IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

COMPUTER AIDED DIAGNOSIS OF LUNG DISEASES USING KNN CLASSIFIER

Anita Titus [1], Nirmaladevi R [2]

[1,2] Department of Electronics and Communication Engineering, Agni College of Technology, India

DOI: 10.5281/zenodo.1283126

ABSTRACT

Lung diseases like pneumothorax and pleural effusion are detected and diagnosed using KNN classifier. This paper proposes a real-time detection algorithm to classify the lung diseases. Initially, a real-time image is collected from the public database. Since the collected images are prone to noise this is removed by the preprocessing techniques. The preprocessed image is further segmented by thresholding method. The inner white region and outer white region of the CT image is removed using image clear border. Then the centroid and the heights of each lung are found. The segmented image is overlapped with the original gray image. The texture features are extracted from the segmented images. The extracted features are energy, contrast, correlation, homogeneity and area which help to classify the CT images. KNN classifier is used to classify the CT image either as normal lung or lung affected with Pneumothorax or Pleural Effusion. Experimental analysis is performed using normal lung images and diseased lung images using KNN, RF (Random Forest) & SVM (Support Vector Machine) classifier. The accuracy and precision of KNN classifier is 90% & 100%, RF classifier is 81% & 80% and SVM classifier is 72% & 75%.

KEYWORDS: K-Nearest Neighbor, Support Vector Machine, Random Forest.

I. INTRODUCTION

This article focuses on classifying the lung disease pneumothorax and pleural effusion using KNN classifier. Pleural effusions (PE), buildups of fluid within the pleural cavity, are usually a symptom of a greater illness such as congestive heart failure, pneumonia, or metastatic cancer. They have also been identified as prognostic indicators, for example, for acute pancreatitis. PE can be formed in two ways: transudates where the fluid is pushed into the plural space from elsewhere due to changes in hydrostatic pressure, and exudates where the fluid is created by the pleural surface itself. The pleural cavity is the space between the visceral pleura and the parietal pleura. The visceral pleura covers the lungs and the parietal pleura run along the inside of the chest wall. As fluid accumulates in the cavity, it compresses the adjacent lung, undermining lung function. Both the general size and location of the effusion can be diagnostically significant and important to patient care. The CT images of normal lung and lung with pleural effusion is shown in the figure 1 & 2.

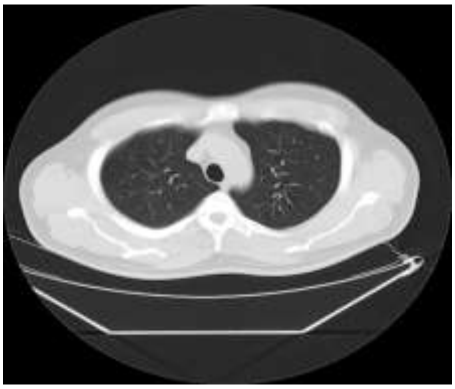

Figure 1. CT image of Normal Lung

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

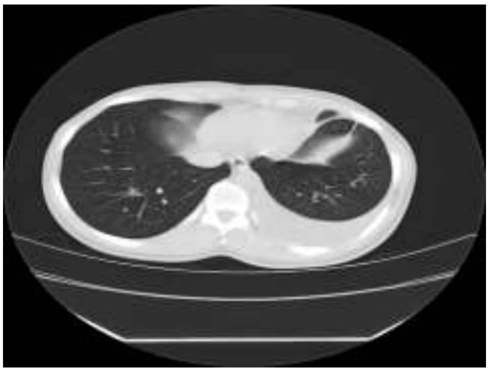

Figure 2. CT image of Pleural Effusion

Pneumothorax is a pleural disorder resulting in the accumulation air in the pleural space. As air is less dense compared to the lung parenchyma, the pneumothorax region will also take the shape of the lungs and the lung cavity and will commonly occupy the upper regions of the lungs. The grayscale value of air corresponds to values close to zero (black) and that of water or fluid is close to 128. Hence the entire pleural effusion region in the CT slice has a grayscale value close to 128. So also, the whole pneumothorax district in the CT cut has the same grayscale esteem as air, which is near zero (dark). Most of the division conspires in the writing have focused on the extraction of the lung parenchyma, so a strategy is proposed for the division and extraction of pleural radiation and pneumothorax, both of which are available in the pleural locale outside the lung parenchyma. The CT pictures of pneumothorax is appeared in the figure 3.

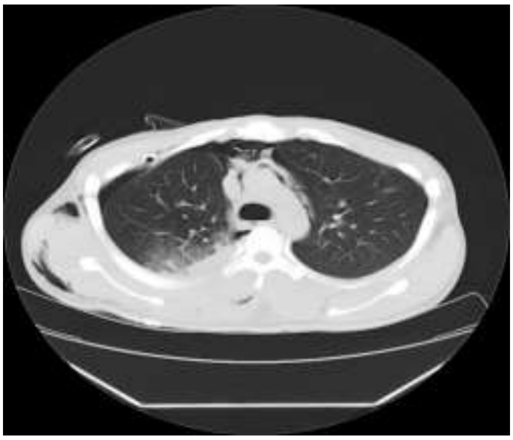

Figure 3. Pneumothorax

II. LITERATURE SURVEY

Titus et al [1] has proposed a CAD system for the detection of pleural effusion and pneumothorax, which affect the pleural membranes of the lungs. The accuracy of a CAD system is largely dependent on the efficiency and preciseness of its segmentation system. The segmentation technique used here extracts the lungs and the regions affected by pleural effusion using conventional thresholding techniques like Otsu's and iterative thresholding, followed by morphological operations. The chest CT slices are initially preprocessed to remove the gaussian noise by using a Gaussian filter. Morphological operations are then applied, to segment the lung parenchyma as well as to extract the region of interest (ROI). Texture features are next extracted from the ROIs, which are used to compute the feature vectors. These are used to train a neural network classifier. After the training process is complete, the query images are given to the system for classification. The CAD system achieved an accuracy of 95.458%, sensitivity of 88.92% and specificity of 97.918%.

Melendez et al [2] have proposed a strategy that utilizes surface-based component extraction and Support Vector Machines (SVM) to arrange chest strange radiograph designs to be specific pleural emanation, pneumothorax, cardiomegaly and hyperaeration. A comparative past endeavor prototyped the grouping framework that accomplished 97% and 87.5% exactness for pleural emission and pneumothorax utilizing histogram esteems, while achieving 70% and 73.33% for cardiomegaly and hyperaeration utilizing picture handling plans. This work intended to build the execution in ordering the said lung designs, particularly for cardiomegaly and hyperaeration. