International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

The Robust Computer Aided Diagnostic System for Lung Nodule Diagnosis

Popuri Ramesh Babu, Inampudi Ramesh Babu, S.V.N.Srinivasu

Abstract**:** ***Lung cancer has been the numerous dangerous among all other variations of cancer. The fast detection of cancer is conjectured to improve the persistence rate of people living with cancer. Our objective is to present an adequate Computer-Aided Diagnosis (CAD) for the identification of lung nodules from the parenchyma area of the lung and yield the nodule into except cancerous or non-cancerous. In this suggestion, A new Hybrid Classifier method has been Prescribed to detect lung nodules based on numerous image processing and machine learning approaches. The construction of this hybrid system is the combination of unsupervised Enhanced Fuzzy C-Mean (EFCM) clustering and Weighted supervised support vector machine (WSVM). The suggested process includes the subsequent operations: i) the image used is magnified originally. Then the area of concern is cropped, where the user can choose the area to be cropped.***

ii) The morphological process is implemented to overcome the blood vessels and magnify the nodules. iii)Nodules are distinguished by labeling.iv)Those classified nodule's characteristics are obtained.v)Neural networks are performed as the classifiers that work primarily based on the features chosen. And also, this proposed flexible computing system was associated with the various well-known learning Techniques. The WSVM for analysis is exhibited in this paper, where the execution of misclassification for each practice sample is unusual. The proposed work was capable of detecting the lung nodule that appears near the lung wall. The Provisional results intimate that the recommended method defeats the impact of outliers and yields higher classification speed than previous algorithms.

Keywords **: Lung Nodule, Computer Aided Diagnosis, CT image, Segmentation, WSVM..**

I. INTRODUCTION

Most common Lung cancers that usually occur in women and men. According to are port presented with the help of the American Cancer Society in 2003, most lung cancers can provide about 13% of all diagnoses for most cancer sand28%ofall cancer deaths. The maximum survival rate for lung cancer analyzed at 5 years is better than 15%. If the disorder is determined at the same time due to mileage, this load will increase to 40-9%. However, the most 15% benefit of specific lung cancers is at this early stage. The survival rate for most people with cancer can be improved by detecting most cancers at previous levels. Early detection can be performed in a population; The most unusual detection of types of lung cancer uses low-dos ex-rays or CT scans. In the first plates that do lung cancer, it was noted that low-dose CT scans are more expensive than conventional chest radiography to detect pulmonary nodules [1].

Nodule sizes will vary widely: Typically, the diameter of nodules can take any value between a few millimeters and several centimeters.

Nodules exhibit a large variation in density Therefore, X-ray vision (some nodules are only more effective than the surrounding lung tissue,while this density is calcified).Since no dules can appear anywhere within the pulmonary discipline, they can darken through the ribs, mediastinum and brown beneath the diaphragm, resulting in a significant variation within the variation with inheritance. To overcome these problems, the author suggested a computer-Aided diagnostic system(CAD)for a trip through pulmonary nodules.The screening machine for most lung cancers is tested in Figure 1. This log is applied with brilliant image processing techniques consisting of plane-level cutting algorithms, corrosion, medium cleaning, stretching, contour, lung extraction area and Flood Fill Algorithms for lung extraction zone. Then, for segmentation, a set of Fuzzy possible C -Means (FPCM) rules is used for learning and the type of Extreme Learning Machine (ELM).

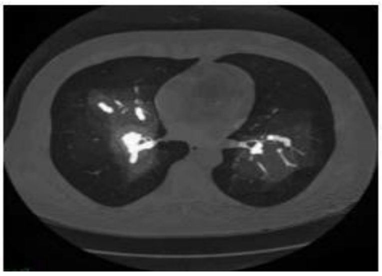


Fig.1 The Lung CT Image

There are several contracted strategies as a good way to diagnose most lung cancers, such as phlegm cytology, MRI scan and chest radiography (X-rays). Computed tomography (CT) should meet as mall node that cannot be seen on the easy chestx-rays.In addition,computed tomography (CT) plays an important position in assessing cancer staging. However, all the above techniques were used to detect cancer in progressive grades, where the probability of patient survival is low and essential.

