1207/min.

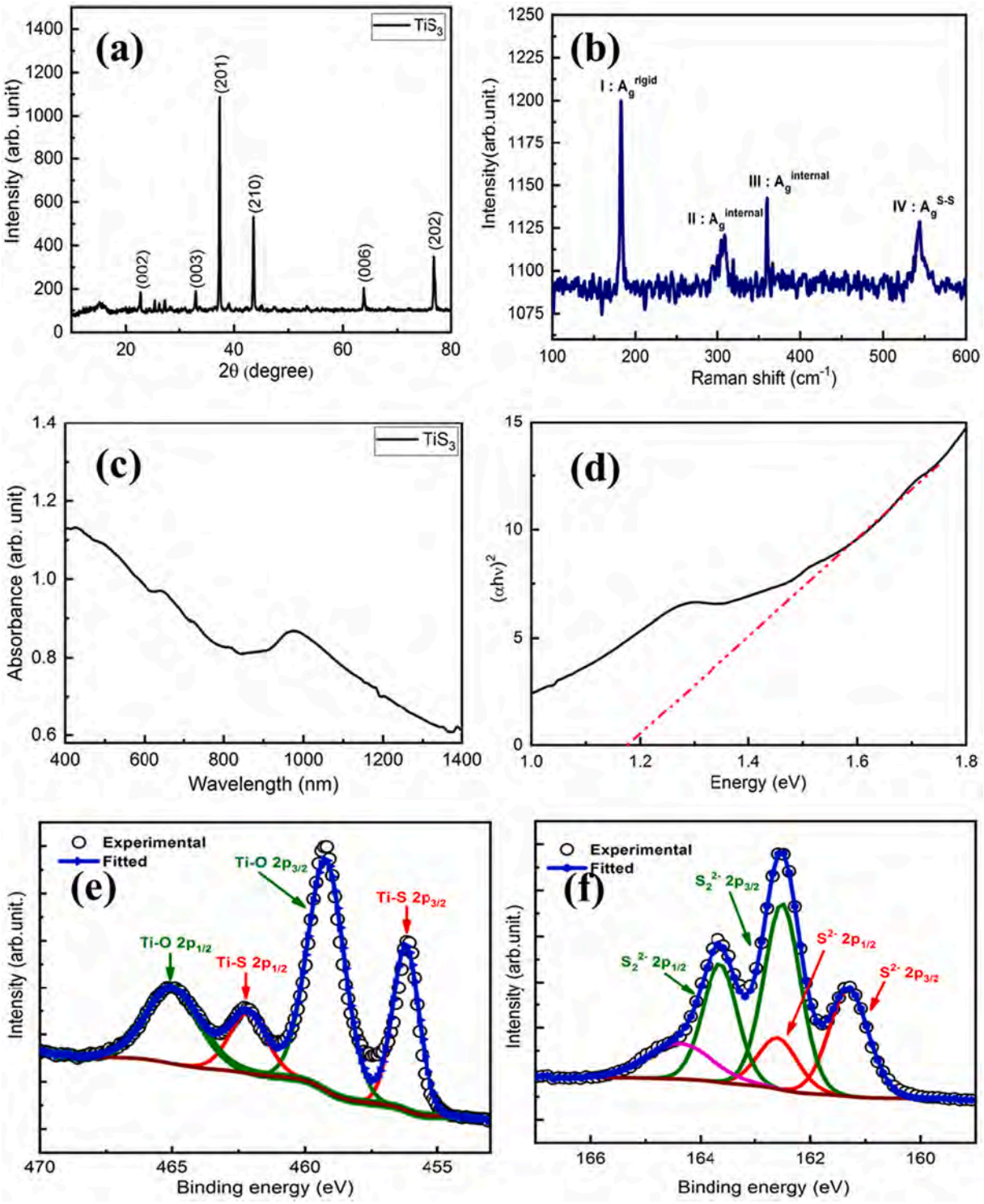


Fig. 2. Shows the various spectra for basic characteristics of TiS_3. (a) X-ray diffraction pattern of TiS_3 nanoribbons, (b) Raman spectrum of TiS_3 nanoribbons sample, (c) Optical absorption spectrum of TiS_3 nanoribbons, (d) Tauc plot for TiS_3 nanoribbons sample, (e) Gaussian fitted XPS spectrum of Ti 2p for TiS_3 nanoribbons, (f) Fitted S2p XPS spectrum of TiS_3 nanoribbons sample.

To ensure the stability of the TiS_3 nanoribbons, they were tested in reusable photocatalytic runs that are shown in Fig. 3(f). Photocatalytic reusability tests indicate that a photocatalyst based on TiS_3 nanostructures displays consistent photodegradation performance for MO pollutant solution after four cycles. The photodegradation of TiS_3 nanoribbons can be elucidated with the help of chemical reactions as below. Initially, pollutant molecules were adsorbed on the surface of TiS_3 nanoribbons (Eq. (2)). Under NIR exposure light, TiS_3 nanoribbons generated electrons and holes in the conduction band (CB) and valence band (VB) sequentially (Eq. (3)). After that, the photo-generated electrons in the CB of TiS_3 nanoribbons produce the superoxide radicals ($\bullet O_2$) from oxygen molecules (Eq. (4)). Similarly, holes present in the VB transform the H_2O molecule in to the hydroxyl radicals ($\bullet OH$) (Eqs. (5), (6)). These highly reacted radicals further react with the organic pollutants molecules and decompose the in to carbon dioxide and water molecules (Eqs. (7), (8)). The scavenger studies concluded that hydroxyl

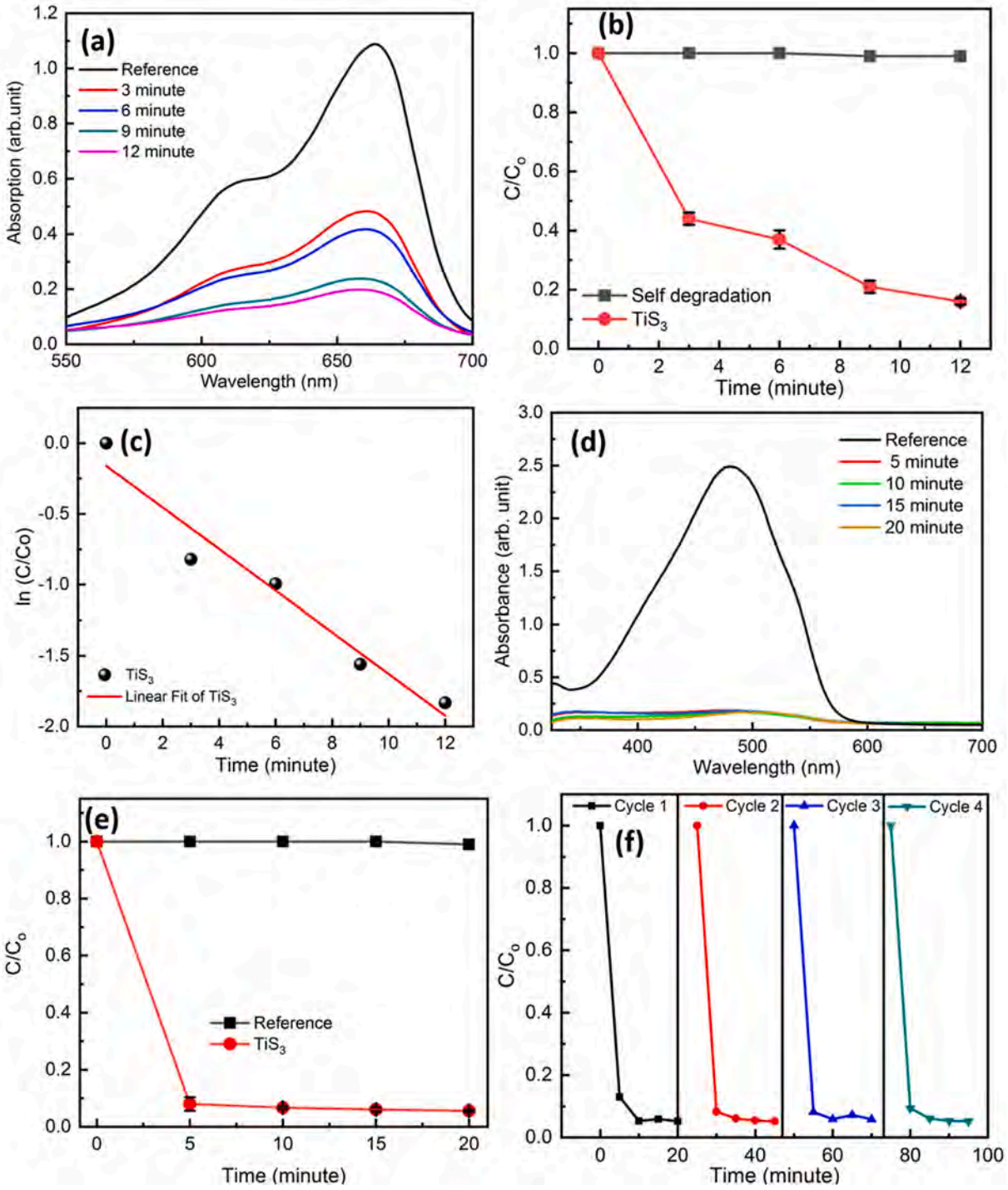


Fig. 3. (a) Optical absorption curves depicting the decomposition of MB pollutant solution using TiS_3 photocatalyst, (b) Rate kinetics for MB molecules decomposition using TiS_3 nanoribbons, (c) ln(C/Co) vs exposure time graph indicating the rate constant value (k), (d) UV–visible absorption spectrum revealing the reduction in the concentration of MO molecule solution through TiS_3 nanoribbons, (e) Photodecomposition rate kinetics of MO molecules solution using TiS_3 nanoribbons sample, (f) Reusability test indicates the stability of the TiS_3 nanoribbons.

and holes are majorly contributed in to the photocatalysis reaction (see Fig. 3 SI).

$$Pollutant\ molecule + TiS_3 \rightarrow Adsorbed\ pollutant\ molecule\ on\ TiS_3 \quad (2)$$

$$(NIR)\ hv + TiS_3 \rightarrow e^-_{CB} + h^+_{VB} \quad (3)$$

$$e^-_{CB} + (O_2)_{ad} \rightarrow \bullet O_2^- \quad (4)$$

$$h^+_{VB} + (H_2O)_{ad} \rightarrow H^+ + \bullet OH \quad (5)$$

$$h^+_{VB} + OH^- \rightarrow \bullet OH \quad (6)$$

$$\bullet OH + \text{Adsorbed molecule} \rightarrow CO_2 + H_2O \quad (7)$$

$$\bullet O_2^- + Adsorbed\ molecle \rightarrow CO_2 + H_2O \quad (8)$$

The schematic diagram explained the NIR-induced

photodegradation capabilities of TiS_3 nanoribbons sample (Fig. 4). Under NIR light, electrons are stimulated and transferred to the conduction band of narrow bandgap semiconductor TiS_3 nanoribbons, whereas holes develop in the valence band. Because TiS_3 has a modest bandgap value, NIR light is strong enough to excite electrons from VB to CB.

A high density of electrons in the CB interacts with the surface oxygen molecules and produces superoxide radicals, whereas holes in the VB convert water molecules into hydroxyl radicals, resulting in the breakdown of MO and MB pollution molecules. The decomposition of MB pollutant molecules because of the interplay among with the highly unsaturated hydroxyl radicals can be briefly elucidated as below [31]. Firstly hydroxyl radical attack on the MB molecule consequently cleavage formed among C–S + =C bond and C-S(=O)-C. Thus central aromatic ring is open between S and N. atoms In the next step, hydroxyl radicals react with the sulfoxide group two times and formed sulfone and sulfonic acid sequentially. Due to the continuous attacking of hydroxyl radicals give rise to the formation of SO_4^{-2} ions. After that the cleavage of N=C bond arises along with the cleavage of $C–S^+ = C$. After that photoexcited electrons creates the ·H radicals which starts to saturation of the amino bonds. This amino group creates the phenols and originates an $\cdot NH_2$ radical. After that ammonia and ammonium ions are formed due to the substitution reaction with hydroxyl radicals. In the next step, hydroxyl radicals interact with the one methyl group react and two symmetrical dimethylphenyl-amino groups and formed an alcohol and then aldehyde and afterward CO_2. The phenyl-methyl-amine radical decompose because of the sequential interaction with hydroxyl radicals and providing the whole decomposition of MB molecule.

To explore the active species responsible for the enhanced photocatalytic activity of TiS_3, charge trapping experiment was performed and presented in the Fig. 4(SI). Charge trapping experiment indicates that hydroxyl radicals are majorly active and contributed in to the photodegradation reaction. Apart from this holes and electrons also have higher contribution as compared to superoxide radicals.

Furthermore, sulfur defect states help by enabling efficient charge separation via electron trapping. These defects were activated by adsorption under NIR light, which was also useful in controlling the recombination rate. As a result, TiS_3 nanoribbons successfully breakdown the MO and MB pollutants solution and have been demonstrated to be an ideal option for environmental remediation applications. Due to their one-dimensionality, TiS_3 nanoribbons may efficiently transport photoinduced charge carriers. Aside from enhanced optical absorption, TiS_3 nanoribbons can also help photodecomposition. One of the most recent developments in photocatalysis is the fabrication of photocatalyst 2D nanostructures that work excellently in the near-infrared range.

Chen et al. [31] described the production of In_2S_3 nanostructures through a surfactant-assisted hydrothermal technique and their use in NIR-induced photocatalytic activity. In their photodecomposition reaction, they utilized 20 mg of the photocatalyst to degrade a 20 mg/L MO pollutant solution over the course of 180 min of NIR illumination. Okeil et al. [30] synthesized titanium chalcogenide (TiS, TiS_2, TiS_3, $TiSe_2$, and TiTe) based nanostructures by chemical vapor transport method and investigated their photothermal catalytic activity. They concluded that the presence of Ti^{+3} and sulfur vacancies in non-stoichiometric TiS, TiS_3 implies increased photocatalytic efficiency towards the methyl orange molecule. In their work, they used 1 mg/mL concentration of photocatalyst to degrade a 0.02 mg/mL solution of MO in 180 min of simulated sunshine.

In our work, TiS_3 nanoribbons are used to degrade a 50 μM solution of MO and 10 μM solution of MB pollutants in 20 min and 12 min, respectively, under NIR irradiation. These findings are both fascinating and important in the realm of advanced photocatalysis. Table 1 shows a comparison of photocatalytic capabilities of TiS_3 nanoribbons with latest reported data. It has been observed that TiS_3 nanoribbons shows better performance in all aspects of photocatalytic reaction as compared to other reported work. Moreover, TiS_3 nanoribbon based photocatalytic studies have been rarely reported in the literature [30]. This research paves the way for TMTC-based nanostructures to be used in advanced environmental applications.

4. Conclusions

Excellent photocatalytic performance of TiS_3 nanoribbons is presented in this paper. Novel chemical vapor transport method was used at well optimized parameters to fabricate TiS_3 nanoribbons in a monoclinic phase with a high degree of crystallinity and purity. State of the art characterization techniques have been used to validate the structural, morphological and chemical characteristics of the synthesized TiS_3 samples. Tauc plot method confirmed a calculated bandgap of 1.18 eV for synthesized material. Photocatalytic property of the prepared material was investigated by the photodegradation of water pollutants. It is observed that TiS_3 nanoribbons turn out to be highly efficient photocatalyst. It decomposes the 50 μM solution of MO and 10 μM of MB pollutants solution under NIR exposure in 20 min and 12 min respectively. The narrow bandgap of TiS_3 nanoribbons-like structures supposedly facilitates effective production of electron-hole pairs under near-infrared region. The outstanding photodegradation nature could be ascribed to the superior optical absorption in the visible and NIR region. For the time being, the findings seem to be novel of their type and it may offer fresh insights into the development of NIR-based TMTC

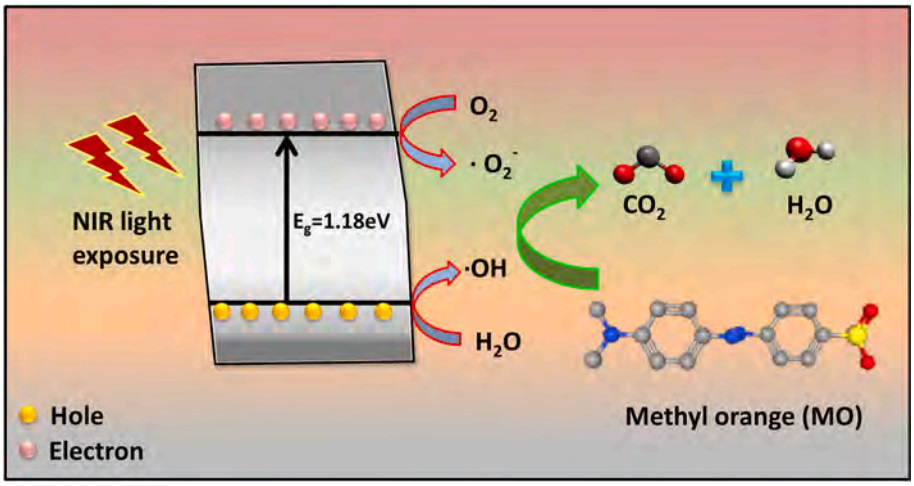


Fig. 4. Schematic representation of the improved sunlight-driven photodecomposition behavior of the TiS_3 nanoribbons sample.

Table 1
Comparison of as-developed photocatalysis materials to that of reported ones.

S. No.	Photocatalyst	Pollutant	Amount of Pollutant	Catalyst Amount	Degradation Time	Light source	Decolorization efficiency %	Decolorization rate per 1 mg of catalyst (mg h^{-1})	Ref.
1	Graphene oxide-TiO_2 composites	MO	12 mg/L, 50 mL (0.6 mg MO)	50 mg	180 min	Visible light	36%	1.44×10^{-3}	[32]
2	TiS_3 powder	MO	20 mg L − 1, 50 mL (1 mg MO)	1 mg	180 min	Sunlight	34%	0.11	[30]
3	Hollow Bi_2WO_6/ RGO spheres	MO	0.031 mmol /L	1000 mg/L	480 min	Sunlight	78.5%	–	[33]
4	ZnO/Fe_3O_4/pumice	MB	0.106 mmol /L	50 mg/L	60 min	Visible light	85.5%	–	[34]
5	Fe_3O_4/CuO/ZnO	MO	0.213 mmol /L	200 mg/L	120 min	UV	89%	–	[35]
6	Hollow Bi_2WO_6/ RGO spheres	Phenol	0.106 mmol /L	1000 mg/L	480 min	Sunlight	66.5%		[31]
7	TiS_3 Nanoribbons	MO	50 μM	0.5 mg	5 min	NIR	90.3%		Our work
8	TiS_3 Nanoribbons	MO	50 μM	0.5 mg	15 min	NIR	94.4%	1.207×10^{-1}	Our Work
9	TiS_3 Nanoribbons	MB	10 μM	0.5 mg	12 min	Sunlight	–	1.468×10^{-1}	Our Work

based photocatalysts, as well as many other NIR-induced applications, such as a photodetector, in the future.

Credit authorship contribution statement

Jaspal Singh: Formal analysis, Writing - original draft; Nishant Tripathi, Prabhash Mishra, Vladimir Pavelyev, Valentyn S. Volkov and Aleksey V. Arsenin: Supervision and Writing - review & editing, Formal analysis, Conceptualization, Investigation & Funding acquisition; P. M. Z. Hasan, Ahmed Alshahrie and Reem Darwesh: Data curation & Formal analysis; R.K. Soni, Vladimir Platonov, Prachi Sharma, Daria Shishkina, Anastasiia Rymzhina, Elisey A. Boltov, Rishikesh Singh, Mohammad Talib, Samrah Manzoor, Debosmita Banerjee and Maria A. Anikina: Data curation, Formal analysis, Data curation, Resources & Investigation.

Declaration of Competing Interest

Author doesn't have any conflict of interest.

Acknowledgments

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-606), and the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia under grant no. (KEP-1-194-41). The photo-sensitive nanomaterial has been synthesized by using grant of project funded by Russian Science Foundation (Grant No. 21-79-00272).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.catcom.2021.106381.

References

[1] P. Chowdhary, A. Raj, R.N. Bharagava, Environmental pollution and health hazards from distillery wastewater and treatment approaches to combat the environmental threats: a review, Chemosphere 194 (2018) 229–246.
[2] J. Singh, S. Kumar, R.K. Soni, Synthesis of 3D-MoS_2 nanoflowers with tunable surface area for the application in photocatalysis and SERS based sensing, J. Alloys Compd. 849 (2020), 156502.
[3] S.K. Loeb, P.J. Alvarez, J.A. Brame, E.L. Cates, W. Choi, J. Crittenden, D. D. Dionysiou, Q. Li, G. Li-Puma, X. Quan, D.L. Sedlak, The technology horizon for photocatalytic water treatment: sunrise or sunset? Environ. Sci. Technol. 53 (2018) 2937–2947.
[4] J. Luo, S. Zhang, M. Sun, L. Yang, S. Luo, J.C. Crittenden, A critical review on energy conversion and environmental remediation of photocatalysts with remodeling crystal lattice, surface, and interface, ACS Nano 13 (2019) 9811–9840.
[5] C. Gao, J. Low, R. Long, T. Kong, J. Zhu, Y. Xiong, Heterogeneous single-atom photocatalysts: fundamentals and applications, Chem. Rev. 120 (2020) 12175–12216.
[6] H. Xu, S. Ouyang, L. Liu, P. Reunchan, N. Umezawa, J. Ye, Recent advances in TiO_2-based photocatalysis, J. Mater. Chem. A 2 (2014) 12642–12661.
[7] W. Gao, W. Liu, Y. Leng, X. Wang, X. Wang, B. Hu, D. Yu, Y. Sang, H. Liu, In_2S_3 nanomaterial as a broadband spectrum photocatalyst to display significant activity, Appl. Catal. B Environ. 176 (2015) 83–90.
[8] F. Horani, E. Lifshitz, Unraveling the growth mechanism forming stable γ-In_2S_3 and β-In_2S_3 colloidal nanoplatelets, Chem. Mater. 31 (2019) 1784–1793.
[9] J. Singh, R. Singh, P. Duhan, Y. Maan, A.K. Manna, D. Banerjee, R.K. Soni, Morphology dependent effective charge separation process in nanostructured MoS_2 thin films for enhanced photodegradation behavior, J. Phys. D 54 (2021), 374103.
[10] W. Cui, M. Shao, L. Liu, Y. Liang, D. Rana, Enhanced visible-light-responsive photocatalytic property of PbS-sensitized $K_4Nb_6O_{17}$ nanocomposite photocatalysts, Appl. Surf. Sci. 276 (2013) 823–831.
[11] W. Cui, Y. Qi, L. Liu, D. Rana, J. Hu, Y. Liang, Synthesis of PbS–$K_2La_2Ti_3O_{10}$ composite and its photocatalytic activity for hydrogen production, Prog. Nat. Sci. 22 (2) (2012) 120–125.
[12] W. Cui, D. Guo, L. Liu, J. Hu, D. Rana, Y. Liang, Preparation of $ZnIn_2S_4$/ $K_2La_2Ti_3O_{10}$ composites and their photocatalytic H_2 evolution from aqueous Na_2S/ Na_2SO_3 under visible light irradiation, Catal. Commun. 48 (2014) 55–59.
[13] P.T. Yin, S. Shah, M. Chhowalla, K.B. Lee, Design, synthesis, and characterization of graphene–nanoparticle hybrid materials for bioapplications, Chem. Rev. 115 (2015) 2483–2531.
[14] S. Manzeli, D. Ovchinnikov, D. Pasquier, O.V. Yazyev, A. Kis, 2D transition metal dichalcogenides, Nat. Rev. Mater. 2 (2017) 1–15.
[15] A.A. Tedstone, D.J. Lewis, P. O'Brien, Synthesis, properties, and applications of transition metal-doped layered transition metal dichalcogenides, Chem. Mater. 28 (2016) 1965–1974.
[16] H.S.S. Ramakrishna Matte, A. Gomathi, A.K. Manna, D.J. Late, R. Datta, S.K. Pati, C.N.R. Rao, MoS_2 and WS_2 analogues of graphene, Angew. Chem. Int. Ed. 49 (2010) 4059–4062.
[17] H.S. Nalwa, A review of molybdenum disulfide (MoS_2) based photodetectors: from ultra-broadband, self-powered to flexible devices, RSC Adv. 10 (2020) 30529–30602.
[18] N. Tripathi, V. Pavelyev, P. Sharma, S. Kumar, A. Rymzhina, P. Mishra, Review of titanium trisulfide (TiS_3): a novel material for next generation electronic and optical devices, Mater. Sci. Semicond. Process. 127 (2021), 105699.
[19] G. Sun, Z. Wei, N. Chen, G. Chen, C. Wang, F. Du, Quasi-1D TiS_3: a potential anode for high-performance sodium-ion storage, Chem. Eng. J. 388 (2020), 124305.
[20] A. Khatibi, R.H. Godiksen, S.B. Basuvalingam, D. Pellegrino, A.A. Bol, B. Shokri, A. G. Curto, Anisotropic infrared light emission from quasi-1D layered TiS_3, 2D Mater. 7 (2019), 015022.
[21] F. Ghasemi, R. Frisenda, E. Flores, N. Papadopoulos, R. Biele, D.P. de Lara, H.S. J. van der Zant, K. Watanabe, T. Taniguchi, R. D'Agosta, J.R. Ares, C. S'anchez, I. J. Ferrer, A. Castellanos-Gomez, Tunable photodetectors via in situ thermal conversion of TiS_3 to TiO_2, Nanomater. 10 (2020) 1–14.
[22] C. Huang, E. Zhang, X. Yuan, W. Wang, Y. Liu, C. Zhang, J. Ling, S. Liu, F. Xiu, Tunable charge density wave in TiS_3 nanoribbons, Chin. Phys. B 26 (2017), 067302.
[23] A.S. Pawbake, J.O. Island, E. Flores, J.R. Ares, C. Sanchez, I.J. Ferrer, S.R. Jadkar, H.S. Van Der Zant, A. Castellanos-Gomez, D.J. Late, Temperature-dependent Raman spectroscopy of titanium trisulfide (TiS_3) nanoribbons and nanosheets, ACS Appl. Mater. Interfaces 7 (2015) 24185–24190.
[24] P. Gard, C. Sourisseau, O. Gorochov, Effect of metal substitution in ZrS_3 and TiS_3 compounds. electronic, Raman, and resonance Raman study of $Zr_{1-x}Ti_xS_3$ ternary phases, 0< x< 1, Phys. Status Solidi B 144 (1987) 885–901.
[25] K.K. Mishra, T.R. Ravindran, J.O. Island, E. Flores, J.R. Ares, C. Sanchez, I.J. Ferrer, H.S. Van Der Zant, A. Pawbake, R. Kanawade, A. Castellanos-Gomez, Raman

fingerprint of pressure-induced phase transitions in TiS_3 nanoribbons: implications for thermal measurements under extreme stress conditions, ACS Appl. Nano Mater. 3 (2020) 8794–8802.
[26] W. Liu, M. Liu, X. Liu, X. Wang, H. Teng, M. Lei, Z. Wei, Z. Wei, Saturable absorption properties and femtosecond mode-locking application of titanium trisulfide, Appl. Phys. Lett. 116 (2020), 061901.
[27] M.E. Fleet, S.L. Harmer, X. Liu, H.W. Nesbitt, Polarized X-ray absorption spectroscopy and XPS of TiS_3: S K-and Ti L-edge XANES and S and Ti 2p XPS, Surf. Sci. 584 (2005) 133–145.
[28] D. Gonbeau, C. Guimon, G. Pfister-Guillouzo, A. Levasseur, G. Meunier, R. Dormoy, XPS study of thin films of titanium oxysulfides, Surf. Sci. 254 (1991) 81–89.
[29] H. Yi, T. Komesu, S. Gilbert, G. Hao, A.J. Yost, A. Lipatov, A. Sinitskii, J. Avila, C. Chen, M.C. Asensio, P.A. Dowben, The band structure of the quasi-one-dimensional layered semiconductor TiS_3 (001), Appl. Phys. Lett. 112 (2018), 052102.
[30] S. Okeil, S. Yadav, M. Bruns, A. Zintler, L. Molina-Luna, J.J. Schneider, Photothermal catalytic properties of layered titanium chalcogenide nanomaterials, Dalton Trans. 49 (2020) 1032–1047.
[31] J. Chen, W. Liu, W. Gao, Tuning photocatalytic activity of In_2S_3 broadband spectrum photocatalyst based on morphology, Appl. Surf. Sci. 368 (2016) 288–297.
[32] C. Chen, W. Cai, M. Long, B. Zhou, Y. Wu, D. Wu, Y. Feng, Synthesis of visible-light responsive graphene oxide/TiO_2 composites with p/n heterojunction, ACS Nano 4 (2010) 6425–6432.
[33] Y. Zhang, C. Shao, X. Li, N. Lu, M. Zhang, P. Zhang, X. Zhang, Y. Liu, Controllable synthesis and enhanced visible photocatalytic degradation performances of Bi_2WO_6–carbon nanofibers heteroarchitectures, J. Sol-Gel Sci. Technol. 70 (2014) 149–158.
[34] M.J. Muñoz-Batista, A. Kubacka, A.B. Hungría, M. Fernández-García, Heterogeneous photocatalysis: light-matter interaction and chemical effects in quantum efficiency calculations, J. Catal. 330 (2015) 154–166.
[35] M. Shekofteh-Gohari, A. Habibi-Yangjeh, Novel magnetically separable Fe_3O_4@ZnO/AgCl nanocomposites with highly enhanced photocatalytic activities under visible-light irradiation, Sep. Purif. Technol. 147 (2015) 194–202.

Performance analysis of cost effective multi-hop Time Sensitive Network for IEEE 802.1Qbv and IEEE 802.1Qbu standards

Performance analysis of cost effective multi-hop Time Sensitive Network for IEEE 802.1Qbv and IEEE 802.1Qbu standards

Asha G Hagargund[1,2] Muralidhar kulkarni[1] and Hariram S Satheesh[3]

[1] Department of Electronics and Communication Engineering, National Institute of Technology Karnataka, Surathkal, India.
[2] Department of Electronics and Communication Engineering, B.M.S Institute of technology and management, Bangalore,India.
[3] ABB Global Industries and Services Private Limited,Bangalore,India.

E-mail: ashagh@bmsit.in, mkul@nitk.edu.in , hariram.satheesh@in.abb.com

Abstract. Time-Sensitive Networking (TSN) is an emerging technology, which enables advancements in applications like industrial automation, automatic vehicle-to-vehicle communication, etc. which hosts various time-critical applications, ensuring bounded latency. The novel idea of this paper is to present OMNET++ simulation-based complex multi-hop TSN network using the native VLAN concept to bring out a cost-effective model for inter-TSN and Intra-TSN domains. This paper investigates the performance of hybrid IEEE standards, ie.IEEE 802.1Qbu and IEEE 802.1Qbv standards. The simulation results show that the combination of these standards, when effectively scheduled in switches will reduce the latency by 3.3 µseconds in time-critical applications. Further, it is observed that in Best effort traffic, frame loss is also very less in the range of 2-5 frames out of 1385 frames. These results certainly will be of great value in more complex TSN deployments.

1. Introduction

Various Ethernet-based industrial 4.0 use cases [1], emphasize the need for efficient communication methods. The Time-Sensitive Networking(TSN) standards assert the protocols and mechanisms to scale and control the Ethernet network[2][25]. TSN finds applications in Industrial automation, Electrical power generation and distribution, Automatic vehicle-to- vehicle communication, Professional audio-video studios, Cellular radio, Multimedia communication [3]. The standard Ethernet network fails to provide the bounded latency and may result in data loss due to legacy scheduling mechanisms when both Operation Technology (OT) and Information Technology(IT) data are involved. OT data is high-priority data that involves hardware and software to monitor and control events with in an industry. IT data is noncritical enterprise-level data within industry. Hence to handle both IT and OT data IEEE has proposed various standards such as IEEE 802.1AS (gPTP, an extension of PTP protocol), IEEE 802.1Qbu (Frame pre-emption), IEEE P802.1Qci(traffic filtering and policing), IEEE 802.1CB redundant frame transmission, IEEE 802.1Qcc (configuration aspects), etc. [4][24][26]. In TSN, based on the type, data is classified into various classes and handled by algorithms specific to traffic classes. Data is classified into 8 classes, which correspond to the 8 priorities set by the

PCP value of the VLAN tag. The details of traffic classes and the priorities assigned to the class are presented in [5][6][7]. To simulate TSN topology and check the performance of the network simulation tools like OMNET's CoRE4INET [8][9], Riverbed [10], NeSTiN g[11][5][27], NS-3[12], TSNS[7],DeNS[13]are available. Few papers show the validation of simulation work with hardware setup[28]. We choose to use the NeSTiNg framework of OMNET++ 5.5.1 and INET 4.1.2. We have extended the work carried in the NeSTiNg simulator by considering a complex network with multiple switches and implementing Virtual Local Area Networks(VLANs) to create TSN domains.

1.1. Details on IEEE 802.1Qbu

This standard is also known as frame pre-emption mechanism and it specifies the port utilization and low latency data transmission of real-time and time-sensitive data. The pre-emption mechanism enables the switch to pre-empt the non-time critical data and transmit the high priority time- critical data when they are to be transmitted through a common port. This may introduce delay in low priority data but guarantees the bounded latency of time critical data. Each port in a TSN switch supports 8 queues and hence 8 traffic classes with different priority levels. The traffic classes are Scheduled traffic(ST), Audio Video Bridging (AVB) traffic with bounded latency and Best-Effort(BE) data with default priority. In our work,scheduled traffic with PCP value 7 is considered for which the schedule is generated offline in Gate Control List(GCL) [7][14][15][16], sensor data with the priority of 6, video data of priority 5 and the BE traffic with default priority or PCP value 0 is considered Which has the lowest priority. Each Ethernet traffic class has a priority and is mapped to either express or preemptible interface. Traffic class mapped to express Queue cannot be pre-empted by other traffic classes. Whereas non-express or preemptible class of traffic can be preempted by high priority frames. For a frame to be pre-emptible the minimum frame size should be 102 Bytes excluding header [7]. Recent studies show that a non- pre-emptive switch introduces the delay of 2336 µsecond by a less priority frame and with pre-emption, the same network experiences the delay of 1016 µseconds per 1000 Mbps switch[7].

1.2. Details on IEEE 802.1Qbv

This standard is also called time aware shaper. It mentions the GCL which enables the highest priority control data in the industrial applications to be delivered to with bounded latency. When the arrival of periodic time-critical data is known the GCL pattern can be efficiently designed. As the nodes and bridges in the TSN network can be time-synchronized using IEEE 802.1AS [5][17][18][19][20][21], the 802.1Qbv can be effectively used for critical data delivery. We have assumed that each port of a TSN switch consists of 8 queues. Each port can be scheduled based on the traffic class. In a single port if GCL is 10000000 then the gate corresponding to Queue 7(when Queue numbers are 0 -7) is open and gates corresponding to 0-6 are closed. Hence for the entire duration of gating where GCL has this pattern the frame of queue 7 can only be sent through the switch and all the remaining data frames are halted in the switch. GCL changes periodically and this period is called the gating cycle. The same cycle repeats throughout the simulation duration. The gating mechanism ensures that only one queue gets access to the egress(exit) port. However, the transmission of non-ST data will stop before the time slot of ST data. This time slot or window is known as a protected window. The switch internals for GCL:10000000 is as shown in figure 1.

The time slot where non-time-critical data frames are allowed through the egress(exit) port is called the unprotected window and also an additional time slot or window during which no frame transmission is allowed and the egress port is in idle state is known as a guard window. The details on all these time window slots and switch gating cycle is shown in figure 2.

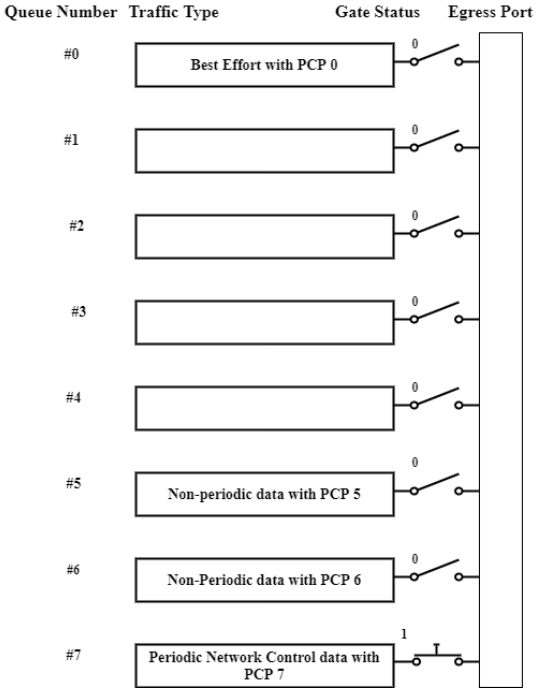


Figure 1. Switch ports and gates.

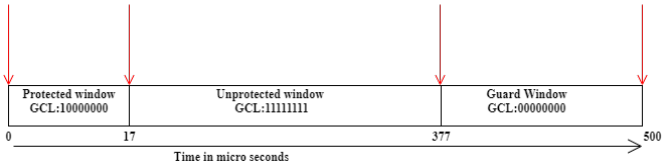


Figure 2. Gate cycle with Protected window, Unprotected window and Guard window.

2. Literature review

Bello[4], this paper provides a detailed introduction of TSN standards which are highly relevant to industrial communication, and also provides the future direction of research. According to the author future scope for research are as follows: While looking at the configuration TSN network, scaling the network schedules with fault tolerance is to be addressed, there is a need for improvement in deploying the applications. Concerning the security of TSN applications, there is a need of detecting the threats early and proposing mitigation strategies. Simulations, Prototypes, Performance studies, use case demonstrations are highly required for the TSN solutions maturity.

Haris Suljic[5], does the performance analysis of TSN by considering IEEE 802.1Qbv, IEEE 802.1Qu, and Credit-based shaper. The simulation was carried out in OMNET++ 5.6.1 and INET 4.2.1 with the NeSTiNg framework. The author has considered a simulation duration of 3 minutes and a simulation cycle of 500 µseconds. Protected window of 17 µseconds and unprotected window of 360 µseconds and guard window of 123 µseconds. Parameters considered for simulation are end-to-end delay and link efficiency. 5 scenarios are considered for performance analysis. The first one is IEEE 802.1Qbv, here the BE traffic experience the packet delay and time-critical data has deterministic data delivery. The Second one is IEEE 802.1Qbu and it is used along with credit-based shapers. The conclusion of this simulation is Audio-video traffic is delayed by credit-based shaper, so frame pre-emption does not reduce the latency of frames with higher priority. The third scenario is using both frame preemption and time aware shaper. The Combination of these 2 reduces the delay of BE traffic whereas delay of deterministic data transmission is unchanged because of using proper gating mechanism. In scenario 4, Gating, Pre-emption, and Credit- based shapers are used. The combined result is frame pre-emption provides reduced end-to-end latency for event-triggered data and BE data experiences more delay compared to the previous scenarios. Scenario 5 is about an industrial in-car network with 14 end nodes and 2 switches. In this scenario frame pre-emption and gating mechanisms are

used. Gating results in an unaltered delay in the data delivery and pre-emption will result in increased delay for BE traffic. The conclusion of this experiment is to look at efficiently and dynamically designing the GCL, also look at various combinations of these standards with various configurations to check the performance of the network in a more efficient way.

In [7], the aspects of Pre-emption are discussed in detail. The performance of the TSN network when there are multiple express data (The data that pre-empts less-priority data) is explained. The author develops a new simulation tool known as TSNS. This tool is used for simulation of the TSN switch behavior according to the IEEE 802.1Q amendments. The environment used is Microsoft's visual studio and code is developed using C language.

In [10], the author develops a tool to check how TSN configuration can be automated. The tool was developed to make use of OMNET's NeSTiNg platform. Researchers can use this extension to automate the configurations, extract the flows and check the performance of the network by using the GUI which is provided by the author. This helps us save time in configuring the parameters through XML files.

Time-aware shaper schedules the frames according to the predefined GCL list.In [15] Ramon Serna Oliver, et al, proposes a method of synthesizing the GCL based on an SMT solver. Aslo discussion on Time aware shaper, Pre-emption, and Credit-based shaper is done [15]. But, still, there is a need to do more analysis by taking various case studies into account. We need to look at trade-offs between various performance parameters and efficiently design the system looking at particular applications. Another motivation behind this paper is to explore various options that are part of the OMNET++ simulation tool for Time-Sensitive Network simulation and look at the options available, appreciate and also look for various improvements in the existing NeSTiNg simulation framework.

Escola[22], The objective of this work is to implement and validate the designed system that allows automatically configuring the TAS's scheduling parameters of the switches of a TSN network. For configuring TAS, IEEE 802.1Qcc proposes NETCONF, RESTCONF, and SNMP from which the author has chosen NETCONF. NETCONF enables us to install, delete and manipulate the configuration of network devices. NETCONF is a network management protocol, but it does not specify how to model the configuration. To tackle this issue, the TSN standards suggest using YANG. YANG is a data modeling language which is used by the NETCONF protocol. YANG is described in the RFC 6020.

Anna arestova[29], In this paper the analysis of Frame preemption(FP) and Time aware shaper(TAS) is carried out and it is compared with Strict priority scheduling(ST) by NeSTiNg framework of OMNET++ simulator. For simulation, a multi-hop, tree topology with multiple star configurations are considered which is comparable to an industrial automation network. The result shows that the TAS produces a large configuration overhead but performs the best and it is suitable for ultra-low latency requirements. FP can be combined with TAS for the efficient band usage. ST shows the least configuration overhead and is suitable for real-time applications with fewer millisecond cycles. The transmission offsets can be effectively scheduled to reduce the interference in the egress queue of a switch when ST and FP standards are used.

3. Proposed Work

3.1. What is the objective of the proposed work?

The objective of this work is to simulate a complex multi-hop TSN network and implement TSN domains using VLAN concepts. Also check the performance of inter-TSN and intra-TSN communication for IEEE 802.1Qbu and IEEE 802.1Qbv when they are applied together.

3.2. What is the need of TSN domains?

Many upcoming industrial automation use cases require TSN domains. A TSN domain is a collection of devices, their ports, and their LANs which use TSN standards to transmit time-

sensitive data [1]. In the proposed work, TSN domains(VLANs) are created based on the functionality of the nodes.Here TSN domain and VLAN are interchangeably used.

3.3. Specification of software tools used for simulation.
Details of the software versions used

i. Operating System - Ubuntu 18.0
ii. OMNET++ - 5.5.1
iii. INET - 4.1.2
iv. NeSTiNg - First version available from omnetpp.org

3.4. What is the need for choosing topology as in figure 5 and figure 6?
We have assumed a real-time scenario where a factory in which the following floors are present.

i. Ground floor: Contains switch "A" and 4 nodes.
ii. First floor: Contains switch "B" and 4 nodes.
iii. Second floor: Contains switch "C" and 4 nodes.
iv. Third floor : Contains switch "D" and 4 nodes.

If on each floor there is a device that belongs to a specific TSN domain then we should go for topology mentioned in figure 5 for intra-TSN and figure 6 inter -TSN communication.

The proposed inter-TSN and intra-TSN communication network are as depicted in figure 3 and figure 4 respectively. In the block diagram the color codes are used for identifying the domains. The devices with same color belong to one domain. We see such '4' domains in figure 3 and '5' domains in figure 4.

As mentioned in the literature survey, according to [5], there is a need to have further look into the configuration aspects of TSN protocols for more efficient data transmission. According to [4] lot of simulations, prototypes verification are required in TSN because many applications require improved versions of existing standards. Hence, we have checked how the combination of both these protocols will increase the efficiency of the communication instead of using them separately.

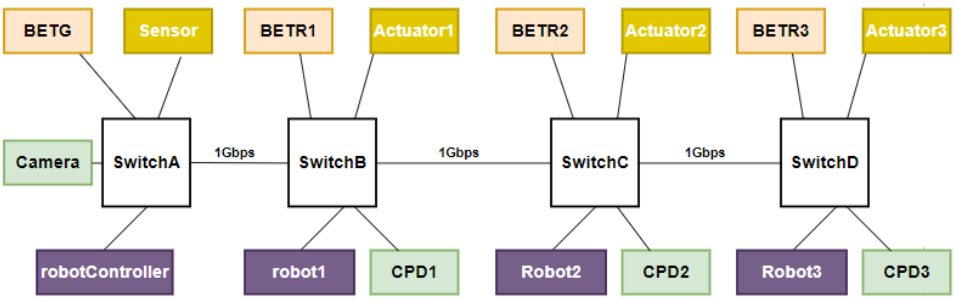


Figure 3. Block diagram of Intra-TSN domain.

3.5. Intra-TSN domain/Intra-VLAN communication
Figure 3 shows the block diagram of the intra-TSN domain network which has 4 TSN-enabled switches and each switch has 4 devices physically connected to them. Each device generates a specific type of traffic and we have scheduled the switch in such a way that each traffic class is handled effectively in the presence of TSN protocols such as Pre-emption and Gating. The configuration of switches, their ports, type of schedule, types of the algorithms used is mentioned in tables 2 and 3.

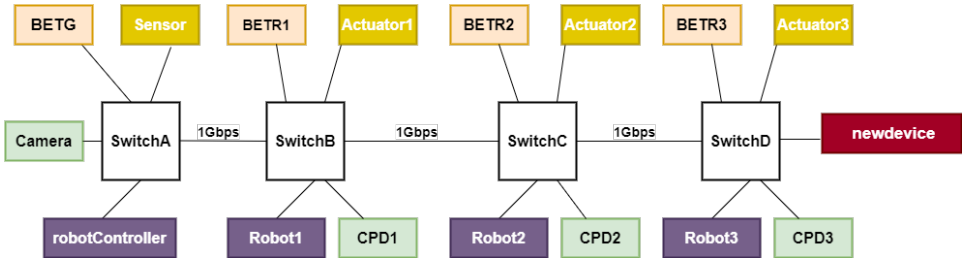


Figure 4. Block diagram of Inter-TSN domain.

The details on the TSN domains and corresponding nodes are mentioned in table 1. These TSN domains are the broadcasting domains or VLANs. 4 VLANs have been created and programmed the network such that the broadcast from a node in a VLAN is limited to only the devices present in that particular VLAN.

3.5.1. Details of TSN domains.

i. TSN Domain 1 is implemented by VLAN 20 which has a robotController that generates time-critical periodic data and broadcast the data to only robot1, robot2 and robot3.

ii. TSN domain 2 is implemented by VLAN 30 and includes a camera which generates a non-periodic data with priority 5 and broadcasts the data to CPD1(Camera Processing Device 1), CPD2, and CPD3.

iii. TSN domain 3 is implemented by VLAN 40 which includes the sensor as the source node that produces nonperiodic data with priority 6. The sensor data is broadcast to only the nodes in that specific VLAN, i.e. to actuator1,actuator2, and actuator3.

iv. TSN domain 4 is implemented by VLAN 60 and BETG as the BE Traffic Generator and broadcasts the data only to BETR1(BE Traffic Receiver), BETR2 and BETR3.

3.5.2. Details on traffic generated by all the nodes and frame sizes.

The robotController transmits minimum sized, network control and periodic traffic hence a frame size of 102 Bytes is chosen. This is the minimum size requirement for frame which is preemptable[7]. Similarly a maximum frame size of 1500 bytes is chosen for BETG and Camera which are similar to Best effort traffic. As the sensors used in real time applications generate the traffic lesser than BE traffic and greater than control traffic hence the sensor traffic chosen is 500 Bytes. Number of robotController frames considered during the simulation period is 138 frames. The number of frames can be increased by increasing the simulation duration. The number of frames sent and received by all the nodes are depicted in Figure 9 and Figure 10.

In the simulation process, the scheduling configuration for switches is written in an XML file and all the VLAN configurations are done in the .ini file of the NeSTiNg framework. The types of data generated by the nodes in the network, size of each data, periodicity, and their priorities are mentioned in table 2.

3.6. Inter-TSN domain/Inter VLAN communication

A new node named as newdevice is now added to Switch D. This belongs to a new VLAN named as VLAN 50.

In the inter VLAN communication we have assumed that there is a communication from robotController that belongs to VLAN20 to newnode that belongs to VLAN 50. Generally when communication between 2 VLANs is implemented by using an L3 router. But, here we

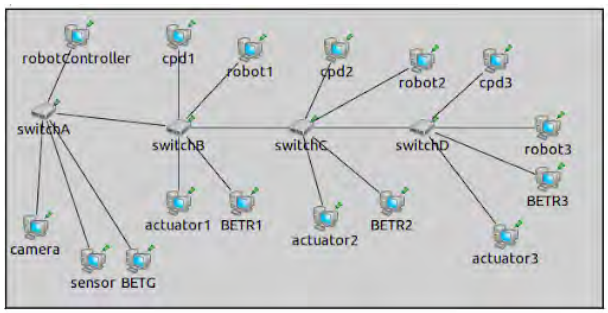


Figure 5. Intra-TSN topology simulated in NeSTiNg.

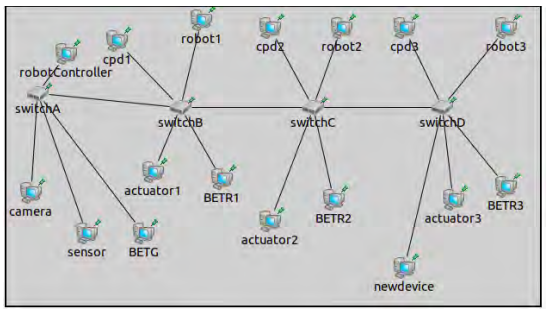


Figure 6. Inter-TSN topology simulated in NeSTiNg.

use the native VLAN concept with the trunk port. This helps us avoid a dedicated router to have communication between 2 VLANS.

figure 7 and figure 8 explain the flowchart of data flow when the native VLAN concept is implemented for inter-TSN domain communication.

We have leveraged the native VLAN concept for inter-TSN domain communication. [23]

We have explicitly tagged the native VLAN with a specific ID in which our desired TSN node is present.

Example: Inter TSN domain traffic L2 routing from VLAN 20 to VLAN 50

Assume that a device is connected to switchD.eth[6] and it is mapped to VLAN 50. The egress traffic of switchC.eth[6] port with VLAN tag 20 will be mapped to VLAN tag 50. This is a native VLAN for this switch-D. Hence the Node connected to switchD.eth[6] gets the TSN traffic which demonstrates inter-TSN domain communication.

Table 1. TSN domains.

TSN Domain	Corresponding Nodes
TSNdomain1	robotController,robot1,robot2,robot3
TSNdomain2	sensor,actuator11,actuator2,actuator3 .
TSNdomain3	camera,CPD1,CPD2,CPD3
TSNdomain4	BETG,BETR1,BETR2,BETR3
TSNdomain5	newdevice

Table 2. Traffic generated by all classes of nodes.

Node	Frame size	Periodic or Not?	PCP value
robotController	102 Bytes	Periodic	7
sensor	500 Bytes	Nonperiodic	6 .
camera	1500 Bytes	Nonperiodic	5
BETG	1500 Bytes	Nonperiodic	4

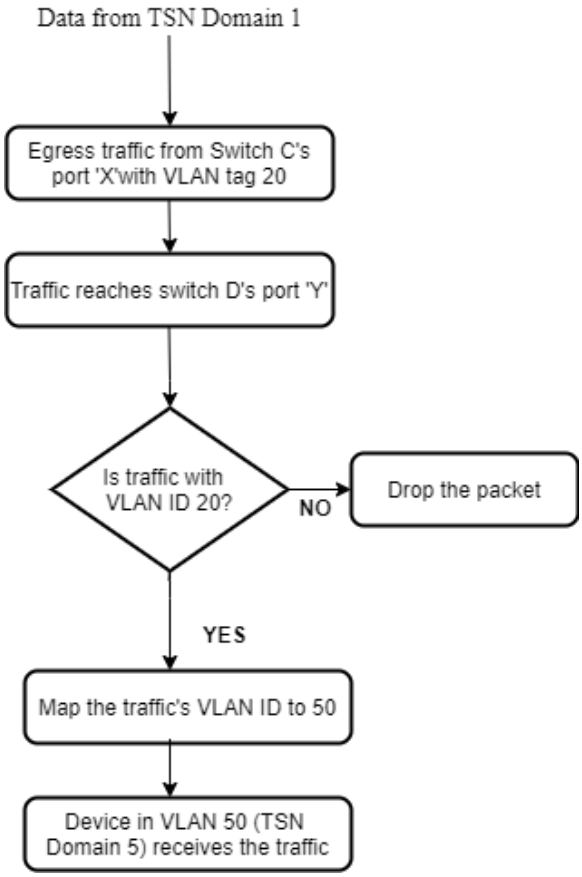


Figure 7. Flowchart demonstrating native VLAN routing for inter-TSN domain.

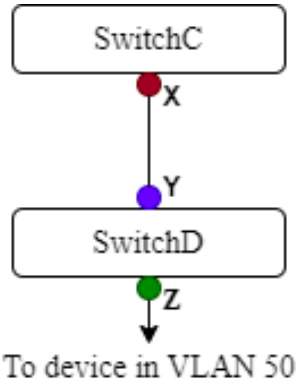


Figure 8. Reference for ports of Switch C and Switch D.

Table 3. Switch parameters used while simulation.

Parameter	Value
`Switch`processing delay	5 µsecond .
`Queue`buffer capacity	363360b
`Ethernet`cable parameters	delay = 0.1 µsecond , datarate = 1Gbps
`TSA`algorithm	Strict priority
`Queue`0 -6 of all switches	Express Queue = False
`Queue`7 of all switches	Express Queue = True

4. Results

According to figure 3 where the communication is within the VLAN. i.e. robotController to robot1,robot2,robot3. Or Sensor to Actuator1, Actuator2, Actuator3, and so on pre-emption along with Gating is applied. 17 µseconds protected window, 360 µseconds unprotected window and, 123 µseconds guard window is considered in one switch cycle. The latency in transmitting time-critical data is constant throughout the simulation time as shown in figure 11. This is possible because, in an entire switching cycle, the robotController transmits the data at a fixed time, and also it is periodic. By knowing this factor GCL vector can be efficiently designed to maintain constant delay in the case of robotController. Looking at the latency charts the

deterministic data delivery is guaranteed for the destinations present after multiple hops also. The latency in transmitting sensor data to all its nodes varying with simulation time as shown in figure 12. This variation in the latency is due to the buffer capacity of queues in switch as well as other high-priority data transmission. Latency for transmitting camera data and BETG data to all the corresponding nodes is periodic as shown in figures 13 and 14 respectively.

The graph of frames sent Vs received by the source nodes to destination nodes of single-hop is shown in figures 9 and 10 for both cases. The robot1, camera1, and sensor1 nodes have 0 frame loss whereas for BETR1 loss is only 2/1385 frames.

In the Inter-TSN network, the high priority robotController data has the constant latency of 13.42 µseconds for the first hop(to robot1), 19.56 µseconds for the second hop(to robot2) and 25.7 µseconds for the third hop(to robot3). This is similar to the latency of the Intra-TSN network. The latency chart for all the remaining source nodes to their corresponding destination nodes in the Inter-TSN network follows the same periodic pattern as intra-TSN latency charts with small variations in the values.

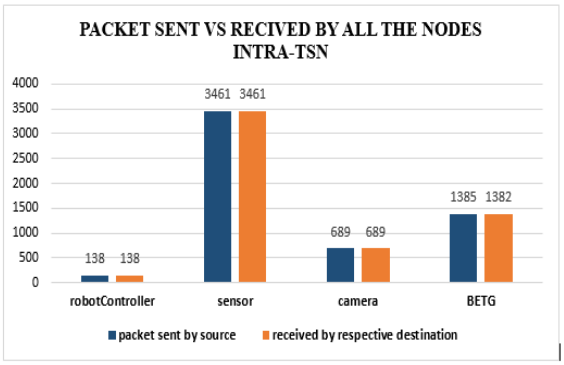


Figure 9. Frames sent Vs received in single hop intra-TSN communication.

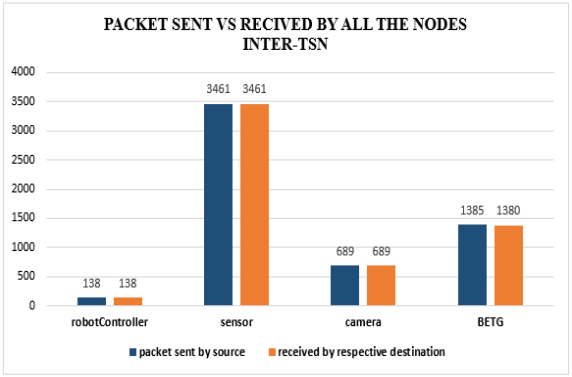


Figure 10. Frames sent Vs received in single hop inter-TSN communication.

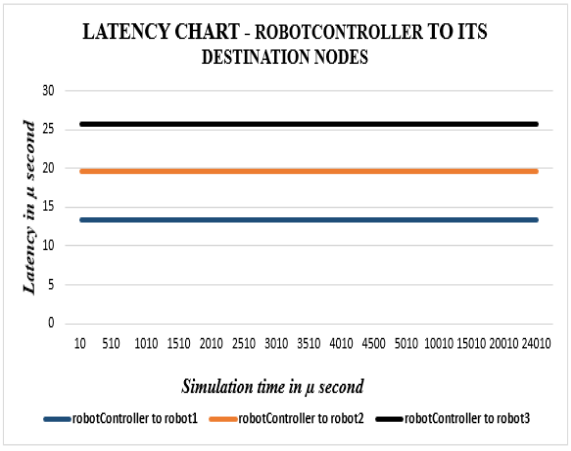


Figure 11. Simulation time vs latency from robotController to all its destinations.

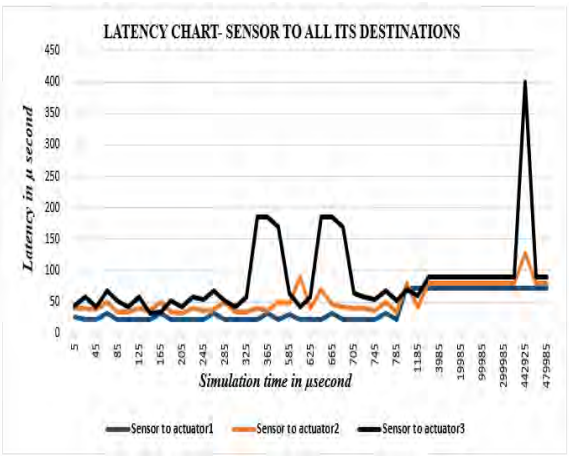


Figure 12. Simulation time vs latency from sensor to all its destinations.

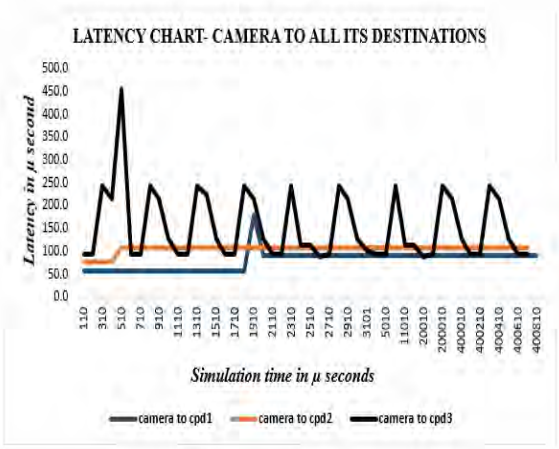


Figure 13. Simulation time vs latency from camera to all its destinations.

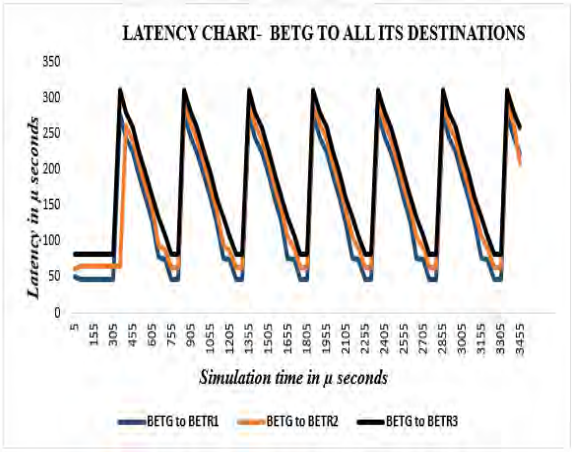


Figure 14. Simulation time vs latency from BETG to all its destinations.

5. Conclusion and future work

In this work, the hybrid IEEE standards 802.1Qbv and 802.1Qbu are used for scheduling the traffic in the switch. By this configuration, we can guarantee deterministic data delivery of time-critical applications and make sure that non-periodic BE traffic will have very little frame loss of 2-5 frames out of 1385 frames. This paper implements a new configuration of TSN domains and also a concept of native VLAN whenever Inter-TSN domain communication is required. This avoids the usage of a dedicated router interconnecting multiple TSN domains. Hence the approach we followed here is a cost-effective solution. The advantage of making multiple broadcast domains in a single network significantly reduces CPU cycles of switches. This avoids unnecessary broadcast of the traffic which is not destined to the nodes. As future work, the hardware validation of complex networks is necessary along with IEEE 802.1AS(Time synchronization) standard. Also proposing new scheduling algorithms to have efficient communication between time-sensitive nodes is required.

References

[1] https://www.ieee802.org/1/files/public/docs2018/60802-industrial-use-cases-0918-v13.pdf
[2] https://1.ieee802.org/tsn/iec-ieee-60802
[3] N. Finn, "Introduction to Time-Sensitive Networking,"2018 *in IEEE Communications Standards Magazine*, vol. **2**, no. 2, pp. 22-28, JUNE 2018, doi: 10.1109/MCOMSTD.2018.1700076.3
[4] L.Lo Bello and W. Steiner, "A Perspective on IEEE Time-Sensitive Networking for Industrial Communication and Automation Systems,2019 " *in Proc. of the IEEE*, vol. 107, no. 6, pp. 1094-1120, June 2019, doi: 10.1109/JPROC.2019.2905334.
[5] Haris Suljic, Mia Muminovi "Performance study and analysis of time sensitive networking" *Thesis for the Degree of Master of Science in Computer Science -Embedded Systems.* Malardalen University School of Innovation Design and Engineering Vasteras, Sweden.June 14, 2019.
[6] Ieee standard for local and metropolitan area networks–bridges and bridged networks,"IEEEStd 802.1Q-2014 (Revision of IEEE Std 802.1Q-2011), pp. 1–1832, 2014
[7] Lejla Murselovi´" Performance analysis of the preemption mechanism in TSN", *Thesis for the Degree of Master of Science in Computer Science*,Malardalen University School of Innovation Design and Engineering Vasteras, Sweden,June 9, 2020.
[8] https://omnetpp.org/download-items/Core4INET.html
[9] J. Jiang, Y. Li, S. H. Hong, A. Xu and K. Wang, "A Time-sensitive Networking (TSN) Simulation Model Based on OMNET++,"2018 *IEEE Int. Conf. on Mechatronics and Automation (ICMA)*, 2018, pp. 643-648, doi: 10.1109/ICMA.2018.8484302.
[10] M. Pahlevan, J. Schmeck and R. Obermaisser, "Evaluation of TSN Dynamic Configuration Model for

Safety-Critical Applications," 2019 *IEEE Intl. Conf. on Parallel Distributed Processing with Applications, Big Data Cloud Computing, Sustainable Computing Communications, Social Computing Networking (ISPA/BDCloud/SocialCom/SustainCom)*, Xiamen, China, 2019, pp. 566-571, doi: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00086

[11] Falk, Jonathan, et al. "NeSTiNg: Simulating IEEE time-sensitive networking (TSN) in OMNeT++."2019 *Int. Conf. on Networked Systems (NetSys)*. IEEE.

[12] T. Wan and P. Ashwood-Smith, "A Performance Study of CPRI over Ethernet with IEEE 802.1Qbu and 802.1Qbv Enhancements," 2015 *IEEE Global Communications Conference (GLOBECOM)*, San Diego, CA, 2015, pp. 1-6, doi: 10.1109/GLOCOM.2015.741759

[13] Addanki, Vamsi, and Luigi Iannone. "Moving a step forward in the quest for Deterministic Networks (DetNet)." 2020 *IFIP Networking Conf. (Networking)*. IEEE.

[14] P. Pop, M. L. Raagaard, S. S. Craciunas and W. Steiner, "Design optimisation of cyber-physical distributed systems using IEEE time-sensitive networks," *in IET Cyber-Physical Systems: Theory Applications*, vol. **1**, no. 1, pp. 86-94, 12 2016, doi: 10.1049/iet-cps.2016.0021.

[15] R. Serna Oliver, S. S. Craciunas and W. Steiner, "IEEE 802.1Qbv Gate Control List Synthesis Using Array Theory Encoding,2018" *IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*, Porto, 2018, pp. 13-24, doi: 10.1109/RTAS.2018.00008.

[16] A. M. Kentis, M. S. Berger and J. Soler, "Effects of port congestion in the gate control list scheduling of time sensitive networks," 2017 *8th Int. Conf. on the Network of the Future (NOF), London*, pp. 138-140, doi: 10.1109/NOF.2017.8251236.

[17] G. M. Garner, A. Gelter and M. J. Teener, "New simulation and test results for IEEE 802.1AS timing performance," 2009 *Int. Symp. on Precision Clock Synchronization for Measurement, Control and Communication*, Brescia, pp. 1-7, doi: 10.1109/ISPCS.2009.5340214.

[18] D. Shrestha, Z. Pang and D. Dzung, "Precise Clock Synchronization in High Performance Wireless Communication for Time Sensitive Networking,"2018 *in IEEE Access*, vol. 6, pp. 8944-8953, doi: 10.1109/ACCESS.2018.2805378.

[19] M. D. Johas Teener and G. M. Garner, "Overview and timing performance of IEEE 802.1AS," 2008 *IEEE International Symp. on Precision Clock Synchronization for Measurement, Control and Communication, Ann Arbor, MI*, pp. 49-53, doi: 10.1109/ISPCS.2008.4659212.

[20] M. D. Johas Teener and G. M. Garner, "Overview and timing performance of IEEE 802.1AS," 2008 *IEEE Int. Symp. on Precision Clock Synchronization for Measurement, Control and Communication, Ann Arbor, MI*,pp. 49-53, doi: 10.1109/ISPCS.2008.4659212.

[21] M. Pahlevan, B. Balakrishna and R. Obermaisser, "Simulation Framework for Clock Synchronization in Time Sensitive Networking," *2019 IEEE 22nd Int. Symp. on Real-Time Distributed Computing (ISORC)*, Valencia, Spain, 2019, pp. 213-220, doi: 10.1109/ISORC.2019.00046

[22] Albert Bergstrom "Automatic generation of network configuration in simulated time sensitive networking(TSN) applications", *Thesis for the Degree of Master of Science in Computer Science with Specialization in Embedded Systems*,Malardalen University School of Innovation Design and Engineering Vasteras, Sweden.June 9, 2020.

[23] https://techhub.hpe.com/eginfolib/networking/docs/switches/K-KA-KB/15-18/atmg/content/ch01s27.html

[24] C. von Arnim, M. Drăgan, F. Frick, A. Lechler, O. Riedel and A. Verl, "TSN-based Converged Industrial Networks: Evolutionary Steps and Migration Paths," 2020 *25th IEEE Int. Conf. on Emerging Technologies and Factory Automation (ETFA)*,pp. 294-301, doi: 10.1109/ETFA46521.2020.9212057.

[25] https://www.ni.com/en-in/innovations/white-papers/18/time-sensitive-networking–tsn–frequently-asked-questions.html

[26] Zhao, Lin, et al. "Comparison of Time Sensitive Networking (TSN) and TTEthernet." 2018 *IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*. IEEE.

[27] A. C. T. d. Santos, B. Schneider and V. Nigam, "TSNSCHED: Automated Schedule Generation for Time Sensitive Networking," 2019 *Formal Methods in Computer Aided Design (FMCAD)*, pp. 69-77, doi: 10.23919/FMCAD.2019.8894249.

[28] Jiang, Junhui, et al. "A Simulation Model for Time-sensitive Networking (TSN) with Experimental Validation." 2019 *24th IEEE Int. Conf. on Emerging Technologies and Factory Automation (ETFA)*. IEEE, 2019.

[29] A. Arestova, K. -S. Jens Hielscher and R. German, "Simulative Evaluation of the TSN Mechanisms Time-Aware Shaper and Frame Preemption and Their Suitability for Industrial Use Cases," 2021 *IFIP Networking Conference (IFIP Networking)*, 2021, pp. 1-6, doi: 10.23919/IFIPNetworking52078.2021.9472830.

Assorted noise identification and removal technique to an obtain enhanced image using CLAHE technique

SHILPA HIREMATH, Dr. A. SHOBHA RANI

Department of ECE, BMS Institute of Technology & Management, Bangalore, Karnataka, India

Abstract

The evolution in multimedia technologies has brought about an abundance of data in the form of images, video, audio which are used in a variety of analysis in various fields including medical field, satellite imagery, image data forensics and many more. However, digital images are mostly corrupted by noises during the process of acquisition, transmission or storage leading to degradation of the image and eventually loss of information. Hence in order to restore the quality of images, suitable image denoising techniques need to be used. Greatest challenge for a denoising algorithm is to enhance visual presentation of image, while protecting pertinent features like edges, during denoising. Further, after denoising, image enhancement techniques can be applied for better perception of the denoised images.

Keywords: Bilateral Filter, CLAHE, Gaussian noise, Median Filter, PSNR, Salt and Pepper noise, Speckle noise, Wiener Filter.

1. Introduction

Digital images are degrading due to the transmission of images from one place to another. These degraded images cannot be used for further image processing. Because it will not give the desirable results after processing. As these images are susceptible to noises under low illumination, high temperature and during long distance transmission. To overcome this various denoising techniques are used. There are various methods that help to remove noise from digital images.

Digital image processing is popularly used in prime fields like medical imaging for recognition of diseases, in face recognition for security grounds, and many more. Consequently, it's mandatory that videos and images must be free of noise for processing. Images are often corrupted by additive or multiplicative noises[11]. Videos and images are deformed by noise like salt and pepper noise, speckle or gaussian noise.

Images are corrupted by different noises during the acquisition process, transmission or storage process. Denoising and enhancement is often important and those are first steps to be taken before any information is extracted from images data. It is necessary to apply an efficient and correct denoising technique [12] to avoid faulty data Image noise which may be caused by sensors and environmental conditions which are unavoidable situations. Identifying noises in an image and applying the correct denoising filter is complex work. Techniques used for denoising images are removing noise by Median Filter, Wiener Filter, Bilateral filter etc. The quality of the image is improved after applying these methods.

2. Literature Survey

The researchers in the below mentioned papers worked on image enhancement techniques:
N. Mohanapriya and B. Kalaavati [1] worked on medical images where they considered spatial domain enhancement techniques and image quality was analyzed based on their performance. Researchers have worked on various techniques to enhance ethe image using the tool called MATLAB. Noise was removed from an image maintaining the image quality.
S.S. Bedi and Rati Khandelwal [2] showed an outline of various techniques in image enhancement considering spatial domain as well as frequency domain. Researchers also showed the short comes and the need in this field of research.
R. Chanana, Er. P Kaur Randhwa, Er. N. Singh [3] worked on Scanned Electron Microscope (SEM) images and presented the results on Contrast Stretching, Histogram Equalization, Local Enhancement Method and Histogram Statistics enhancement techniques.
Adin R. Rivera, B. Ryu, O. Chae [4] came up with an algorithm which considers the entire image and enhances

it accordingly like edges are sharpened, fetches the details hidden in the textured region etc. The enhanced image is obtained by considering the mapping function for each pixel.

R. Garg, B. Mittal, S. Garg [5] worked on rice image and used various enhancements algorithms like contrast limited Adaptive Histogram Equalization algorithm, Equal area dualistic sub image histogram equalization algorithm, and Dynamic Histogram equalization algorithm. Also, researches have provided the experimental results using the measuring factors like PSNR, AMBE, contrast and visual quality.

K. R. Hole, V. S. Gulhane, N. D. Shellokar [6] worked on canonical genetic algorithm and showed a better result after image enhancement and image segmentation techniques were implemented. This work was carried out without the prior knowledge of how the image was degraded.

Arun R, Madhu S. Nair, R. Vrinthavani and R. Tatavarti [7] projected a novel technique by which low contrast images can also be enhanced and the method is alpha rooting. The researchers have converted the image using discrete cosine transform worked on it by applying alpha rooting to magnitude coefficient, converted back to spatial domain and the results show that this approach leads to better enhanced image.

N. R. Mokhtar, N. H. Harun, M. Y. Mashor, H. Roseline, N. Mustafa, R. Adollah, H. Adilah, N. F. Mohd Nasir [8] worked on leukemia images by applying techniques of image enhancement in contrast stretching like local, global, partial, bright and dark contrast stretching. The result shows that image quality will be improved by partial contrast stretching.

A. Roy, S. Bandopadhaya, S. Chandra and A. Suhag [9] worked on real time still and streaming images where the authors worked on detecting impulse noise in an image and applied fuzzy filter to restore the original pixels. Results shown by them have proven to be improved in terms of peak signal to noise ration.

H. K. Rafsanjani, H. Noori and N. Naseri [10] worked on impulse noise, by considering ENI operator they could differentiate between noise and noiseless values. The results are shown in both measures quality and quantity wise which shows the good performance.

3. Methodology of the Proposed Model

3.1 Process flow

Noise is present in different forms in different images; hence we need to identify appropriate noise. Following which correct filter/transform needs to be selected based on type of noise in image. Noise removal/reduction sometimes leads to loss of edge information; hence it is necessary to preserve edges and remove only noise. Various enhancement techniques like histogram equalization, CLAHE and others can be added to improve the interpretability of images.

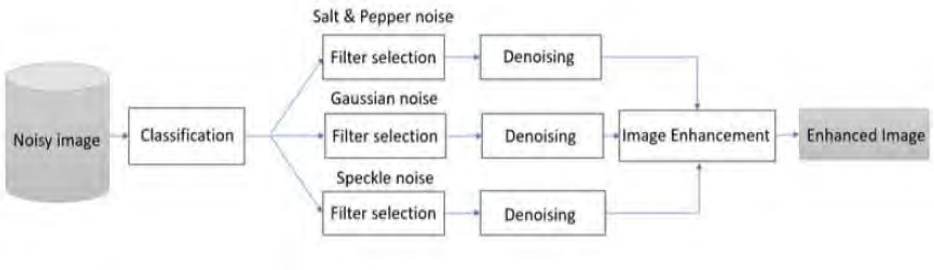


Figure 1: Block diagram of proposed model

3.2 Classification of noisy images

For image classification, we have used a pretrained neural network, Googlenet. Classification of images based on noise present in them. A datastore of images is created. That allows us to read and process data stored in multiple files on a disk, or a database as a single entity. Next the images are split into two groups one for training and the other for testing. Data augmentation is performed to make the images suitable for neural network. Next, we train Googlenet on our dataset.

3.3 Types of noises in an image

We considered three types of noise which are commonly found in images. They are Gaussian, Salt and Pepper and Speckle noise. Gaussian noise is one of the most common noise in the images. Due to the sensor noise caused by poor illumination Gaussian noise appears in the image and also due to high temperature and transmission. Salt and Pepper noise is also called as impulse noise. This noise is used by white and dark pixels i.e., 0's and 1's disturbances distributed all over the image signal. Speckle noise is the granular kind of pattern present in the radar coherent images. Under laser light illumination due to interference of scattered light results in the formation of speckle noise in images. It is found in the airborne images and Synthetic aperture Radar image (SAR). If speckle noise present in SAR oceanographic images it will be difficult for image interpretation for further process in image processing.

3.4 Noise removal filters

The noise present in an image has to removed using a suitable filter. Depending upon the noise present in an image, one among the following filtering technique is applied to the image:

- **Median Filter**
- **Wiener Filter**
- **Wiener Bilateral Filter**

Median filter: The median filter is used to remove the salt and pepper noise present in the image and this filter also used to preserve the edges in the image. This filter also enhances and smoothens the image. A sliding window of window size 3*3, 5*5, 7*7, is moved all over the image pixel by pixel and the center pixel value is replaced with the median value of all its neighboring pixel values. This filter is also called as non- linear filtering technique.

Wiener filter: Wiener filter is one of the most important images denoising techniques for noise such as gaussian noise, which along with denoising removes blur in images caused by linear motion or unfocused optics. Wiener denoises images based on estimating statistics for each neighbor pixel.

Wiener Bilateral Filter: This filter may be a combination of Wiener and bilateral filter. It gives simpler results than individual filters. This filter gave best results for speckle noise. The bilateral filter helps in edge preserving, noise reducing and also in smoothening the image. In this filter each pixel intensity value is replaced by weighted average of pixel values nearby.

3.5 Image enhancement technique

After Image denoising using appropriate filters image enhancement technique is applied as a part of next stage of the proposed methodology. We used CLAHE technique for image enhancement. Histogram equalization works well when the histogram of the image is confined to particular region. If histogram is applied all over the image the lighter pixel values become much lighter and we may lose most of the information due to over brightness. So, to overcome this problem we used adaptive histogram equalization. Here images are divided into blocks called "tiles" and then histogram equalization is applied on each of these blocks if there is a noise present in block it will get over amplified to avoid this contrast limited adaptive histogram equalization (CLAHE) is used which clips the noise more than the threshold or the clip limit. These excess values are accumulated and redistributed in the image. The resulting new tiles are stitched using bilinear interpolation to produce an output with more contrast. We comparatively studied the histogram equalization and CLAHE technique on denoised images. But we got better PSNR values and a better pleasing image for CLAHE technique compared to histogram equalization.

4. Results and Discussions

The below figures show all the three types of considered noises i.e salt & pepper noise, gaussian noise and speckle noise without and with enhanced images. Figure 2 shows for salt & pepper noise before and after applying enhancement technique for grey image. Same is shown for color image in figure 3. Figure 4 & 5 are shown for Gaussian and Speckle noise respectively.

Figure 2: Original gray image with salt and pepper noise (left), enhanced image (right)

Figure 3: Original color image with salt and pepper noise (left), enhanced image (right)

Figure 4: Original color image with Gaussian noise (left), enhanced image (right)

Figure 5: Original color image with Speckle noise (left), enhanced image (right)

The below tables 1, and 2 show experimental results of various PSNR values obtained for different noise levels in

each type of noise. The best filter for a particular noise is highlighted in yellow color in the table. The same data is plotted on a bar graph (figure 1 and figure 2) for better visualization. And also, the performance analysis of CLAHE technique on images is also obtained which is shown in tables 3, 4 and 5 and plotted used a bar graph which can be seen in figures 8, 9 and 10. As the window size of filter increased, we got the better PSNR values and thereby the image quality also enhanced.

