ISSN2515-8260 Volume10, Issue 02, 2023

# LUNG CANCER DETECTION USING MACHINE LEARNING ALGORITHMS

**Dr.M.Rajaiah,**Dean Academics & HOD, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

**Mr.D.V.VaraPrasad,** Associate Professor ,Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

**Mr.T.Siva,** UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

**Mr.T.Chaitanya,** UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

**Ms.S.Bhavana,** UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

**Ms.Ch.Praveena,** UG Scholar, Dept of CSE, Audisankara College of Engineering and Technology, Gudur.

**ABSTRACT:**

Lung Cancer detection making use of medical imaging is still a challenging task for radiologist. The objective of this researchis to classify the types of lung tumours for extracted and selected features using learning algorithms. In this paper, an experimental study is conducted on 100 cases of lung cancer to evaluate the performance of learning classifiers (DNN, SVM,Random Forest, Decision Tree, Naïve Bayes) with different medical Imaging (DICOM) features to identify the two types ofLung cancer (Benign and Malignant). The proposed methodology intends to automate the entire procedure of diagnosis byautomatically detecting the tumor, measuring the required values such as diameter, perimeter, area, centroid, roundness,indentations and calcification. Experiment is conducted in to two phases: In the first phase, identify the most significantfeature used in lung cancer analysis by CT scan and perform the mapping to computer related format. In the second phase,feature selection and extraction is performed to machine learning algorithms. To evaluate the performance of classifiers interm of classification accuracy and improving the false positive rate, every stage of evolution is divided into four differentphases: single phase module, single slice testing, series testing and testing of learning algorithms. Experimental resultsshow significant improvement in false positive rate up to 30% for both Benign and Malignant. Whereas, Deep NeuralNetwork (DNN) demonstrate high values in term of classification accuracy in comparison with other classifiers. The proposed methodology for lung cancer

detection system having a potential to reduce the time and cost of diagnosis procedureand use for early detection of lung cancer.

## 1.INTRODUCTION

Artificial intelligence can enable the computer to think. Computer is made much more intelligent Artificial by AI. Machine learning is the subfield of AI study. Various researchers think that without learning, intelligence cannot be developed. There are many types of Machine Learning Techniques that are shown in Figure 1. Supervised, Unsupervised, Semi Supervised, Reinforcement, Evolutionary Learning

### Different Algorithms and Techniques For Cancer Detection

**Supervised Learning:** Offered a training set of examples with suitable targets and on the basis of this training set, algorithms respond correctly to all feasible inputs. Learning from exemplars is another name of Supervised Learning. Classification and regression are the types of Supervised Learning.Classification: It gives the prediction of Yes or No, for example, "Is this tumor cancerous?", "Does this cookie meet our quality standards?"Regression: It gives the answer of "How much" and "How many".

**Unsupervised learning**: Correct responses or targets are not provided. Unsupervised learning technique tries to find out the similarities between the input data and based on these similarities, un-supervised learning technique classify the data. This is also known as density estimation. Unsupervised learning contains clustering [1] .Clustering: it makes clusters on the basis of similarity.

**Semi supervised learning**: Semi supervised learning technique is a class of supervised learning techniques. This learning also used unlabeled data for training purpose (generally a minimum amount of labeled-data with a huge amount of unlabeled-data). Semi-supervised learning lies between unsupervised-learning (unlabeled-data) and supervised learning (labeled-data).

**Reinforcement learning:** This learning is encouraged by behaviorist psychology. Algorithm is informed when the answer is wrong, but does not inform that how to correct it. It has to explore and test various possibilities until it finds the right answer. It is also known as learning with a critic. It does not recommend improvements. Reinforcement learning is different from supervised learning in the sense that accurate input and output sets are not offered, nor sub- optimal actions clearly précised. Moreover, it focuses on on-line performance.

**Evolutionary Learning:** This biological evolution learning can be considered as a learning process: biological organisms are adapted to make progress in their survival rates and chance of having off springs. By using the idea of fitness, to check how accurate the solution is, we can use this model in a computer [1] .6) Deep learning: This branch of machine learning is based on set of algorithms. In data, these learning algorithms model high-level abstraction. It uses deep graph with various processing layer, made up of many linear and nonlinear transformation.

a) **Evalutionary computation for Classification:**EML techniques have been widely used for classification. The aim of classification algorithms is to learn a model/classifier that can correctly classify unseen instances (test data) by observing a set of given instances (training data).

b) **Evalutionary computation for Regression:**Regression is a major ML task that attempts to identify and express the underlying relationship between the input features/variables and the target variable(s). Regression analysis is utilised for forecasting in widespread areas, e.g., finance, traffic, medicine, and biology (Glaeser and Nathanson 2017).