Manuscript published on November 30, 2019.
* Correspondence Author
Popuri Ramesh Babu*, Research Scholar, Dept. of CSE, AcharyaNagarjuna University, Guntur, A.P.
Dr Inampudi Ramesh Babu, Professor, department Professor, Dept. of CSE, AcharyaNagarjuna University, Guntur, A.P.
Dr SVN Srinivasu, Professor, Dept. of CSE, Narasaraopeta Engineering College, Narasaraopet, Guntur, A.P.

About 80% of patients at intermediate or advanced levels are effectively identified for most cancers. If this is determined at the initial levels, the risk of staying charged may be excessive.A very crucial but difficult task for any radiologistis to detect and monitor breast cancer nodules. Few nodules will not be detected, as they will be hidden by anatomical structure or due to the decline offirst-order images.In addition,mileage is enhanced by the criteria of self-selection and variable used by radiologists [2]. Therefore, there may be a desire at the first level for the following techniques to address lung cancer in its early stages. Because unmarried CT testing generates many statistics that will be evaluated manually, image processing and statistics extraction strategies serve as a prominent tool for improving the analysis-oriented approach. Therefore, A CAD method can be a very useful tool for radiologists to attack and diagnose anomalies once at previous levels.

II. RELATED WORK

The Computer-Aided Diagnostic system (CAD) It is used for the early detection of most lung cancers by studying CT scans. The basic concept of CAD development is not to authorize the analysis of a device, but instead, the machine algorithm acts as a resource for radiologists and indicates the location of suspicious elements, in order to increase general sensitivity.CAD systems have four basic wishes: to improve accuracy and good prognosis, increase compliance with treatment by early detection of most cancers, avoid useless biopsies, and reduce the time required to interpret radiologists [3].

A. *Lung CT Images*

CAD can be further improved for pulmonary nodule detection and assertive and quantitative evaluation through vessels characterized by the use of CT. Most cancers are an unusual cause of cancer death worldwide. The subsequent duration of a computerized tomography experiment of tens of hundreds and many humans can be a fundamental task for radiologists to interpret. To fill this gap,computer-assisted detection algorithms can also be the most promising solution.The first step in comparing the detection effects of most lung cancers is the use of CAD, lung nodule detection,which can also form lung cancer at initial levels.Several investigations have shown that images can expect high Lung nodules. The panel has become in-depth in developing the most stringent strategies for implementation.Image processing is the most important step in the category of lung cancer.About screening by Muhammad Naumanetal.(2004)The main problem in image dispensation that you want to resolve is the nodule detection and segmentation of lungs, where nodule functions include depth,size,geometric properties,and image shape.Where,as in previous image processing, S APatil et al. (2010) Apply medium filters to eliminate noise or redundant information from images.According to J.Dehmeshkietal(2007),a method has been proposed to match the molds of the Genetic Algorithm Template-Matching (GATM) based on the spherical nodule detection model. During Segmentation of lung cancer images, fragmentation of the lungs from below is very difficult. They were able to achieve a 72% sensitivity with 31 false positives per survey. Anindya Gupta et al. (2015) Flood filling algorithm (FFA) methods are designed to increase sensitivity and Ranging in automatic cleavage and detection of pulmonary nodules in CT. The algorithm was sensitively discovered 82%. Orozco et al [8]. Use the GLCM matrix with four different angles to extract the properties. A non-parametric workbook called Transport Support Machines (SVM) is used as a workbook and has a reliability rate of 84%.

III. METHODS AND MATERIALS

The dataset used in these panels is obtained from LIDC(Lung Database Imaging Consortium)[9]. LIDC is a collection of lung image data that can be made publicly available through the National Cancer Institutes photo archive.Inthissection,wewillbrieflydiscussthetechniqueofcollecting statistics with the help of LIDC. The dataset comes with CT photographs and XML logging. CT images are designed by Digital Imaging and Communications in Medicine and are designed to manage,store,print and transmit statistics on scientific images.TheXMLlog provides records of information about the nodules it contains, spatial proximity coordinates of nodules, size classifications, and radiological functions. LIDC has 4 in-board radiologists. The LIDC statistics collection can be made based on the stages of a particular blind consultation and non-lethal reading lasses.In the first component,each radiologist criticizes CT independently. In these condpart, he

met dozens of 4 radiologists and introduced themselves to radiologists for a second overview,. allowing radiology to refine his earlier opinion based on the review of others. Thereafter, the effects of each radiologist were subsequently pooled to form the last assessment without blinding