Table 1: PSNR comparison for different denoising techniques for Gaussian noise

GUASSIAN NOISE	NOISE LEVEL1 - 10%	NOISE LEVEL2 - 15%	NOISE LEVEL3 - 20%
Wiener	24.32	24.73	25.36
Median	20.79	20.99	21.3
Wavelet	22.75	23.25	24.06
Bilateral Median	20.85	21.04	21.35
Bilateral Wiener	23.84	24.21	24.73
Bilateral Wavelet	24.19	24.98	25.79
Wiener Bilateral	23.29	23.63	24.14

Table 2: PSNR comparison for different denoising techniques for Speckle noise

SPECKLE NOISE	NOISE LEVEL1 - 1%	NOISE LEVEL2 - 3%	NOISE LEVEL3 - 5%
Wiener	26.71	27.69	26.02
Median	24.37	25.69	23.81
Wavelet	25.08	28.25	26.22
Bilateral Median	24.56	25.86	23.94
Bilateral Wiener	26.2	27.06	25.34
Bilateral Wavelet	26.13	29.03	27.14
Wiener Bilateral	26.09	27.04	25.18

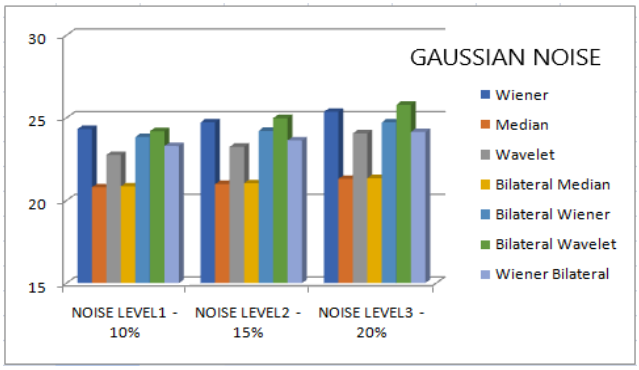


Figure 6: Performance of image denoising filter techniques under Gaussian noise

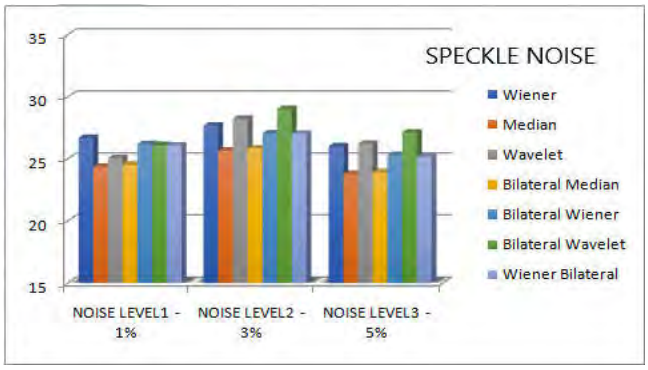


Figure 7: Performance of image denoising filter techniques under Speckle noise

Table 3: PSNR comparison before and after enhancement using CLAHE for salt & pepper noise

SALT & PEPPER NOISE	NOISE LEVEL1 10%	NOISE LEVEL2 15%	NOISE LEVEL3 20%	NOISE LEVEL4 40%	NOISE LEVEL5 60%
MEDIAN (bef clahe)	18.4	17.3	16.44	9.05	7.26
CLAHE	50.86	50.85	50.84	52.8	52.22

Table 4: PSNR comparison before and after enhancement using CLAHE for gaussian noise

GUASSIAN NOISE	NOISE LEVEL1 10%	NOISE LEVEL2 15%	NOISE LEVEL3 20%	NOISE LEVEL4 40%	NOISE LEVEL5 60%
WIENER (bef clahe)	20.9	21.02	21.21	22.65	26.6
CLAHE	43.53	44.13	44.67	46.37	47.39

Table 5: PSNR comparison before and after enhancement using CLAHE for speckle noise

SPECKLE NOISE	NOISE LEVEL1 1%	NOISE LEVEL2 2%	NOISE LEVEL3 3%	NOISE LEVEL4 4%	NOISE LEVEL5 5%
WIENER BILATERAL(bef clahe)	26.11	23.68	22.31	21.32	20.57
CLAHE	41.99	41.98	41.97	41.95	41.93

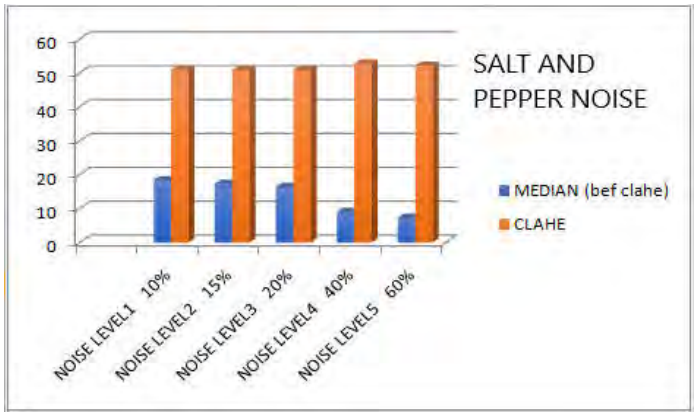


Figure 8: Performance of CLAHE technique after denoising the image under Salt&pepper noise.

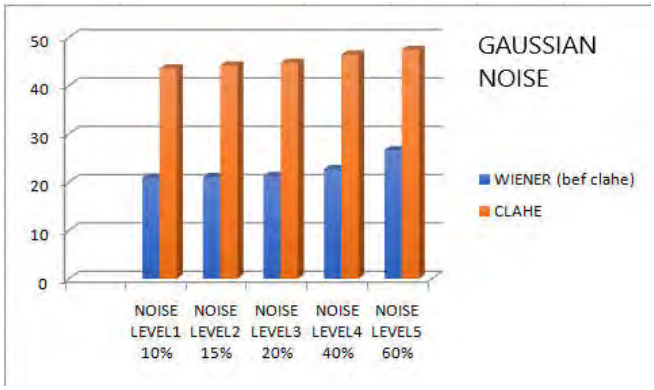


Figure 9: Performance of CLAHE technique after denoising the image under Gaussian noise

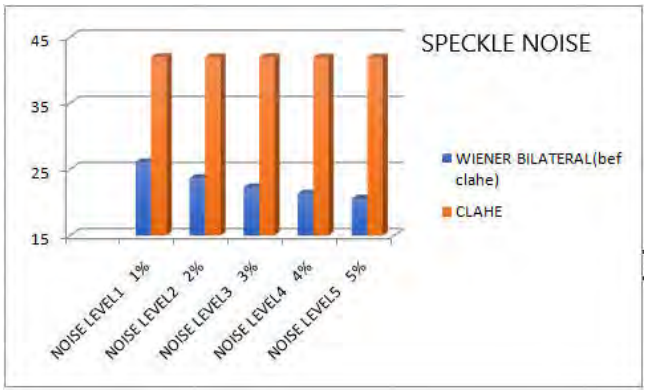


Figure 10: Performance of CLAHE technique after denoising the image under Speckle noise.

5. Conclusion

Classification of noisy images employing a pretrained neural network (NN) by means of transfer learning is employed. In our work, we have considered three kinds of noise; Gaussian, Salt & Pepper and Speckle noise. For transfer learning googlenet is being used. Images are classified based on different noises present in them. Several well-known image denoising techniques like Wiener filter, Median filter, Wavelet Transform, Bilateral filter were implemented and analyzed on the classified images and their performance was comparatively studied. The quality is evaluated by using different noise Levels and PSNR. From the experimental result, it was found that Wiener filter gives better PSNR value for images with Gaussian noise. Its performance was shown to be competitive with or exceeding the performance of other algorithms. Similarly, the Median filter gave better PSNR value for images with Salt & Pepper noise. Bilateral filter worked best for speckle noise images. Image enhancement was implemented using CLAHE technique.

References

[1] Mohanapriya, N. and Kalaavathi, B., 2013. Comparative study of different enhancement techniques for medical images. *International Journal of Computer Applications*, *61*(20).
[2] Singh, G. and Mittal, A., 2014. Various image enhancement techniques-a critical review. *International Journal of Innovation and Scientific Research*, *10*(2), pp.267-274.
[3] Chanana, R., Parneet, E., Randhawa, K., Navneet, E. and Randhawa, S., 2011. Spatial domain-based image enhancement techniques for scanned electron microscope (SEM) images. *IJCSI International Journal of Computer Science Issues*, *8*(4).
[4] Rivera, A.R., Ryu, B. and Chae, O., 2012. Content-aware dark image enhancement through channel division. *IEEE transactions on image processing*, *21*(9), pp.3967-3980.
[5] Garg, R., Mittal, B. and Garg, S., 2011. Histogram equalization techniques for image enhancement. *Int. J. Electron. Commun. Technol*, *2*(1), pp.107-111.
[6] Hole, K.R., Gulhane, V.S. and Shellokar, N.D., 2013. Application of genetic algorithm for image enhancement and segmentation. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *2*(4), p.1342.
[7] Arun, R., Nair, M.S., Vrinthavani, R. and Tatavarti, R., 2011. An alpha rooting based hybrid technique for image enhancement. *image*, *9*(10), pp.1-10.
[8] Mokhtar, N., Harun, N.H., Mashor, M.Y., Mustafa, N., Adollah, R. and Mohd Nasir, N.F., 2009. Image enhancement techniques using local, global, bright, dark and partial contrast stretching for acute leukemia images.
[9] Roy, A., Bandopadhaya, S., Chandra, S. and Suhag, A., 2022. Removal of impulse noise for multimedia-IoT applications at gateway level. *Multimedia Tools and Applications*, pp.1-18.
[10] Rafsanjani, H.K., Noori, H. and Naseri, N., 2022. Diffusion based method for impulse noise removal using residual feedback. *Computers & Mathematics with Applications*, *107*, pp.45-56.
[11] Hiremath, S. and Rani, A.S., 2020. A Concise Report on Image Types, Image File Format and Noise Model for Image Preprocessing.
[12] Kalpana, H.S., 2015. Review Paper: To Study the Image Denoising Techniques. *International Research Journal of Engineering technology (IRJET)*, *2*.
[13] Ramadhan, A., Mahmood, F. and Elci, A., 2017. Image denoising by median filter in wavelet domain. *arXiv preprint arXiv:1703.06499*.
[14] Fan, W.Q. and Xiao, W.S., 2019. Image denoising based on wavelet thresholding and Wiener filtering in the wavelet domain. *The Journal of Engineering*, *2019*(19), pp.6012-6015.

[15] Zhang, K., Zuo, W., Chen, Y., Meng, D. and Zhang, L., 2017. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, *26*(7), pp.3142-3155.

Smart Agriculture and Smart Farming using IoT Technology

Smart Agriculture and Smart Farming using IoT Technology

Dankan Gowda .V[1] , Sandeep Prabhu M[2], Ramesha. M[3], Jayashree M Kudari[4] and Ansuman Samal[5]

[1]Department of Electronics and Communication Engineering, B.M.S Institute of Technology and Management, Bangalore, Karnataka, India.
[2]Department of Electronics and Communication Engineering, Canara Engineering College, Benjanapadavu, Bantwal, Mangalore, Karnataka, India.
[3]Department of Electronics and Communication Engineering, GITAM School of Technology GITAM(Deemed to be University), Bangalore, Karnataka, India.
[4]Department of Computer Science and Information Technology, JAIN Deemed to be University, Bangalore, Karnataka, India.
[5]Faculty of Hospitality and Tourism Management, Siksha O Anusandhan Deemed to be University, Bhubaneshwar, Odisha, India.
Email: dankan.v@bmsit.in

Abstract. It has become easier to access agriculture data in recent years as a result of a decline in digital breaches between agricultural producers and IoT technologies. These future technologies can be used to boost productivity by cultivating food more sustainably while also preserving the environment, thanks to improved water use and input and treatment optimization. The Internet of Things (IoT) enables the production of agricultural process-supporting systems. Referred to as remote monitoring systems, decision support tools, automated irrigation systems, frost protection systems, and fertilisation systems, respectively. Farmers and researchers must be provided with a detailed understanding of IoT applications in agriculture as a result of the knowledge described above. This study is about using Internet of Things (IoT) technologies and techniques to enhance agriculture. This article is meant to serve as an introduction to IoT-based applications in agriculture by identifying need for such tools and explaining how they support agriculture.
Keywords: Internet of Things (IoT), Internet of Lighting, Fertilization, Agriculture, Remote, Communication.

1. Introduction

The advancement of science and technology, the global GDP has risen consistently. As a result, the presence of this reality has encouraged the development of smart farming, which use sensors and irrigation systems to manage crops as they grow. With sensor-based computer applications, more accurate information about the crop, soil, and environment may be gathered. It promotes high-quality process and raw materials throughout the entire product process. This is because utilising the Internet of Things in smarter agriculture makes it more competitive than traditional methods. Combined with IoT-based smart agriculture technologies, organic agricultural agriculture and family farming may see a benefit. It is the case that between agricultural producers and IoT technology, a digital breach has occurred, thus farmers are not vulnerable to IoT assaults. Sustainable use of water and input and treatment optimization will allow farmers to produce more food while also preserving the environment. To include agriculture in smart use of natural resources, usage of technologies such as remote control, decision support tools, automated irrigation systems, frost avoidance, and fertilisation

is required. These activities are supplied by IoT technologies, which provide devices like as hardware, intelligent apps, integration platforms, control procedures, operating systems, and cloud computing. The benefits of IoT and the Internet may be gained through the Cloud of Things, which combines IoT with cloud computing. Another requirement for the IoT is for it to provide society with information transparency. This work summarizes the current IoT-based agricultural tools and applications, which are broken down into distinct areas [1,2]. The aim of this paper is to describe all these topics in detail, as well as discuss the advantages of using IoT-based technologies in agriculture.

The yields obtained with less capital and labor have improved over time, with significant innovations having been made in human history. However, as long as the population rate is high, the demand and supply cannot balance, regardless of the periods. The numbers that have been estimated say that in 2050, the world population will be around 25 percent larger than it is now. In the vast majority of the growth predicted for the developed countries, it is expected that much of it will occur. Urbanization is expected to increase even more rapidly in the future, and about 70 percent of the world's population will be urban by 2050 (currently 49 percent). A further factor contributing to food demand is that income levels will be multiple times what they are now, which will lead to further increases in food demand particularly in developing countries [3,4].

Consequently, dietary preferences will change from wheat and grains to legumes and then to meat. To meet the increased demand for food due to an increasingly urbanized and wealthy population, food production must increase by a factor of two by the year 2050. This prediction is particularly important because of the current prediction of approximately 2.1 billion tons of annual cereal production and the prediction of a rise in meat production of over 200 million tons to meet the predicted demand of 470 million tons. Crop processing is playing an increasingly important role in industrial economies as well. On top of that, the bio energy demand began to increase in the food crops-based bio energy market. From the start of the twentieth century until now, only the manufacture of ethanol has used 110 million tons of coarse grains (approximately 10 percent of the world production). Food protection is in danger due to the rise in industrial and other uses of food crops for bio-fuel production. the scarcity of agricultural capital is making the demands even more onerous.

2. Literature Survey

The deployment of IoT in agriculture has the potential to affect our society and the rest of the world. Nowadays, we see weather, soil, and water drying up as land that's critical to agriculture declines, making it harder and harder to produce food. Agriculturalists will benefit from using Internet of Things (IoT) technology, which will help them cut down on generated wastelands while also improving production. This figure could stem from the number of missions the farmhouse automobiles have performed, or from the amount of compost used in the composting process. A smart agriculture system may be defined as a food system that utilizes emerging nutrition that is uncontaminated and is accessible to a wide number of people. With the expansion of the entire Farming system with the addition of the Smooth Agricultural IoT platform, the Internet of Things (IoT) plays a larger role in agriculture. Although the Internet of Things (IoT) is utilized in Farming, it has saved not only the time of agriculturalists but also massive quantities of liquid and power, thanks to the interconnectedness of devices and services. It is able to preserve frequently encountered topographies including moisture, high temperature, soil, etc. and offers real-time surveillance through the crystal-clear map. In agriculture, embracing Internet of Things (IoT) will yield various benefits. For example, the farmhouse automobiles have accomplished numerous missions. Smart agriculture is thus basically an integrated, uncontaminated method of emerging nutrition that supports crowds. The smooth agricultural system extends the farming system by not only watching the soil, but also physically monitoring it [5,6]. Even though individual devices and networking aren't directly saving the agriculturalists' time, the Internet of Things (IoT) is negatively affecting wasteful spending on assets such as Liquid and Power. The overall goal of this model is to preserve frequent topographical features like moisture, temperature, soil, and other information, and provide a real-time crystal-clear

surveillance. In addition to the advantages described above, agriculture will benefit from implementing Internet of Things (IoT).

Tunable Fields: Precision agriculture is a way or practice that makes the farming process more correct and managed to raise live stocks and grow crops. The use of IT and objects such as sensors, self-supporting cars, computer hardware, control systems, robotics and many others. The main additives are in this technique. Precision farming using IoT is presented in figure.1. Precision agriculture has become one of the most well-known agricultural IoT initiatives in recent years and this technique has begun to be used by a large number of organisations.

Figure 1. Precision farming using IoT [1]

Data Analytics: The predictable database system now has not enough parking space to store the facts from the IoT sensors. In the clever agriculture machine the cloud based mainly facts garage and a stop-stop IoT platform plays an important part. These structures are expected to play a vital role in finishing higher sports. In the IoT universe, sensors are the main source of massive facts. The numbers are analyzed and the use of analytical equipment converted into meaningful facts. The analytics of records helps to evaluate temperature, farm animal conditions and crop situations. The statistics collected use the technical advances and make better decisions for this purpose. You can understand the plants' real-time reputation using IoT devices, by collecting the facts from sensors[7,8]. You will gain a perception using predictive analytics to make better decisions on harvesting. The fashion analysis allows farmers to understand future climatic conditions and vegetation harvesting. IoT enabled farmers to maintain the quality of the vegetation and the fertility of the land in the agricultural industry, thus improving and improving the amount of products exceptional.

Climate Conditions: Climate plays a vital role for agriculture. And mistaken climate know-how deteriorates the quantity and the first class of crop production considerably. But IoT answers allow you to know the weather situation in real time. Within and outside the agricultural fields, sensors are mounted. They collect environmental statistics that are used to select the right plants that can grow and sustain in precise climatic situations. The entire IoT atmosphere consists of sensors that can locate conditions such as humidity, precipitation, temperature and more accurately in real time. There are various sensors that have to be hit on each of these parameters and configured to fit your clever agricultural needs. These sensors reveal the situation and climatic conditions of the crops around them. If any troubling climatic conditions are established, the ship is an alert. The lack of physical presence in troubling climatic conditions, which eventually increases productivity and helps farmers to gain higher agricultural authorisations, is eliminated[9,10].

Smart Greenhouse: Greenhouse agriculture is a technique complementing crop yields, greens, end results etc. Greenhouses handle environmental parameters in two ways, either manually or by a proportional control mechanism. However, these approaches are much less successful because

manual intervention has risks, including production loss, energy loss and labor prices. A smart, IoT-based greenhouse now does not simply track but regulates the environment. There is a need for human action. Various sensors that are in line with plant requirements are used to monitor the environment in an intelligent greenhouse. A cloud server then creates a computer for remote access when associating IoT usage. The cloud server confidentially enables records to be processed and manages the flow. This design offers farmers the best and most effective solutions with minimum and almost no manual intervention.

Agricultural Drones: Scientific advances have almost revolutionized agricultural operations and there is trend disturbance with the advent of agricultural drones. Ground and aerial drones are used for the fitness assessment, crop inspection, planting, spraying of crops and field assessment. With the right approach and preparation based on real reality, the drone generation has given the agriculture industry a strong push and overhaul. Drones with thermal or multi-spectral sensors select areas where irrigation changes are necessary. When plants begin to grow, sensors indicate their health and measure the index of their plants. Clever drones have eventually reduced the environmental impact. The technology based Smart farming using Drones is presented in figure.2. The consequences were such that there was a great reduction in the chemical effects of groundwater and Livestock Internet monitoring correspondences allow farmers to obtain materials about their livestock's neighborhood, proper life and welfare [11,12]. This calculation allows them to recognize their livestock's position. Such as identifying animals that are uncomfortable in order to divide the herd, preventing the disease from developing for the whole animal. The feasibility of farmers with Internet of Things (IoT)-based sensors to locate their farm animals helps them to transportation of depressed hard work charges by a significant amount.

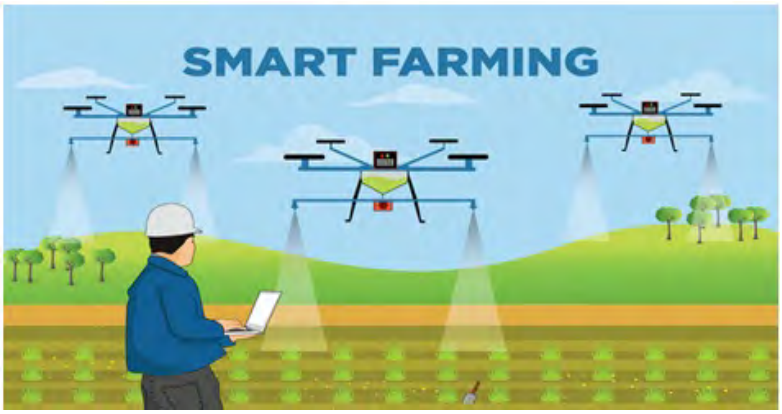


Figure 2. Smart farming using Drones [2]

3. IoT Irrigation System

Figure 3 illustrates a typical IoT Irrigation system. Crop docking and air temperature measurement sensors are connected to the network gateway by means of a Wireless Sensor Network (WSN). ZigBee is a popular choice with many different applications since it is simple to implement and personalise. The connection to the Internet will be at a lower cost since this software only requires a limited quantity of data capacity. 4G LTE mobile communications are being used to connect to the Internet wirelessly. Mobile network data is sent to the cloud-based web services that are subscribed to, and the data is made available to other cloud-based web services[13,14]. Software that has the ability to integrate all kinds of data and combine it with CWSI models to create irrigation index values is able to analyse agricultural land data and, based on this information, can apply CWSI models to assess water need. Weather service and satellite imagery are other relevant sources of information, and the results are irrigation index values for each area. The data provided to the network gateway is sent to a

controller for irrigation management, after which it is forwarded back to the gateway. Both data and results, as well as making changes to terminals, may be seen via specialised web applications, and farmers may also use these web apps to provide other farmers and professionals access to their data and outcomes.

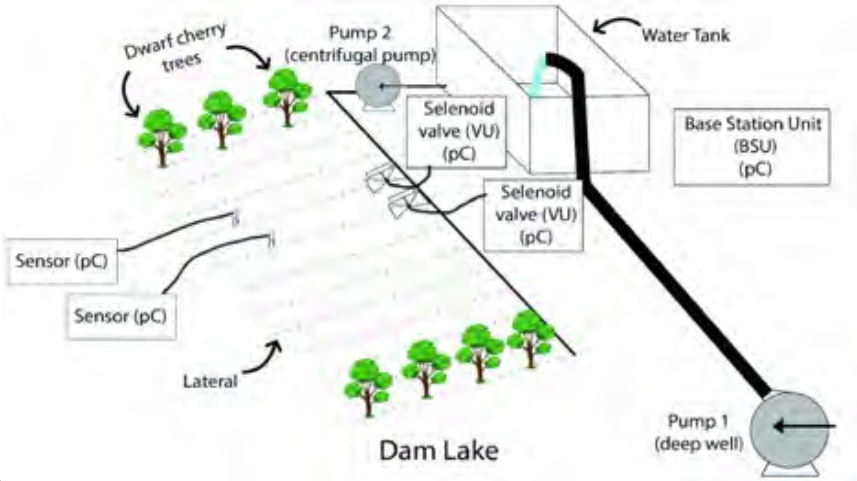


Figure 3. IoT Irrigation System [3]

4. Smart Cow Farm

In addition to cattle farming, IoT technology is also used in livestock farming. With the use of IoT sensors on the cows and the use of smartphones or tablets, farmers can easily find cows and detect significant animal welfare issues like rumination and lameness. As farmers use Internet of Things (IoT) technology to convert their animals into the so-called Internet of Cows, connected cows, or intelligent cows, their animals will come to be known as Internet of Things cows. As far as cattle farming and/or ranching is concerned, the Internet of Things paradigm may help by improving water, power, nutrition, and other resource production while also preserving animal welfare. Additionally, it gives farmers the ability to organise information, create reports, categorise cows, and track the life cycle of each animal [15,16]. IoT has given farmers tough issues like identifying cow estrus, which has long been a challenge for the dairy industry. When it comes to cows, the farmer will spend anywhere from 20 to 30 minutes four to five times a day in the stables to check whether the cow is in heat, which is a symptom of estrus. In figure.4. Presented an technology base IoT Smart Cow Farm system. The majority of instances occur at night while the farmer is asleep. There are a number of additional examples, such the animal disease "cattle lameness" which has a major impact on cows' productivity in terms of output, fertility, and lifespan. The importance of these problems has diminished because of the use of IoT technology. Cows equipped with IoT sensor tags are connected to their ears, necks, or legs to track their daily labour and overall well-being every 24 hours. After gathering data, the results are analysed using complex statistical and empirical models to determine whether a cow is in heat. As a consequence, the IoT solution detection rate may reach up to 95%, whereas conventional methods achieve detection rates in the range of 55 to 70%. Fully integrated IoT health monitoring systems are designed to provide accurate and dynamic diagnostics of lameness. In the smart cow farm, both animal sensors and other nearby sensors will be used to collect varied data. Data is used for automated equipment control, as well as animal status monitoring and user decision recommendations. Diversified sensors are miniaturising, reducing prices, and increasing performance, which will lead to more specialised IoT solutions to assist cow farmers.

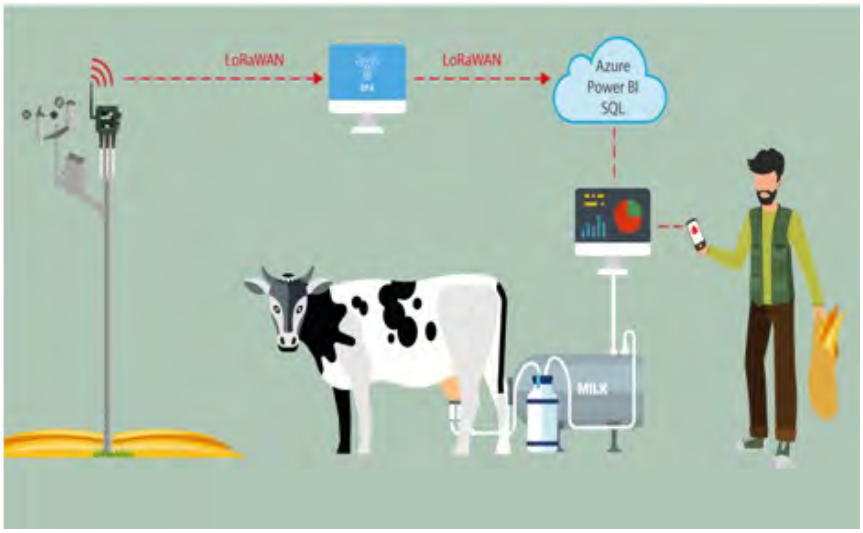


Figure 4. IoT Smart Cow Farm [4]

5. IoT Can Improve Agriculture

The Internet of Things (IoT) is used by farmers to implement certain IoT methods for enhancing their agricultural. Concentrated penalties for these practises may persuade farmers to abandon them. 1. D ata; loads of information such as weather, good soil, crop development, or animal health gathered by smart agricultural sensors These data may be utilised to help your business enhance the smartness of its country and its workers' overall performance, productivity, and efficiency. 2. Reduced production risks due to improved switching over internal techniques. If you are able to anticipate the outcome of your manufacturing, you can plan ahead and be better prepared to deliver your goods on time. If exactly how much undergrowth you are going to cultivate, you might persuade your products that their sale is increasing. 3. Increased business efficiency by process mechanization. You will master various technologies through the building period, e.g. drainage, composting or pest control by applying smooth strategies. 4. Budget organization and unused decrease awareness of the manufacturer by the increased controller. If you are smart to see irregularities or control fitness in harvest, you can moderate the dangers behind your products. 5. Improved dominance and capacity growth. Improved regulator completed construction development and established principles of production excellence and increasing mechanization completed with volume[17].

6. Challenges of IoT for Agriculture

The is a All smart farming responses must start with data analysis. If you can't make sense of the information you've gathered, it'll be of no use. As a result, you'll need strong data analytics, predictive algorithms, and devices to analyze the data and derive operational insights [18,19]. the equipment is made of of Selecting the sensors for your instrument is critical for internet of things agriculture (or create a custom one). The manner in which you look for information and how you make decisions all influence your choice. Whatever the situation may be, it is possible to discern the efficacy of your product based on the quality and consistency of the data you gather. Keeping up with Hardware maintenance is an important project in agriculture Internet of Things products, since sensors are commonly used in the topic and can be easily destroyed. As a result, you must make certain that your hardware is both reliable and simple to maintain. Then you'll have to update your sensors more often than you'd like. The revolution is underway, In figure .5. shows the various Smart farming tools. In order to be used in the field, intelligent agricultural applications must be adapted [20,21]. To access the information on the website, a company owner or farm administrator must be able to use a mobile phone or a personal computer anywhere in the globe. Furthermore, each linked tool must be self-

contained and have sufficient wireless diversity to connect to other devices and transmit data to the central server. The services available You'll need a robust internal infrastructure to ensure that your smart farming application runs smoothly (and that the load of records can handle it). Furthermore, the internal systems must be pleasant to use. Failure to make our system more user-friendly only increases the appeal of someone who interrupts, steals your information, or even uses your self-satisfying tractors.

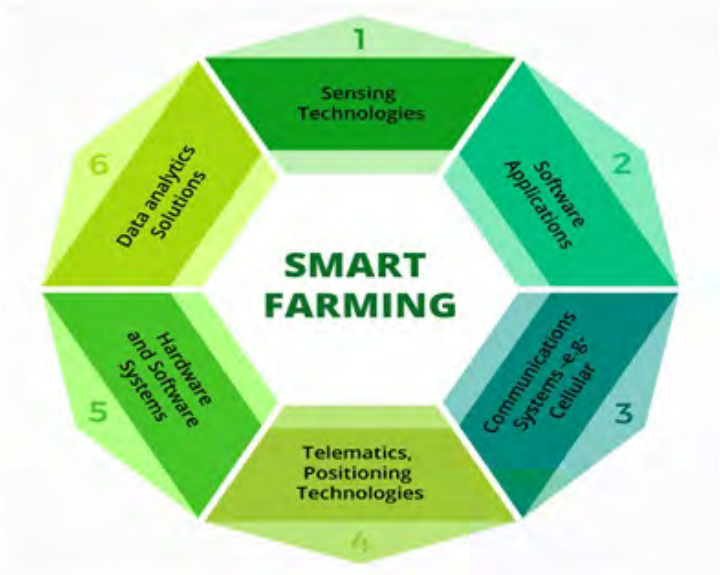


Figure 5. Smart farming Tools [5]

Smart Farming is a cultured understanding of management using current devices to improve the quantity and excellence of sophisticated properties. In the 21st period farmers have access to GPS, ground browsing, data management and the internet of machinery for stuff. By measuring difference within a field with confidence and familiarizing the method, farmers can significantly increase the efficiency and more selective use of pesticides and stimulants. Smart farming is a virtual global call for these days [22,23]. Intelligent farming offers many possibilities including nice water, healthy plants. Smart agriculture is an idea for managing agriculture, with a view to controlling advanced know-how which includes extensive knowledge, the cloud and Internet of Things (IoT) to follow, look, mechanize and compare approaches.

7. Conclusion

Agriculture facilitated by the Internet of Things has helped implement current technical responses to time explored understanding. This allowed the distance between production and pleasant production and quantity to be combined. Statistics The fact that more than one instrument are obtained and measured for real time use or garage are placed in a database guarantees rapid intervention and significantly less harm for vegetation. With smooth cease-fire and better execution of business processes, production becomes quicker and affects supermarkets in the wildest period. IoT farming applications are production of expressive statistics that farmers and farmers can possibly collect. Large owners of land and smallholder farmers need to appreciate the potential IoT demand for agriculture by linking intelligent know-how to increasing their manufacturing attractiveness and sustainability. In this paper we research the IoT application for agriculture and how farmers can expand by using the

Internet of Agriculture Stuff. This article examines the job opportunities of the Internet of Things (IoT).

References

[1] M. Ayaz, M. Ammad-Uddin, Z. Sharif, A. Mansour and E. -H. M. Aggoune, "Internet-of-Things (IoT)-Based Smart Agriculture: Toward Making the Fields Talk," in *IEEE Access,* vol. 7, pp. 129551-129583, 2019, doi: 10.1109/ACCESS.2019.2932609.

[2] I. Mat, M. R. Mohd Kassim, A. N. Harun and I. M. Yusoff, "Smart Agriculture Using Internet of Things," *2018 IEEE Conference on Open Systems (ICOS),* 2018, pp. 54-59, doi: 10.1109/ICOS.2018.8632817.

[3] R. Dagar, S. Som and S. K. Khatri, "Smart Farming – IoT in Agriculture," *2018 International Conference on Inventive Research in Computing Applications (ICIRCA),* 2018, pp. 1052-1056, doi: 10.1109/ICIRCA.2018.8597264.

[4] C. Yoon, M. Huh, S. Kang, J. Park and C. Lee, "Implement smart farm with IoT technology," *2018 20th International Conference on Advanced Communication Technology (ICACT),* 2018, pp. 749-752, doi: 10.23919/ICACT.2018.8323908.

[5] S. R. Prathibha, A. Hongal and M. P. Jyothi, "IOT Based Monitoring System in Smart Agriculture," *2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT),* 2017, pp. 81-84, doi: 10.1109/ICRAECT.2017.52.

[6] I. M. Marcu, G. Suciu, C. M. Balaceanu and A. Banaru, "IoT based System for Smart Agriculture," *2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2019, pp. 1-4, doi: 10.1109/ECAI46879.2019.9041952.

[7] D. V. Kishore, Shivashankar, A. C. Ramachandra, and C. Pandurangappa, "Optimization of motorcycle pitch with non linear control," in *2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016 - Proceedings*, 2017, pp. 1656–1660, doi: 10.1109/RTEICT.2016.7808114.

[8] Kishore D.V, Shivashankar, and S. Mehta, "MANET topology for disaster management using wireless sensor network," in *International Conference on Communication and Signal Processing, ICCSP 2016*, 2016, pp. 0736–0740, doi: 10.1109/ICCSP.2016.7754242.

[9] A. C. Ramachandra, M. N. Thippeswamy, C. Pandurangappa, and P. Ramesh Naidu, "Modelling and performance evaluation of anti-lock braking system," *Journal of Engineering Science and Technology*, vol. 14, no. 5, pp. 3028–3045, 2019.

[10] M. Penna, J. J. Jijesh, and Shivashankar, "Design and implementation of automatic medicine dispensing machine," in *RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings*, 2017, vol. 2018-Janua, pp. 1962–1966, doi: 10.1109/RTEICT.2017.8256941.

[11] A. C. Ramachandra, M. N. Thippeswamy, C. Pandurangappa, and P. Ramesh Naidu, "Synthesis and modeling of antilock braking system using sliding mode controller," *Journal of Advanced Research in Dynamical and Control System*, vol. 10, no. 12, pp. 208–221, 2018.

[12] C. A. Varun, Shivashankar, M. Sahana, R. S. Varun, and T. Rajesh, "Implementation of swarm intelligence in obstacle avoidance," in *RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings*, 2017, vol. 2018-Janua, pp. 525–528, doi: 10.1109/RTEICT.2017.8256652.

[13] S. B. Sridhara, K. B. Naveen, M. Ramesha, and G. N. Pai, "Internet of things: Internet revolution, impact, technology road map and features," *Advances in Mathematics: Scientific Journal*, vol. 9, no. 7, pp. 4405–4414, 2020, doi: 10.37418/amsj.9.7.11.

[14] P. Ramesh Naidu, N. Guruprasad, "Design and implementation of cryptcloud system for securing files in cloud," *Advances in Mathematics: Scientific Journal*, vol. 9, no. 7, pp. 4485–4493, 2020, doi: 10.37418/amsj.9.7.17.

[15] M. Ramesha, Sridhara S.B, Naveena Pai G, "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*, vol. 8, no. 7, pp. 3893–3897, 2020, doi: 10.30534/ijeter/2020/158872020.

[16] P. Ramesh Naidu, N. Guruprasad, "A High-Availability and Integrity Layer for Cloud Storage, Cloud Computing Security: From Single to Multi-Clouds," *Journal of Physics: Conference Series 1921(1),* https://doi.org/10.1088/1742-6596/1921/1/012072

[17] S. B. Sridhara, M. Ramesha, "Recent advances in graph theory and its applications," *Advances in Mathematics: Scientific Journal*, vol. 10, no. 3, pp. 1407–1412, 2021, doi: 10.37418/amsj.10.3.29.

[18] Ramesha, M., Sridhara, S.B., Pai, G.N., Patil, S.K., "Design of Antilock Braking System Based on Wheel Slip Estimation," *Journal of Physics:Conference Series 1706(1),2020, https://doi.org/10.1088/1742-6596/1706/1/012216.*

[19] Pai, G.N., Sridhara, S.B., Shashidhara, K.S., Gangadhara, "Signal Analysis and Filtering using one Dimensional Hilbert Transform," *Journal of Physics:Conference Series 1706(1),2020,* https://doi.org/10.1088/1742-6596/1706/1/012107.

[20] M. Ramesha, K. Jeevan and B. M. Sathisha, "Implementation of IoT Based Wireless Electronic Stethoscope," *2020 Third International Conference on Multimedia Processing, Communication & Information Technology (MPCIT)*, 2020, pp. 103-106, doi: 10.1109/MPCIT51588.2020.9350476.

[21] Puneeth Kumar M. V, Naveena Pai G, Vinay Kumar B. C, "Dynamic Analysis and Control Strategies of an Anti-lock Braking System," *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2020, pp. 1677-1682, doi: 10.1109/ICECA49313.2020.9297642.

[22] G. Naveena Pai, M. Swathi Pai, M. Shruthi and B. Naveen K, "Internet of Things: A Survey on Devices, Ecosystem, Components and Communication Protocols," *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2020, pp. 611-616, doi: 10.1109/ICECA49313.2020.9297458.

[23] M. Ramesha, Jeevan K.M, "Study and analysis of bted error correction codes for cryptography applications," *International Journal of Advanced Trends in Computer Science and Engineering,* vol. 9, no. 5, pp. 8938–8942, 2020, doi.org/10.30534/ijatcse/2020/293952020.

IoT Enabled Smart Lighting System for Smart Cities

IoT Enabled Smart Lighting System for Smart Cities

Dankan Gowda .V[1] , Arudra Annepu[2], Ramesha. M[3], Prashantha Kumar K[4] and Pallavi Singh[5]

[1]Department of Electronics and Communication Engineering, B.M.S Institute of Technology and Management, Bangalore, Karnataka, India.
[2]Department of Computer Science and Engineering, Rajiv Gandhi Institute of Technology, Bangalore, Karnataka, India.
[3]Department of Electronics and Communication Engineering, GITAM School of Technology GITAM(Deemed to be University), Bangalore, Karnataka, India.
[4]Department of Civil Engineering, NMAM Institute of Technology, Nitte, Udupi (District) Karnataka, India.
[5]Department of Electronics and Communication Engineering, Hindustan Institute of Technology and Science, Chennai, Tamilnadu, India.
Email: dankan.v@bmsit.in

Abstract. The pace of urbanisation has risen tremendously during the last few decades. To provide a higher quality of life, urban dwellers will require a greater variety of improved services and apps. The term "smart city" refers to integrating contemporary digital technology in the setting of a city to improve urban services. There are possibilities to create new services and connect disparate application areas with each other as a result of the use of information and communication technologies in the smart city. However, to make sure the services in an IoT-enabled smart city environment remain running without depleting valuable energy resources, all of the apps have to be maintained using energy resources that are kept at a minimum. IoT can enhance a city's lighting system since it uses more energy than other municipal systems. An intelligent city integrates lighting system sensors and communication channels with enhanced intelligence features for a Smart Lighting System (SLS). To control lighting more efficiently, SLS systems are built to be autonomous and efficient. We cover the SLS and evaluate several IoT-enabled communication protocols in this article. Furthermore, we evaluated several use scenarios for IoT enabled indoor and outdoor SLS and generated a report detailing the energy consumption in different use cases. By using IoT-enabled smart lighting systems, our research has shown that energy savings are possible in both indoor and outdoor settings, which is equivalent to a forty percent reduction in energy usage. Finally, we went through the SLS in the smart city research plans.
Keywords. Internet of Things (IoT), Internet of Lighting, Wi-Fi, Smart Lighting System (SLS), Communication, Interface.

1. Introduction

The phrase smart city is a fairly new word that has had a high rate of dissemination in the last few years. The arrival of the new paradigm has fostered cooperation among academia, industry, governments, and organisations, with people joining in as well. In [1], the authors use a smart city as an example of a well-defined geographical area, in which a range of technologies such as ICT, logistics, energy production, and more work together to help people achieve things like overall well-being, inclusion, and participation, while also ensuring that the environment is clean and healthy.
Nowadays, along with practical implementations, the smart city idea has been blamed for frequently being solely technology-driven, and pushed only by the interests of technology firms. Meanwhile, the

municipality and citizens have been given very little attention. As a result, this has necessitated a more sustainable methodology.

Sustainability has been well-established throughout time and enjoys widespread support. It is built on three essential elements: social well-being, environmental well-being, and financial well-being. Recently, a new definition [2,3] has been proposed to refer to a "sustainable city." It defines these cities as those that are able to absorb the inflow of materials and energy, as well as properly dispose of waste, without overextending the city's ecosystem. In other words, if a city wants to conserve its natural resources, then the amount of resources used inside the city should be equal to or less than the quantity of resources given by the environment (e.g., soil, water, or energy resources). Finally, since the city's activities have the potential to greatly impact the environment's ability to provide resources to citizens and other members of the ecosystem, pollution levels resulting from those activities should not overwhelm the environment's capacity to supply resources to citizens and other members of the ecosystem. While the idea of sustainability is quite basic and obvious, it has been attacked since in certain instances it does not align with contemporary societal trends, such as a rise in the amount of digital activity.

The ideas' development is therefore leading to a new wave of academic debates proposing a new paradigm: Smart Sustainable City. In further depth, this paradigm strives to create a "smart city" by concurrently considering urban sustainability and smartness. Consequently, understanding how to apply ideas such as these will influence the day-to-day activities of people. The information used to create this strategy is derived from the latest wave of technological progress, namely the increase in the number of IoT-enabled devices and entities.

The IoT is at the core of technological change and transformation in many situations and environments by creating and managing a network of linked devices that gather information about the physical world and modify their behaviour based on the ever-changing context in which they "live." Since IoT innovation is being introduced, smart sustainable cities will be able to enhance various elements of their urban administration, for instance, public transit, public lighting, e-governance, public safety, security, environmental monitoring, and mobility. The use of IoT technology is predicted to enable all of the available resources, including electricity, soil, water, people, and more, to be monitored, controlled, and managed [4,5]. Connection is crucial for organisations' ambitious goals; the most effective approach is to provide dependable connectivity to encourage efficient sharing of information. As such, due to the variety of different city scenarios that include varying communication technologies and network architectures, most instances necessitate heterogeneity of technologies and architectures, since they depend on the characteristics of specific services that need to be implemented or operational constraints, such as the availability of a power source. Connectivity, the backbone of a smart city, enables the implementation of services of interest to people and institutions.

2. Literature Survey

In order to build a smart city ecosystem, technology has to be a critical component as well as look at factors such as social and human capital. Most cities now use bespoke systems and solutions to meet their unique requirements, however these approaches are not appropriate for other cities across the world, and occasionally just a subset of different elements are required [6,7]. These results synthesise many of the literature's findings on smart cities and their major obstacles, problems, and open challenges. These four problems may be split into four distinct categories, and these categories can be described in more depth as citizens, mobility, government, and the environment. Smart cities should take into consideration the quality of life of the inhabitants while also factoring in the privacy issue, particularly when their personal information and household-level data are concerned. Because people may be concerned about the introduction of new technology, or see it as invasive, this is noteworthy. In addition, it is critical to focus on equality, which means that everyone in the community should benefit from improvements in smart city technology, and in particular, no metropolitan regions should be left behind [8,9]. The need for a change in government models stems from the wider idea of the smart city concept, which involves combining institutional policies with bottom-up initiatives in order

to be more flexible and to improve the strength of community relations, fostering collaboration, and promoting communication among various entities to prevent the formation of multiple similar initiatives, which would not work together the most efficiently [10]. Mobile city deployment includes the provision of a sustainable, inclusive, and efficient mobility system for both products and people. Still another domain of smart city design that has not been thoroughly researched, and thus may hold answers to previously unanswered questions, is the incorporation of the environment into city services. For instance, sustainable resource management (such as water and energy), pollution, and the impact of urban activities can all be explored. The Key areas to deal with in a smart city are shown in Figure.1.

Interoperability, as now deployments are more frequently based on private and isolated solutions, is currently thought of as a potential obstacle to smart city development [11,12]. To achieve affordable scale and maximise the outcomes, open standard-based devices must be utilised at all levels. In order to effectively coordinate data collecting and analytic operations across many systems, further coordination also is needed between the systems.

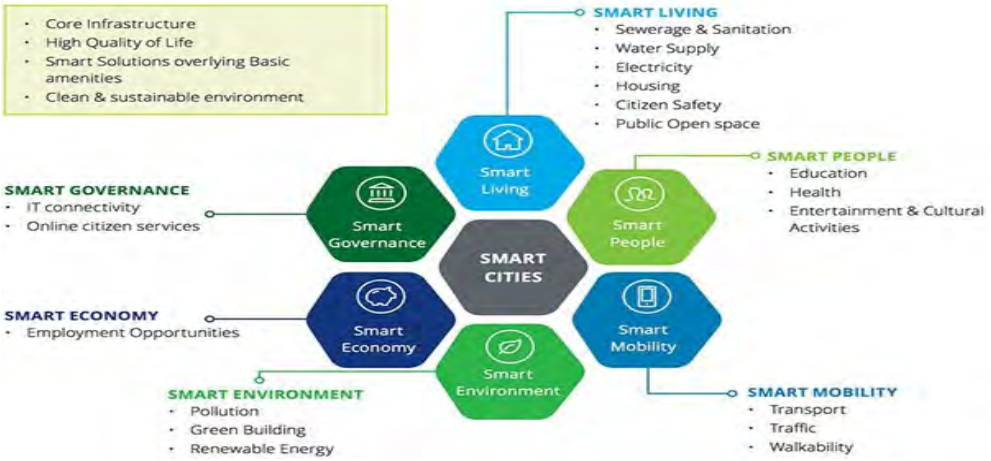


Figure 1. Key areas to deal with in a smart city [1].

3. IOT Communication Protocols Used In Smart Lighting Systems

Size and scalability are important in SLS. An SLS's main focus is on how these different components may logically interact. The LU and CC should be able to communicate data via an IoT communication protocol stack. Various IoT-enabled SLS communication protocols are found here. Using an SLS requires two different ways of communication. long-distance communication A LCSU and a CC unit are both examples of long-range communication in SLS. SLS is typically composed of multiple LCUs and a central CC. Following the LU study, local LCUs link the data to a CC. LCUs are also connected to exchange data. A few hundred metres to several hundred miles are CC distances since each LCU is accessible across the city. A long-range communication protocol is required to connect LCUs and CC. A variety of protocols are used to link LCUs to the CC. Short-range communication, or communication between devices within visual range, is a wide phrase. A SLS is a short range (under 100m) between LUs and connected LCUs. SLS uses short-range protocols to communicate between LCUs and LUs. Some short-range protocols (e.g., DALI [13,14]) are wired or wireless (e.g., ZigBee [15], JenNET-IP [16], 6LoWPAN [17]). This article will concentrate on IoT-enabled SLS communication protocols, since IoT use in SLS is growing. Hundreds of LUs are placed around the city in an IoT-enabled SLS, forming LCUs. A communication protocol should be able to communicate with a large number of LUs and LCUs while maintaining battery life, low cost, low data rate, and low complexity. Short-range IoT communications may use both wired and wireless methods. Wired solutions are less expensive than wireless solutions since they utilise existing infrastructure and do not need cables or complicated connections. suggesting wireless outdoors lights while favouring wired

inside lighting (for SiLS). Wireless networking standards IEEE 802.15.4 is the IoT wireless communication standard (Low Rate WPAN). PHY and MAC layers of the OSI Model have been successfully modified for low-power Internet of Things devices. Among the IEEE 802.15.4-based wireless protocols used for these activities are ZigBee, 6LoWPAN, and JenNET-IP. Each IEEE 802.15.4-based protocol has its own protocol stack. 1) 802.15.4-based protocols, especially ZigBee, are in high demand for IoT devices and apps. The ZigBee Alliance sets the standard. As a mature communication protocol for usage in low- and medium-range wireless sensor networks, the ZigBee Alliance has maintained its commitment to support, innovation, and development. 802.15.4 implements the MAC layer and provides additional services like encryption, authentication, association (only valid nodes may be connected to the topology), and routing protocol (AODV, a reactive ad hoc protocol). End device nodes are defined in ZigBee. The three flavours (colours) of SLU in a generic SLS are CC, LCU, and LU. To build a ZigBee SLS, the SLS must have one-to-one communication. "ZigBee is suggested by Leccese et al. Their approach uses XBee modules to implement the ZigBee protocol. 2) IPV6 over Low-Power Wireless Personal Area Networks (WPAN). 6LoWPAN is built on the Internet Protocol Suite for smaller IoT devices. 6LoWPAN provides versatility to SLSs through data transfer and control. 6LoWPAN data packets allow sensor data and control message transfer. This protocol stack acknowledges each successful packet delivery. Combining wired and wireless networks may significantly reduce 6LoWPAN installation costs [10]. In this way, 6LoWPAN outperforms other wireless protocols (for example, ZigBee, JenNET-IP). The benefits aside, 6LoWPAN is a basic Internet of Things application platform that connects existing sensor networks or other IoT devices through IP. 6LoWPAN allows developers to build new applications such as temperature control and weather monitoring. 3) Jennic created JenNET-IP, a wireless protocol stack based on IEEE 802.15.4. 6LoWPAN is an improved version of JenNET-IP. For CC-to-LUC communication, the SLS uses an IPV6 network. JIP and JenNet are new layers in 6LoWPAN and JenNet-IP. The JIP layer provides application-level device access. Control the system by stacking this layer on top of the preceding one. Developing apps that utilise this protocol layer allows for more data transmission. The JenNet protocol offers multi-hop capabilities [18,19]. JenNet is used to manage the network and safeguard outbound communications. Using JenNet-IP in the SLS allows for more nodes to be connected than other IoT-enabled protocols. JenNet-IP says the system can handle over 1,000 LUs, allowing for enormous network building. In addition, JenNet-IP, an upgraded version of 6LoWPAN, offers a sophisticated application development platform. An SLS may also create additional procedures. Z-Wave is a comparable IEEE 802.15.4 protocol to ZigBee. They are low-cost, low-power communication devices. In contrast, Z-wave allows for the creation of a mesh network rather than a single point-to-point link. The range of the devices is increased, allowing delivery even if the LCU fails. Z-Wave was private until 2016, however now that the specs have been made public, anybody may build their own Z-Wave device. In recent years, long-range wide-area networks like LoRaWAN have emerged. The LoRaWAN platform uses LoRa technology developed by the LoRa Alliance. Various protocols are discussed here, including Bluetooth Low Energy (BLE). BLE is best used for one-to-one communication, such as monitoring exercise equipment, computers, and peripherals. An inexpensive option for simple sensor networks, this network may accommodate a number of network topologies. This enables for many-to-many communication, which is unique to Bluetooth mesh. also known as wired protocols Indoor lighting (in homes, schools, and businesses) is part of a smart city's SLS. While wired protocols exist to link an SLS to a CCS, wireless methods should be used instead. DALI [21] and Power Line Communication (PLC) [20] are herewired methods for transmitting an SLS across short distances. PLI: Power line infrastructure is used for indoor and outdoor networking and communication. Using PLC in SLSs is intended to save costs by using pre-built networks. PLC-based lighting systems include two primary hardware components: a microcontroller and a PLC modem. A PLC microcontroller receives, processes, and transmits control signals to and from a PLC modem. The PLC modem modulates and demodulates data before to transmission to reduce the impacts of noise and interference. Serial connection allows for data transmission rates up to 500 Mbps between LUs. In a PLC modem with an RF transmitter, a

microcontroller receives data. PLC integrated in power wires transmits most messages and controls the whole lighting system. DALI is a lighting standard.

IEC standard, DALI uses a proprietary protocol to link lighting equipment through a bus or star network. Digital circuitry is used to set up an SLS for DALI. A Manchester-coded frame connects each LU in a DALI-enabled SLS. Sensors such as motion sensors and light sensors provide and monitor data for the command to regulate motion as well as the reaction to that command. The DALI-compatible devices must be connected to the two DALI terminals. DALI's 6-bit addresses limit it to 64 nodes. DALI benefits SLSs. DALI allows the control of devices from a variety of manufacturers. Because it does not need several processes for various goods, this light is more efficient. Less electricity is utilised, saving money. DALI's total LU capacity is 64, making it useless for street lighting. DALI data transmission speed and interoperability with wireless sensor networks have been improved by Yuan Ma et al. NRZ and MPE (Manchester Phase Encoding) are used at 9600 baud. DALI has created a new transfer layer with sensors to enhance the lighting system's utility.

4. Elements of Smart Lighting System

Sensors are the most prevalent, followed by algorithms, with everything else in between. Lighting control systems may evaluate the day, light spectrum, or occupancy to decide the final reaction. Algorithms may operate inside devices or systems to manage workloads or tasks given to them. They may also be operated on the cloud, eliminating the need to transmit command messages [22]. Algorithms may refer to many cutting-edge technological solutions that constantly shift colours, such as tunable lights, techniques that control colour response, real-time colour adjustments, and real-time techniques that help reduce energy use. Circadian cycles are often used to create aesthetically complex lighting patterns. The initial lighting design schematic presented in Fig.2 represents the main components of the design. Rather than following rigid input design requirements, autonomous algorithms are taught to react to user choice and gender. The biological clock that controls our circadian cycles, as well as numerous other systems, including as hormone release, body temperature, and circadian awareness, have all been shown to be influenced by lightweight in the last decade. Since circadian cycles rely on luminescence rather than colour correction, the spectrum of light frequencies from red to blue is more significant. Because the system may affect and control many physiological characteristics, expanding can also imply gaining power. The museum benefits from complete spectrum management in gardening, fine arts, and public gathering spaces, among other disciplines.

In a network, the physical and logical layers interact at the system device level and the device hierarchy. Using different physical topologies, such as a loop, stars, or a combination of both star and bus, increases reliability and opens up expansion options. Traditional communication networks may be placed over wireless ones utilizing cables or wireless connecting covers in physical installations. You may connect to the network through wired or wireless connections.

Lighting products have been linked with IoT networking technologies to better serve a broader variety of applications. 0-10V, DALI, Digital Multiplexer (DMX), Local Area Network (LAN), and Power Line Communication (PLC) are the primary wired interfaces used for networking. They also utilise wireless technologies like as infra-s GHz, Bluetooth, and infra-shred optical Lumina[23]. 6LoWPAN, for example, uses a physical layer focused on short range and rapid network connection.

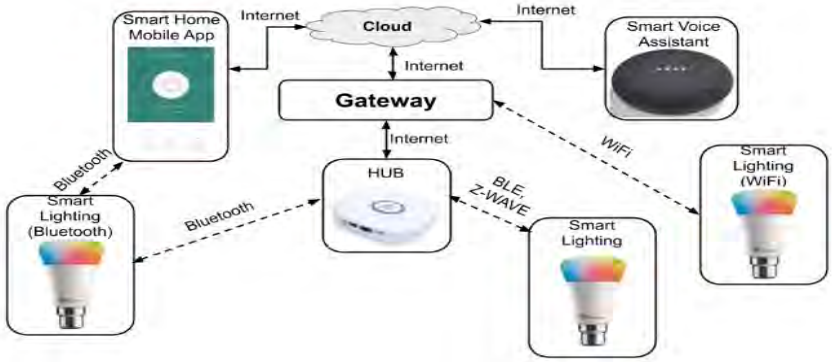


Figure 2. Key elements for smart lighting systems [2]

5. Sensors for Smart Lighting Platforms

A wide range of working sensor technologies and communication techniques is what makes a smart lighting device excellent, if not what keeps it inexpensive. In contemporary IoT applications, digital sensors are used to alter lighting mechanisms to aid in adaptive operation[16,17]. If you know about low-intensity light sensors and photodiodes, can they alert you when it becomes dark? Red, green, and blue sensors are used for LED and CFL recolour (fluorescent) lighting to detect their primary colour, producing RGB material for indoor environments and optical connections, but for wireless applications, visible light communication (VLC) is the most important photo diode works technology.

Creating white light by mixing individual red, green and blue LEDs as shown in figure.3. If you know about photo resistors and photosensitive cells that react to lesser amounts of light to alter the luminous flux depending on user activity. Spectroscopy is a unique and creative characteristic that allows us to utilise these devices to collect light in the visible spectrum.

Figure 3. Concept of colour mixing individual red, green and blue LEDs [3]

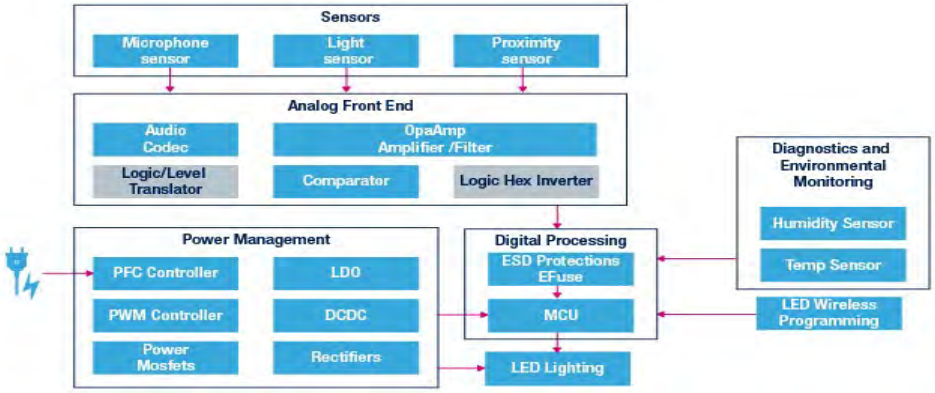


Figure 4. Sensor technologies embedded in smart lighting [4]

Since any type of light spectrum and any point-range is generated by a circadian lighting device, the control ability is critical to tuning the Color Rendering Index (CRI) and correlated color temperature (CCT) in real time is crucial. Sensor technologies embedded in smart lighting is presented in figure.4. As well as optimizing the lighting efficiency, it is known that sensors and controls which use LED technology may lose Luminescence levels increase with age and/ decrease over time.

6. Communication Interfaces for Smart Lighting

A professional lighting system must have universal connection, either wired or wireless. DALI, PLC, and Ethernet are required for critical infrastructure and street lighting. Non-critical infrastructure lighting applications and systems may utilise Wi-Fi, ZigBee, or a commercial traffic control system. For various reasons, these indoor LED lighting systems prefer IEEE 802.15.4, VLC, Bluetooth low energy, or sub-GHz protocols. Both ZigBee Light Link and 6WAN utilise the IEEE802.15.4 Media and Super Access Point (AP) Layers.15 Most of them have their own code names (encoding standards) to indicate how secure they are [20,21].

Manufacturers may build systems to connect with goods using simplified standards, data centre management, or automation (IPv4 and IPv6). Although based on the IoT platform, the Open AIS project (from Europe) is creating a framework for diverse lighting systems. A standardized framework for lighting interface and extendable APIs enable the light system to be utilized in a broad variety of building systems and independent of specific cloud services.

The answer to this interoperability problem is to standardize lighting protocols, which are helpful in situations where open ecosystems with conflicting protocols are acceptable and closed ecosystems are unfeasible. In this instance, utilizing lightweight protocols like Universal Plug and Play (UDP) and TCP may extend compatibility.

7. Future Research on Smart Lighting System

As the IoT-enabled smart lighting system continues to grow, the rate of development is increasing. This part details open issues connected to the execution of smart city initiatives and the security of SLSs. A lot of issues still need to be fixed in order to further increase the efficiency of SLS. This has been around for ages. Connecting all of an SLS's components with an IoT-enabled protocol is very necessary. The Low-power Wide-Area Network (LoRaWAN) makes it possible for low-data-rate connection to be established over large geographical areas, with many different IoT devices. Several SLSs, each using a different protocol, need to communicate in order to create a centralized lighting system. Looking at an issue from many angles To link various application domains, the concept of a Smart City is essential. SLS systems allow new services to be delivered to urban regions, making SLS more versatile. In conjunction with SoLS, low-cost autonomous solutions may be made available to traffic managers via smart traffic management. Additionally, weather systems using environmental sensors, including rain sensors, temperature sensors, and humidity sensors, may be used in SLS settings. Municipal services that use SLS may be less expensive and more efficient if used with other applications. System security system As they are centrally controlled, an attacker may target SLSs because this gives them access to other connected services. This also granted the attacker complete control over the city's lighting system, which might lead to even more severe attacks or allow them to totally rule the city. Sensors may be used to alter behavioral patterns that have been anticipated. With the implementation of future smart lighting standards, future researchers will confront difficulties. Unfortunately, at this time, there is no trustable method for granting and cancelling keys. A privacy mechanism may be employed to safeguard the user's privacy, but a complete security system may not be utilised. While low-power and low-cost end devices are available, security and efficiency will be sacrificed for them. A large and rigorous security system may slow down the functioning of the system and result in an increased installation cost. Implementation of inadequate security measures may have catastrophic results.

8. Conclusion

Energy efficiency is a critical problem in an IoT-enabled smart city setting. This is a serious issue, given the anticipated population growth in urban regions over the next few decades. We have devoted this article to discussing IoT-enabled Smart Indoor and Outdoor Lighting Systems (SiLS, SoLS) in the context of a smart city, where energy consumption may be reduced and operations made more intelligent via the use of sensors and actuators. In terms of power consumption, connection, and reliable administration, a variety of Internet of Things-enabled communication protocols may be utilised to construct a successful smart lighting system. Finally, we computed and provided the power consumption for SiLS and SoLS in a variety of use cases and situations. Energy consumption in indoor and outdoor settings may be decreased by up to forty percent when IoT-enabled SLS is used instead of conventional lighting systems. Finally, we addressed the benefits of SiLS and SoLS, as well as research difficulties for those who are interested in furthering their study in these areas.

References

[1] K. Sikder, A. Acar, H. Aksu, A. S. Uluagac, K. Akkaya and M. Conti, "IoT-enabled smart lighting systems for smart cities," *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, 2018, pp. 639-645, doi: 10.1109/CCWC.2018.8301744.

[2] N. Adnan, N. Kamal and K. Chellappan, "An IoT Based Smart Lighting System Based on Human Activity," *2019 IEEE 14th Malaysia International Conference on Communication (MICC)*, 2019, pp. 65-68, doi: 10.1109/MICC48337.2019.9037601.

[3] F. J. Montalbo and E. Enriquez, "An IoT Smart Lighting System for University Classrooms," *2020 International Symposium on Educational Technology (ISET)*, 2020, pp. 3-7, doi: 10.1109/ISET49818.2020.00011.

[4] Mattoo A., Saxena K., Kumar S., Bagwari N. (2021) "Design of IoT-Based Smart Illumination System in Smart Cities". In: Abraham A., Castillo O., Virmani D. (eds) Proceedings of 3rd International Conference on Computing Informatics and Networks. Lecture Notes in Networks and Systems, vol 167. Springer, Singapore. https://doi.org/10.1007/978-981-15-9712-1_44.

[5] Sharma A., Deopa A., Ahuja A., Goel V., Nath V. (2020) Internet of Things Based Smart Lighting System Using ESP8266 and Zigbee. In: Nath V., Mandal J. (eds) Nanoelectronics, Circuits and Communication Systems. NCCS 2018. Lecture Notes in Electrical Engineering, vol 642. Springer, Singapore. https://doi.org/10.1007/978-981-15-2854-5_34.

[6] Praveen Kumar R., Smys S., Raj J.S. (2021) Ingenious Lighting System (ILS) for Smart Cities Using IoT. In: Raj J.S. (eds) International Conference on Mobile Computing and Sustainable Informatics. ICMCSI 2020. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-49795-8_14

[7] D. V. Kishore, Shivashankar, A. C. Ramachandra, and C. Pandurangappa, "Optimization of motorcycle pitch with non linear control," in *2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016 - Proceedings*, 2017, pp. 1656–1660, doi: 10.1109/RTEICT.2016.7808114.

[8] Kishore D.V, Shivashankar, and S. Mehta, "MANET topology for disaster management using wireless sensor network," in *International Conference on Communication and Signal Processing, ICCSP 2016*, 2016, pp. 0736–0740, doi: 10.1109/ICCSP.2016.7754242.

[9] A. C. Ramachandra, M. N. Thippeswamy, C. Pandurangappa, and P. Ramesh Naidu, "Modelling and performance evaluation of anti-lock braking system," *Journal of Engineering Science and Technology*, vol. 14, no. 5, pp. 3028–3045, 2019.

[10] M. Penna, J. J. Jijesh, and Shivashankar, "Design and implementation of automatic medicine dispensing machine," in *RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings*, 2017, vol. 2018-Janua, pp. 1962–1966, doi: 10.1109/RTEICT.2017.8256941.

[11] A. C. Ramachandra, M. N. Thippeswamy, C. Pandurangappa, and P. Ramesh Naidu, "Synthesis and modeling of antilock braking system using sliding mode controller," *Journal of Advanced Research in Dynamical and Control System*, vol. 10, no. 12, pp. 208–221, 2018.

[12] C. A. Varun, Shivashankar, M. Sahana, R. S. Varun, and T. Rajesh, "Implementation of swarm intelligence in obstacle avoidance," in *RTEICT 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Proceedings*, 2017, vol. 2018-Janua, pp. 525–528, doi: 10.1109/RTEICT.2017.8256652.

[13] S. B. Sridhara, K. B. Naveen, M. Ramesha, and G. N. Pai, "Internet of things: Internet revolution, impact, technology road map and features," *Advances in Mathematics: Scientific Journal*, vol. 9, no. 7, pp. 4405–4414, 2020, doi: 10.37418/amsj.9.7.11.

[14] P. Ramesh Naidu, N. Guruprasad, "Design and implementation of cryptcloud system for securing files in cloud," *Advances in Mathematics: Scientific Journal*, vol. 9, no. 7, pp. 4485–4493, 2020, doi: 10.37418/amsj.9.7.17.

[15] M. Ramesha, Sridhara S.B, Naveena Pai G, "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*, vol. 8, no. 7, pp. 3893–3897, 2020, doi: 10.30534/ijeter/2020/158872020.

[16] P. Ramesh Naidu, N. Guruprasad, "A High-Availability and Integrity Layer for Cloud Storage, Cloud Computing Security: From Single to Multi-Clouds," *Journal of Physics: Conference Series 1921(1), https://doi.org/10.1088/1742-6596/1921/1/012072*

[17] S. B. Sridhara, M. Ramesha, "Recent advances in graph theory and its applications," *Advances in Mathematics: Scientific Journal*, vol. 10, no. 3, pp. 1407–1412, 2021, doi: 10.37418/amsj.10.3.29.

[18] Ramesha, M., Sridhara, S.B., Pai, G.N., Patil, S.K., "Design of Antilock Braking System Based on Wheel Slip Estimation," *Journal of Physics:Conference Series 1706(1),2020, https://doi.org/10.1088/1742-6596/1706/1/012216.*

[19] Pai, G.N., Sridhara, S.B., Shashidhara, K.S., Gangadhara, "Signal Analysis and Filtering using one Dimensional Hilbert Transform," *Journal of Physics:Conference Series 1706(1),2020, https://doi.org/10.1088/1742-6596/1706/1/012107.*

[20] M. Ramesha, K. Jeevan and B. M. Sathisha, "Implementation of IoT Based Wireless Electronic Stethoscope," *2020 Third International Conference on Multimedia Processing, Communication & Information Technology (MPCIT)*, 2020, pp. 103-106, doi: 10.1109/MPCIT51588.2020.9350476.

[21] Puneeth Kumar M. V, Naveena Pai G, Vinay Kumar B. C, "Dynamic Analysis and Control Strategies of an Anti-lock Braking System," *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2020, pp. 1677-1682, doi: 10.1109/ICECA49313.2020.9297642.

[22] G. Naveena Pai, M. Swathi Pai, M. Shruthi and B. Naveen K, "Internet of Things: A Survey on Devices, Ecosystem, Components and Communication Protocols," *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2020, pp. 611-616, doi: 10.1109/ICECA49313.2020.9297458.

[23] M. Ramesha, Jeevan K.M, "Study and analysis of bted error correction codes for cryptography applications," *International Journal of Advanced Trends in Computer Science and Engineering,* vol. 9, no. 5, pp. 8938–8942, 2020, doi.org/10.30534/ijatcse/2020/293952020.

Smart Villages: IoT Technology Based Transformation

Smart Villages: IoT Technology Based Transformation

Pankaj Mudholkar[1], Megha Mudholkar[2], Puneeth Kumar B S[3], Dankan Gowda V[4] and S. Srinivasulu Raju[5]

[1],[2]Assistant Professor, Department of Master of Computer Applications.
Thakur Institute of Management Studies, Career Development and Research, Mumbai, Maharashtra (State), INDIA.
[3]Assistant Professor, Department of Computer Science.
St.Joseph's College (Autonomous), Bangalore, Karnataka (State), INDIA.
[4]Assistant Professor,Department of Electronics and Communication Engineering.
B.M.S. Institute of Technology and Management, Bangalore, Karnataka (State), INDIA.
[5]Department of Electronics and Instrumentation Engineering.
Velagapudi Ramakrishna Siddhartha Engineering College,Vijayawada, , Andhra Pradesh(State), INDIA.

E-mail: researchr08@gmail.com

Abstract. In recent years, large-scale urbanisation has been on the rise with cities being the hubs for growth, but rural business continues to play an important role in any country's overall development. A recent report reveals that almost 69 percent of India's population resides in their villages, accounting for almost 50 percent of the nation's GDP. This agricultural region varies from small towns with less than 500 residents to small towns. There are also similar situations in western nations. Despite their economic contribution, smaller villages seem to earn fewer in terms of infrastructural expenditure. The major problems addressed in these small towns include shortage of adequate public transit, emergency care and limited knowledge on federal subsidies for rural areas. Our vision is to expand the Smart city to Smart village by allowing use of recent technical advances and giving more attention to the problems in rural areas. The approach in this chapter is to bring IoT technology to Villages by literally showing a network of linked sensors and knowledge dissemination devices, controlling energy use and ensuring infrastructure protection. It gives a wide-ranging vision of enhancing the standard of living in villages and encourages them to meet the essential needs of domestic villagers.

1. Introduction

There are many obstacles in the realization of a rural development that incorporates and tracks all of the village infrastructure and services to their fullest potential and seeks to harness the collective intelligence In order to benefit from the Internet of Things (IoT), systems, which involves the creation of a cloud-based network that can provide virtual process and storage as well as well as monitoring and visualization software, it is essential to also be developed. the integration of Information Technology (IT) and Operaional Technology (OT) objectives in the realm of energy management, aimed at using smart billing and data analytics Enhanced cloud-based IoT services seem to increase the system's efficiency by allowing the management of different collection plans and routes to be created on the fly and across the day as they

are required. providing residents with access to environmentally friendly and affordable energy services acts as a key driver of intelligent community services and renewable resources serves as a basic tool for accomplishing those objective[1] and also the most of rural areas is an integrated systems; basically, this means that most parts of rural America are non-decentralized and/ donas an effect, most of rural America is non-centralized These clusters provide opportunity for growth, as well as having economic benefits. Rurban is based on creation of these initiatives which offers infrastructure growth-inducing opportunities in rural communities, as well as fostering their socio-economic growth. For each cluster, you should be able to foresee the following additional components: This chapter looks to deal with the transportation problems first in our healthcare system, expands on the sanitation problems, considers the problem of an expanded education system, describes methods for garbage management, and finalizes our highway issue lighting control,Fleet management, Mostly, which expands the economy by incorporating agricultural and food processing, agri-food supply and distribution, agricultural supply and distribution mainly IoT-enabled smart village can help to reduce costs by increasing performance, allowing more assets to be utilized, and generating a greater return on process resources. Utilizing sensors and communications that allow it to update and assess the data in real time and information and real-on-enabled analytics tools enables smart decisions to be implemented for sensor trackers. What is being said about the internet's expansion is being the growth and integration of data, systems, and with this, it, opportunities for individuals, companies, and possibly even industries will be increased. Information Technology and infrastructure would be necessary for rural growth, though financial investment is necessary for the process. thousands of networking equipment and computing devices are installed in this complex network, as well as numerous sensors are being used in it with all these pieces of the equipment on real time, there will be an increase in operational and repair costs because it is vital to achieve excellent reliability and efficiency preferably sensors and reliable data control units are placed in every single irrigation field to ensure proper information for the application of intelligent irrigation systems, broad area data network and IoT technology an application of the influence of more sophisticated tools and technologies in the community which caters to targeted people and issues with added functions that lead to dependable data transmission. By focusing on conventional farming and turning it into an IoT business opportunity, it enables the latter and creates new opportunities for both product and technology growth for the IoT industry. The chapter presents a forum and methodology for conducting intelligent agricultural ecosystem monitoring in a sensors-based system that assists with cultivation-dependent research. It is made up of three subsystems, the first of which is the GSM module, the sensor units, and the M2M Cloud Computing More than two-thirds of India's population resides in rural areas, rural life predominates in India[2]. It is necessary for India to focus on villages in order to provide a central communication system, but because of the communication issue, India's development would be hindered. For this reason, villagers know nothing about Ration status, the government's scheme and funds, and those who have raised dairy cattle or goats do not understand why they have not received funding, and those who haven't received funding can't tell their neighbors, Applications on and knowledge from which they are still only provided are able to come out, can expand into the system of villagers and help the problem be overcome and find solutions for it. a smartphone app that functions as a platform for farmers to introduce new farm workers and register new farm-workers The above-mentioned android framework is a centralized management system[3]. Most of the Indians live in a small towns and villages, which means that it is imperative to include their views in order to really understand India. Now that people in rural areas are discovering how helpful cell phones can be, it is important to show them that these are modern times and there are applications that will benefit them. Information technology is a very simple to communicate to adjust their views as well as the idea of a Smart Village, as a character. An intelligent community expects individuals to know their role and demonstrates how to take charge of it

by speaking about it. The amount of capital needed in order to meet the demand for digital needs for everyone in our country is considerable, and therefore a large number of people unable to meet that demand[4]. This chapter was formed after months of research where they are realized the advantages of smart phones and how they can help people who live in rural areas It is vital to have smart people in a community because they plan their various facets of the community through their imagination and relationships with the rest of the population. It is difficult to identify a large number of intelligent people in Indian villages, so a handful of intelligent individuals have to be found for each group. They are need to be responsible for furthering the education of future generations.

2. NEED FOR SMART VILLAGE

By this, the author means each country has contributed to the advancement of a city's global reputation as a project to build a local knowledge-based economy in the largest metropolitan areas[5,6]. To a great extent, rural areas are in need of critical infrastructure including highways, running water, and electric power. An effort was made to build world megacities that are more related to one another, however, these cities struggled to provide much attention to the large population that was currently residing in them. Often for the benefit of small towns and rural areas, rather than the entire nations, developing and growing cities should expand rural communities. There would be better employment opportunities for youths in the country and thus discouraging rural-to-based youths from moving to the city. The most significant goal of the reform will be to remunerate farmers[7,8] and work with them to maximize potential compensation, along with mentoring and helping them to learn the ways to improve their long-term financial position, as an occupation. The class chairmanship chairmanship extends coverage such as crop insurance, soil health, and environmental practices to the base of the organization, as well as he or pesticide subsidies. In this context: Vending machine (in the village) neighborhoods: ensure economic sustainability and embrace cultural sensitivity in developing neighborhoods and also face the same problems because of direct entry to the global market has been an issue with several layers of intermediaries and a shortage of qualified workers. Still unable to change the overall economic condition of a large villages, even though people are trying to focus on enhancing their wealth and improving their access to basic necessities. Most of the rural area people don't have a facility to access the daily basic needs like electricity and a fresh water supply, including irrigation. In order to succeed here, a number of strategies can be used: make improvements to people's innate abilities, including, though not limited to, using technology to augment these skills Make sure to have adequate digital and informational Technology (IT) literacy a major initiatives, as well as connect to the digital market for sure Expanding local development allows for sustainable energy resources to be made available, such as access to educational and healthcare facilities, a decent and secure living, water use, waste water avoidance, and gender equality, and a power to be attained, thus paving the way for economic growth. The IoT is an association of networking devices which is woven together through the web. It consists of mechanical gizmos, sensors, machines, cars, and so forth apart from the workplace. Portables[9,10] as well as PCs are included in this category. Some of these devices are designed so that they can transmit information to various other devices on the web-capable gadgets. Anything that has an Internet connection (gadgets) and a sensor now has a possibility of forming a network. In other words, IoT is commonly referred to as the IoT. The technique of interfacing various physical objects to the web is more and more commonly known as Middleware Interface Matching (MIMO). The concept of web of a IoT was introduced by Kevin Ashton in the year 1999, and referred to as such A society full of social network with each member knowing their own unique address has been defined as a IoT, because each thing in society is more connected to others now than ever. The items found in IoT identified as articles have the ability to gather and transport data through the IoT[11,12] capable of measuring and

moving information that are distinct from that found in person or cars have this capability. It is by and far majority rule commonly utilized in all kinds of different business organizations. it motivates them to work together, showcasing the upgraded management to a client, and increasing efficiency while conveying the upgraded administration to the company, which also strengthens the client's organizational vision and strategy. In general, it has evolved from microservices, distributed and MEMS (microelectromechanical systems) innovation [13]. People interacting with a machine solely by automated means (or through using computer programs) without human interferences are moving from M2M (machine-to-to-to-machine expansion to M2M expansion. The unit is already primed to expand into the cloud and ready to take on knowledge Devices are the goal of the IoT allowed. The specific use cases discussed in Fig. 1 are the standard function of these products and systems is made clear in the expression "embedded everywhere," which has been outlined.

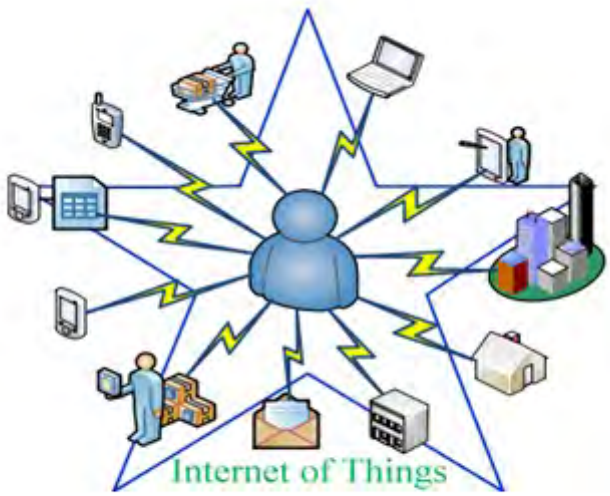


Figure 1. Internet of Things [7]

Various conditions have been used to get information about the point (and retrieve it) such as sensors, processors, and mechanisms to use in processing and delivering the data The components that are part of the IoT entry point or gateway are responsible for reporting and accessing data are also part of the network that exchange and collect it. This feature can be expanded to include the integration of other features [14] as well, though it may not directly follow on from the other features. Autonomous devices doodian devices only operate when they are working on their own, without help from the operator. They can pair their devices with the computers to make setting up and gathering data simple and straightforward. There are three distinct but complementary processes. These are to collect information, organize information, and summarize information, and then look at it, analyze information gathering and activities.

2.1. Collecting and Sending the Information

So far, this information has been captured and transmitted through the grid. This has largely been finished thanks to the use of sensors. With regards to moisture, one notices that various sensors have been used for example, some claim it appears to be a temperature sensor, others say it is a moisture sensor, even others that it's a movement sensor, and so on. Once these

sensors are attached, this collects and summarizes data from a variety of different objects to give an overall impression of the decision making process and then brings about a conclusion. in addition, for instance, on the chance that ranchers can discover information about the dampness on their homestead, they will know that water should be applied to it when their premises has enough water, ranchers may make decisions about when to do it and how much water to apply. it causes the ranchers to ensure that the water is provided in an appropriate measure for their cattle as opposed to offering too little to the homesteads and buildings Making effective use of available resources is an important part of the problem-solving process, so setting objectives and keeping them firm [15,16] is also helps to protect them from being wasted. The use of this device enables ranchers to do more and give back to the community, both economically and for the world's people. The sensor makes it possible for the machines to sense the environment.

2.2. Receiving and Acting

Once you have gathered the data, it also expands on it to allow you to keep a constant eye on trends and development To our knowledge, the computers are also very remarkable in their capabilities. We'll find out as soon as the information is accessible, they are able to expand on it. For example, the printer obtains the data and afterwards, the data is sent to the printer to be printed. when the data is tracked and traced, the complexity of the network grows. There is no human intervention required to gather the information with sensors, as long as the sensors are used. The expand task includes getting and delivering data, collecting data, following up on the data, and compiling the results.

3. COMPONENTS OF IoT

3.1. Sensor

The sensor is the film that associates the IoT contraption to the surface condition or individual. As the name proposes, it recognizes the progressions and sends information to the cloud for handling. These sensors persistently gather information from the environment and communicate the data to the resulting layer. A model likes pressure sensors, temperature sensors, light force indicators. Gateway: Gateway Promotes stream the executives and convention layer to move information starting with one gadget then onto the next. It deciphers the system conventions for gadgets and gives encryption to the information streaming in the system [17]. It resembles a layer among cloud and gadgets that channel away from the digital assault and unlawful access to information.

3.2. Cloud

IoT frameworks send gigantic information from gadgets and this information should be overseen effectively to create significant yield. To store this tremendous measure of information, the IoT cloud is utilized. It gives apparatuses to gather, procedure, and store information. Information is promptly accessible and distantly available through the web. It likewise gives a stage to examination. IoT cloud is an advanced elite system of workers to perform rapid handling of gigantic measure of information.

3.3. Analytics

Analytics is the way toward changing over crude information into some important structure. IoT analysis supports continuous research, which grabs continuous changes and irregularities[18]. The information is then transferred over into an organization that is straightforward by the end client. Clients or business can investigate the patterns appeared in reports, foresee the market, and plan ahead for effective usage of their thoughts.

3.4. User Interface

The User interface is a noticeable, substantial piece of IoT frameworks. It is that piece of the framework that connects with the end client. The data can be made accessible either in report design or as certain activities like trigger an alert, a notice, and so on. The client can likewise decide to play out certain activities. It is imperative to make an easy to understand interface that can be utilized absent a lot of exertion and specialized information. The simpler the UI is the more effective the item.

Figure 2. Componets of IoT

4. CHARACTERISTICS OF IoT

After experiencing the Introduction to IoT, presently we will find out about the attributes. Attributes are the purpose for the accomplishment of IoT[10]. There are six principle qualities. Every trademark includes a lot of abilities that make IoT a conquest.