**U-Net Convolutional Network:** The U-Net Convolutional Network is used for biomedical image segmentation. It takes an input image and an output mask of the region of interest. It first generates a vector of features typically in a convolutional neural network, and then use another upconvolutional neural network to predict the mask given by the vector of features [20][21][22]. This is a binary classification task using morphological and radiological features extracted from the images and masks. The features are continuous and numerical, but can be discretized into categories.Exploring different methods to diagnose lung cancer will be a prime aim in this paper. Computed tomography can be used to capture images of lungs across various dimensions so that a 3D image of the chest can be formed. This 3D image can be used to detect tumors present. Normally a doctor or any field expert uses a CT image to detect cancer. Due to the large number of CT images, it is difficult for a doctor or radiologist to detect cancer quickly and accurately. But with the advancement in technology, Computer-Aided Diagnosis (CAD) can be utilized to complete this duty efficiently and in considerably less time. This process has two separate processes i.e. first to identify all the nodules present in the CT image and second to classify the detected lung nodules. In general, a CAD system comprises the following steps which are shown below in figure 1.

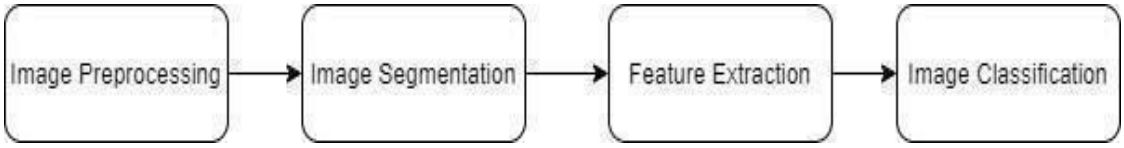


**Fig 1**:Basic Steps Involved in CAD System

it is applied to small regions of the images known as tiles. Bilinear interpolation is used to combine the different enhanced parts/regions of the image. Wiener filters are used to reduce the noise by a significant amount. Region extraction plays an important role to get the desired region. Morphological operations such as closing were used to get the desired region i.e. region having lung lobes and leaving behind the blood vessels,

bronchi, and all other internal parts. The structuring element of disc shape was used in the closing operation. While in the feature extraction process, texture-based features were concentrated as intensity value is not the right parameter to extract features. The classification of the pre-processed image is done using FCM. FCM is chosen as it retains important features of the image. FCM classifier is based on unsupervised learning. Figure 4 represents the process flow of the technique proposed in this paper


**Fig2:**Process Flow Of System

Deep Learning trained neural network and Improved Profuse clustering are used in this study. CT images contain lowquality images and it has noise, so to remove all this, CT image pre-processing is done. For improving the image quality, Image histogram techniques are used as it is a very efficient method on different images Segmentation of cancer affected regions is done with the help of improved CT image using IPCT. The improved profuse clustering technique is applied to segment cancer influenced parts from the improved lung CT image. For detecting inconsistency in the image pixels, two procedures of improved profuse techniques work as it checks the image pixel and puts the similar superpixel in the same group. Predicting the similitude of data using the pixel eigenvalue is done during the process of segmentation when the pixels are continuously examined. Different features of spectral that are standard deviation, 3rd-moment skewness, mean, and 4th-moment kurtosis are derived from the region which are segmented and which is forwarded for the feature extraction stage as it is very effective to spot lung cancer which has connected features. 98.42 accuracyensured.