A. *LUNG SEGMENTATION*

The Segmentation based on Full threshold splitting was used in particular CT image classification. The stroke method divides images by developing a binary screen with deference to the density in the image [11]. Areas that resemble the lung area have been removed, while areas that are not considered lungs have been deleted. A binary image is implemented with admiration to the threshold rate so that each pixel with gray bands above the threshold price is at least one and the rest of the pixels are set to 0.

$$(X,Y) = \{ \begin{matrix} 1 & > \\ 0 & if(X,Y) \leq ? \end{matrix} \} \quad (1)$$

It is calculated by the equation. 1, where T is the stroke value and enter of the image. The fill process identifies the pixels in the slots, automatically converts the value of those detected pixels from 0 to 1 and fills the slots in that binary image. Then the shape of the binary image is exposed and the crowded image is subtracted and then we get the pulmonary subject.

B. FEATURE EXTRACTION

Different functions constitute the basic factors in the category. Geometrical characteristics similar to region, diameter, perimeter and irregular index of diagnosed pulmonary nodules were estimated. The following parameters are extracted from the detected nodules.

1) *Area*

It is Calculated within the location of interest by adding pixels to the binary image containing the price 1. This standard load provides the true set of all pixels that are near the recognized object.

2) *Perimeter*

All of this is expanded pixels placed within the perimeter of the found region. It is calculated by measuring the distance between each pair of non-stop pixels located within the perimeter of the selected area.

3) *Mean*

The mean is demarcated because the quantity of the pixel values is divided by the help of a whole set of pixels within the location described in the Eq.2, where, P(i,j) is the pixel intensity value in the factor (i,j) and M x N is the image dimension.

$$\mu = \frac{1}{M - N} \sum_{I=1}^{M} \sum_{J=1}^{N} (i,j) \qquad (2)$$

A. *4. Equivalent Circle Diameter*

$$Equiv.diameter = \sqrt{\frac{4 \times Area}{\pi}} \qquad (3)$$

It is the diameter of the circle, which would equivalent area of that object. It is given by Eq. 3

a) *NODULE DETECTION*

Various nodules are located along the external edge of the lung or in the blood vessels were deleted by many different techniques described above. A new method has been developed to detect nodules present in those areas.First,the binary image threshold is calculated using morphological processes. Then, the border is thrown from the original image and part fragmented hobby. In this way,the gray level along the lung wall inside the newly formed image will be identical to the inner lung and make it easier to identify the nodules along the lung wall. The nodes are then classified in the use of connected additions[10].

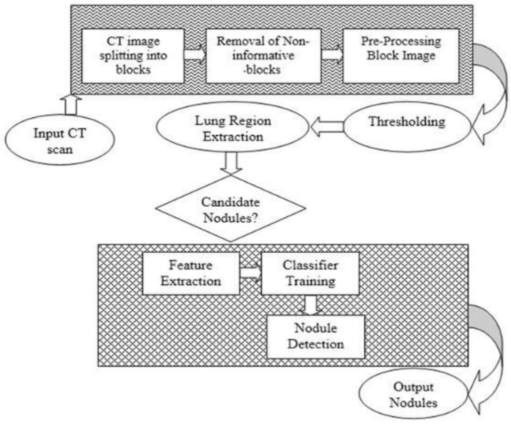


Fig.2 Typical computer-aided diagnosis (CAD) system for lung cancer

IV. PROPOSED ENHANCE FUZZY MEANS CLUSTERING

Enhanced Fuzzy C-Mean (EFCM) Clustering is an unsupervised algorithm provided in scientific images. This technique relies mainly on the average of each group and aggregates the values of the associated information with in the corresponding groups. Confusion generally persists in many gray-colored medical images of many tissues. EFCM is one of the aggregate approaches to scientific image segmentation. Several FCM assembly programs were implemented to divide the MRI into several parts of the frame.

Let $X = \{x_1, \ldots \ldots . , x_n\}$ be a data set where $x_i \in R^d$ and assume there are k clusters and c_j is the centroid of cluster j. Then, we have

Let w be a weight matrix Each value belongs to each set with a specific value

$w \quad w$

$$w = \begin{bmatrix} w_1 & w_1 \\ w_{1,} & w_{n,k} \end{bmatrix}, w_{i,j} \in [0,1]$$

--- (4)

This Proposed methodology has the following two restrictions

1. $\sum_{j=1} w_{i,} = 1k$
2. $0 < \sum_{j=1}^{n} w_{i,} < 1$

---------- (5)

We have provided preliminary information about EFCM. The EFCM algorithm is explained below.

Algorithm:

Initialize a fuzzy partition and set the weight W (for all W_{ij})

Repeat

a) Calculate the centre of clusters using the fuzzy partition.

(b) Update the fuzzy partition, i. e. , W_{ij}

(3) Until the centres do not change.

In this algorithm, c_j is

$$c_j = \frac{\sum_{i=1}^{n} w_{ij}^{p} \cdot x_i}{\sum_{i=1}^{n} w_{ij}^{p}},$$

It is an extended version of the centroid formula that is used in the means k. The variance is only the degree of membership per pint belongs to each group. Weights are defined as

$$w_{ij} = \frac{(1/(dist(x_i,c_j))^2)^{\frac{1}{1-p}}}{\sum_{j=1}(1/(dist(x_i,c_j))^2)^{\frac{1}{1-p}}} \quad \text{--------- (6)}$$

Some of square errors which regulate the condition of repeat loop is

$$= \sum_{j=1}^{k} \sum_{i=1}^{n} w_{ij}^{p} \qquad d(x_i, c_j)^2 \quad \text{-------- (7)}$$

Where p is an element that determines the effect of weights and $p \in [0, \ldots.. \infty]$, if $p > 2$, the force goes to 0. This results in the end result that weights tend to be at least one / k. If p goes to one, electricity increases the weight of the club from the points approached by the group. When p moves to 1, the club has a trend to 1 for the nearest group and has a tendency to 0 for the other groups (e.g. K-means). The set of EFCM bases has been used effectively in clinical images and has proven useful for producing strong effects in the case of bad or corrupt pixels. Produces fast and reliable MRI results with confined human interaction.

Weighted Support Vector Machine (WSVM)

It is a grouping of Lung nodules within the container tree establishment using WSVM with error statistics. False positives pass through the elegance of powerful nodules once the idle threshold is applied in the previous step. Therefore, the performance selection of RSFS features can be demoted because the grace of the minority can pass unattended with the problem. To evade issues effected by class in equality, the sub-sample system that recognizes false positives near selection boundaries has changed to be approved. False positives away from the selection threshold are more conducive to proper classification, whereas false positives near the restriction group, as an example that has been removed near the actual nodules, may be incorrect. Therefore, the provisions of greater differentiation of the characteristic group have been identified to distinguish appropriate nodules from false positives. The method of using a smaller sampling method combined with WSVM has become an unbalanced dataset applied to this technique.

Gmean has followed to avoid accurate terrible prediction of minority elegance, as proven

$$Gmean = \sqrt{Sensitivity \times Specifity} \quad (8)$$

$$Sensitivity\,(TPR) = \frac{TP}{TP+FN} \quad (9)$$

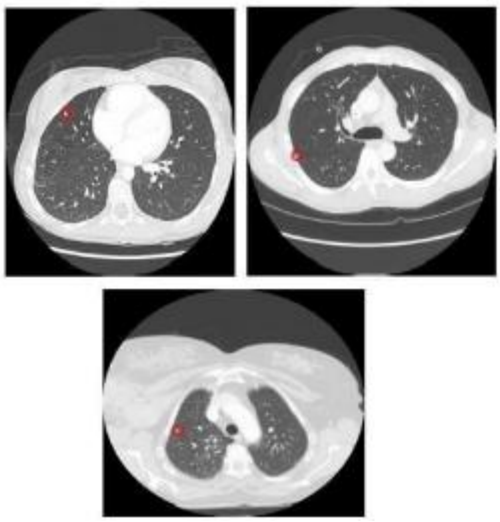
$$Specifity(TNR) = \frac{TN}{TN+FP} \quad (10)$$

In the previous equations, TP,TN,FP and FN for mature fine and true negative, positive and false respectively. The two-layer approval check has been changed to be used in these investigations.