4.1. Intelligence

IoT frameworks are widely used in the market as a result of their knowledge. A blend of calculations and PC empowers the framework to advise a change in condition and take fitting activities. For instance Systems are sufficiently astute to detect an abrupt spike in temperature and trigger an alert for fire.

4.2. Connectivity

Connectivity is the principle normal for IoT as it empowers the framework to send information and remain associated with different gadgets. It gives a framework to arrange availability and capacity cooperatively.

4.3. Expressing

IoT is tied in with connecting insightfully with the external condition and people. Communicating empowers this intelligence. Communicating permits us to show yield into this present reality and contribution from individuals and the earth.

4.4. Sensing
Sensitivity implies mindful of the progressions around us. Sensor advancements furnish us with the way to make the experience that mirrors attention to changes in the physical world and individuals in it. It helps in communicating[19]. This structures the contribution of the IoT framework and gives a superior comprehension of the unpredictable world around us.

4.5. Energy
Everything in this world is driven by vitality. IoT frameworks are made savvy enough to combine vitality from the external condition and monitor it. It is likewise made vitality effective to work for a more extended term.

4.6. Security
Wellbeing and security is the most significant element of any framework. If the framework isn't make sure about to digital assault and illicit mediation, no one will utilize it. IoT frameworks manage individual information that is the reason it's a commitment that all security measure ought to be taken consideration in this framework. All IoT frameworks are sufficiently secure to manage individual information.

5. APPLICATIONS OF IoT
There are various applications of the IoT including buyer-provided IoT, startup Internet Things, manufacturing Internet Things, and mechanical uses. Additionally, these applications can be mobile, electric, vitality, and other. smart-savvy homes and hot-blooded computers, warm-willed appliances can be aided by PCs and mobile phones (PC and phone regulated ventilation and climate) Clients will be able to have sensors with which to collect and analyze their own information, and which will transmit the data to additional developments, in order to make their lives simpler. These are just a few of the models that may include being sophisticated, being shrewd a vehicle that is meant to make money, a program that creates revenue, or even being wearable. Due to the fact that the systems are well-designed, the costs are kept low and temperatures can fluctuate, making life in the region more bearable[20]. an extremely efficient cultivating and extensive growth-regulating system enable growers to discover and control whether or not specific traits, such as mugginess, dampness, and yields, are developing This encourages the talented people to aspire to be enlightened, clever meters, as well as city infrastructure such as streetlights and disinfection.

5.1. Scope
Its extent is wide and enormous in this day and age as everything is associated with the Internet. It associates the gadgets in the different frameworks to the Internet and the gadgets or articles speak to themselves can be controlled from anyplace. It helps in accomplishing more information and spots, more methods of expanding the effectiveness, and improving wellbeing and security. It causes the association to expand the presentation through IoT investigation and security to accomplish great outcomes. There are numerous associations like oil and gas, protection, producing, transportation, foundation, and other retail parts that can get advantages from it and some areas of now accomplishing the advantages from it. The IoT stage, for the most part, empowers the gadget or article to an onlooker and to recognize and comprehend the circumstance without getting the assistance of individuals or some other human intercession. In this way, the extent of the IoT is incredible later on and its previously demonstrating the outcomes also.

5.2. IoT Security
The IoT is associated with billions of gadgets with web and an immense number of information focuses are engaged with gathering, moving, and sending the data. Because of its long structure

and extended surface, the security and protection of IoT are the immense worries for the various associations. As these gadgets are firmly associated, the assailant can misuse one weakness to control the entire information in one go and the gadgets that are not refreshed normally are the primary explanations behind these sorts of assaults. Other individual data like age, address, and Mastercard subtleties, and so on are likewise being given by the client when the gadgets are associated with one another by means of different methods. These gadgets are acquired to sell the clients very own data. It likewise represents the hazard to foundation, power, transportation and budgetary help, and so on.

5.3. Audience for Learning

The right audience for learning these is engineers, venture chiefs, as well as others who value creativity, entrepreneurs, would be individuals who are into business specialists, as well as business investigators who're eager to discover IoT. The challenge for entrepreneurs is to gain awareness of new ideas that are appearing on the market, how these ideas interconnect, and use this knowledge to bring their own invention to the world and gain an added value. It lessens the amount of manual work and time it takes. Although this requires additional resources to implement, there are fewer variables that can easily be accounted for in the costs of expansion. To be sure, there are many steps to be taken. The IoT include proficiency with the ability to acquire internet friends and acquaintances through an easy-to-use medium, whether it be online or not [21]. The process of developing recordings into objects is highly effective, but there is still some use that can be made of the recorded materials. Workers are drawn to joining professional associations because they expect things to change and to be developed through networking with their peers and also because these relationships help their company, according to research. All inside work and plans have been approved and will be carried out by the organizations in support of their members and customers to locate the criticality of the IoT.

6. THE IDEA OF SMART VILLAGE

People now live in big cities in greater than in proportion to their population percentage of the population. This may be the reason that both researchers and governments focus on building sustainable cities, since the idea is that they have their resources with them, and are efficient enough to remain so Using sound and prudent fiscal planning, these cities may apply their resources in a useful and well-directed manner The same concept can be applied to the rural towns as well. The economy of the majority of which is dependent on agriculture is rural is like in India, in which the population has a strong growth potential Compared to the people in the big cities, life in small towns is not as easy, and it's generally harder in villages as well[13,14]. When you work to help the growth of the cities and villages, you have to also strive to improve the quality of life for the people who reside in them. With a certain amount of effort, specific concepts from smart city can be put into small towns. Even, as an example, the use of cameras and sensors in streets, such as traffic surveillance and healthcare among others However, on the other hand, it is necessary to develop such solutions and introduce new methods of use, including cattle/livestock farming, which can work only with simple ideas. Various aspects of the villages are considered in the subsequent sections; next will consider how well villages are doing with the use of IoT and the quality of life, leading to a detailed look at Smart Village practices. To create a Smart village, one must first determine all the artifacts that interact with each other and next identify the possibilities for integrating them. Then sensors, cameras, switches, and other devices will be placed that are much larger and serve other purposes, such as fixed alarm buttons will be added[15]. This array of sensors and related equipment will be connected to the internet, which will generate a massive amount of data that can then be stored and processed in the cloud servers. Even after taking care of all the logistical issues, this data can be mined for finest details with Big Data analytics like Hado. At the end of the day, the goal is to achieve

smart homes, climate control, security, and class management are also sought after.

6.1. Smart Buildings

Smart homes and buildings uses sensors and cameras to track and collect information about all aspects of life within them. Constant data will be generated, which means you will have constant feedback about what is happening with your project. for example, sensors can be found in a house, whooshing the sprinklers to fight fire, if they sense smoke is nearby, and starting to extinguish. Also, the sensors track energy use and disable and light switches allow home and building automation to be turned off when there is no longer a need for them. Access to a greater energy capacity is the critical need in rural areas where you can't have power at all times. According to the existing security measures, the building's security can be closely monitored with cameras and corrective measures can be implemented to counter any irregularities. Water levels and pressure can be assessed in the water tanks and used to refill them when there is a problem.

6.2. Smart Weather and Irrigation

Making sure that accurate weather information is accessible to the villagers is an asset. The majority of rural populations subsist on a steady-state subsistence economy, in which agriculture is their main source of income. In theory, the use of environmental sensors helps farmers in many ways, but it does not fully and completely describe the weather since sensors can only detect conditions and not predict it. a large number of farming operations, including planting, watering, and harvesting all rely on the weather. With the ability to use sensors in the fields and to measure the flow of water, it is possible to implement effective irrigation methods that allow full use of available resources. If it's going to rain the next day, it is highly recommended that the crops be watered on an extra day to provide them with sufficient water. Every bit of information on farming techniques, from environmental conditions to economic forecasts, can be supplied to farmers via cell phones. proximately gaugedometers in the water-related projects and for example, water levels in the dams and reservoirs can be measured to identify how much the water projects may need to expand.

6.3. Smart Farming

As the primary production is a fundamental necessity for any community, farmers must be able to reap the greatest value from the systems of IoT and Smart village. In order to have complete food traceability, the produce needs to go from the field to the table. sensors and other data can be used to keep track of anything from the whole activity sequence of activities. Many who are interested in the process are the growers, purveyors, business people, packers, warehousers, and transport firms, as well as well as the dealers, transporters, in addition to retailers. can help the farmers make informed decisions on what type of crops to plant, the sensors can monitor the germination progress of each week in the fields, and guide farmers using sensors to decide what type of application methods, including row-specific or nutrient-drip-dependent, should be used, can advise on the need for water application, and perform yield predictions based on the soil or climate With remote sensor and satellite imagery, it is possible to predict which kinds of diseases and pesticides the farmers are most likely to have in their fields. Their farmers' cell phones is state of the art in order to keep them up on what's new in agriculture. if the device identifies an emergency, an emergency, a safe procedure may be automatically followed and take appropriate action[16]. Caleb understands this because his father and grandfather farmed wheat for many decades before the time of mechanization, but also because he was the victim of an earlier example which has been previously discussed. For example, a small sparks can set entire farms on fire, which makes bigger farmers very angry. There are environmental sensors that can sense smoke at the start of a fire and automatically begin to flow water to extinguishing

it before it can get out of control. This also pertains to sensors, which can detect ripening of vegetables and fruits and the moment they are ready to ship, helping the produce companies prevent delays before they occur. Once you've chosen the produce to be sold, the appropriate locations in the market, suitable plans should be developed to facilitate sales.

6.4. Smart Dairy

Dairy has enabled secondary activity, and business to increase. Having the use of sensors and cameras in the barn or herdery will help farmers in better supervising their work. Any possible variation is shown as a warning message and any identified changes are reflected by procedures are brought into play. Smart devices may be used to regulate the animals' temperature in the ideal range. As in the above statement, it is possible to track the food, water, and health necessities of cattle in the same way. You run the risk of someone or something getting sick if you graze livestock in the open fields without an adult. Sensors in the fields will enable farmers to cut their staff numbers by eliminating the need for human oversight, which means they can operate more efficiently.

6.5. Smart Healthcare

Additional healthcare and medical services are required in the rural areas in order to raise the quality of life. The community pharmacies and healthcare facilities in the dispensaries need to be connected to each other in order to supply their requirements. in order to allow the sensors to record various medical devices to communicate with the patient, movement, tissue wounds' rate of circulation, and body temperature, among other things, are typically installed in the patient's beds. Multiple machines including X-rays, CT scans, ultrasound, and other reporting devices send data to the doctor instantly expand on these reports to become more detailed. such basic services will contribute to the overall health of the health of the villages.

6.6. Smart Surveillance System

As the population densities are lower in rural villages are more remote, the police stations are far away, and less crime is being reported, security is likely to be less effective Because of these reasons, small communities demand better surveillance is needed in villages. Additionally, sensors and cameras are going to help monitor the area surveillance and relay data, which will be helpful in identifying the entire village locations where emergency situations may arise. During a burglary or robbery, you can use the alarm buttons in case of which send out a signal to the police department nearest to you to warn the others of the situation. The cameras will assist in catching the thief after they record and reveal his/demonstrate where he stole the item. to ensure that events like this don't occur in the future, data can be processed to include additional variables.

6.7. Smart Education

A more general way of expressing the importance of education as a means to implement all of development It is easier to introduce new innovations as more people are aware of them. This phenomenon has an inverse relationship with population density: In dense cities, built-up areas, digital divide seems to exist more than in villages and small towns. Smart village are built around its components, and this construction involves a serious commitment on the part of the people living occupants. Leveraging members may become stakeholders, which means that they may take part in any endeavor in the community, enrich its economy, and also lead the villagers' well-being. Dealing with children and adolescents becomes less of a struggle when we approach education in an exciting and interactive way. Video games have the benefit of engaging children because they're more likely to learn in an interactive way than a lecture style, but lectures are

still great for helping them to absorb information. A IoT is made up of various technologies like the Internet, smartphones, as well as intelligent devices, which assists in the process of education. The use of LCD screens and video games will promote children's learning and appreciation for things, as well as encourage interaction with others who are adults. The homes and businesses in the Smart village can be used to teach those in the villages about a variety of skills, such as about how to use the services and systems, tools and how to solve problems. Classrooms in the local villages can be provided with Internet access to new and the computers and school can then serve as a more interesting and engaging educational experience.

7. Conclusion

Back in the day, people used to regard the IoT and Smart city as two very speculative ideas, the latter of which was envisioned as the possible future. Today's rapid growth in the use of technology, however, has unfortunately, meant that it is becoming a reality. This chapter is written to provide solutions for the small towns and villages thus reflecting the various requirements found in these other types of communities. Interests in various aspects of this process have been investigated, as well as several suggestions have been given. Various advancements in fields of technology have made it necessary for various services to be provided with a sophisticated delivery. To be competitive, demand response, new equipment should be installed in modern locations that represent current priorities in rural areas. Making the irrigation easy and accurate while minimizing water consumption is part consumption will be the focus of the effort in expanding the use of the automated system. Waste-related technology can also presents the population and changing lifestyle needs, where an Expanded phrase: When the population and use increases, cloud-based resource management and hygiene technologies have the capability to be put to use in waste disposal. In addition to this project's efforts to alleviate poverty, which are also focused on environmental issues, it has been recognized for boosting the rural economy by using different techniques. To meet the basic design requirements, consider all possible energy resources, instruments, tools, and the entire financial economics of the system must be taken into consideration.

References

[1] M. A. Alzain, B. Sohć: *MCDB: Using Multi-clouds to Ensure Security in Cloud Computing.* , 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing, Sydney, NSW. (2011), 784–791.

[2] D. Agrawal,ć, El Abbadić: *Database Management as a Service: Challenges and Opportunities, Data Engineering.* , IEEE 25th International Conference . (2009),1709–1716.

[3] R. Buyya,ć, C. S. Yeoć: *Cloud computing and emerging IT platforms: Vision hype and reality for delivering computing as the 5th utility.* , Journal of Future Generation Computer Systems.25(6) (2009), 599–616.

[4] P. Rathod,ć, S. Sapkalć: *Audit service for data integrity in cloud.* , International Journal of Advanced Research in Computer Science and Software Engineering.4(4) (2014), 288–292.

[5] M.Pennać, J.J. Jijeshć: *Design and implementation of automatic medicine dispensing machine.* , 2017 International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT). (2017), 1962–1966.

[6] Ramesh Naidu.Pć, Guruprasad. Nć: *A High-Availability and Integrity Layer for Cloud Storage, Cloud Computing Security: From Single to Multi-Clouds.* ,Journal of Physics: Conference Series. (2021), 1921(1), 012072.

[7] Ramesh .M, S. B. Sridharać: *Recent advances in graph theory and its applications.*, Advances in Mathematics: Scientific Journal. **10**(3) (2021), 1407-1412.

[8] Ramesh .M,ć, S. B. Sridharać: *Design of Antilock Braking System Based on Wheel Slip Estimation.*,Journal of Physics: Conference Series. (2020), 1706(1),012216.

[9] Pai G.N,ć, S. B. Sridharać: *Signal Analysis and Filtering using one Dimensional Hilbert Transform.*,Journal of Physics: Conference Series. (2020), 1706(1),012216.

[10] Ramesh .MĆ, Jeevan K.MĆ: *Implementation of IoT Based Wireless Electronic Stethoscope.* , MPCIT 2020 - Proceedings: IEEE 3rd International Conference on "Multimedia Processing, Communication and Information Technology. (2020),103106, 9350476.
[11] Pai.G.NĆ, Kumar P.MĆ: *Dynamic Analysis and Control Strategies of an Anti-lock Braking System.* , Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020. (2020),16771682, 9297642.
[12] Pai.G.NĆ, Swathi PaiĆ: *Internet of Things: A Survey on Devices, Ecosystem, Components and Communication Protocols.* , Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020. (2020), 611616, 9297458.
[13] Ramesha .MĆ, Sridhara S.BĆ: *Study and analysis of bted error correction codes for cryptography applications*, International Journal of Advanced Trends in Computer Science and Engineering. **9**(5) (2020),89388942.
[14] V. SahanaĆ, C. A. VarunĆ: *Implementation of swarm intelligence in obstacle avoidance.* , 2017 International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT). (2017), 525–528.
[15] Ramesha .MĆ, Sridhara S.BĆ: *FPGA Implementation of Low Power High Speed BTED Algorithm for 8 Bit Error Error Correction in Cryptography System*, International Journal of Emerging Trends in Engineering Research. **8**(7) (2020), 3893–3897.
[16] D. V. KishoreĆ, S. MehathaĆ: *MANET topology for disaster management using wireless sensor network.* , 2016 International Conference on Communication and Signal Processing (ICCSP). (2016), 0736–0740.
[17] Ramchnadra A.CĆ, D. V. KishoreĆ: *Optimization of motorcycle pitch with non linear control.* , 2016 International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT). (2016), 1656–1660.
[18] Thippe swamy M.NĆ, A.C. RamachanrdaĆ: *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.
[19] Thippe swamy M.NĆ, A.C. RamachanrdaĆ: *Modelling and Performance Evaluation Of Anti-Lock Braking System* , Journal of Engineering Science and Technology. **14**(5) (2019), 3028–3045.
[20] Ramesha.M, S. B. SridharaĆ: *Internet of Things: Internet Revolution, Impact, Technology Road Map and Features.*, Advances in Mathematics: Scientific Journal. **9**(7) (2020), 4405-4414.
[21] Ramesh Naidu.P, N. GuruprasadĆ: *Design and Implementation of CryptCloud System for Securing Files in Cloud.*, Advances in Mathematics: Scientific Journal. **9**(7) (2020), 4485-4493.

SCOPUS INDEXED CONFERENCES

2021-22

Big EEG Data Images for Convolutional Neural Networks

S. Thundiyil[1], M. Thungamani[2], S.A. Hariprasad[3]

1. Department of ECE, BMS Institute of Technology and Management, Bangalore, India
2. Department of Computer Science, CoH, University of Horticultural Sciences, Bangalore, India
3. Faculty of Engineering and Technology, Jain Deemed to be University, Bangalore, India
saneesh@bmsit.in, thungamani_k@rediffmail.com, sa.hariprasad@jainuniversity.ac.in

2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB) | 978-1-6654-2897-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/SPMB52430.2021.9672272

An open image database of Electroencephalogram (EEG) plays a vital role in developing deep learning algorithms for conducting research in EEG signal processing. This abstract presents the generation of "Big EEG Data" images for researchers to develop and test deep learning algorithms, especially using convolutional neural networks. The image database created in this work uses the Temple University Hospital (TUH) Electroencephalography Seizure Corpus (TUSZ) [1]. We have used 'TUSZ v1.5.2' database for the creation of the images [2].

An EEG signal in its general form is considered as a time series data where the temporal information and spectral information play a crucial role in understanding the underlying characteristics of the signal. There are several deep learning techniques proposed for the analysis and classification of time series data. Apart from the prominent architectures in recurrent neural networks, researchers propose novel algorithms to address time-series data classification and prediction issues [3]–[5]. The studies suggested by Lotte et al. [6] reveal that the researchers are applying deep learning methodologies to detect and classify epileptic seizures. However, there is a need to improve the accuracy. Convolutional neural networks are very well matured to classify images into multiple classes. In this work, we propose to use the convolutional neural networks for the classification of EEG signals. The challenges here are the time series representation of the EEG signal itself. It may be noted that standard time-series deep learning architectures such as LSTM, RNN, etc., can work directly on the time series data, whereas CNN cannot be directly used on such data. Hence, we suggest converting the EEG signal to a set of images by extracting three characteristics of time-sliced EEG signals. The methods proposed in this abstract do not directly represent EEG time-domain signal in the image format, but the image representation of the values obtained while performing three transformation techniques as explained below.

The conversion of time-series data such as EEG into an image is performed by obtaining properties such as mutual information and correlation between a set of successive samples in EEG signals and by obtaining the discrete cosine transform and hence spectrogram of EEG signals. Obtaining the mutual information and correlation coefficient values are done after converting the EEG signal of long duration into short-duration signals through time slicing based on ECG rhythms.

The methodology adopted to convert the time series data into images is illustrated in Figure 1. The steps involved in the creation of an image database are:

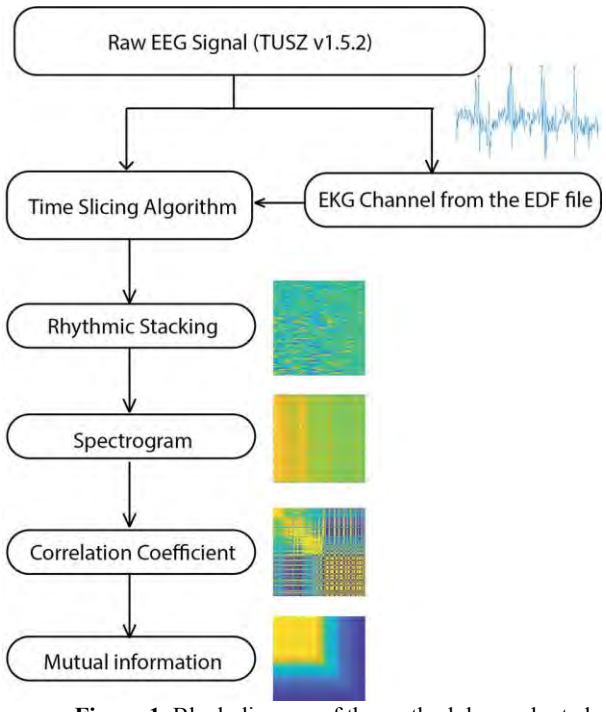


Figure 1. Block diagram of the methodology adopted

1. Time-slicing of EEG data based on ECG rhythms;
2. EEG image creation using rhythmic stacking;
3. Images based on mutual information available in the successive block samples;
4. Images based correlation information of time slices;
5. Images based on spectrogram information of time slices.

The underlying principle of our approach is the time-slicing of the EEG recordings based on the R peaks of the ECG signal. Different studies by Ako.M et al. [7] and Bhavsar et al. [8] provide an insight into the correlation between EEG activities and heart rate variability. Hence a time-slicing approach based on *'r peaks'* was chosen to study the EEG patterns. The first algorithm makes the simple stacking of EEG patterns based on the ECG rhythms. The EEG recording available in the TUH Seizure Corpus is in the European data format (edf). Most of the recordings in this corpus contain 21-32 channels per recording. In our algorithm, we have separated the ECG channel by extracting the label EKG-REF or equivalent label from *'edf'* file. Later the sample index corresponding to *'r peaks'* was obtained after finding the peaks. Since the time duration of all the channels is the same for one recording, we have sliced the EEG channels based on the *'r peaks'* of the ECG waveform. Each such slice is now stacked together to form image data. The slicing procedure and sample of a resultant image is shown in Figure 2. To make the size (width) of the image the same, we introduced zero padding if there is a mismatch in the length of each segment due to variation in R-R interval.

In the second algorithm, we used the entire channel recording without time slicing to obtain the spectrogram images. We use one image for each channel as indicated in Figure 3(a). The spectrogram uses the Short Time Fourier Transform, which is a useful transformation technique since it preserves both time and frequency domain characteristics [9]. The total number of images generated is based on the number of channels available in each *'edf'* file.

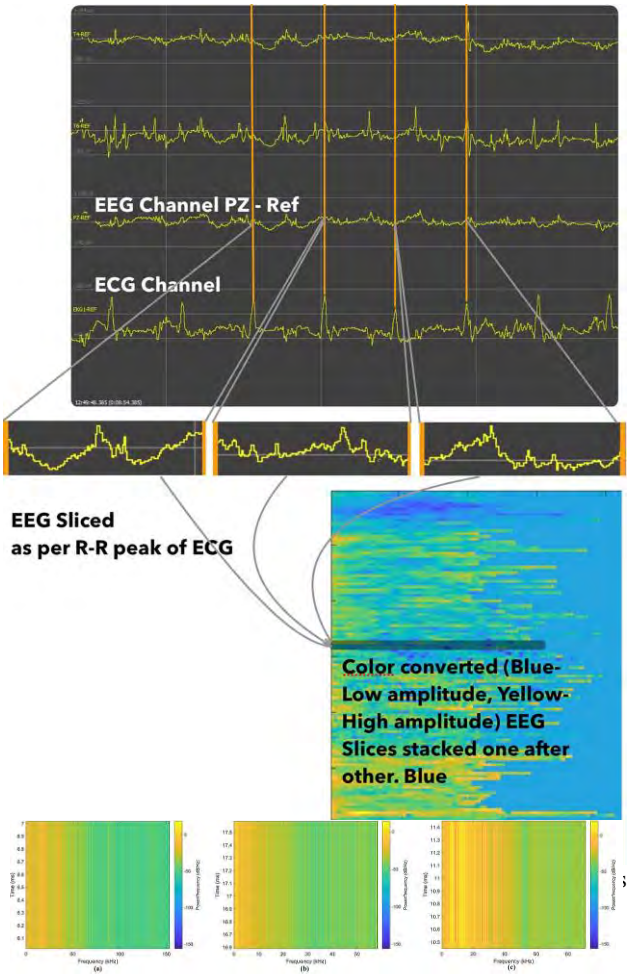


Figure 3(a). Sample images of spectrogram of entire channel

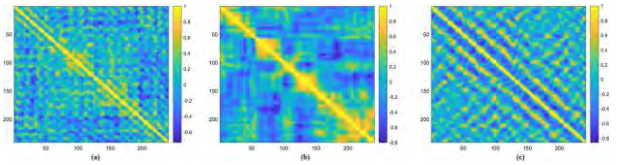


Figure 3(b). Sample images based on correlation coefficient between every EEG slices

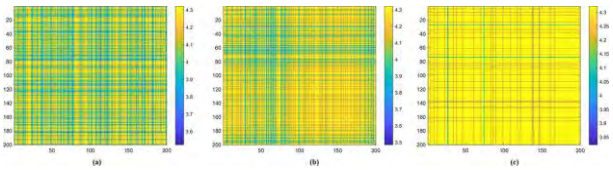


Figure 3(c). Sample images based on mutual information between every EEG slices

In the third algorithm, the correlation coefficient between each time-sliced segment was obtained. Cross-correlation is a technique that estimates the correlation between two signals. Cross-correlation accounts for time delays by shifting one of the two signals [10]. In our approach, the correlation is calculated between two slices, and a matrix is created using the correlation values between each slice. Hence for one channel,

if the number of slices is M, we get M correlation coefficient values for each slice. We get a total of $M \times M$ matrix of correlation coefficients. This matrix is further converted to image format. Hence for each recording, we got the number of images equivalent to the number of EEG channels. Similarly, in the fourth algorithm, the mutual information between every slice of each channel is obtained [11]. The mutual information matrix corresponds to each channel is converted to images. Figure 3(b) and Figure 3(c) shows the sample images generated using correlation coefficient and mutual information based algorithms.

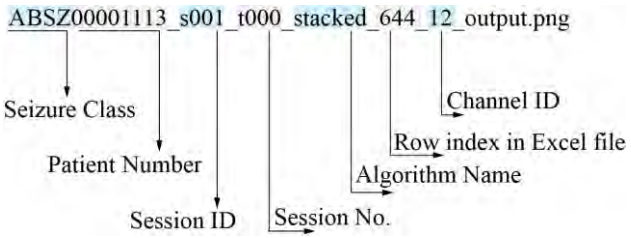


Figure 4. Description of filename

Table 1. Number of images generated under various classes

	Seizure Category	Stacked images		Spectrogram Images		Correlation based images		Mutual information-based images	
		Train	***Dev***	***Train***	***Dev***	***Train***	***Dev***	***Train***	***Dev***
Image Data Bank 1	FNSZ	26179	2286	42577	4679	2627	1029	232	32
	GNSZ	6415	1644	10964	1592	356	438	121	104
	SPSZ	1179	96	1298	93	151	130	92	34
	CPSZ	4119	256	8797	1111	488	258	164	117
	ABSZ	1056	192	1573	192	84	100	257	111
	TNSZ	279	95	486	589	248	498	215	99
	TCSZ	494	409	900	506	273	348	278	101
	MYSZ	64	32	64	32	64	32	125	10
	Total	**39785**	**5010**	**66659**	**8794**	**4291**	**2833**	**1484**	**608**
Image Data Bank 2	FNSZ	1600	1568	672	1568	1582	-	704	-
	GNSZ	7250	2680	6416	2622	3001	-	526	-
	SPSZ	38155	9156	38565	8787	32515	-	950	-
	CPSZ	11199	4880	9052	4751	7741	-	786	-
	ABSZ	64	-	-	32	64	-	64	-
	TNSZ	1462	96	1277	93	1433	-	117	-
	TCSZ	907	550	420	537	907	-	315	-
	MYSZ	540	1408	485	1364	540	-	247	-
	Total	**61177**	**20338**	**56887**	**19754**	**47783**	**0**	**3709**	**0**

All the images obtained from these four algorithms have the same naming conventions as indicated in Figure 4 except for the algorithm name field. We have used complete TUEZ v1.5.2 resources for the creation of two master databases. In the first database, all the images were formed by implementing all four algorithms to entire EEG recordings irrespective of the seizure occurrence. The second database is created by implementing all four algorithms to the portion of the recording where seizure is identified. This was done by extracting the starting and ending time of occurrence of seizure. The details about the time record are provided in the documentation available in the TUSZ corpus.

Table 2. Folder names corresponds to each class

Folder Name	Event Name
FNSZ	Focal Non-Specific Seizure
GNSZ	Generalized Non-Specific Seizure
SPSZ	Simple Partial Seizure
CPSZ	Complex Partial Seizure
ABSZ	Absence Seizure
TNSZ	Tonic Seizure
TCSZ	Tonic Clonic Seizure
MYSZ	Myoclonic Seizure

Table 3. Training options selected for Alexnet architecture

Solver	sgdm
Initial learning rate	0.0001
Validation frequency	50
Epochs	30
Minimum batch size	128
L_2Regularization	0.0001
Gradient Threshold Method	L_2Norm
Momentum	0.9

The number of images generated against each category is indicated in Table 1. The images created are categorically saved to different folders, corresponding to eight different classes, as shown in Table 2. The top-level folder contains the training and development folders named exactly as per the TUSZ naming conventions. Each of these folders has four subfolders which is named based on the image creation method adopted, such as Stacked, Spectrogram, Cross_Correlation, and Mutual_Information. Each of these folders contains seven subfolders which is named as per Table 2. Each of these folders is a seizure class. The typical filename in each of these folders is in the format as mentioned in Figure 4. The patient number, session ID, and session number fields are the same as in the original EEG file available in the EDF format in the TUSZ database. Using the same naming convention makes it easy for researchers to compare the deep learning algorithms with works that

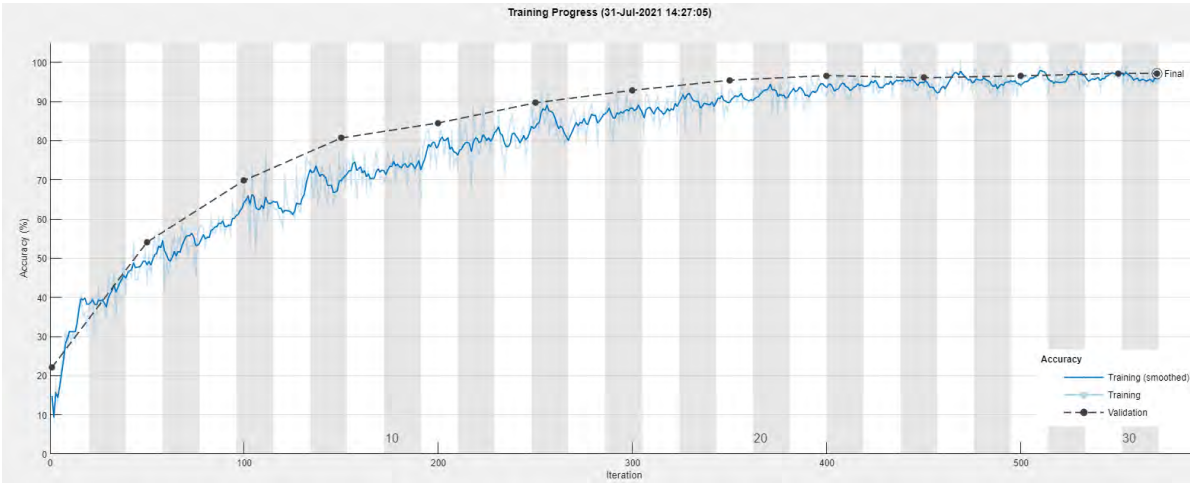


Figure 5. Training Progress of mutual information based images with AlexNet architecture

directly use the TUSZ data. The images are cropped to remove the axis information which is generated while implementing the algorithms.

Authors have tested these images for classification accuracy using convolutional neural networks architectures. Since images pertaining to each class is available in separate folders, standard Alexnet architecture, Resnet18 architecture, and GoogleNet architecture were used for testing the classification accuracy [12][13]. We have implemented the architecture for multiclass problems (7-Class) by changing parameters in the fully connected layer of CNN architecture, such as the number of output classes. As indicated in Table 1, the number of images per class is not balanced. This is due to the insufficient recordings available in the corpus for a particular type of seizure (e.g., Myoclonic Seizure). Due to the imbalance in the data, other than validation accuracy, we have calculated F1-Score and Specificity. The values of training options (hyperparameters) used for training the Alexnet architecture are listed in Table 3. Similar training options were provided for the other two architectures as well. Figure 5 shows the training progress upon training Alexnet architecture with mutual information based images. The confusion matrix obtained for the same database (mutual information based images) are shown in Figure 6. Similarly, all four image databases were tested with three CNN architectures. The results obtained in each case are tabulated in Table 4.

The performance of these methods is compared with similar works performed on the TUH database by other authors. Golmohammadi et al. used gated recurrent units (CNN/GRU) and LSTM models on TUH Seizure data for

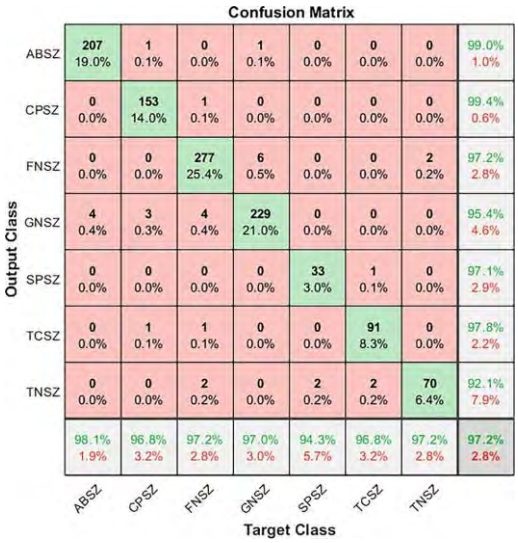


Figure 6. Confusion matrix for mutual information image based classification

Table 4. Performance parameters

Spectrogram Images (AlexNet)	F1 - Score	97.47
	Accuracy	97.15
	Specificity	99.51
Correlation Coefficient based Images (ResNet-18)	F1 - Score	93.64
	Accuracy	93.45
	Specificity	98.88
Mutual information based Images (ResNet-18)	F1 - Score	97.95
	Accuracy	97.89
	Specificity	99.62
Stacked Images (ResNet-18)	F1 - Score	95.64
	Accuracy	95.50
	Specificity	99.24

classification and achieved specificity of 97.1 % and 91.49 %, respectively [14]. L. Wei and A. Mooney presented the XGBoost-based method to detect seizures from TUH Corpus. They have achieved an accuracy of 67.01% while training and 58.85% during validation [15]. For the abnormal dataset from the TUH EEG corpus, S Roy et al. applied 1D-CNN-RNN and could achieve an accuracy of 82.27% [16]. Yildirim et al. proposed a 1D CNN model to classify the normal and abnormal EEG signal from the TUH EEG corpus and achieved an F1-Score of 78.92 and an accuracy of 79.34% [17]. In several performance parameters, our proposed methods have shown better results for the training dataset.

In this abstract, we have discussed the generation of an image database from TUSZ v1.5.2 using four techniques. The EEG time-domain signal extracted from the *'edf'* file is converted into images based on four properties of the time-sliced EEG signal. The image database is tested for classification accuracy using convolutional neural networks. The highest accuracy of 97.89% has been achieved when trained ResNet architecture with images based on mutual information. The accuracy achieved is for seven classes. The image database serves as a valuable resource for training deep learning networks.

REFERENCES

[1] A. Harati, S. Lopez, I. Obeid, J. Picone, M. P. Jacobson, and S. Tobochnik, "The TUH EEG CORPUS: A big data resource for automated EEG interpretation," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium (SPMB),* Dec. 2015, pp. 1–5, doi: *10.1109/spmb.2014.7002953.*

[2] V. Shah *et al.*, "The Temple University Hospital Seizure Detection Corpus," *Front. Neuroinform.*, vol. 12, p. 83, 2018, doi: *10.3389/fninf.2018.00083.*

[3] P. Lara-Benitez, M. Carranza-Garcia, and J. C. Riquelme, "An Experimental Review on Deep Learning Architectures for Time Series Forecasting," *International Journal of Neural Systems*, vol. 31, no. 3, p. 213001, 2021, doi: *10.1142/S0129065721300011.*

[4] Gamboa, John Cristian Borges. "Deep learning for time-series analysis," 2017, *arXiv preprint,* arXiv: *1701.01887.*

[5] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Min. Knowl. Discov.*, vol. 33, no. 4, pp. 917–963, 2019, doi: *10.1007/s10618-019-00619-1.*

[6] Lotte, Fabien, et al. "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of Neural Engineering* 15.3, 2018, 031005, doi: *10.1088/1741-2552/aab2f2*.

[7] M. Ako *et al.*, "Correlation between electroencephalography and heart rate variability during sleep," *Psychiatry Clin. Neurosci.*, vol. 57, no. 1, p. 59—65, Feb. 2003, doi: *10.1046/j.1440-1819.2003.01080.x*.

[8] R. P. Bhavsar, "Assessing variability of EEG and ECG/HRV time series signals using a variety of non-linear methods," University of Hertfordshire, 2020, doi: *10.18745/th.22617.*

[9] Y. Yuan, G. Xun, K. Jia, and A. Zhang, "A multi-view deep learning framework for EEG seizure detection," *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 1, pp. 83–94, 2019, doi:*10.1109/jbhi.2018.2871678*.

[10] Tsiouris KM, Pezoulas VC, Zervakis M, Konitsiotis S, Koutsouris DD, Fotiadis DI. A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals. *Comput Biol Med.* Aug 2018,vol. 99, pp. 24-37. doi: *10.1016/j.compbiomed.2018.05.019*.

[11] Ince, Robin AA, et al. "A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula," *Human brain mapping* 38.3, 2017, 1541-1573, doi:

10.1002/hbm.23471.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105. Available: *https://dl.acm.org/doi/10.5555/2999134.2999257*.

[13] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778, doi: *10.1109/CVPR.2016.90*.

[14] M. Golmohammadi *et al*., "Gated recurrent networks for seizure detection," *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2017, pp. 1-5, doi: *10.1109/SPMB.2017.8257020*.

[15] L. Wei and C. Mooney, "Epileptic Seizure Detection in Clinical EEGs Using an XGboost-based Method," in *2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2020, pp. 1–6, doi: *10.1109/SPMB50085.2020.9353625*.

[16] S. Roy, I. Kiral-Kornek, and S. Harrer, "Deep Learning Enabled Automatic Abnormal EEG Identification," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 2756–2759, doi: *10.1109/EMBC.2018.8512756*.

[17] Ö. Yildirim, U. B. Baloglu, and U. R. Acharya, "A deep convolutional neural network model for automated identification of abnormal EEG signals," *Neural Comput. Appl.*, pp. 1–12, 2018, doi: *10.1007/S00521-018-3889-Z*.

Big EEG Data Images for Convolutional Neural Networks

S. Thundiyil[1], M. Thungamani[2], S.A. Hariprasad[3]

1. Department of ECE, BMS Institute of Technology and Management, Bangalore, India
2. Department of Computer Science, CoH, University of Horticultural Sciences, Bangalore, India
3. Faculty of Engineering and Technology, Jain Deemed to be University, Bangalore, India

ABSTRACT

- An open image database of Electroencephalogram (EEG) plays a vital role in developing deep learning algorithms for conducting research in EEG signal processing.
- This abstract presents the generation of "Big EEG Data" images for researchers to develop and test deep learning algorithms, especially using convolutional neural networks.
- The image database created in this work uses the Temple University Hospital (TUH) Electroencephalography Seizure Corpus (TUSZ). We have used 'TUSZ v1.5.2' database for the creation of the images.

NEED OF IMAGE REPRESENTATION

- An EEG signal is generally represented as a time-domain. The temporal information and spectral information play a crucial role in understanding the underlying characteristics of the signal.
- It may be noted that standard time-series deep learning architectures such as LSTM, RNN, etc., can work directly on the time series data.
- Convolutional neural networks are very well matured to classify images into multiple classes. In this work, we propose to use the convolutional neural networks directly for the classification of EEG signals. The challenge is that CNN cannot be used on time series data.
- We suggest converting the EEG signal to a set of images by extracting three characteristics of time-sliced EEG signals.

TIME SERIES TO IMAGES

- The conversion of time-series data such as EEG into an image is performed by obtaining properties such as mutual information and correlation between a set of successive samples in EEG signals and by getting the discrete cosine transform and hence spectrogram of EEG signals. Obtaining the mutual information and correlation coefficient values are done after converting the EEG signal of long duration into short-duration signals through time slicing based on ECG rhythms.
- The methodology adopted to convert the time series data into images is illustrated in Figure 1.

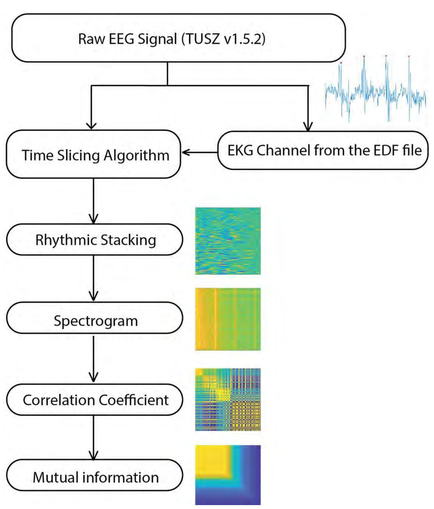


Figure 1 . Steps involved in timeseries to image conversion

SEGMENTATION OF EEG SIGNALS

The underlying principle in this approach is the time-slicing of the EEG recordings based on the R peaks of the ECG signal. The ECG channel is separated by extracting the label EKG-REF or equivalent label from 'edf' file. Later the sample index corresponding to 'r peaks' was obtained after finding the peaks. Since the time duration of all the channels is the same for one recording, we have sliced the EEG channels based on the 'r peaks' of the ECG waveform.

Each such slice is now stacked together to form image data. The slicing procedure and sample of a resultant image is shown in Figure 2.
To make the size (width) of the image the same, zero-padded the sequence if there is a mismatch in the length of each segment due to variation in R-R interval. The first algorithm created this image database which is a direct representation of amplitude values in the color-coded form.

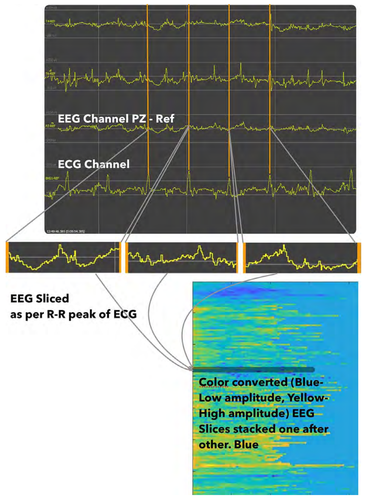


Figure 2. Slicing of EEG based on ECG rhythms

IMAGE CREATION USING SPECTROGRAM, CORRELATION COEFFICIENT AND MUTUAL INFORMATION

In the second algorithm, the entire channel recording without time-slicing is used to obtain the spectrogram images. The spectrogram uses the Short Time Fourier Transform, which is a useful transformation technique since it preserves both time and frequency domain characteristics. The total number of images generated is based on the number of channels available in each 'edf' file. In the third algorithm, the correlation coefficient between each time-sliced segment was obtained. Cross-correlation is a technique that estimates the correlation between two signals. Cross-correlation accounts for time delays by shifting one of the two signals. In our approach, the correlation is calculated between two slices, and a matrix is created using the correlation values between each slice. Hence for one channel, if the number of slices is M, we get M correlation coefficient values for each slice.

We get a total of M×M matrix of correlation coefficients. This matrix is further converted to image format
Similarly, in the fourth algorithm, the mutual information between every slice of each channel is obtained. The mutual information matrix corresponds to each channel is converted to images.

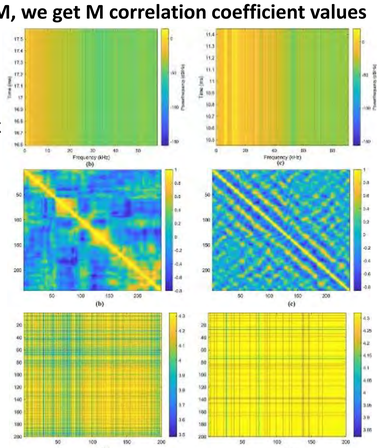


Figure 3 . Sample images generated based on three algorithms

EEG IMAGE DATASTORE

The number of images generated against each category is indicated in the table below. The images created are categorically saved to different folders, corresponding to eight different seizure classes, as shown in Figure 4.

The typical filename of each image is in the format mentioned in Figure 5. The patient number, session ID, and session number fields are the same as in the original EEG file available in the EDF format in the TUSZ database. Using the same naming convention makes it easy for researchers to compare the deep learning algorithms with works that directly use the TUSZ data. The images are cropped to remove the axis information which is generated while implementing the algorithms. Total number of images generated in the current version is listed in Table 1.

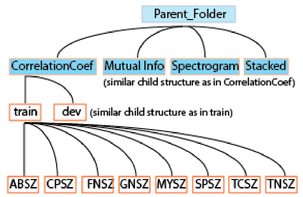


Figure 4. Structure of image folder

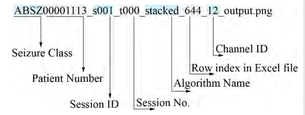


Figure 5. File name information of each image

Image Data Bank	Seizure Category	Stacked images		Spectrogram Images		Correlation based images		Mutual information-based images	
		Train	*Dev*	*Train*	*Dev*	*Train*	*Dev*	*Train*	*Dev*
1	FNSZ	26179	2286	42577	4679	2627	1029	232	32
1	GNSZ	6415	1644	10964	1592	356	438	121	104
1	SPSZ	1179	96	1298	93	151	130	92	34
1	CPSZ	4119	256	8797	1111	488	258	164	117
1	ABSZ	1056	192	1573	192	84	100	257	111
1	TNSZ	279	95	486	589	248	498	215	99
1	TCSZ	494	409	900	506	273	348	278	101
1	MYSZ	64	32	64	32	64	32	125	10
1	**Total**	**39785**	**5010**	**66659**	**8794**	**4291**	**2833**	**1484**	**608**
2	FNSZ	1600	1568	672	1568	1582	-	704	-
2	GNSZ	7250	2680	6416	2622	3001	-	526	-
2	SPSZ	38155	9156	38565	8787	32515	-	950	-
2	CPSZ	11199	4880	9052	4751	7741	-	786	-
2	ABSZ	64	-	-	32	64	-	64	-
2	TNSZ	1462	96	1277	93	1433	-	117	-
2	TCSZ	907	550	420	537	907	-	315	-
2	MYSZ	540	1408	485	1364	540	-	247	-
2	**Total**	**61177**	**20338**	**56887**	**19754**	**47783**	**0**	**3709**	**0**

Table 1. Number of images created under each class of seizure

IMAGE VALIDATION USING CNN

The images were tested for the classification accuracy using three CNN architectures; Alexnet, Resnet18, and GoogleNet. All three pre-trained networks were tuned to match the multiclass requirement (7-Class) by changing parameters in the fully connected layer in each architecture. As indicated in Table 1, the number of images per class is not balanced. This is due to the insufficient recordings available in the corpus for a particular type of seizure (e.g., Myoclonic Seizure). Due to the imbalance in the data, other than validation accuracy, we have calculated F1-Score and Specificity.
The values of training options (hyperparameters) used for training the Alexnet architecture are listed in Table 2. Similar training options were provided for the other two architectures as well.

Solver	sgdm
Initial learning rate	0.0001
Validation frequency	50
Epochs	30
Minimum batch size	128
L_2Regularization	0.0001
Gradient Threshold Method	L_2Norm
Momentum	0.9

Table 2. Training options

Figure 6 shows the training progress upon training Alexnet architecture with mutual information-based images. The confusion matrix obtained for the same database (mutual information-based images) is shown in Figure 6. Similarly, all four image databases were tested with three CNN architectures.

RESULTS AND DISCUSSIONS

Figure 6 shows the training progress upon training Alexnet architecture with mutual information-based images. The confusion matrix obtained for the same database (mutual information-based images) is shown in Figure 6. Similarly, all four image databases were tested with three CNN architectures.

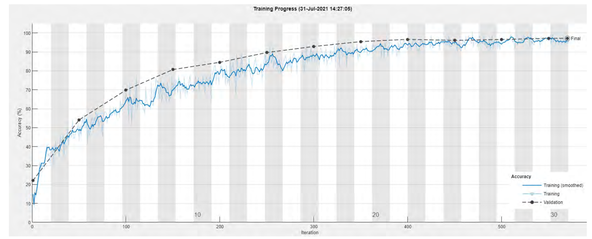


Figure 6. Training Progress of mutual information-based images with AlexNet architecture

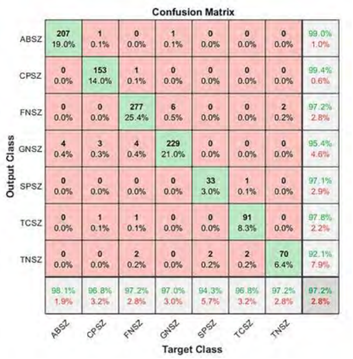


Figure 6. Confusion Matrix

The performance of our approach is compared with similar works performed on the TUH database by other authors. Golmohammadi et al. used GRU and LSTM models on TUH Seizure data and achieved a specificity of 97.1 % and 91.49 %. L. Wei and A. Mooney presented the XGBoost-based method to detect seizures from TUH Corpus and achieved an accuracy of 67.01% while training and 58.85% during validation. For the abnormal dataset from the TUH EEG corpus, S Roy et al. applied 1D-CNN-RNN and could achieve an accuracy of 82.27%. Yildirim et al. proposed a 1D CNN model to classify the normal and abnormal EEG signal from the TUH EEG corpus and achieved an F1-Score of 78.92 and an accuracy of 79.34%.

Spectrogram Images (AlexNet)	F1 - Score	97.47
	Accuracy	97.15
	Specificity	99.51
Correlation Coefficient based Images (ResNet-18)	F1 - Score	93.64
	Accuracy	93.45
	Specificity	98.88
Mutual information based Images (ResNet-18)	F1 - Score	97.95
	Accuracy	97.89
	Specificity	99.62
Stacked Images (ResNet-18)	F1 - Score	95.64
	Accuracy	95.50
	Specificity	99.24

Table 3. Summary of performance parameters

CONCLUSION

In this abstract, we have discussed the generation of an image database from TUSZ v1.5.2 using four techniques. The images are tested for classification accuracy using CNN. The highest accuracy of 97.89% has been achieved when trained ResNet architecture with images based on mutual information. The accuracy achieved is for seven classes. The image database serves as a valuable resource for training deep learning networks.

ACKNOWLEDGEMENT

This research is supported by the research laboratory of the ECE department at BMS Institute of Technology and Management. The database used in this work is provided by The Neural Engineering Data Consortium (NEDC) at Temple University.

Design and Simulation Analysis of a Piezoresistive Cantilever Beam for Low Pressure Detection

M.Lakshmi Prasanna [1] and Dr. Anitha V R [2]

[1]Research Scholar, IEEE Student Member,
Department of Electronics and Communication Engineering, JNTUA, Ananthapuramu, Andhra Pradesh, India.
[2]Associate Professor, IEEE Senior Member,
Department of Electronics and Communication Engineering, BMS Institute of Technology and Management,
Yelahanka, Bangalore-560064, Karnataka, India.

[1]lakshmi.mungara9@gmail.com and [2]anithavr@bmsit.in

Abstract. Many studies have shown that most MEMS devices use light beams whose deflection response is the main focus before choosing it for specific applications. This paper deals simulation of the MEMS-based cantilever beam for low pressure detection. COMSOL 5.4 Multiphysics FEM model is used to study the behavior of beam and compared with the conventional cantilever beam. Both are analyzed with two different materials SiO2 and SU-8. The change in the deflection has been analyzed with various pressure values applied on the cantilever beam. Through simulation and analysis results, it is observed that the deflection of the proposed beam is good at lower pressure values.

Keywords: Micro cantilever, deflection, rectangular beam, FEM, COMSOL 5.4

1 Introduction

In the past few decades, MEMS system attracts researchers due to its small size, quicker response, low power consumption, easy fabricating, low cost of production and high sensitivity. Among many MEMS devices micro cantilevers became more popular in the field of Bio-MEMS where low-mass load of biomolecules draws more attention on MEMS devices. Microcantilever sensors offer a wonderful means that to notice single cell or molecules within the atmosphere. Highly functionalized surface of the microcantilever absorbs specific targeted biomolecules and a surface-stress has been induced thereby leading a deflection inflicting a downward bend at the free end. In addition to a single micro-cantilever beam structure, different micro-cantilever beam arrays can also be easily manufactured. Therefore, sensors based on micro cantilever beams have been potentially applied in various fields [1-4]. Recently, research on cantilever beams used in chemical and biological sensors has developed rapidly in the medical, food, and agricultural industries. Therefore, the targeted range of the micro cantilever expands with the passage of time, and an in-depth analysis of the micro cantilever should be carried out.

In the field of sensors, sensitivity is one of the most important factors of cantilever beam analysis. Geometry, shape and resonance mode play a key role in the sensitivity of the cantilever sensor, because these parameters greatly affect the resonance frequency, quality issues and

* No academic titles or descriptions of academic positions should be included in the addresses. The affiliations should consist of the author's institution, town/city, and country.

deflection of the cantilever [7,8]. So far, various researchers have given different cantilever structures with microcantilever beams of different sizes to produce better deflection and sensitivity. Researchers have used different materials like SiO2, SiN3 and polymers like PDMS, SU-8 etc [5, 9]. Most of the researchers are interested to go with SU-8 material because of its improved Young's modulus and yield strength [10].

Piezoresistive sensing is one of the major integrated technologies into micro-cantilever sensors which are compatible with external electronic devices [1]. A piezoresistor is placed at the one end of the cantilever beam with some doping concentration. When the target analyte is bonded to the cantilever surface it encounters a surface stress resulting the change in resistance of the piezoresistor. Generally piezoresistor is connected in the form of Wheatstone bridge format as shown in Figure 1.

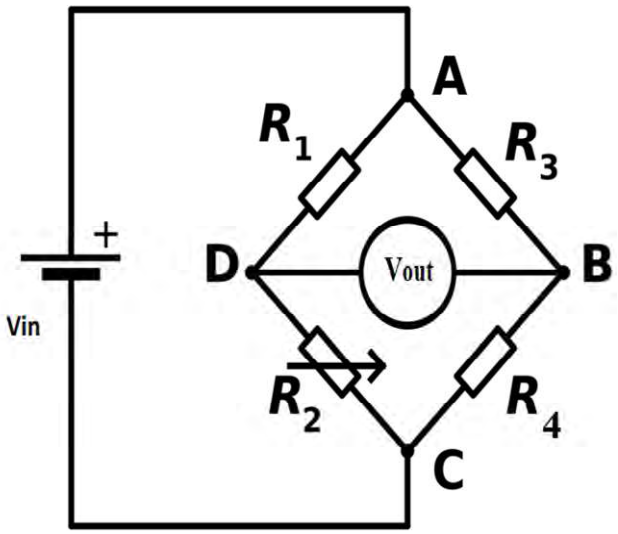


Figure.1: Wheatstone bridge

Here the change in resistance leads to a change in output voltage and is given by the equation

$$V_{out} = \left[\frac{R_2}{R_1+R_2} - \frac{R_4}{R_4+R_3}\right]V_{in} \qquad \text{-------(1)}$$

.

2 Design Parameters and Proposed design of a Cantilever

In this considered low pressure bimolecular detection for BioMEMS application. Figure 1 shows the schematic designs of conventional rectangular beam and our newly proposed high sensitive cantilever beam design for low pressure application, with identical piezoresistor at the fixed end of cantilever beam. The design of the proposed cantilever beam consists of four rectangular bar structures having dimensions 10µm X 50µm X 0.5µm, with equally spaced at the flexible end of the cantilever beam. In this work proposed cantilever beam is going to design with two different materials: SiO_2 and SU-8 polymer. The specifications of the materials are listed in the Table 1.

Table1: Material Properties of SiO_2 and SU-8

Material Property	Value	
	SiO_2	SU-8
Young's Modulus	70GPa	4.02GPa
Poisson's Ratio	0.17	0.26
Density	2200kg/m^3	1218kg/m^3

The proposed cantilever sensor has a piezoresistor which is connected in the form of Wheatstone Bridge setup. P-type polysilicon is used for the piezoresistor. The dimensions of piezoresistor are taken as 12µm X 16µm X 0.15 µm. A terminal voltage of 3V is applied to the piezoresistor. The piezoresistive layer doping density is chosen as 10G [1/m3]

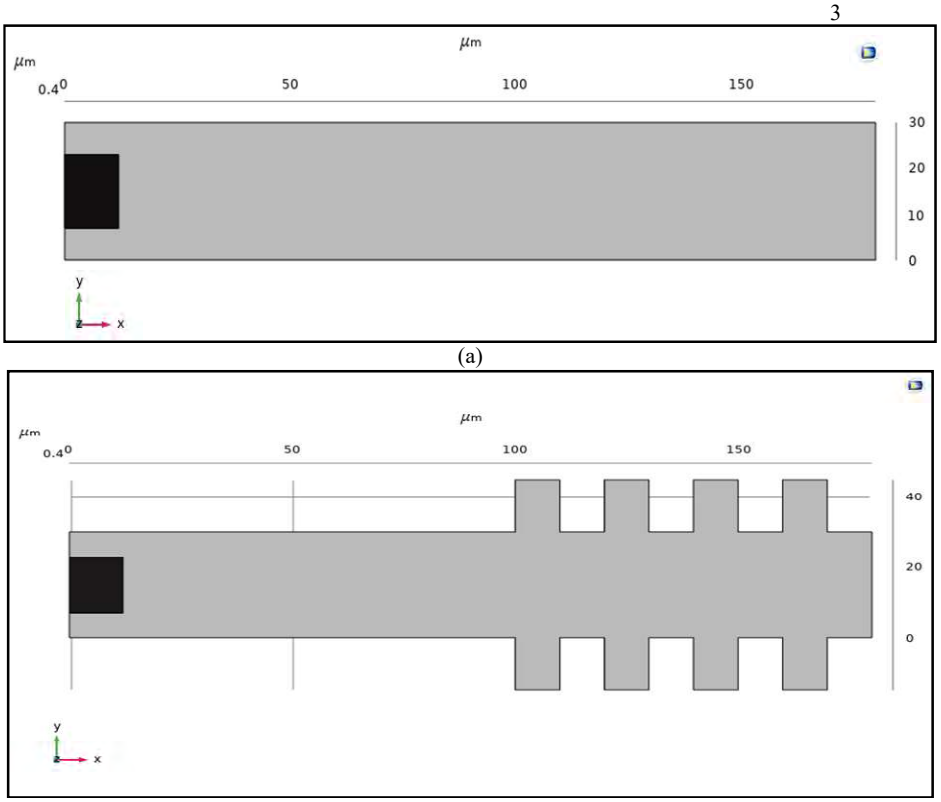
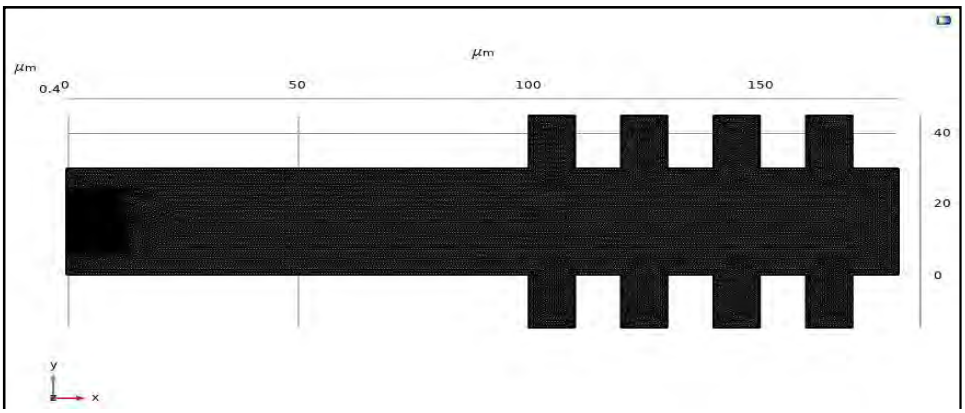


Figure.1: (a) Normal Rectangular cantilever beam (b) Proposed cantilever beam both having thickness of 0.5µm

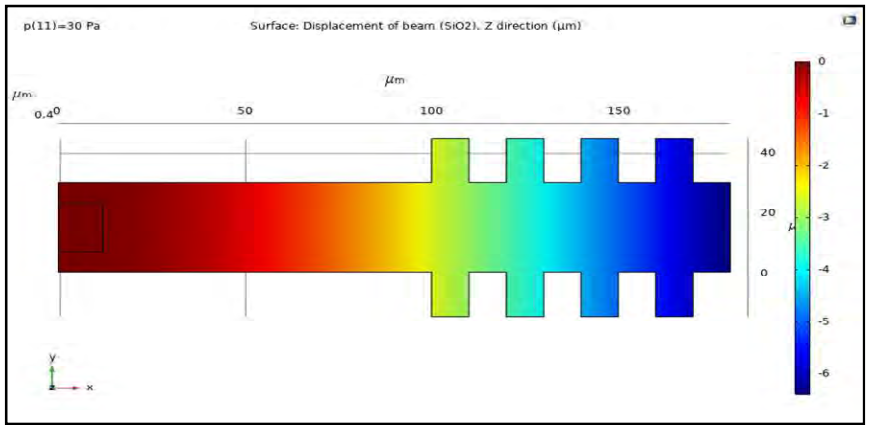
3 Simulation

COMSOL 5.4 Multiphysics has been used to design, simulate, and analyze proposed cantilever beam. The piezoresistivity-domain currents physics and stationary study has been used for modeling. The cantilever beam is fixed at one end, and the boundary load is applied in the form of the target biomolecule concentration, defined as pressure, and applied in the z direction. Free tetrahedral meshes with ultra-fine element sizes are used for high-quality analysis. The mesh result is shown in the Figure. 2. A parameter sweep has been defined for the specified pressure value and thickness value to analyze and compare the proposed cantilever beams for the materials SiO2 and SU-8.

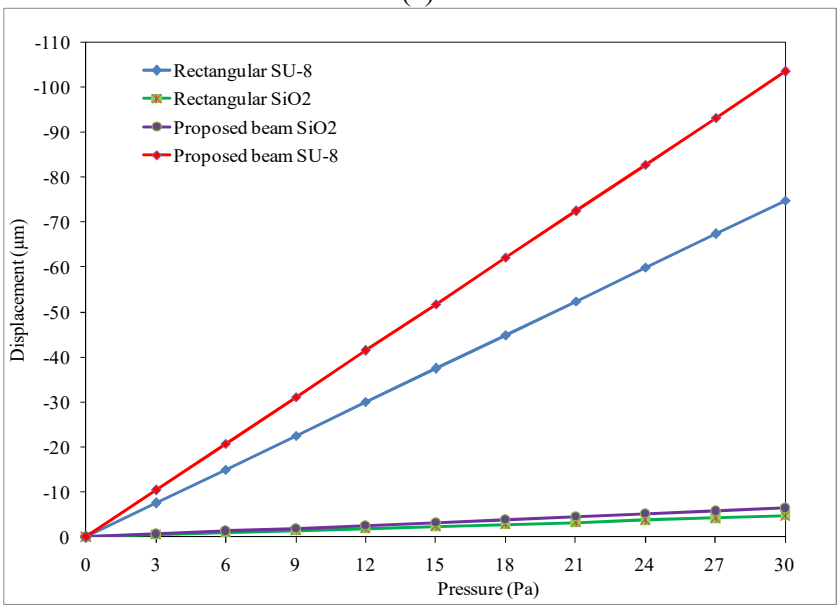


4 Results and Discussion:

The proposed cantilever beam design has been simulated using COMSOL Multiphysics 5.4 software. When the beam is exposed to some pressure, it results in mass loading effect on the cantilever beam surface. Because of this pressure, cantilever will undergo some deflection, shown in Figure.3 (a). A comparison plot of deflection between normal rectangular beam and proposed beam with two materials are shown in Figure. 3(b) for different pressure values.



(a)



(b)

Figure.3: (a) a Sample of displacement for the proposed cantilever beam and (b) a Comparison plot of displacement for the conventional rectangular beam and proposed cantilever beam.

The deflection the beam undergo with mechanical stress. This mechanical stress results in change of resistance in piezoresistor which results in change in resistance and the change in resistance is shown in Figure.4. Deflection of beam has been measured with various thickness values by applying 10Pa of pressure. It is plotted in figure.5.

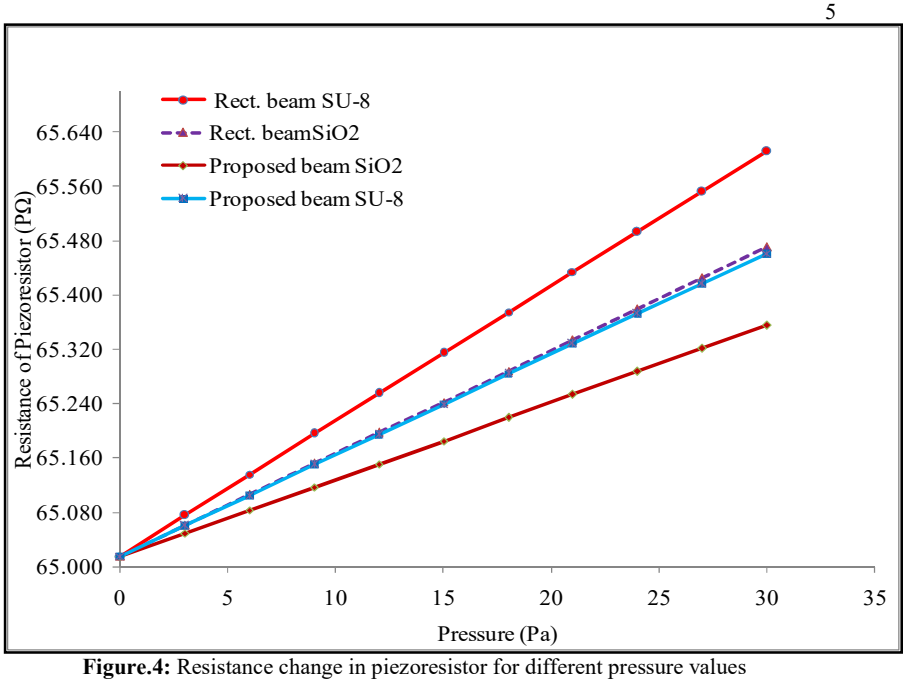


Figure.4: Resistance change in piezoresistor for different pressure values

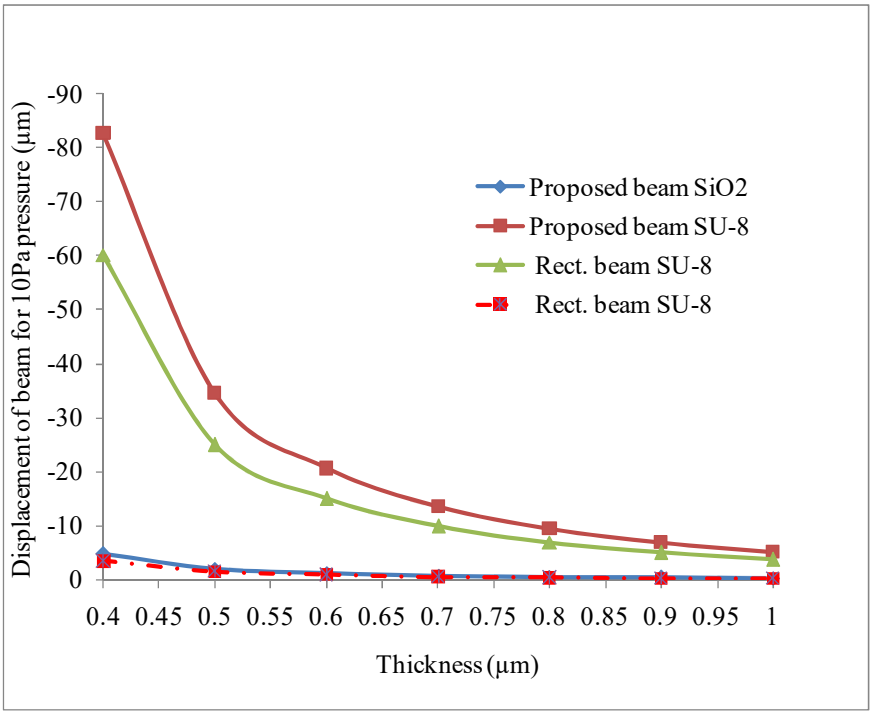


Figure.5: Deflection of beam for different thickness values

5 Conclusions

This study proposes a high-sensitivity MEMS piezoresistive micro-cantilever beam for studying low pressures in the field of BioMEMS. The characteristics of the proposed cantilever beam are compared with a conventional rectangular micro-cantilever with the same piezoresistor. The result shows that the deflection of the proposed micro-cantilever beam is better than the traditional rectangular micro-cantilever beam. The mechanical stress caused by the biomolecule

cantilever beam with both SiO2 and SU-8 material cantilevers provide better displacement. The linear relationship between the pressure and change in resistance of piezoresistor indicates that the proposed design of cantilever beam works well at lower pressure levels. Also for optimum thickness levels the results are good.

Acknowledgment. A Special thanks to the National MEMS Design Centre, Sree Vidyanikethan Engineering College, A.Rangampet, Tirupathi, for providing the research facility.

References (in Basic)

1. Guo, Kai & Jiang, Bo & Liu, Bingrui & Xingeng, Li & Wu, Yaping & Tian, Shuang & Gao, Zhiyue & Zong, Lijun & Yao, Shuo & Zhao, Mingzhe & Mi, Chunxu & Zhu, Gengzeng. (2021). Study on the progress of piezoelectric microcantilever beam micromass sensor. IOP Conference Series: Earth and Environmental Science. 651. 022091. 10.1088/1755-1315/651/2/022091.
2. Ali, Shaik & Bhuvaneswari, H. & Kumar, B.. (2020). Design and modeling of a large deflection microcantilever using rectangular SCR. AIP Conference Proceedings. 2281. 020037. 10.1063/5.0027942.
3. Rotake, Dinesh & Darji, Anand. (2020). Design and Reliability Testing of Microcantilever-Based Piezoresistive Sensor for BioMEMS Application. 10.13140/RG.2.2.25052.92803.
4. Ashok, Akarapu & Nighot, Rohit & Sahu, Nagesh & Pal, Prem & Pandey, Ashok. (2019). Design and analysis of microcantilever beams based on arrow shape. Microsystem Technologies. 25. 10.1007/s00542-019-04555-4.
5. Gharge, Bhagyashri & Upadhye, Vaishali & Bodas, Dhananjay. (2015). Design and simulation of microcantilevers for detection of pathogens. 10.1109/ISPTS.2015.7220122.
6. Lamba, Monica & Singh, Dr. Kulwant & Chaudhary, Himanshu. (2020). Design analysis of polysilicon piezoresistors PDMS (Polydimethylsiloxane) microcantilever based MEMS Force sensor. International Journal of Modern Physics B. 34. 2050072. 10.1142/S0217979220500721.
7. Passian, Ali & Thundat, Thomas & Thanihaichelvan, Murugathas. (2017). Microcantilever Sensors. 10.1016/B978-0-12-803581-8.10525-9.
8. Rotake, Dinesh & Darji, Anand & Singh, Jitendra. (2020). Ultrasensitive Multi-Arm-Microcantilever-Based Piezoresistive Sensor for BioMEMS Application. 10.1109/VDAT50263.2020.9190249.
9. Parsediya, Deep. (2016). Deflection and stresses of effective micro-cantilever beam designs under low mass loading. 111-114. 10.1109/ICEPES.2016.7915915.
10. Mathew, Ribu & Sankar, Ravi. (2018). A Review on Surface Stress-Based Miniaturized Piezoresistive SU-8 Polymeric Cantilever Sensors. Nano-Micro Letters. 10. 10.1007/s40820-018-0189-1.

Proceedings of the 11th International Conference on Computer Engineering and Networks pp 504–515

A Hierarchical Machine Learning Frame Work to Classify Breast Tissue for Identification of Cancer

J. Anitha Ruth, Vijayalakshmi G. V. Mahesh, R. Uma & P. Ramkumar

Conference paper | First Online: 12 November 2021

Part of the Lecture Notes in Electrical Engineering book series (LNEE,volume 808)

Abstract

In this work a study is conducted on the Breast tissues using machine learning algorithms to identify the Breast cancer. The paper proposes a hierarchical two stage classification frame work utilizing the features provided by the Breast tissue dataset of UCI machine learning database

with SVM_RBF, kNN and Decision tree algorithms. The experimental results and their analysis at stage-1 binary classification indicated better performance of the method. Further the analysis under multi class classification specifies the robustness of the methodology by providing reduction in false alarm rates.

Keywords

Hierarchical classification **Electrical Impedance Spectrscopy**

Breat tissue **Machine learning** **Breast cancer**

An ensemble classification based approach for breast cancer prediction

An ensemble classification based approach for breast cancer prediction

Vijayalakshmi G V Mahesh[1], Mohan Kumar M[2]

[1]Department of Electronics and Communication Engineering, BMS Institute of Technology and management, Bangalore, India

[2]Department of Electronics and Communication Engineering, Yenepoya Institute of Technology, Mangalore, India

E-mail: vijayalakshmi@bmsit.in, mohankumar@yit.edu.in

Abstract. Breast cancer is the second major reason for deaths in women. Early detection of the breast cancer and receiving the appropriate treatment can reduce the death rates as survival becomes hard in the higher stages of the tumor growth. Application of machine learning in healthcare play a key role in aiding the clinical experts to detect the disease at a early stage and perform precise assessment. This paper proposes a pattern recognition methodology that uses breast cancer biomarkers as the attributes and ensemble classification approach for accurately detecting the presence of cancer. The proposed method was evaluated for the samples of the breast cancer Coimbra dataset by fusing the decisions of naive Baye's , radial basis function neural network and linear discriminant analysis classifiers based on majority voting rule. The experimental results demonstrated the enhanced performance of the system with fusion of classifiers as compared to the single classifiers.

1. Introduction

An uncontrolled way of cell division leads to cancer and these cells often invade into other healthy tissues. Breast cancer is the most common and is the leading cause of death in women. It can be invasive or non invasive: invasive cancer spreads from the ducts or glands to other parts of the breast while noninvasive cancer does not spread from the original tissue. Based on this, the following types of breast cancer are identified (i) Ductal carinoma in situ(DCIS):It is non invasive where the cancer cells are confined only to the ducts of the breast, (ii) Lobular carinoma in situ(LCIS): A non invasive condition where the cancer cells are limited to the milk producing glands of the tissue, (iii) Invasive ductal carcinoma(IDC): This is invasive type of cancer. The cancer cells develop in the milk ducts of the breast and then invade to the nearby tissues and organs. IDC is the most common type of breast cancer and (iv) Invasive lobular carinoma(ILC): This type of cancer begins in the lobules of the breast and spreads to the surrounding tissue.

Based on the size of the tumor and the extent of its spread into nearby tissues and organs the doctors recognize the different stages of cancer as stage 1, stage 2, stage 3 and stage 4 breast cancer. Stage 1 is a primary tumor with its size around 2cms while stage 4 is advanced where the tumor can have any size and the cancer cells have spread to nearby tissues and organs. Early detection of the cancer i.e.,

in the initial stages improves possibility of survival. If the cancer reaches stage 4, the chances of survival comes down. Thus early detection of cancer is very crucial and regular checks and screening can aid in detecting the symptoms earlier. Clinical experts diagnose the cancer after routine screening. The use of biomarkers and imaging tests helps the clinicians to identify the breast cancer. The most widely used image screening methods are Mammogram, ultrasound and Magnetic resonance imaging(MRI). Further researchers are trying to find new breast imaging methods to improve the detection and treatment of the cancer and have come up with Scintimammagraphy(molecular breast imaging), Positron emission mammography(PEM), Electrical impedance imaging(EIT) and Elastography

Though imaging methods play an important role in decision making followed by treatment. They require expertise to deliver significant and accurate information . For example, extraction of feature descriptors from images using mathematical techniques to detect and differentiate the tumor and non tumorous region. Also the reliability of the detection method needs to be assessed based on (i) Accuracy, (ii) Precision, (iii) Correlation to the disease and (iv) Usability. Biomarkers[1], on the other hand are gaining lot of importance in accurately detecting and diagnosing the cancer as a result of advancements in bioinformatics tools. Recognition of significant biochemical changes in cancer has facilitated towards advances in detection and its treatment. They can be relied upon for early detection and accurate evaluation. Breast and prostate cancers are diagnosed comparatively earlier as compared to other cancers. Additionally, the cancer biomarkers are significant in detecting the disease for asymptomatic persons at the early stage. Ideally a portion of a protein or a protein in the blood or urine samples is considered as a marker for tumor detection. Thus this paper proposes a pattern recognition system that use biomarkers as features or descriptors combined with machine learning algorithms to effectively identify the presence or absence of cancer. Accordingly, this is a classification task where the classifier predicts the class label of the sample. The performance of the model is later quantitatively evaluated using the metrics derived from the confusion matrix.

The rest of the paper is organized as follows, Section 2 presents a comprehension of related work. Section 3 describes the methodology of the works. The experimental results and discussion are provided in section 4. Finally, section 5 concludes the paper.

2. Related work

Biomarkers in cancer are the characteristics that are measured and they describe the pathological state of an living being. Cancer biomarkers are categorized into predictive, diagnostic and prognostic based on how they are handled. The first cancer biomarker was recognized from the blood sample of a patient with colon cancer [2]. Later different biomarkers were developed to identify other cancers. The standards and measurements of breast cancer biomarkers are provided in [3] and a overview is found in[2].Further, the advancements in technology and data analytics is seen to solve various health related issues to improve the quality of health care and provide effective treatment. Recent years have witnessed a biggest transformation of Artificial intelligence in health care domain powered by machine learning algorithms that can assist the clinicians/experts in diagnosing the disease and build up effectual solutions. Thus the acquired dataset is fed to the algorithms that can discover the patterns and anomalies to make right decisions about the disease diagnosis.

Many researchers have put forward their ideas in using cancer biomarkers with machine learning algorithms to improve the precision of detection and analysis. Li and Chen[4] investigated the adaption of cancer attributes in detecting the disease which can reduce the probability of death rate. They used Wisconsin Breast cancer and Breast Cancer Coimbra datasets with Decision tree(DT), Random forest(RF), Support Vector Machine(SVM), Neural Network(NN) and Logistics

Regression(LR) Classifiers and compared their performances . The experimental study showed that the Random forest algorithm provided better results in terms of prediction accuracy, F -score and Area Under the Curve(AUC) as compared to other classifiers. Farah Sardouk et al.,[5] studied the predictors of Breast Cancer Coimbra dataset to estimate the presence or absence of the cancer. They created six machine models using Adaboost, Regression , Random Forest, Jrip, RBFNN and J48 classifiers in WEKA and MATLAB environments. The models were assesses using True positive rate, False positive rate, precision, recall, F1-score MCC and AUC. The results proved that the machine

learning algorithms were efficient in predicting the outcome. Miguel Patricio et al., [6] at university of Coimbra collected clinical attributes from 166 individuals based on anthropometric data. The measured data or features were then fed to logistic regression, Support vector machine and random forest machine learning algorithms to evaluate the probability of using the anthropometric data as biomarkers for cancer prediction. which were given to the developed a model that can be used as a biomarker. The evaluation of the classifier models for the metrics sensitivity, AUC and specificity indicated the better performance of SVM for the following predictors: Glucose, Resistin, BMI and age. Thus they can be efficient biomarkers of breast cancer.