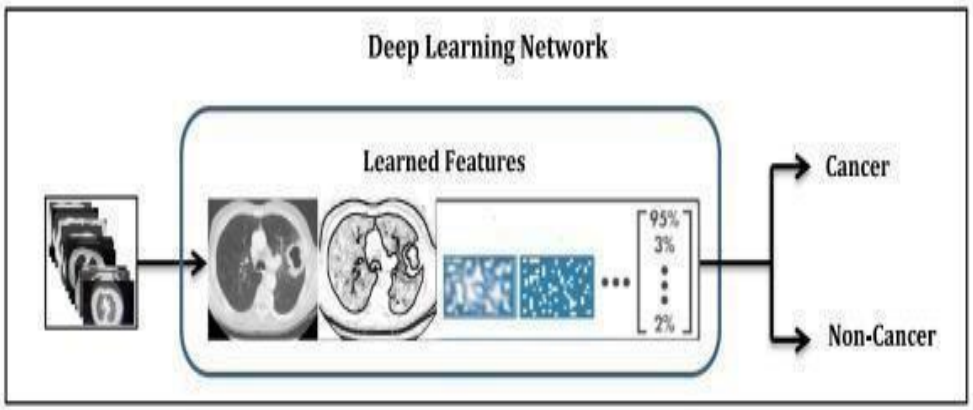


**Fig:**Deep Learning training process structure

CT images from the Cancer Image Archive Database were used in DICOM format. These images were then pre-processed using various image enhancement techniques such as Median Filtering, Smoothening, and Contrast Adjustment to remove noise and

improve image quality. Further Morphological opening operations were performed after transforming the grayscale image into a binary image for image segmentation. In the feature extraction method features like area, perimeter, and eccentricity (roundness) are evaluated. Using these features classification of images is done into normal and abnormal using SVM supervised learning classifier. The proposed methodology as said by the authors detects cancer in the early stages accurately. Figure 6 shows the cancerous lung CT image and its histogram representation.

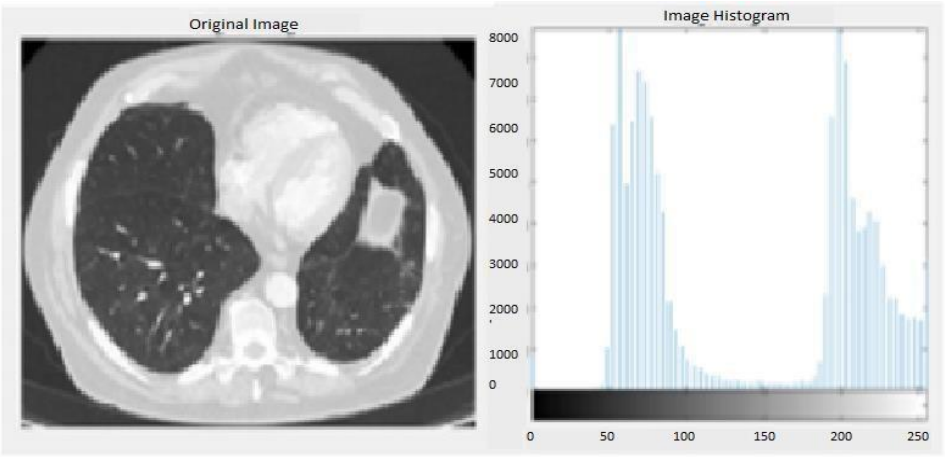


**Fig:**Cancerous lung CT image and histogram of the image

**Artifical Intelligence role in cancer diagnosis**: Artificial intelligence is improving the detection of lung cancer.Machine learning systems for early detection could save lives. After years of helping to train an artificial-intelligence (AI) system to find the early stages of lung cancer, Mozziyar Etemadi was thrilled when the computer found tumours in scans of patients more accurately than trained radiologists did1. He was even more excited when his team gave the system old computerized tomography (CT) scans of the chests of people who later developed lung cancer. No doctor had seen anything amiss in these early scans, but the machine did. Early cancer diagnosis and artificial intelligence (AI) are rapidly evolving fields with important areas of convergence.

**Literature Survey:**

| Study | Method | Dim | No of cases | Accuracy (in %) | Sensitiv (in % |
|---|---|---|---|---|---|
| Roozgard et al.[55] | Warwershed Algorithm,MedianFilter,Morphological Dilation and Erotion Filter | 2D | 12 | -- | -- |
| Si Guang-lei et al.[45] | Thresholding,Morphological Processing | 2D | 85 | 90.5 | -- |
| Takahiro et al.[19] | Thresholding,Watershed Algorithm | 2D | 31 | 96.9 | -- |
| Vijaya Kishore et al.[19] | Thresholding Watershed Algorithm.Region Growing Segmentation | 2D | -- | -- | -- |
| Maria Evelina et al.[54] | 3D Region Growing Method Wavefront Algorithm Morphological poerations | 2D | 138 | -- | 80 |

| | | | | | |
|---|---|---|---|---|---|
| Amal Farag et al.[54] | Template Matching | 2D | 50 | -- | -- |
| Mickias Assefa et al.[49] | Template Matching | 2D | 165 | 81.21 | -- |
| Amal Farag et al.[52] | Template Matching Registration | 2D | 50 | -- | -- |
| Hiram Madero et al.[66] | Template Matching,Watershed Agorithms | 2D | 61 | 82.66 | 96.15 |
| Amal A Farag et al.[51] | Variational Level Set Segmentation. Template Matching | 2D | 50 | 70 | -- |

**Methodology:**

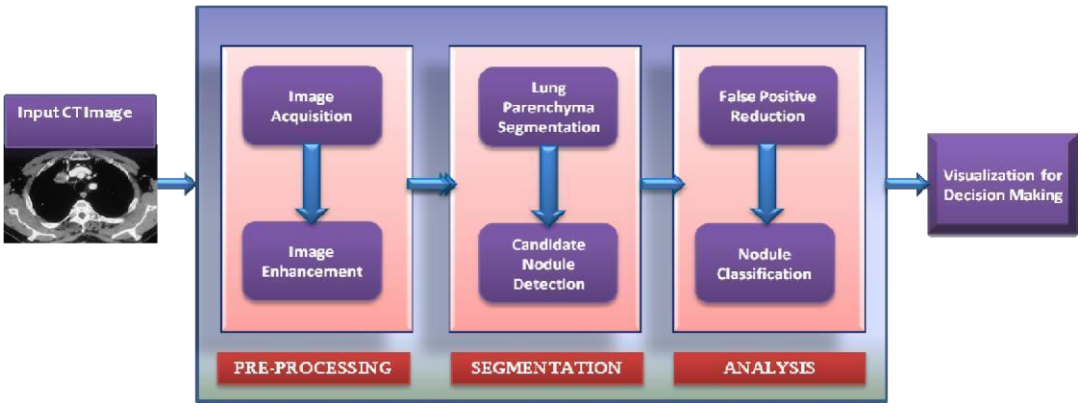


**IMAGE ACQUISITION**

To design a CAD for lung nodule detection system, CT images are well preferred as offers visualization of low contrast or small volume nodules by diminishing the slice thickness. Lung CT images can be acquired from publicly available databases namely Early Lung Cancer Action Program (ELCAP) [5], Lung Image Database Consortium (LIDC) [6] or Medical Image Database [7]. As per the literature there are other imaging techniques and many private databases used by the researchers, which are obtained from private hospitals.

## PREPROCESSING

Since the Digital Imaging and Communications in Medicine (DICOM) images with noise uses the reconstruction method to enhance resolution denoising these original images is necessary. Preprocessing of an image refers to the procedure of enhancing the quality and interpretability of the input lung image by reducing the noise and unwanted artefacts. In lung CT scans, preprocessing improves the visibility of pulmonary nodules. Types of filters used in preprocessing phase are Laplacian of Gaussian filter (LOG), Ring Average filters, Median filters, Morphological filters, Selective Enhancement Filter and so on.

### Selective Enhancement filters

Various types of selective enhancement filters are used to enhance blob like structures and to suppress vessel like structures by [8] [9] [10] [11] and [12] recommended a selective enhancement filter to enhance dot like objects and to repress lung vessels. Cylindrical and spherical filters were combined for a better visualization of nodules by [13].

### Log filter

The Laplacian of Guassian (LoG) filter is preferred in enhancing blob like structures whose intensity is differs from that of background. [14] used LoG filters to enhance the input image. [15] and [16] recommended LoG filter for enhancement.