Performance of nodule classification. Candidates for nodules within a non-vessel tree company that survived the above disclosure approach were classified as nodules. At the same time, applicants for rnodules within the pottimber foundation were identified as actual high quality with the help of the WSVM classifier. Below Figure 3 gives

examples of nodules finding using the proposed system

a) *b)*



c)

Fig.3 Examples of finding nodules under the suggested methodology, marked with red circles

V. RESULT AND DISCUSSION

In this part, the recommended schema representation is expected to verify the effectiveness of the system. Overall performance became prior to the use of the LIDC database and was compared to various previous methods.CT lung images collected from the L1DC database are provided as inputs into trained neural networks. A total of 16 patient data were taken and 279 images were processed on CT. The proposed approach was able to capture seventy-two nodules, and the false quality rate became 0.9% lower. The overall accuracy of the proposed technique is 92%. The results are shown in Figure 4, which specifies that the set of rules can be familiar with nodules within the vessel.

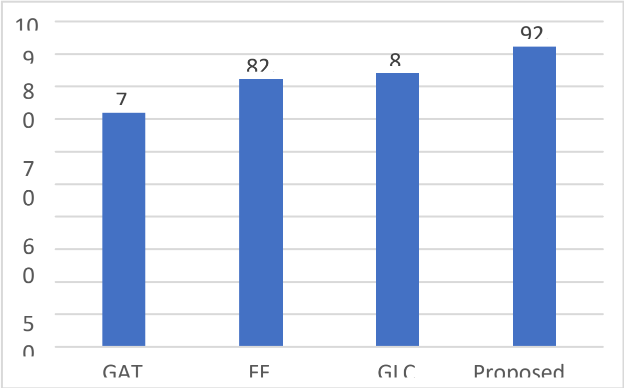


Fig 4: Detection accuracies of algorithms

A. *EVALUATION METHODS*

The performance of the proposed EFCM system was assessed using the three parameters: specificity, sensitivity, and accuracy.

$$ens\ tiv \qquad P\ /\ (TP + FN)$$

$$Specificity = TN\ /\ (TN + FP)$$

$$Accuracy = (TP + TN)\ /\ (TP + FP + TN + FN)$$

The true positives (TPs) Identified through this approach was 102, therefore, sensitivity was reported at 82.93% (102/123). The number of false positives (FP) identified was 31, and therefore, privacy was reported at 82.48% (146/177). Accuracy was estimated at 82.67% (248/300). Specific TPs show that this fragmentation method can identify some nodules whose size is greater than or equal to 2 mm.

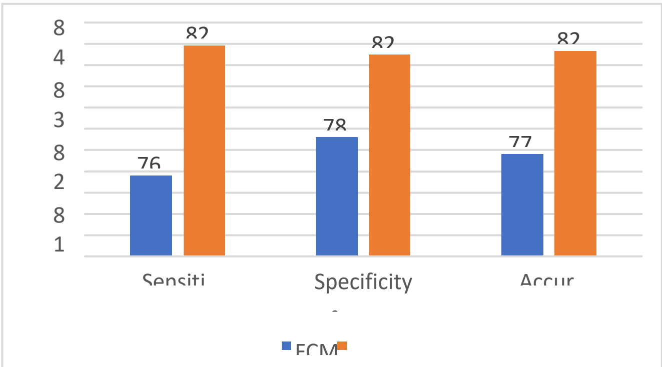


Fig.5 Performance measures with FCM and EFCM

Fig.5 shows,Comparison of different performance Measures of the proposed segmentation system using FCM and EFCM.It is clearly observed that the CAD system with proposed EFCM has High accuracy than the CAD

System with the FCM Scheme.The results also showed that EFCM works better than FCM, EFCM is used for fragmentation in the following proposed approach

VI. CONCLUSION

The CT images of lungs are measured for the needed detection of lung nodules. Initially, the images are preprocessed where the contrast levels are connected, and the user is permitted to crop the image to choose the ROI. Then segmentation is made, and nodules are distinguished from the ROI. A new Hybrid Classifier has been Proposed to detect lung nodules in this approach. The hybrid method is the mixture of unsupervised Enhanced Fuzzy C-Mean (EFCM) clustering and Weighted supervised support vector machine (WSVM). The proposed approach has decreased false positives significantly in nodule applicants by using the most discriminative texture features. The experimental results present an indication that the proposed method can efficiently classify nodules. By the proposed method, a total accuracy of 92.2% and a very low false-positive rate of 0.9%wereachieved.Among the clustering methods,Enhanced FuzzyC-Mean(EFCM)hasgiven greater accuracy for various medical images to compare segmentation results, and different validation techniques have been described