Jin Yue et al.,[7] carried out statistical analysis, spectral analysis and least squares fitting analysis on the Breast Cancer Coimbra dataset to check the interclass variation. The analysis was able to clearly indicate the significant distinction between the two classes: cancerous and non cancerous. Yolanda D. Austria et al.,[8] investigated 11 classification algorithms and their variants : LR, k Nearest Neighbor(kNN), SVM, DT, RF, Gradient boosting and naive Baye's for predicting the breast cancer on the Coimbra dataset. The performance of the models were evaluated using accuracy and prediction time. The evaluation displayed an higher accuracy of 74.14% with Gradient boosting classifier and non linear SVM was fast in predicting the label without any delay. The work also identified the anthropometric data body mass index(BMI) as the principal predictor. Ratula Ray et al.,[9] explored the use of DT, kNN, RF and Guassian naive Baye's machine learning algorithms in aiding the pathologists to make precise decision about the presence of the disease considering both image and numeric(biomarkers) databases. Authors used breast histopathology images from Kaggle and breast Coimbra dataset for the evaluation of classifier models. The analysis indicated the better performance of random forest classifier which provided precision , recall and F1 score of 71%, 71% and 70% respectively for the numeric dataset. Similarly it presented precision , recall and F1 score of 91%, 92% and 92% respectively. The study also implies the importance of resistin and BMI biomarkers that influence the decision process.

From the evaluation of the contributions it is found that majority of the methods used various machine learning algorithms independently to improve the success rate of cancer detection. This paper presents a model that makes use of a ensemble classification where the classifiers are fused at the decision level to enhance the performance of the system.

3. Methodology

The methodology of the proposed system to predict the presence of breast cancer is shown in Figure1.The frame work includes (i) Feature extraction (ii) Classifier/Machine learning model creation (iii) Classification.

The extraction of suitable features or descriptors are essential in deciding the accuracy of the system. The anthropometric data measured from the input samples or the biomarkers form the features. These features are given to different classifiers for models creation which can predict the class label of the given sample. To enhance the presentation of the system, the decisions of the individual classifiers are combined based on decision level fusion strategy. The created models are evaluated to validate their performance. The details of the process is described in subsequent sub sections.

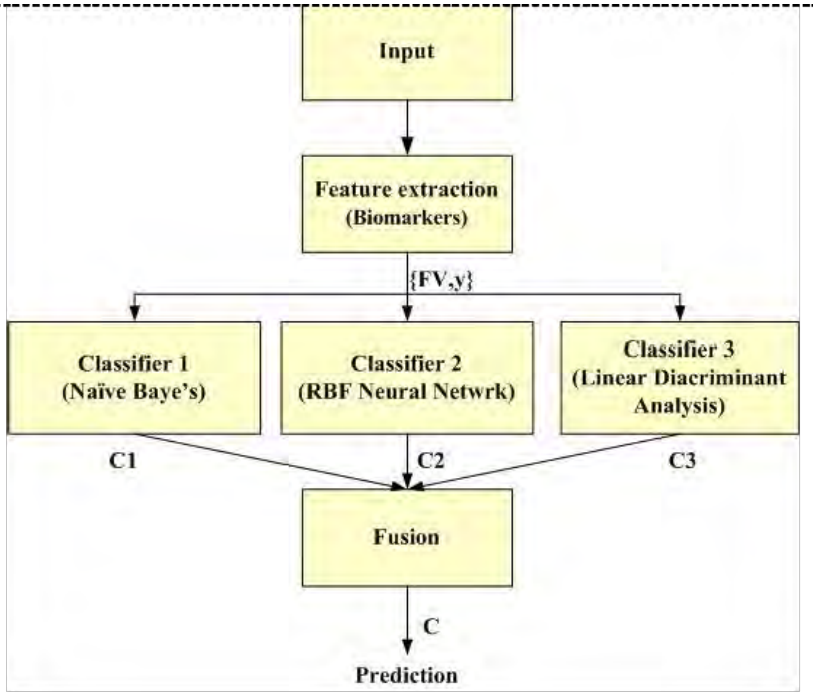


Figure 1. Methodology of the proposed system

3.1 Dataset and Feature extraction

The work considered Breast Cancer Coimbra dataset[6] collected from 116 persons. From each participant nine clinical features were measured as displayed in Table 1.

Table 1. Anthropometric data(Features)

Feature	Description	Unit
A	Age	Years
BMI	Body Mass Index	Kg/m2
G	Glucose	Mg/dL
I	Insulin	μU/mL
HOMA	Homeostasis Model Assessment index	--
L	Leptin	Ng/mL
Ad	Adiponectin	μg/mL
R	Resistin	Ng/mL

These clinical measures are the anthropometric data collected during procedural blood analysis. Out of 116 input samples, 64 samples are with breast cancer and the remaining 52 are healthy samples. An observation of the features of dataset as illustrated in Figure 2 identifies the distinction between the cancerous and non-cancerous samples that clearly indicate that they are potential biomarkers for the detection of the pathology .

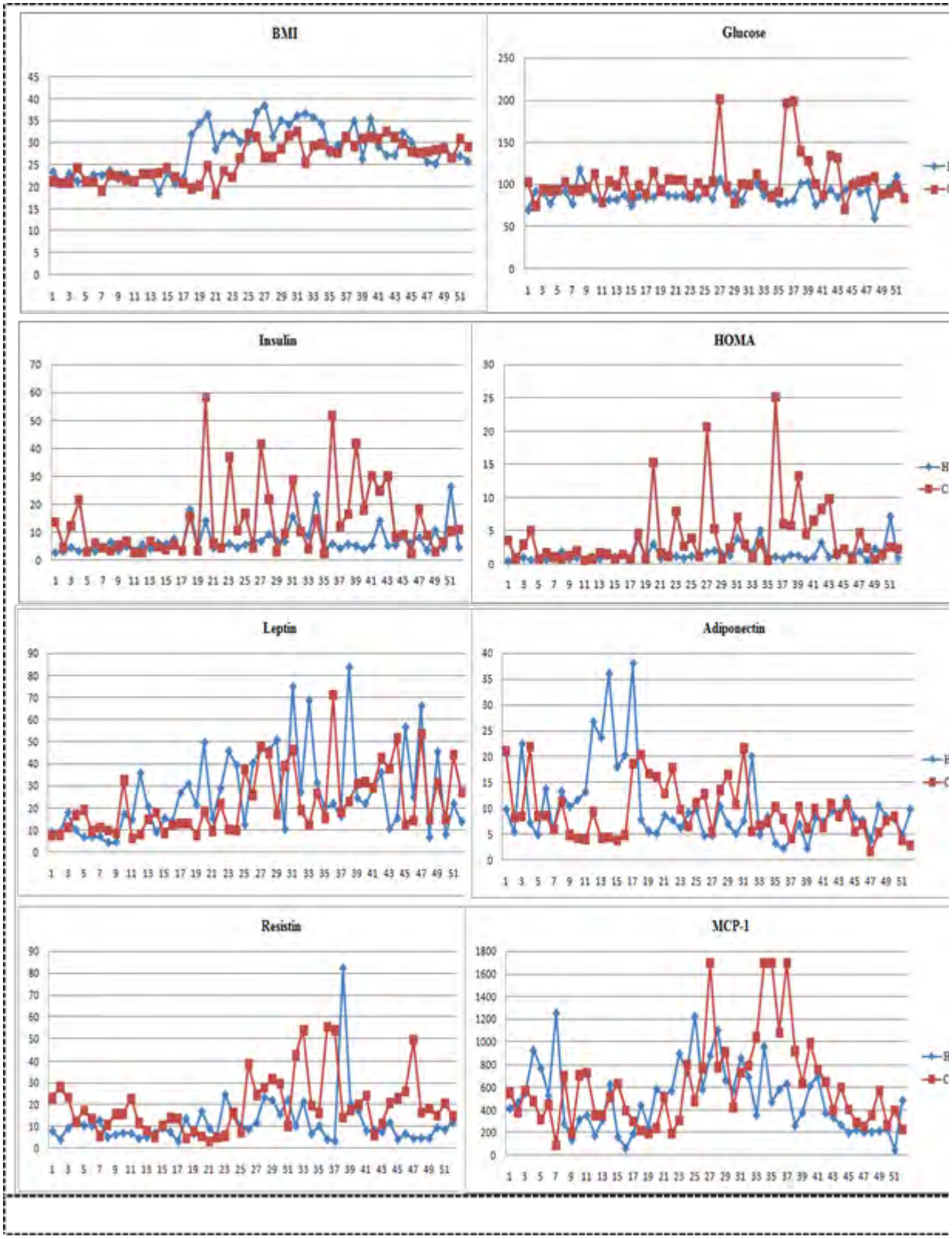


Figure 2. Features of Healthy and Cancerous samples

The features are combined to form a feature vector as shown,

FV=[A, BMI, G, I, HOMA, L, Ad, R, MCP-1] (1)

which is provided to the classifiers for training and testing. To validate the proposed system, the dataset with its features was divided into two non overlapping training and testing sets with a ratio of 75:25. The training set is labeled using the class labels y={'1' , '0'}, where '1' indicates healthy and '0' indicates Cancerous. The labeled training set {FV,y} is passed to the classifier for creating the model. The model created is then tested using the samples from the testing set to predict its outcome C ϵ y.

3.2Machine Learning Algorithms

The classifier predicts the outcome of an unseen sample 'Q' as a function of labeled training set and the learning parameters of the algorithm .

$$C=f(FV_Q, \{[FV,y]: \text{training parameters}\}$$

where FV is the feature vector of the training samples and FV_Q is the feature vector of the query/unseen sample.

The proposed system used three classifiers: Naive Baye's(NB), Radial Basis Function Neural network(RBFNN) and Linear Discriminant Analysis(LDA).

3.2.1 Naive Baye's classifier

It is a simple probabilistic classifier which is based on Baye's theorom[10]. It requires fewer number of training samples for prediction. Baye's rule can be formulated as

$$P(y/X)=\frac{P(X/y).P(y)}{P(X)} \quad (2)$$

where X signify the features and y is the label representing the category of the input sample. In the context of the work presented here, the above equation can be rewritten by expanding the feature set as,

$$P(C/f_1,f_2,.....,f_9) = \frac{P(f_1/C).P(f_2/C).P(f_3/C).....P(f_9/C)\ P(C)}{P(f_1)P(f_2).....P(f_9)} \quad (3)$$

where $(f_1, f_2, ... f_9)$=FV

For all the elements of the dataset, the denominator is constant, thus we get

$$P(C/f_1,f_2,.....,f_9) \text{ proportional } P(C) \prod_{i=1}^{9} P(f_i/C) \quad (4)$$

Here C can be one of the class label belonging to y. Thus the outcome C is found based on maximum probability, i.e.,

$$C=\text{argmax}_C\ P(C)\ \prod_{i=1}^{9} P(f_i/C) \quad (5)$$

The classifier finds its applications in biometrics(face recognition), healthcare, weather prediction, security, emotion recognition, sentiment analysis etc.

3.2.2 Radial Basis Function Neural network(RBFNN)

RBFNN[11] is one of the variants of feed forward neural network that uses radial basis functions as the activation function in the hidden layer. The network converges faster to reach the performance goal. The NN architecture has three layers: Input layer, Hidden layer and Output layer. The number of neurons in the input layer correspond to the dimension of the feature vector and the neurons in output layer correspond to the number of class labels. Each layer is associated with its activation function. The activation functions in the hidden layer are implemented as Guassian functions.

The labeled training set {FV,y} is provided to the network for training . During training the parameters such as the network weights, RBF activation function centres and their distributions are tuned reach the performance goal: Mean squared error(MSE)

For a given input FV, the output at a node of output layer is computed as,

$$C_i = \sum_{j=0}^{J} w_{ij}\, \varphi_j\, (\|FV - M_j\|) \quad (6)$$

Where, FV is the input to the network/Feature vector

w_{ij} is the weight from hidden neuron j to output neuron i

||.|| L_2 / Euclidean norm

φ_j is the RBF in the hidden neuron j with center M_j.

3.2.3 Linear Discriminant Analysis(LDA)

LDA classifier proposed by Fisher [12]is based on the concept of reducing the within(intra) class variation and maximizing the between(inter) class variation. This classifier finds class separation by minimizing the spread within class and maximizing the distance between each class. Thus it is based on the statistical measures mean and variance. For the given dataset and class labels, LDA works by projecting it on to hyperplane to find the scalar, of all the possible lines, the line that gives maximum class separability is chosen.

The best projection vector is selected by maximizing the Fisher's criteria given by,

$$J(W) = \frac{W^T S_B W}{W^T S_W W} \quad (7)$$

here J(W) indicates the measure of difference between means of the classes

S_B represents the *within class* scatter matrix S_W represents *between class* scatter matrix W is the projection matrix.

Maximizing J(W) yields
which is Fisher's Linear Discriminant

$$W = \text{argmax}(J(W)) = S_W^{-1}(\mu 1 - \mu 2), \quad (8)$$

where, $(\mu 1 - \mu 2)$ is the distance between the projected means for two classes.

4. Results and Discussion

The experiment was conducted to identify the presence of the breast cancer for the given sample based on fusion of classifiers. The proposed method is validated with hold out method where the dataset with 116 samples was separated into training and testing sets with the ratio of 3:1. Thus the training set has 87 samples and testing set 29 samples.

As a first step, all the training samples were passed to feature extractor that provided the feature vector FV with 9 dimensions. Each feature of the FV has different ranges, so the data is normalized using Z- score normalization that gives normalized FV with zero mean and standard deviation of one as shown

$$FV_n = \frac{FV - \mu}{\sigma} \quad (9)$$

Later the FV_n was labeled to form training data set $Z_i = \{FV_n, y\}_{i=1}^{87}$. The training data set was accepted by the Naive Baye's(NB), Radial Basis Function Neural network(RBFNN) and Linear Discriminant Analysis(LDA) classifiers for creating the models. During training, the learnable parameters of the classifiers were modified based on the performance function. The RBFNN was optimized for a spread constant of 3.5 and for NB classifier the prior probabilities was set to [0.2 0.2]. The trained models were tested with the testing data set $Z_{iQ} = \{FV_n\}_{i=1}^{29}$, where Q represents the Query or unseen sample.

The results of the individual classifiers were observed and analyzed using the parameters Receiver Operating Characteristics(ROC), Classification accuracy(CA), Precision, Recall(Sensitivity), Specificity and F1 score which are computed using the elements of confusion matrix[13] as shown,

$$\text{Classification accuracy(CA)} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (10)$$

$$\text{Precision} = \frac{(TP)}{(TP+FP)} \quad (11)$$

$$\text{Recall(Sensitivity)} = \frac{(TP)}{(TP+FN)} \quad (12)$$

$$\text{Specificity} = \frac{(TN)}{(TN+FP)} \quad (13)$$

$$\text{F1 score} = 2 * \frac{(Precion*Recall)}{(Precision+recall)} \quad (14)$$

where TP - Number of true positives FN- Number of false negatives FP- Number of false positives TN- Number of true negatives

The ROC obtained for all the classifiers is displayed in Figure 3 and other metrics are shown in Table 2.

The analysis of ROC and the performance measures indicate the poor presentation of the classifiers when used individually. The highest CA, precision and specificity was obtained from RBFNN that provided 69%, 70% and 81.25%respectively.

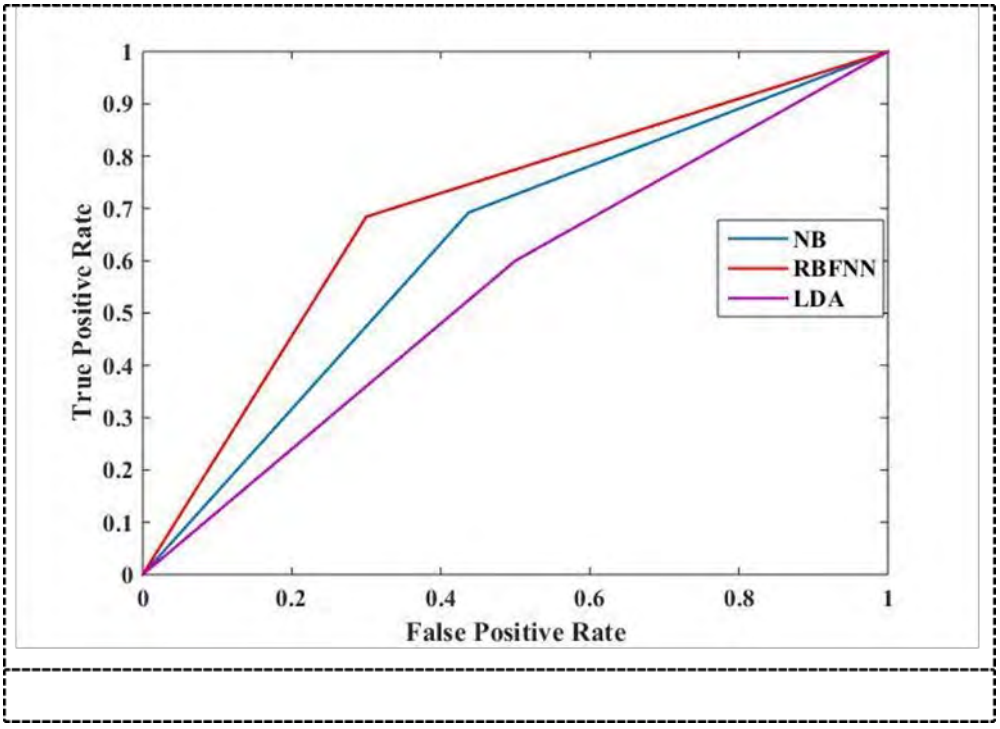


Figure 3. ROC obtained from the testing set

Table 2. Classifier results

Classifier	NB	RBFNN	LDA	Ensemble classification
CA(%)	62.06	69	58.17	75.86
Precision(%)	56.25	70	53.84	75
Recall(%)	69.23	53.84	53.84	69.23
Specificity(%)	56.25	81.25	62.5	81.25
F1-score	0.6206	0.6086	0.5383	0.7199

To improve the performance of the system, the decision of the individual classifiers can be fused[14] which is applied in the proposed frame work. A review of the literature finds different fusion strategies based on max, min, sum, product, mean, median and majority vote rules. This work utilizes binary classifiers with class label vector y={'1', '0'}. Let C_1, C_2 and C_3 be the outputs of NB, RBFNN and LDA classifiers such that Ci ϵ y. The outputs are fused as shown,

$$C= F(C1, C2, C3) \quad (15)$$

here F signifies the fusion rule

Now based on the fusion, we arrive at the final decision by opting for majority voting rule which requires odd number of single classifiers. This is the simple and a sensitive approach that assigns the sample with a most frequent class labels. Finally to demonstrate the significance of ensemble classification, the final decision was analyzed using the performance metrics and the results are tabulated in Table 1. The results show an 75% precision with ensemble classification that represents a low false positive rate . The analysis of F1-score which is the harmonic mean of precision and recall is advantageous in our work as the dataset considered has uneven class distribution(52 healthy and 64 cancerous samples). The range of F1-score is [0 1]. Thus a value of 0.7199 is better as compared to individual classifier results.

It is observed from the results that, the ensemble classification provided the results greater than the average of the single classifiers indicating a enhanced performance of the system subjected to CA, precision, recall, specificity and F1-score.

5. Conclusion

This paper proposes a pattern recognition system with ensemble of classifiers for classifying a sample into two classes viz presence of breast cancer and healthy. The work utilized anthropometric data /biomarkers as the features for data representation . These features or descriptors were provided to the NB, RBFNN and LDA machine learning algorithms for classifier model creation and evaluation. The experimental study on breast cancer Coimbra dataset with the proposed framework revealed the efficacy of the biomarkers as better features. Further the analysis of results obtained from fusing the decisions of single classifiers indicated improved performance measures in terms of classification accuracy, precision, recall, specificity and F1-score. This justified the deployment of ensemble classifiers in better discrimination of classes leading to enhanced performance of the system

6. References

[1] Kim SH, Hoffmann U, Borggrefe M, Akin I, Behnes M. Advantages and limitations of current biomarker research: from experimental research to clinical application. Current Pharmaceutical Biotechnology. 2017 May 1;18(6):445-55.

[2] Chatterjee, Sabarni K., and Bruce R. Zetter. "Cancer biomarkers: knowing the present and predicting the future." (2005): 37-50.

[3] Cole KD, He HJ, Wang L. Breast cancer biomarker measurements and standards. PROTEOMICS–Clinical Applications. 2013 Jan;7(1-2):17-29.

[4] Li Y, Chen Z. Performance evaluation of machine learning methods for breast cancer prediction. Appl Comput Math. 2018 Oct 18;7(4):212-6.

[5] Sardouk F, Duru AD, Bayat O. Classification of breast cancer using data mining. American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS). 2019 Jan 4;51(1):38-46.
[6] Patrício M, Pereira J, Crisóstomo J, Matafome P, Gomes M, Seiça R, Caramelo F. Using Resistin, glucose, age and BMI to predict the presence of breast cancer. BMC cancer. 2018 Dec 1;18(1):29
[7] Yue J, Zhao N, Liu L. Prediction and Monitoring Method for Breast Cancer: A Case Study for Data from the University Hospital Centre of Coimbra. Cancer Management and Research. 2020;12:1887.
[8] Austria YD, Jay-ar PL, Maria Jr LB, Goh JE, Goh ML, Vicente HN. Comparison of Machine Learning Algorithms in Breast Cancer Prediction using the Coimbra Dataset. cancer. 2019;7:10.
[9] Ray R, Abdullah AA, Mallick DK, Dash SR. Classification of Benign and Malignant Breast Cancer using Supervised Machine Learning Algorithms Based on Image and Numeric Datasets. InJournal of Physics: Conference Series 2019 Nov (Vol. 1372, No. 1, p. 012062). IOP Publishing.
[10] Rish I. An empirical study of the naive Bayes classifier. InIJCAI 2001 workshop on empirical methods in artificial intelligence 2001 Aug 4 (Vol. 3, No. 22, pp. 41-46).
[11] Mahesh VG, Raj AN. Invariant face recognition using Zernike moments combined with feed forward neural network. International Journal of Biometrics. 2015;7(3):286-307.
[12] Bouzalmat A, Kharroubi J, Zarghili A. Comparative study of PCA, ICA, LDA using SVM classifier. Journal of Emerging Technologies in Web Intelligence. 2014 Feb 1;6(1):64-8.
[13] Fawcett T. ROC graphs: Notes and practical considerations for researchers. Machine learning. 2004 Mar 16;31(1):1-38.
[14] Mi A, Wang L, Qi J. A multiple classifier fusion algorithm using weighted decision templates. Scientific Programming. 2016 Jan 1;2016.

Materials Today: Proceedings 56 (2022) 3755–3759

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Switching, Raman and morphological studies on $Si_{20}Te_{74}Sn_6$ chalcogenide glass

Jagannatha K.B. [a,b,*], Chandasree Das [b]

[a] BMS Institute of Technology and Management, Department of ECE, Bengaluru 64, India
[b] BMS College of Engineering, Department of EEE, Centre for Nanomaterials and Display, Bengaluru 19, India

ARTICLE INFO

Article history:
Available online 15 January 2022

Keywords:
Glass
PCM
Memory switching
XRD
SEM

ABSTRACT

Chalcogenide materials are getting lot of research attention due to their ability to switch from an amorphous to a crystalline phase with lower threshold voltage which is a primary requirement for fabricating phase-change memory. Here we have prepared the $Si_{20}Te_{74}Sn_6$ bulk sample using melt quenching technique. It has been observed that the sample exhibits a memory switching phenomenon with the addition of Sn dopant on Si-Te base glasses. Memory switching is mainly observed due to Te rich host matrix and the metallicity of Sn dopant. The sample exhibits a linear variation of threshold voltage with an increase in thickness, which is compatible with memory switching behaviour. Raman studies on as-quenched $Si_{20}Te_{74}Sn_6$ glass are carried out to study the structural behaviour of the network. The observed Raman peaks are attributed to crystalline tellurium and the vibrational modes of Sn/Si lattice. Morphological studies on switched and un-switched samples using SEM reveal a structural alteration in the switched region, which can also be related to the creation of conducting channel between two electrodes during memory switching.

Selection and peer-review under responsibility of the scientific committee of the First International Conference on Design and Materials (ICDM)-2021.

1. Introduction

The amorphous chalcogenide glasses (CG's) undergo an electrical switching phenomenon which is reported by Ovshinsky in the year 1969 [1] and is a scientifically intriguing and technologically relevant phenomenon. The use of this switching phenomenon in information storage, thermoelectric sensors, energy and so on has long been contemplated; however, in recent years phase change memories (PCM) have become a reality. The advantages are high scalability, multibit capability, lower cost, high endurance, compatibility with CMOS technology etc. PCM is considered the most promising technology for non-volatile random-access memory (NVRAM) [2,3].

In PCM technology, the logic '0' is represented as an amorphous state in which the resistance of the material will be very high and the crystalline state as logic '1', where the resistance of the material is low. There are two forms of switching available in PCM (i) Memory switching and (ii) Threshold switching. In the case of memory type switching the material initially will be in high resistance amorphous state. As the applied electric field increases the material remains to be in an amorphous state, up to a specific voltage called threshold voltage (V_{th}). A threshold voltage material changes to a crystalline state in which the resistance of the material is low. When the applied voltage is removed, the material stays in the crystalline state showing an ohmic behaviour. In case of threshold switching, the material remains to be in an amorphous state as the applied electric field increases, after V_{th} the material changes its state to a crystalline state. When the injected electric field is withdrawn the material returns to its amorphous state again [4].

The CG's are made from group 16 elements of the periodic table, which include Sulphur (S), Selenium (Se) and Tellurium (Te). The development of newer PCM material with a better electrical switching characteristic is critical for the success of PCM over conventional NVRAMs [5]. The research on the electrical switching characteristics of glasses helps in the finding of new glasses for PCM applications.

The Te based CG's are broadly studied over the past decade because of their good electrical properties and the memory switching phenomenon in Te rich glasses. Further, metal-doped Te glasses

* Corresponding author at: Electronics and Communication Engineering, BMS Institute of Technology and Management, Bangalore 560064, India.
E-mail address: jagan@bmsit.in (J. K.B.).

https://doi.org/10.1016/j.matpr.2022.01.014
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the First International Conference on Design and Materials (ICDM)-2021.

can be easily devitrified and they exhibit memory switching [6–8]. Several Si and Te based CG's have been found to exhibit memory switching phenomena Ge-Te-Si [10], Si-Te-Sb [11], Si-Te-Bi [12], Si-Te-Cu [13]. Furthermore, the inclusion of metallic alloying elements to the glasses causes significant alterations in the electrical switching behaviour of Ge-Te-Sn [14], Si-Te-Cu [15], Ge-Te-Sb [16], Ge-Te-In-Ag [17]. The addition of metallic dopant Sn, to Ge-Te samples results in memory switching behaviour of the sample and also Sn doped $Ge_2Sb_2Te_5$ samples have shown lower crystallization transition temperatures and exhibit ultra-fast crystallization (<10 ns). This motivates us to investigate the electrical switching, Raman and morphological studies on $Si_{20}Te_{74}Sn_6$ bulk material and find the feasibility of the prepared material for PCM application.

2. Methodology

A melt quenching process has been used for preparing bulk $Si_{20}Te_{74}Sn_6$ glass. Si, Te, and Sn elements ~99.999% pure of quantity 1gm is weighed using balance meter and transferred to 6 mm diameter ampoule which is cleaned using soap solution and acetone. The sample which is loaded in an ampoule is sealed at 10^{-6} mbar using flash evaporation unit, the sealed ampoule is placed at the centre of horizontal rotary furnace, and the furnace is programmed such that the temperature rises from room temperature till 1100 °C. The ampoule is quenched in a mixture of ice-cold water and NaOH after 36 h. The X-ray diffractometer (XRD) "PANalytical X'Pert3", is operated at 30 mA, 40 kV and the diffraction pattern is recorded from 20° to 80° at a rate of 3°/min.

The switching studies are carried out with LabVIEW 7 (National Instruments) and a Keithley (2410^c) source measure unit (SMU) with maximum current and voltage limits of 1A and 1100 V respectively. When the thickness of the sample is greater than 0.3 mm, power consumption is high, due to which the V_{th} increases. Bulk samples are polished to a thickness of 0.3 mm and switching studies are carried out. By applying an electric field to the sample, the current and voltage developed across the sample are captured.

Raman experiments on bulk materials are carried out using a HORIBA (LabRAM HR 800) in backscattered mode. A sample is illuminated using an argon ion laser with a 532 nm line and 0.1mW power. An acquisition period of about 60 s is used for a better signal-to-noise ratio. Scanning electron microscope (SEM) VEGA3 TESCAN LMU is used for morphological studies and elementary composition. SEM of a sample surface is taken with an acceleration potential of 15 kV, resolution at 20um and magnification of 15 kV.

3. Results and discussion

3.1. XRD analysis

The preliminary investigation of confirming the amorphous nature of prepared chalcogenide glasses is by verifying the resistance of the as-prepared $Si_{20}Te_{74}Sn_6$ bulk sample. In this study, $Si_{20}Te_{74}Sn_6$ bulk sample exhibited resistance in the mega-ohm range. Further, the nature of the prepared sample can be validated using XRD. Fig. 1 illustrates the XRD pattern of the as-prepared $Si_{20}Te_{74}Sn_6$ bulk sample. The void of sharp peaks confirms the non-crystalline nature of the prepared sample. Fig. 2 displays the EDAX spectra and constituents of the as-prepared $Si_{20}Te_{74}Sn_6$ sample. The spectra of the sample are taken at random locations on the sample and the sample composition closely matches that of the original $Si_{20}Te_{74}Sn_6$ bulk sample.

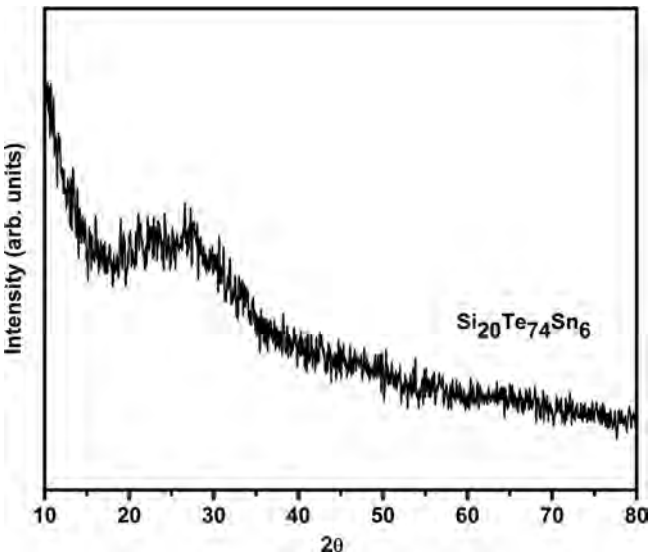


Fig. 1. XRD pattren of as-prepared $Si_{20}Te_{74}Sn_6$ bulk sample.

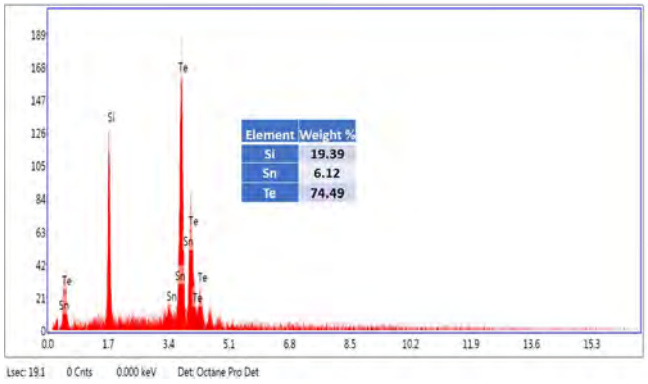


Fig. 2. EDAX spectra of $Si_{20}Te_{74}Sn_6$ bulk sample.

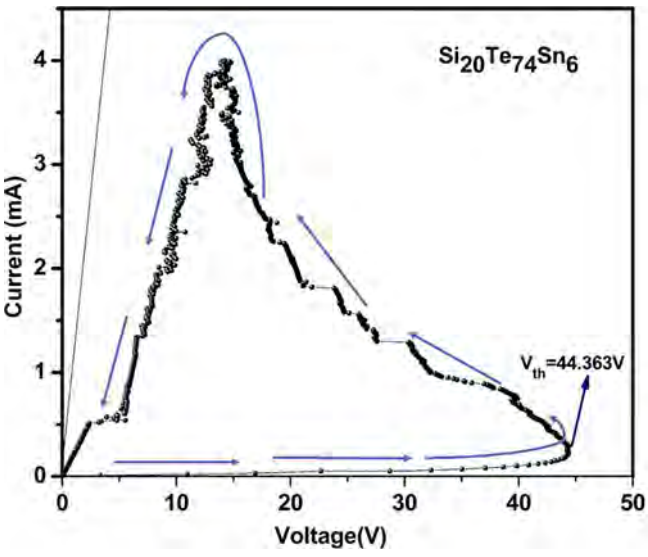


Fig. 3. Current-Voltage characteristics of $Si_{20}Te_{74}Sn_6$ bulk sample.

3.2. Switching analysis

Electrical switching analysis is an important criterion for analysing material for PCM applications. Fig. 3 describes the I-V graph of $Si_{20}Te_{74}Sn_6$ bulk sample; The I-V curve clearly shows that the sample is migrating from high resistance to low resistance state, once the applied electric field is removed the sample remains in logic '1' state and do not return to its previous amorphous state. This observation clearly shows as-prepared $Si_{20}Te_{74}Sn_6$ bulk sample exhibits a memory switching phenomenon at current of 4 mA and above, the switching characteristics are performed on four to five $Si_{20}Te_{74}Sn_6$ bulk samples and V_{th} values are repeatable within a difference of ±4 V [9].

The memory switching in CG's is instigated by the creation of a crystalline conducting channel within the sample and these samples change the phase from amorphous to crystalline easily [18]. Also, Te rich glasses frequently display memory switching phenomenon and the main factor which determines the memory switching in $Si_{20}Te_{74}Sn_6$ may be the influence of metallicity of Sn dopant in Si-Te host matrix.

3.3. Thickness v/s V_{th}

The thickness (D) variation with reference to V_{th} offers insight into the switching mechanism. In the case of memory switching samples, the V_{th} is said to be a square root or linearly related to thickness. The linear correlation is seen for thermally thick slab (TTS) whereas square root for thermally thin slab (TtS) [19]. This dependency is categorised into TTS or TtS as per the behaviour of the sample, which is determined by heat conductivity (λ) which further reflects heat dissipated by sample via the electrode and thermal conductivity (k). If λD is less than 2k, the sample operates as TTS and if λD is greater than 2k then the sample behaves as TtS. In TTS, non-uniform temperature distribution happens due to the large difference in temperature within a material and the boundary of the material. In case of TtS, the temperature drop at the boundary of the sample is more because of the high loss of heat at electrode which gives the uniform temperature distribution within the sample.

Fig. 4 displays the variation of V_{th} of a $Si_{20}Te_{74}Sn_6$ bulk glass with respect to thicknesses. From figure, it is seen that the value of V_{th} linearly increases with increasing thickness 'D'. The thickness (D) of the sample varies from 0.3 mm to 0.5 mm; similar results are seen in As-Te-Si [20], Al-Ge-Te [21] , Ge-Te-Pb [22] glasses also.

3.4. Raman studies

Fig. 5 shows the Raman graph of as-prepared $Si_{20}Te_{74}Sn_6$ glass. The Raman spectra's maximum peaks fall under 150 cm^{-1} which reiterates the non-crystalline nature of the prepared samples. The spectra have two high intense peaks and one low intense peak ranging between 50 and 200 cm^{-1}. Based on peak fitting (Gaussian) the phonon modes are ascribed to approximate peak locations at 102, 126, and 142 cm^{-1} respectively. The Raman spectra of the bulk $Si_{20}Te_{74}Sn_6$ sample closely matche those of prior studies [23–25]. Based on studies by A.S. Pine et.al on crystalline Te [23], the peak at 102 cm^{-1} is credited to A1 mode, peak at 126 cm^{-1} and 142 cm^{-1} can be attributed to the E_{T0} modes of Te-Te chains. The phonon assignment is consistent with prior Raman investigations [26]. The A1 mode is caused by the interaction of Te-Te atoms alone, whereas the ET0 mode might have be caused by the vibration of Si or Sn. The phonon mode at 102 cm^{-1} is assigned to the lower ET0 mode, which corresponds to Raman spectra recorded by Pine et al. [23].

3.5. Structural studies

Morphological studies are carried using SEM to determine the structural changes after switching. A current of appropriate magnitude is used for memory switching which has a thermal origin, the applied current creates a sufficient electric field to change the phase of the material (amorphous to crystalline) and establishes a conducting channel between electrodes causing the material to switch [8]. Fig. 6 (a) and (b) represents the unswitched and switched SEM morphology of $Si_{20}Te_{74}Sn_6$ bulk sample respectively. It is seen that the unswitched sample is smooth with no morphological changes whereas, in SEM image of switched sample a morphological change can be seen which indicates the change of phase at electrode region. Similar morphological changes are observed for memory switching phenomenon in Si-Te-Cu, Ge-Te-Si, Si-Te-Sn and Ge-Te-Sn [10,13,27,28] glasses.

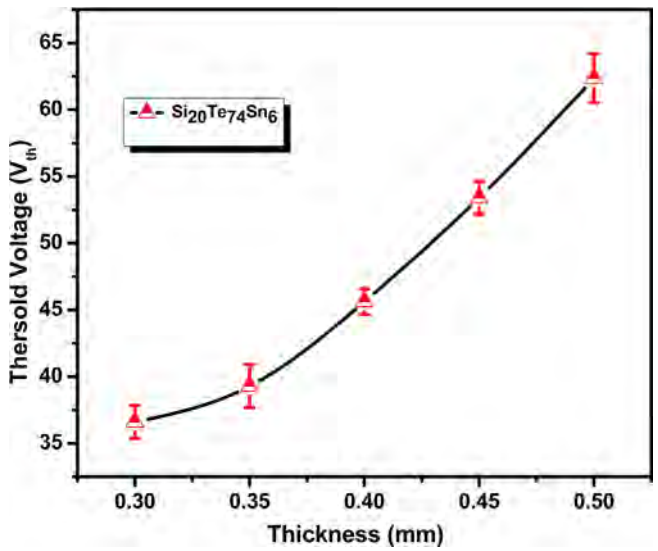


Fig. 4. Variation of V_{th} with the thickness of a $Si_{20}Te_{74}Sn_6$ bulk samples.

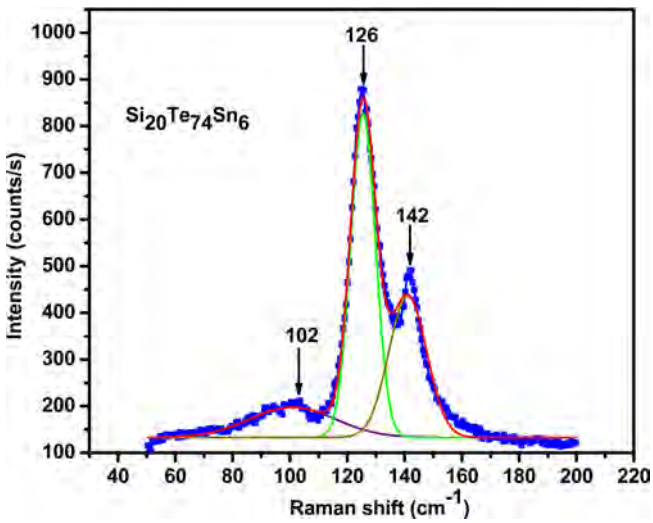


Fig. 5. Raman spectrum of as-prepared $Si_{20}Te_{74}Sn_6$ sample (Gaussian peak fitting).

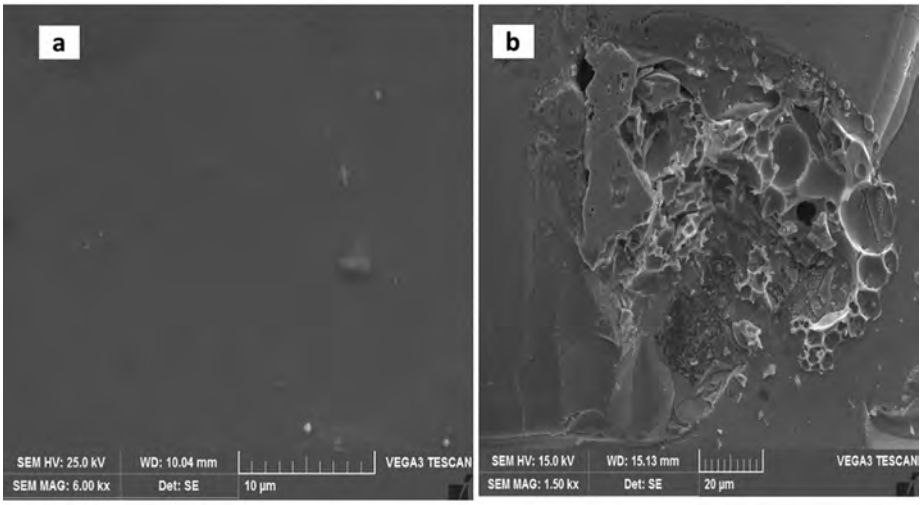


Fig. 6. SEM image of (a) un-switched and (b) switched $Si_{20}Te_{74}Sn_6$ bulk sample.