### Lung parenchyma Segmentation

Image segmentation is a technique of partitioning an image in to multiple regions of the lung refers to the process of extracting the lung region from other anatomical parts of the body in chest CT images. This process plays a vital role in nodule detection by improving accuracy and precision that helps in early diagnosis of lung cancer. An accurate segmentation will reduce the computational cost of detection. In CT image of a lung, the anatomical structures that may require segmentation are lungs themselves, the airways, the vessels, lung lobes. Segmentation is a complex activity due to pulmonary structures of similar densities namely arteries, veins, bronchioles and different scanners used. Numerous publications have addressed the issue of segmentation of lungs. The proposed techniques are measures with respect to accuracy, processing time and level of automation. Mostly used segmentation techniques can be classified as methods based on thresholding, deformable boundaries,

edge detection and shape models. Lung is a sack of air in the body which exhibit as darker regions in CT images compared to other parts of the chest. This fact has motivated the researchers to explore an optimum threshold which separates the lung from other tissues. The majority of the research in the segmentation revolved around techniques such as thresholding, multiple thresholding, optimal thresholding, adaptive thresholding, region growing, graph cuts, active contour model, hybrid segmentation, fuzzy c means clustering, morphological operations and so on.

**Thresholding**

In thresholding techniques selecting the threshold level and in region growing approach selecting a seed point has a greater impact on segmentation outcome. The recent methods use model based and optimization methods as compared to earlier heuristic based approaches. [25] obtained a threshold that separates the lung from other tissues. Region growing followed by Connected Component Analysis method was proposed by [26] to extract the lung region. [27] proposed an accurate segmentation method with four step. (1) Extracting the airway from lung region (2)Removal of pulmonary arterial and venous vessel trees by finding a suitable threshold (3) Using a largest threshold to separate right and left lung (4) Morphological smoothing of lung boundary . [28] performed the segmentation using histogram analysis followed by Connected Component Labeling (CCL) and morphological closing operations to smother the segmented lung. The accuracy of threshold based segmentation depends upon image acquisition protocol and acquisition type. In most of the cases the concentration of pulmonary constructs, say arteries, veins, bronchi are very much near to the concentration of chest tissues. To overcome this inhomogeneity in the densities in the territory of lung, further rigorous post processing is required.

**Candidate Nodule Detection**

A pulmonary nodule is almost spherical shaped opacity measuring less than 3cm in diameter surrounded by lung parenchyma. Its shape can be deformed by the neighbouring vessels or pleural surface. 4 types of nodules identified by [39] were, Nodule detection is a process of identifying the nodules and their location in the lung field. The success of this process heavily depends on the accuracy of the lung parenchyma segmentation and false positive reduction method. Wellcircumscribed nodules detection is relatively easy as they are isolated in nature. But the detecting the other three types is a challenging task as these types generate mostly false positive results.

**Template matching**

Template matching methods to segment the SPNs were used by [47] [48] that could detect the circular /semicircular nodules. [49] developed both circular and semicircular templates to detect nodules residing inside and on the boundaries of lung region. This circular, spherical hypothesis is not enough to portray the actual geometry of nodules. Lesion's geometry can be irregular due to their attachment to the pleural surface or lung vessels. A variational level set segmentation was proposed by [50] [51]

where a signed distance function was used to represent a 2D contour followed by template matching algorithm to extract juxta pleural nodules. [52] proposed an Active appearance Model (AAM) with template matching and registration to detect SPNs. Template based approach for the segmentation of nodules was implemented by [53].

## NODULE CLASSIFICATION

After the candidate nodules are detected and false positives are reduced, the resultant set of nodules must be classified to be benign or malignant ones. Most of the pulmonary nodules are benign but they may represent an early stage of lung cancer. Early detection of a malignant (cancerous) nodule increases the survival rate of the diseased. Many Computer Aided Diagnosis (CADx) systems have been developed that differentiate malignant lesions from benign ones, and also gives the insight into the proximity of detected nodule to be malignant. Receiver operating Curve (ROC) approach is used to evaluate the classifiers. Despite the fact that the benign and malignant nodules have overlapping attributes, certain discriminating features such as morphological features, shape, size, appearance and growth rate of nodules makes the classification of nodule possible. Among these parameters growth rate of SPNs are regarded as an inventive clue for evaluating malignancy. [64] has listed various techniques are used for classification of SPNs such as Rule based classifiers, Markov Random Field, Neural Network, Bayesian classifier so on. The following sections reviews mostly used techniques for the categorization.

## 4.CONCLUSION

One of the most fatal diseases to have existed is lung cancer. This disease unfortunately is extremely tough to treat after having spread upto an extent or reaching a serious stage. ComputerAided Detection (CAD) is one of the constantly growing technologies that help detect cancer by feeding in certain inputs containing patient-related information such as scans like CTScan, X-Ray, MRI Scan, unusual symptoms in patients or biomarkers, etc. SVM, CNN, ANN, Watershed Segmentation, Image enhancement, Image processing are a few methods used to improve the accuracy and aid the process. For training, the most popular datasets used are LUNA16, Super Bowl Dataset 2016, and LIDC-IDRI. By the means of this review paper, we aim to list out all the major researches that have been done over the past years and can be improved upon to achieve better results.

## REFERENCES

[1] American cancer society, Cancer facts and figures, 2013.

[2] Heather R. Sanders, Maher Albitar, “Somatic mutations of signaling genes in non-small-cell lung cancer”, Cancer Genetics and Cytogenetics, 203, Elsevier, pp.7-15, 2010.

[3] M.V. Sprindzuk et al. “Lung cancer differential diagnosis based on the computer assisted radiology: The state of the art”, Polish journal of Radiology, Pol J Radiol, 75(1): 67-80, 2010.

[4] K.Devaki1 and V.MuraliBhaskaran, "Study of Computed Tomography Images of the Lungs: A
Survey", IEEE-International Conference on Recent Trends in Information Technology, ICRTIT,pp.837842, 2011.

[5] Early Lung Cancer Action Program(ELCAP) : http://via.cornell.Edullungdb.html

[6] Lung Image Database Consortium(LIDC): https://imaging.nci.nih.gov/ncia/login.jsf

[7] Medical Image Database: MedPix : http://rad.usuhs.edu/medpix/index.html

[8] Takahiro Miyajima et al., "Classification of Lung Nodules Temporal Subtraction Image Based on Statistical Features and Improvement of Segmentation Accuracy", proc. 12th IEEE International Conference on Control, Automation and Systems, Jeju Island, Korea, pp.1814-1817, 2012.

[9] H. Arimura, S. Katsuragawa, K. Suzuki, F. Li, F. Shiraishi, and S. Sone, et al., "Computerized scheme for automated detection of lung nodules in low-dose Computed Tomography images for lung cancer screening", Acad. Radiology, vol. 11, pp. 617-29, 2004.

[10] Farag, A.E. Baz, G. Gimelfarb, and R. Falk, "Automatic detection and recognition of lung abnormalities in helical CT images using deformable templates", Medical image Computing and
Computer assisted Intervention-MICCAI, Springer, Berlin, vol. 3217, pp. 856-64, 2004

[11] Sun, S.S., Li, H., Hou, X. R., et al., "Automatic segmentation of pulmonary nodules in CT images, IEEE", 1st International Conference on Bioinformatics and Biomedical Engineering (ICBBE), pp. 790– 793. 2007.

[12] Jia, T., Zhao, D.-Z., Yang, J.-Z., et al., "Automated detection of pulmonary nodules in HRCT images", IEEE, 1st International Conference on Bioinformatics and Biomedical Engineering (ICBBE), pp. 833– 836. 2007.

[13] Chen et al., "Pulmonary micro nodule detection from 3D chest CT medical image", in MICCAI, vol. 3217, pp. 821-828. 2004.

[14] Sergei V. Fotin et al., "A multiscale Laplacian of Gaussian filtering approach to automated pulmonary nodule detection from whole-lung low-dose CT scans", Medical Imaging 2009, Proc. of SPIE Vol. 7260, pp. 72601Q-1-8,2009.

[15] Diciotti, S., Picozzi, G., Falchini, M., et al., "3D segmentation algorithm of small lung nodules in spiral CT images", IEEE Trans. Inf. Technol. Biomedical 12, pp, 7–19 2008.

**Author Profiles**

**Dr.M.Rajaiah,** Currently working as an Dean Academics & HOD in the department of CSE at ASCET (Autonomous), Gudur, Tirupathi(DT).He has published more than 35 papers in Web of Science,Scopus,UGC Journals.

**Mr.D.V.Varaprasad**, Currently working as an Associate professer in the department of CSE at ASCET Autonomous),Gudur, Tirupati(DT).

**Mr.T.Siva,** B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur. He has pursuing in computer science and engineering.

**Mr.T.Chaitanya,**B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur. He has pursuing in computer science and engineering.

**Ms.S.Bhavana,**B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur. She has pursuing in computer science and engineering.

**Ms.Ch.Praveena,**B.Tech student in the department of CSE at Audisankara College of Engineering and Technology, Gudur. She has pursuing in computer science and engineering.