REFERENCES

1. Muslim,&Qasim,2018,"Aknowledge-basedimageenhancementand denoisingapproach", pp.1–14.
2. Khan,S.A&Usman,M,2018,"Reliablefacialexpressionrecognitionf ormulti-scaleimages using weber local binary image-based cosine transform features",pp.1133–1165.
3. Khan, S. A, and Shaheen, M, 2018, "Face recognition under varying expressions and illumination using particle swarm optimization",pp.94–100.
4. Muhammad Nauman Mughal and WaseemIkram, 2004, "Early lung cancer detection by classifying chest CT images: a survey", pp.67-72,
5. S.A.Patil and V R Udupi, 2010, "Chest x-ray features extraction for lung cancer classification", pp.271-277.
6. J. Dehmeshki and H. Amin, 2007, "Automated detection of lung nodules in CT imagesusing shape-based genetic algorithm",pp.408–417.
7. Anindya Gupta, Tönis Saar, 2015, "Methods for increased sensitivity and scope in automatic segmentation and detection of lung nodules in CT images",pp.375-380.
8. Madero Orozco, H. Cruz Sanchez, 2013, "Lung Nodule Classification in CT Thorax Images Using Support Vector Machines ", pp.277-283.
9. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2176079/(Lung Image DatabaseConsortium).
10. JunjieBai, Qi Song, 2015, "Learning orientation invariant contextual features for nodule detection in lung CT scans",pp.1135-1138.
11. Yang Chunran ,Guo Yi , 2018, "Automatic Detection and Segmentation of Lung Nodule on CTImages".
12. Hussain, S, and Yang, S, 2018, "An Effective Framework for Driver Fatigue Recognition Based on Intelligent Facial Expressions Analysis",pp.67459–67468.
13. Naqi, S. M, and Yasmin, M, 2018, "Multistage segmentation model and SVM-ensemble for precise lung nodule detection",pp.1083–1095.
14. VenkataTulasiramu Ponnada, S.V.Naga Srinivasu Integrated Clinician Decision Supporting System for Pneumonia and Lung Cancer Detection, International Journal of Innovative Technology and Exploring Engineering, 2019, ISSN: 2278-3075, Volume-8,
15. VenkataTulasiramuPonnada, S.V.Naga Srinivasu, Edge AI System for Pneumonia and Lung Cancer Detection, International journal of innovative technology and exploring engineering, ISSN: 2278-3075, Volume 8, Issue 9, 2019
16. VenkataTulasiramuPonnada, S.V.NagaSrinivasu, Efficient CNN for Lung Cancer Detection, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8 Issue-2, July 2019.
17. VenkataTulasiramuPonnada, S.V.Naga Srinivasu, End to End System for Pneumonia and Lung Cancer Detection Using Deep Learning, International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume 8 Issue-6, August 2019.

AUTHORS PROFILE

Mr.Popouri Ramesh babu ,MTech , Research Scholar, International experience: worked as a asst professor in Jimma University ministry of higher education ETHIOPIA Area of Interest : Machine Learning and , Deep learning.

Professor INAMPUDI RAMESH BABUM, Tech, Ph.D, Department Of Computer Science, ACHARYA NAGARJUNA UNIVERSITY Area of Interest : IMAGE PROCESSING, Achievements: Best Teacher award from Government Of Andhara Pradesh working as chairman board of studies in the department of computer science, ANU Vast experience in various levels and committees like course structures design introducing new circulam and new courses .

Dr.S.V.NSrinivasu, Professor, Computer science and engineering, Narasaraopeta Engineering College, Narasaraopet, Andhra Pradesh 522601, India. He has published more than 58 articles in reputed journals which includes Scopus and other indexed bodies. He is a member of IEEE, MIET. His area of interest includes Software Engineering, Software Testing, Machine Learning, Deep learning, Mobile Networking, Operating systems, Data Mining, Image Processing and Black chain technology.