4. Conclusion

As-prepared $Si_{20}Te_{74}Sn_6$ bulk sample is synthesized using melt quenching technique and it exhibits non-crystalline and memory switching behaviour. The threshold voltage increases linearly with an increase in sample thickness, which agrees with the memory switching phenomenon of $Si_{20}Te_{74}Sn_6$ sample. The Raman analysis confirms the crystalline phase of tellurium and vibrational modes of Sn/Si lattice. The interaction of Te-Te and Si-Te atoms causes the A1 mode, whereas the E_{T0} mode may be caused by the vibration of Te and Sn atoms. SEM analysis reveals that during memory switching, there is a structural alteration between switched and unswitched samples, which may be due to the formation of the channel between two electrodes.

Jagannatha KB and Chandasree Das are equally contributed to this paper

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

Dr. Chandasree Das and Jagannatha KB acknowledge Science and Engineering Research Board (SERB) (File No.: SPG/2021/000359), and TEQIP (II), India for providing the funding support for their research work.

References

[1] S.R. Ovshinsky, Reversible electrical switching phenomena in disordered structures, Phys. Rev. Lett. 21 (20) (1968) 1450–1453, https://doi.org/10.1103/PhysRevLett.21.1450.
[2] R. Bez, A. Pirovano, Non-volatile memory technologies: Emerging concepts and new materials, Mater. Sci. Semicond. Process. 7 (4-6) (2004) 349–355, https://doi.org/10.1016/j.mssp.2004.09.127.
[3] S. Raoux, W. Wełnic, D. Lelmini, Phase change materials and their application to nonvolatile memories, Chem. Rev. 110 (1) (2010) 240–267, https://doi.org/10.1021/cr900040x.
[4] A. Pirovano, A.L. Lacaita, A. Benvenuti, F. Pellizzer, R. Bez, Electronic switching in phase-change memories, IEEE Trans. Electron Dev. 51 (3) (2004) 452–459, https://doi.org/10.1109/TED.2003.823243.
[5] M. Wuttig, Phase-change materials: towards a universal memory?, Nat Mater. 4 (4) (2005) 265–266, https://doi.org/10.1038/nmat1359.
[6] 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 (10) (1974) 1595–1601, https://doi.org/10.1007/BF00540757.
[7] G. Jones, R.A. Collins, Threshold and memory switching in amorphous selenium thin films, Phys. Status Solidi (a) 53 (1) (1979) 339–350.
[8] C. Das, M.G. Mahesha, G.M. Rao, S. Asokan, Electrical switching studies on amorphous Ge-Te-Sn thin films, AIP Conference Proceedings, 1349(PART A) (2011) 633–634, doi: 10.1063/1.3606017.
[9] K.B. Jagannatha, D. Roy, C. Das, Electrical switching and crystalline peak studies on Si20Te80−xSnx ($1 \leq x \leq 7$) chalcogenide bulk glasses, J. Non-Cryst. Solids 544 (2020) 120196, https://doi.org/10.1016/j.jnoncrysol.2020.120196.
[10] 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. 40 (23) (2007) 7515–7518, https://doi.org/10.1088/0022-3727/40/23/040.
[11] R. Lokesh, N.K. Udayashankar, S. Asokan, Electrical switching behavior antimony doped bulk Si-Te chacogenide glasses, 3(2) (2011) 55-60.
[12] B.J. Fernandes, P. Munga, K. Ramesh, N.K. Udayashankar, Thermal stability and crystallization kinetics of Bi doped $Si_{15}Te_{85-x}Bi_x$ ($0 \leq x \leq 2$) chalcogenide glassy alloys, Mater. Today:. Proc. 5 (8) (2018) 16237–16245, https://doi.org/10.1016/j.matpr.2018.05.115.
[13] D. Roy, B. Tanujit, K.B. Jagannatha, S. Asokan, C. Das, Influence of Cu doping in Si-Te-based chalcogenide glasses and thin films: electrical switching, morphological and Raman studies, IEEE Trans. Electron Dev. 68 (3) (2021) 1196–1201, https://doi.org/10.1109/TED.2021.3051925.
[14] 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. Alloy. Compd. 721 (2017) 674–682, https://doi.org/10.1016/j.jallcom.2017.06.070.
[15] D. Roy et al., Electrical switching studies on $Si_{15}Te_{85-x}Cu_x$ bulk ($1 \leq x \leq 5$) glasses, 1966 (2018) 2003, doi: 10.1063/1.5038712.
[16] K. Nishimura, M. Suzuki, I. Morimoto, K. Mori, Ge–Te–Sb based overwritable phase change optical disk, Jpn. J. Appl. Phys. 28 (S3) (1989) 135, https://doi.org/10.7567/JJAPS.28S3.135.
[17] G. Sreevidya Varma, D.V.S. Muthu, A.K. Sood, S. Asokan, Electrical switching, SET-RESET, and Raman scattering studies on Ge15Te80−xIn5Agx glasses, J. Appl. Phys. 115 (16) (2014) 164505, https://doi.org/10.1063/1.4873237.
[18] S.R. Ovshinsky, H. Fritzsche, Amorphous semiconductors for switching, memory, and imaging applications, IEEE Trans. Electron Dev. 20 (2) (1973) 91–105, https://doi.org/10.1109/T-ED.1973.17616.
[19] J.J. O'Dwyer, The Theory of Electrical Conduction and Breakdown in Solid Dielectrics Clarendon, Clarendon Press, Oxford, 1973.
[20] M. Anbarasu, S. Asokan, Electrical switching behavior of bulk As-Te-Si glasses: composition dependence and topological effects, Appl. Phys. A Mater. Sci. Process. 80 (2) (2005) 249–252, https://doi.org/10.1007/s00339-004-3030-2.
[21] J.R. Bosnell, C.B. Thomas, Preswitching electrical properties, 'forming', and switching in amorphous chalcogenide alloy threshold and memory devices, Solid State Electron. 15 (11) (1972) 1261–1271, https://doi.org/10.1016/0038-1101(72)90047-0.
[22] P.Z. Saheb, S. Asokan, K.A. Gowda, Electrical switching studies of lead-doped germanium telluride glasses, Appl. Phys. A 77 (5) (2003) 665–668, https://doi.org/10.1007/s00339-002-1759-z.
[23] A.S. Pine, G. Dresselhaus, Raman spectra and lattice dynamics of tellurium, Phys. Rev. B 4 (2) (1971) 356–371, https://doi.org/10.1103/PhysRevB.4.356.
[24] J. Sun, Q. Nie, X. Wang, S. Dai, X. Zhang, B. Bureau, C. Boussard, C. Conseil, H. Ma, Structural investigation of Te-based chalcogenide glasses using Raman

spectroscopy, Infrared Phys. Technol. 55 (4) (2012) 316–319, https://doi.org/10.1016/j.infrared.2012.03.003.
[25] S.R. Gunti, S. Asokan, Thermal and electrical switching studies on $Ge_{20}Se_{80-x}Bi_x$ ($1 \leq x \leq 13$) ternary chalcogenide glassy system, J. Non-Cryst. Solids 356 (33–34) (2010) 1637–1643, https://doi.org/10.1016/j.jnoncrysol.2010.06.028.
[26] S.R. Gunti, A. Sundarrajan, Observation of high pressure o-GeTe phase at ambient pressure in Si-Te-Ge glasses, AIP Adv. 2 (1) (2012) 012172, https://doi.org/10.1063/1.3696862.
[27] C. Das, G.M. Rao, S. Asokan, The electrical switching and thermal behavior of bulk $Ge_{15}Te_{85-X}Sn_x$ and $Ge_{17}Te_{83}$ - XSnx glasses, J. Non-Cryst. Solids 358 (2) (2012) 224–228, https://doi.org/10.1016/j.jnoncrysol.2011.09.021.
[28] K.B. Jagannatha, D. Roy, B. Tanujit, S. Asokan, C. Das, A composition-dependent thermal behavior of $Si_{20}Te_{80-x}Sn_x$ glasses: observation of Boolchand intermediate phase, J. Non-Cryst. Solids 577 (121311) (2021) 1–2.

2020 3rd International Conference on Energy, Power and Environment: Towards Clean Energy Technologies | 978-1-6654-2536-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICEPE50861.2021.9404519

An Accelerated Approach to Parallel Ensemble Techniques Targeting Healthcare and Environmental Applications

Tejasri Kari*, Leelavani N *, Sayeera Banu A*, DhanuShree R*, Jagannatha K.B* and Santhi Natarajan, IEEE Member†
*Department of Electronics and Communication Engineering, †Department of Artificial Intelligence and Machine Learning,
BMS Institute of Technology and Management, Visvesvaraya Technological University, Belgaum, Karnataka, India 590018
1by16ec036@bmsit.in; 1by16ec045@bmsit.in; 1by16ec106@bmsit.in; 1by16ec108@bmsit.in;
jagan@bmsit.in; santhi.natarajan@bmsit.in

***Abstract*—Ensemble learning techniques adopt comprehensive learning methodologies that produce optimized predictions with reduced variance and bias. The structured Random Forest ensemble learning technique equips a set of weak and diverse decision trees, resulting in an active hybrid learning ensemble. Plagued with high computational complexity, Random Forest Ensemble continues to be the preferred technique when accuracy is of primary importance for learners. Efforts to accelerate the Random Forest Ensembles are in place, however failing to efficiently utilize the data transmission bandwidth between the host and the accelerator hardware. This paper provides an architectural overview of a reconfigurable accelerator based architecture of the Random Forest Ensemble with an efficient data path model for data streaming. The paper also derives the need for an accelerated parallel ensemble method by deriving the results from equivalent sequential software implementations of the algorithm. The validation of the results have been done on healthcare application involving breast cancer classification and environmental applications involving temperature prediction and fuel consumption.**

I. Introduction

A. Ensemble Learning: A Computational Perspective

Ensemble methods bring in random perturbations into the learning techniques and generate multiple variants/models from a single learning set L. Later, we combine the predictions of those models to form the prediction of the ensemble [1].

The Random Forest Ensemble recursively partitions the training vector space resulting in grouping together the samples with the same labels. Consider a set of training vector $\mathcal{X}$ with a corresponding set of labels y. Let N_i be the i^{th} node and d_i be the data contained in N_i. A split is made for each node, based on the split value $SV = \mathbf{F}(f, \theta)$ where f is the feature subset of F and θ is a threshold value. The split divides the data into two subsets: $D_\alpha(\theta)$ and $D_\beta(\theta)$ and the number of training samples into n_α and n_β. N_i is the total number of nodes.

$$\mathcal{X} = \{x_1, x_2, x_3, \dots x_M\} \tag{1a}$$

$$y = \{y_1, y_2, y_3, \dots y_M\} \tag{1b}$$

$$D_\alpha(\theta) = (x, y)|x_f <= \theta \tag{1c}$$

$$D_\beta(\theta) = (x, y)|x_f > \theta \tag{1d}$$

Algorithm 1 Ensemble Learning: Random Forest

// Input: Data Set D = $\{(x_1, y_1),(x_2, y_2), \dots ((x_m, y_m)\}$, Feature Set F, Randomization Factor R, Number of trees T
// Output: Root node of i[th] tree
- -
for $\forall i \in \{1, 2, \dots T\}$ **do**
  $N_i \leftarrow$ Root node of i[th] tree
  if All targets belong to same class i.e y_i or $F \in \emptyset$ **then**
    Return N_i
  end if
  $D_i \leftarrow$ bootstraped sample from D
  for Each node **do**
    $f \leftarrow$ Randomly selected R features from F
    $N_f \leftarrow$ Best Feature from f features
    $N_p \leftarrow$ Best Split based on N_f
  end for
end for
return N_i

The error at the i^{th} node is the measure of misclassification. The most popular functions to calculate the error are the Gini Index, Entropy, or Misclassification. The mean squared error or the mean absolute error can be used for regression [2].

$$H_{gini}(x) = \sum_{\forall y_{pred}} y_{pred}(1 - y_{pred}) \tag{2a}$$

$$H_{entropy}(x) = -\sum_{\forall y_{pred}} y_{pred}(log(y_{pred})) \tag{2b}$$

$$H_{gini}(x) = \sum_{\forall y_{pred}} 1 - max(y_{pred})) \tag{2c}$$

The loss function can be expressed as

$$L(d, SV) = \frac{n_\alpha}{N_i} H(d_\alpha(SV)) + \frac{n_\beta}{N_i} H(d_\beta(SV)) \tag{3}$$

The ensemble learning method is represented in Algorithm 1 and the dataflow is represented in Fig. 1a.

To train the Random Forest Ensemble, the following stepwise approach is used, by which n numbers of trees in steps 1 through 4 for repeating n times.

1) Step 1: Randomly choose the k features from the total m features, where $k \ll m$
2) Step 2: Use the best-split point for node d calculation, among the k features.
3) Step 3: Using best-split, divide the node into daughter nodes.

978-1-6654-2536-0/21/$31.00 ©2021 IEEE

TABLE I: Random Forest Parallel and Accelerated Versions: State of the Art

Work	Features	Platform	Specifications
PARF [4]	A task parallel version of training and learning phase of Random Forest ensemble.	Fortran	Cluster hardware with MPI (Message Passing Interface).
FastRF [5]	A task parallel version with speed and memory advantage of Random Forest ensemble	Weka	Multithreading enables a speedup of 2.3X.
CudaRF [6]	A task parallel GPU version of learning and classification phases of Random Forest ensemble	CUDA	Upto 9.2X speedup achieved on NVIDIA GPUs
PLANET [7]	Parallel Learner	Hadoop	Map-Reduce model.
SPRINT [8]	Parallel R Interface for Random Forest ensemble based microarray analysis	C	MPI and multi-core
Randomjungle [9]	A task parallel version of ensemble classifier for microarray data	C++	Analysis of big Genome Wide Association (GWA) data
FPGA1 [10]	Provides multi-value decision diagrams (MDDs), at order $O(2^N)$ that increases the number of nodes within MDD	FPGA	10.7X speed up over GPU, 14.0X speedup over CPU
FPGA2 [11]	Parellel ensemblefor letter and digit recognition.	FPGA, Co-simulation C / RTL	SoC unit Zynq-7000 is used.
FPGA3 [12]	A task parallel version of Random Forest ensemble on reconfigurable hardware.	FPGA, Co-simulation C / RTL	20X speedup over CPU version
FPGA4 [13]	A compact Random Forest training model.	FPGA/GPU/Multicore	The optimization allows the classifier to be performed in a pipeline or single-instruction type multi-thread (SIMT).

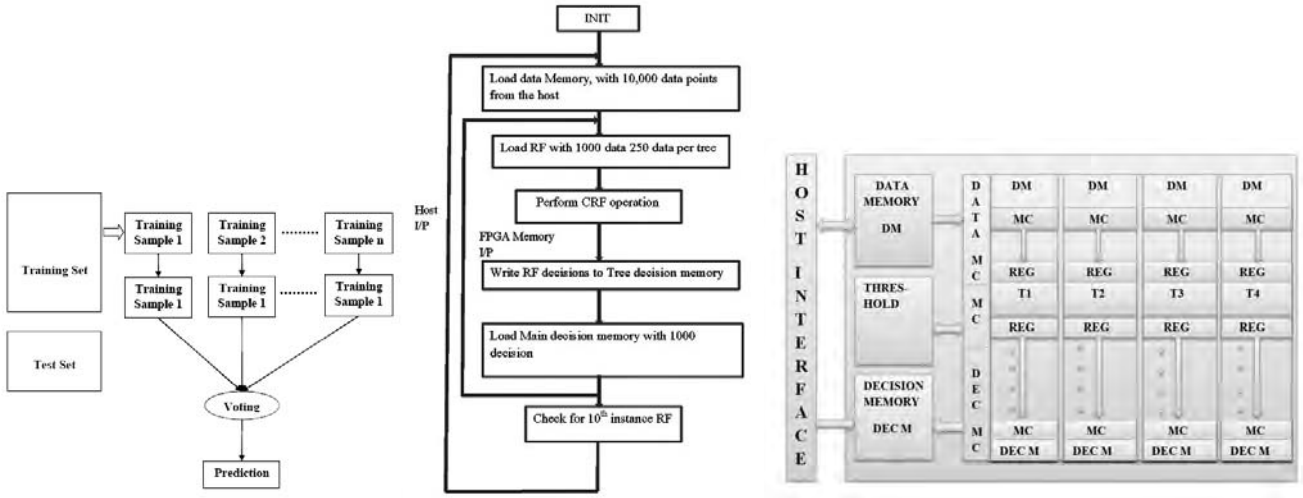


Fig. 1: From Left to Right: (a) The Ensemble Learning Architecture on Reconfigurable Hardware (b) The Ensemble Learning Data Path on Hardware (c) The Ensemble Learning Algorithm

4) Step 4: Repeat steps 1 to 3, until the number of nodes hits l
5) Choose the results of most votes as the final prediction

To predict, we need to pass test features through the rules of each tree by using the trained Random Forest algorithm that have been created at random. The ensemble resorts to majority voting to arrive at the final valid predictions. Here, the class probability estimates are averaged to predict the most likely class as output, thereby minimizing the average zero-one error of individual predictions.

Thus, Random Forest Ensemble presents the following advantages:

1) Categorical values can also be modelled on a Random Forest classifier.
2) For both classification and regression tasks, the same Random Forest algorithm and Random Forest classification can be used.
3) The missing values will be dealt with by the Random Forest classifier.
4) If there is more number of trees in the forest, the Random Forest classifier will not overfit the model.

B. Parellel Ensembles: A Review of the State of the Art

With its inherent concurrent memory access and computing parallelism, Field Programmable Gate Arrays (FPGA) based reconfigurable accelerators are good platforms for accelerating Random Forest Performance. A comparative listing of the existing solutions of parallel implementations of Random Forest Ensemble are provided in Table I. It is desirable to reduce the costs of implementing FPGA between various applications, but the contextual time between applications has a considerable effect on the forest architecture. Random forest, which comprises a large number of decision trees that use FPGA, has high speed of training. Apart from FPGAs, the state-of-the-art designs include Graphic Processing Unit (GPU) based accelerators and other multi-core parallel implementations of Random Forest Ensemble [3].

This work presents an accelerator based architectural model for parallel Random Forest ensemble, exploring features of the Random Forest to accelerate the Random Forest Training step. FPGA shows high memory efficiency, as well as a batch training strategy to fully exploit the high memory bandwidth that the chip-based memory on FPGA devices offers. The

Algorithm 2 Parallel Ensemble Learning: Random Forest

// Input: Data Set D = $\{(x_1, y_1),(x_2, y_2), \ldots ((x_m, y_m)\}$, Feature Set F, Randomization Factor R, Number of trees T C: Number of nodes / Number of cores available in parallel setup
// Output: A parallel Random Forest R
- -
if T is greater than N **then** $K = N/T$
  for $\forall j \in \{1, 2, \ldots K\}$ **do**
    for $\forall i \in \{1, 2, \ldots T\}$ **do**
      $N_i \leftarrow$ Root node of i$^{\text{th}}$ tree
      if All targets belong to same class i.e y_i or $F \in \emptyset$ **then**
        Return N_i
      end if
      $D_i \leftarrow$ bootstraped sample from D
      for Each node **do**
        $f \leftarrow$ Randomly selected R features from F
        $N_f \leftarrow$ Best Feature from f features
        $N_p \leftarrow$ Best Split based on N_f
      end for
    end for
  end for
end if
return N_{i_K}

Algorithm 3 : FindBestSplitRandom($\mathcal{L}_t$, K)

// Purpose: To find the best split s^* that partitions $\mathcal{L}_t$, among a random subset of $K \leq p$ input variables
// Input: $\mathcal{L}_t$, K
// Output: s^*
- -
$\Delta = -\infty$
Draw K random indices j_k from 1 to p
for $k = 1$ *to* K **do**
  Find the best binary split $s^*_{j_k}$ defined on X_{j_k}
  if $\Delta i(s^*_{j_k}, t) > \Delta$ **then**
    $\Delta = \Delta i(s^*_{j_k}, t)$
    $s^* = s^*_{j_k}$
  end if
end for
Return s^*

Algorithm 4 : FindRandomSplit-PERT($\mathcal{L}_t$, X_j)

// Purpose: Draw a random split on X_j that partitions $\mathcal{L}_t$
// Input: $\mathcal{L}_t$, X_j
// Output: s^v_j
- -
$min_j = min\{x_{i,j} | (x_i, y_i) \in \mathcal{L}_t\}$
$max_j = max\{x_{i,j} | (x_i, y_i) \in \mathcal{L}_t\}$
Draw v uniformly at random from $[min_j, max_j]$
return s^v_j

Algorithm 5 : k loops

// Purpose: To find the number of k loops required for Parallel algorithm
// Input: n, p
// Output: k
// Parameters: p - Number of processors, n - Number of trees to be generated
- -
$k = \frac{n}{p}$
if $n = p$ **then**
  $k = 1$
else if $n < p$ **then**
  $k = 1$ (max)
else
  $k > 1$
end if

proposed architecture can not only be applied in classification / regression based problems [14], but also for outer detection, clustering and interpretation purposes. The work also highlights the implemenation of a basic Random Forest ensemble on software (Python and MATLAB) and brings out the need

Algorithm 6 : Cost of Parallel algorithm

// Purpose: To find the cost of Parallel algorithm
// Input: $p(n)$, $W(n)$
// Output: Cost $C(n)$
- -
$C(n) = p(n) \times W(n)$
$= p \times k$
$= n$
$= W^*(n)$

for accelerator and parallel ensemble, based on the errors and computational time in each case.

II. Ensemble Learning on Reconfigurable Hardware: An Algorithm-Architecture Codesign

A. The Random Forest Ensemble: A Parallel Approach

Here, we present the Random Forest Ensemble that adopts a Single Instruction Multiple Data (SIMD) approach, that can parallely build an ensemble of decision trees, as decision trees usually present high variance and low bias and are most expected to take advantage from the averaging process. These trees grow with random perturbations in the tree induction procedure, resulting in an assortment of Random Forest ensembles.

We present a generalized parallel training/learning algorithm in Algorithm 2. The proposed method provides a hybrid approach to parallel ensembles. While having a task parallel implementation in generating parallel Random Forests, the individual trees inducted to the ensemble are also parallelized. The proposed parallelization is applicable both at the learning and prediction/classification phase.

For applications with very large number of input variables, a randomized variant of the tree induction algorithm is used, that searches for the best split at each node over a random subsample of the variables. This is presented in Algorithm 3.

Perfectly fit decision trees are grown, where we randomly choose the split variable and discretization threshold, as designed in Perfect Random Tree Ensembles (PERT). This method is explained in Algorithm 4. At node t, the split variable X_j is drawn at random using Algorithm 3 with $K = 1$. The cut-point v is set midway between two randomly drawn samples. A tree is inducted until all nodes become pure or until it is no longer possible to draw samples of different output values. This approach might sometimes result in larger decision trees grown with minimal randomization.

B. Ensemble Learning on Reconfigurable Hardware: The Datapath Architecture

The architecture supporting the parallel implementation of the Random Forest Ensemble is presented in Fig. 1c. This is designed to run on parallel hardware like FPGAs and GPUs. The focus of this section is on the datapath architecture, enabling the parallel implemenation discussed above. Here, there are three memory stages taken into consideration i.e., Data memory, Threshold memory and Decision memory. The host memory is a space where all datasets are stored. Threshold memory is the place where all the threshold values for the

features in the dataset are stored and Decision memory is used to store the final decisions of the trees. There are three memory controllers i.e., Data memory controller, threshold memory controller, and Decision memory controller. The threshold values from threshold memory and values from the datasets are compared to form the decision trees. From the Host interface randomly for instance, 10,000 datasets are taken into data memory and randomly 1000 datasets are passed to data memory controller as shown in Fig. 1b. Then the datasets get divided into 250 datasets each for a tree which are further passed to data memory and memory controller of the trees which are then stored in a register. The threshold values from threshold memory and values from the datasets are compared to form the decision trees. The final decisions of the trees are stored in tree decision memory through memory controllers. Finally, for each predicted target calculate the votes and regard the highly voted target as the final random algorithm prediction.

C. Ensemble Learning on Reconfigurable Hardware: Performance Analysis

The $SIMD$ model for task parallelism of Random Forest classifier is defined by the parallel basic operation, common to all active decision trees at an instance of time. Worst case complexity $w(n)$ of the proposed parallel model using $p(n)$ processors shall be the maximum number of parallel basic operations over all inputs of size n.

The speedup of the proposed parallel algorithm is:

$$S(n) = \frac{W^*(n)}{W(n)} \tag{4}$$

where, $W^*(n)$ is the smallest worst case complexity over all known sequential algorithms for solving problem in hand.

Cost of the parallel ensemble is given by:

$$C(n) = P(n) \times W(n) \tag{5}$$

For judging the quality of the proposed model in Algorithm 2, the cost is compared with that of sequential version with cost and speed emphasis (Algorithm 1) $(W^*(n) \times 1)$. The parallel model is cost optimal if $C(n) = W^*(n)$.

The worst case sequential time is given by:

$$W^*(n) = f(n) \tag{6}$$

The number of k loops needed for Parallel algorithm is presented in Algorithm 5.

Hence, worst case time of Parallel algorithm is a function of k i.e,

$$W(n) = f(k) \tag{7}$$

From Eq 6 and Eq 7,

$$Speedup = \frac{W^*(n)}{W(n)} = \frac{n}{k} \tag{8}$$

Thus Speed up is always greater than 1.

Hence, Parallel algorithm is cost optimal as proved in Algorithm 6.

Thus, this section highlights the parallel realization of Random Forest Ensemble and the associated performance analysis parameters. In the next session, we present the results of the sequential implementation of the algorithm in Python and MATLAB to derive the importance of the parallel version of the algorithm.

III. Results and Discussions

We started with the sequential implementation of Random Forest Ensemble to derive the need for an accelerated parallel verison.

We have implemented the Random Forest Ensemble in its sequential form in both Python and MATLAB to understand the practical aspects of randomization and depth of inducted decision trees, when they move to the parallel platform. They are then applied to four different benchmark data sets, where we are willing to compromise on the computational cost to obtain high performance while working with mission critical data. The four data sets are for prediction of temperature, iris flower classification, petrol consumption prediction and breast cancer classification respectively. The results from the experimental set up are presented below.

A. Temperature Prediction

The problem predicts the maximum in Seattle tomorrow with one year of the past weather data. A part of the Random Forest is taken and explained in the image Fig. 2a. One interesting note is that, although there are 261 training data points, only 162 samples can be found in the root node. This is because a random subset from the data points with a replacement is applied to each tree in the forest. Random data sampling, together with a random sample of a subset of the features at the tree nodes, is the reasons why the model is termed a Random Forest.

The temperature prediction results for the implemented python code are shown in Fig. 3a, which captures the graph plotted between importance and variable. To quantify the utility of all variables in the whole Random Forest, we see the relative importances of variables. The importances show how much the prediction improves, including a certain variable. Fig. 3b shows the graph plotted between the Maximum temperatures and date, with actual and predicted values. Fig. 3c shows the graph plotted between the Maximum temperature and date which is a slight modification of Fig. 3b. The temperature data set is presented in Fig. 3d.

B. Petrol Consumption Prediction

The tree structure for the forest ensemble is given in Fig. 2b. The Mean Absolute Error value is 47.9825, with Mean Squared Error (MSE) at 3469.7007375, Root Mean Squared Error (RMSE) at 58.9041657058. The MSE is 47.9825 which are less than 10 percent of the average petrol consumption i.e. 576.77. It indicates that we have used enough number of estimators (n_estimators). The n_estimator parameter indicates the no. of trees in the Random Forest. In this, we used 200 n_estimators. The error values decrease with the number of estimators increasing.

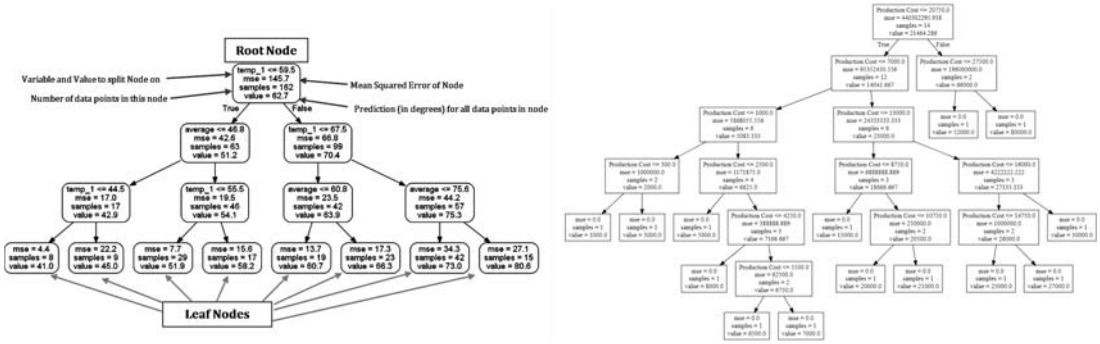
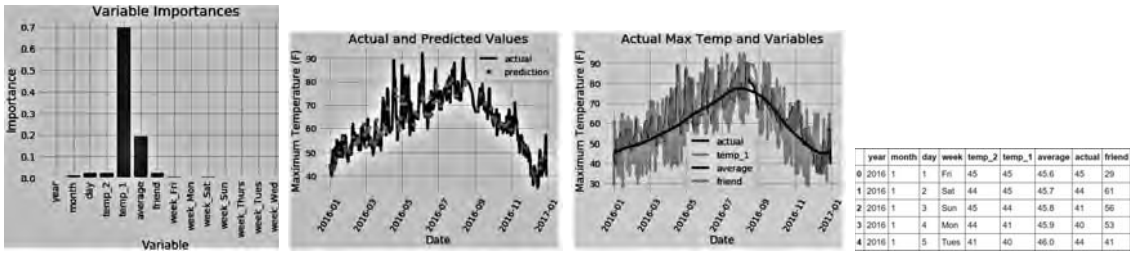


Fig. 2: From Left to Right: (a) The tree structure for temperature prediction (b) The tree structure for petrol consumption prediction



	year	month	day	week	temp_2	temp_1	average	actual	friend
0	2016	1	1	Fri	45	45	45.6	45	29
1	2016	1	2	Sat	44	45	45.7	44	61
2	2016	1	3	Sun	45	44	45.8	41	56
3	2016	1	4	Mon	44	41	45.9	40	53
4	2016	1	5	Tues	41	40	46.0	44	41

Fig. 3: The results of Temperature Prediction: From Left to Right: (a) Variable Importances, (b) Actual versus predicted values, (c) Actual Max Temperature and Variables, (d) Temperature Datasets

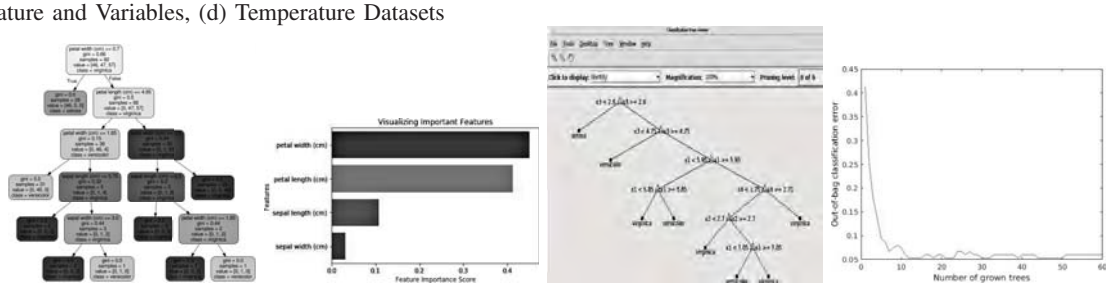


Fig. 4: The results of IRIS Classification: From Left to Right: (a) Tree Structure, (b) Visualizing Salient Features, (c) MATLAB Decision Tree, (d) OOB Errors Versus No of Grown Trees

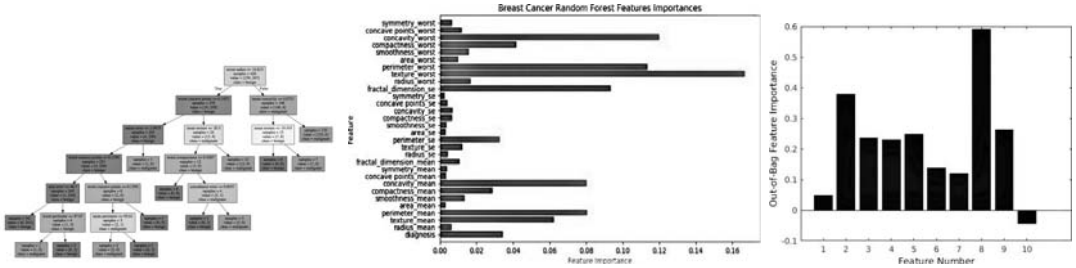


Fig. 5: The results of Breast Cancer Classification: From Left to Right: (a) Decision Tree, (b) Feature Importances, (c) OOB feature importance versus feature number

C. IRIS Classification

A model is taken on the iris flower dataset which includes length of the sepal, width of the sepal, length of the petal, width of the petal, and type of flowers. Three types exist: setosa, versicolour, and virginica. The tree structure for this application is given in Fig. 4a. The results for the iris classification for python implementation is shown in Fig. 4b.

1) MATLAB Results for IRIS Classification: The output shown in the Fig. 4c is one of the decision trees of iris Random Forest. Here the tree predicts classifications based on four predictors x1, x2, x3 and x4 which correspond to sepal length, sepal width, petal length and petal width which are compared against threshold values. Out-of-bag error is an estimation technique which is used to measure the prediction error of Random Forest. When we train each tree in Random Forest, we will not use all the data samples. Those unused samples can be used as test data and are called out-of-bag data samples which are used to find prediction error/ classification error. The graph in Fig. 4d shows the out-of-bag classification errors for certain number of trees.

D. Breast Cancer Classification

The work is with Sklearn breast cancer data sets to build a Random Forest classification model that classifies breast cancer tumours as either “malignant” or “benign”. The tree structure is given in Fig. 5a, with the Worst radius set to <=16.815. The best obtained training set accuracy is 0.9930

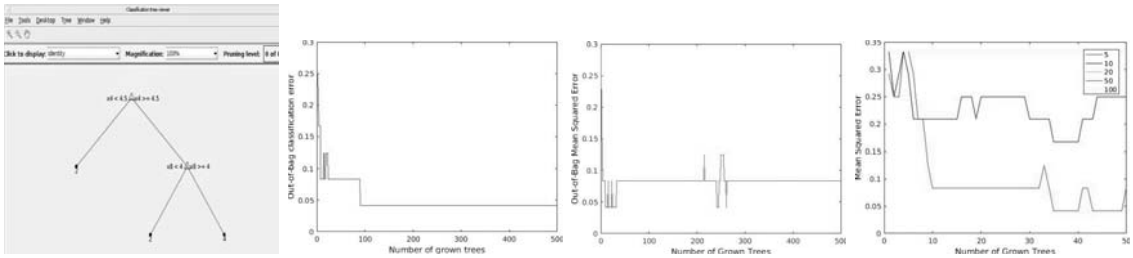


Fig. 6: The results of Breast Cancer Classification in MATLAB: From Left to Right: (a) MATLAB Decision Tree, (b) OOB classification error versus number of grown trees, (c) OOB mean square error versus number of grown trees, (d) Mean Square Error versus number of grown trees

Sl.No	Language	Datasets	Time complexity
1.	Python	Breast Cancer	2.314222
		Iris	2.184154
2.	Matlab	Breast Cancer	0.8048563
		Iris	5.6516504

Fig. 7: Ensemble Learning with RF: Python Versus MATLAB

and testing set accuracy is 0.9441. As shown in the Fig. 5b the features with high importance are selected and with low importance are removed. The out-of-bag feature importance estimates in Fig. 5c indicate more importance features in unused data samples (test data). The bar graph suggests that feature number 8 is more importance feature followed by feature number 2, 9 and 5.

The decision tree is captured in Fig. 6a. The out-of-bag classification errors of number of grown trees is shown in Fig. 6b, where we can see that the out-of-bag error varies from 0.25 to 0.1. After 100 to 500 grown trees, the out-of-bag error remains constant at 0.05. Fig. 6c is the graph showing the mean squared error of out-of-bag data samples for the number of grown trees (The lower MSE value indicate the better model) and the MSE of data sets for number of grown trees is shown in Fig. 6d.

The performance comparison for the python and MATLAB model are given in Fig. 7. As seen in the above results, the performance of the Random Forest Ensemble, in its sequential form, shows high dependence on the splits with the random nature of the learning sets. The high cut point variance seems to be responsible for the significant part of generalization error in the tree structures. An attempt to smooth the decision tree boundaries would be to attempt extreme randomization, as proposed in Algorithm 4.

Extremely Randomized Trees become Totally Randomized Trees, when a single variable X_j and a discretization threshold v are drawn at random at each node. This results in a tree structure that can learn independent of output variable Y. This is a closest fit to Perfect Random Tree Ensembles. This is best realized in the parallel solution offered in Section II. Thus the results and the subsequent discussion here derive the importance of a parallel ensemble realization, to optimize performance. The accelerated version of the parallel implementation is sure to have an edge in performance over the software thread-level parallelization, as we have discussed in the state-of-the-art methods.

IV. Conclusion

This paper highlights the parallel architecture designed for realizing parallel Random Forest Ensemble Model. This is an algorithm-architecture codesign effort, taking advantage of the parallel architecture of the reconfigurable hardware and the inherent parallel nature of inducted trees in the Random Forest ensemble. While the state-of-the-art parallel realizations focus on the task level parallelism, we focus on the micro-level parallelism, achieved at both ensemble level and at the individual tree level. This architecture is applicable to a wide range of applications including healthcare, environment monitoring, market assessment etc.

References

[1] Omer Sagi Lior Rokach*"Ensemble learning: A survey"*, Advanced Review, pp. 1-14, February 2018.

[2] SL. Breiman *"Random forests"*, Machine Learning. 45(1), pp. 5-32, 2001.

[3] Muhsenowaida, HantianZhang, Ce Zhang and Gustavo Alonso, *"Scalable inference of decision tree ensembles:flexible design for cpu-fpga platforms"*, IEEE, 2017

[4] Wojciech K, P J Czochanski, *"A Hybrid MPI/OpenMP Parallelization of K-Means Algorithms Accelerated Using the Triangle Inequality"*, IEEE Access 7, 42280-42297, 2019

[5] M N Wright, A Ziegler, *"A fast implementation of random forests for high dimensional data in C++ and R"*, arXiv preprint arXiv:1508.04409, 2015

[6] Grahn, H., Lavesson, N., Lapajne, M. H., Slat, D. (2011). *"CudaRF: A CUDA-based implementation of Random Forests"*, 2011 9th IEEE-ACS International Conference on Computer Systems and Applications (AICCSA).

[7] B PandaJoshua S. HerbachSugato B J. BayardoRoberto J. Bayardo, *"PLANET: Massively parallel learning of tree ensembles with MapReduce"*, August 2009 Proceedings of the VLDB Endowment 2(2):1426-1437

[8] L Mitchell, Terence M. Sloan, M Mewissen, Peter Ghazal, T Forster, Michal Piotrowski and Arthur Trew, *"Parallel classification and feature selection in microarray data using SPRINT"*, Concurr Comput. 2014 Mar 25; 26(4): 854–865. doi: 10.1002/cpe.2928

[9] Q. Wu, Y. Ye, Y. Liu and M. Ng, *"Stratified Random Forest for Genome-wide Association Study"*, 2011 IEEE International Conference on Bioinformatics and Biomedicine, Atlanta, GA, 2011, pp. 10-15.

[10] Chuancheng, Christos-Savvas Bouganis, *"Accelerating random forest training process using fpga"*, IEEE, 2013

[11] Hiroki nakahara, Akirajinguji, Simpeisato and tsutomusaso, *"A random forest using a multi-valued decision diagram on an fpga"*,2011 IEEE 47th International symposium on multiple-valuedlogic,pp.266-271

[12] Rafal Kulaga Marek Gorgon, *"Fpga implementation of decision trees and tree ensembles for character recognition in vivado HLS"*, 2015 Image processing and communication,vol.119,no.2-3,pp.71-82,ipc-2015-0012

[13] B.V.Essen, C.Macaraeg, M.Gokhale and R.Prenger, *"Accelerating a Random forest classifier :Multi-core, GP-GPU or FPGA"*, it IEEE Annual Int'l Symp. on field- programmable custom computing machines(FCCM), 2012

[14] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen., *"Classification and Regression Trees"*, Chapman and Hall/CRC, Boca Raton, FL, 1984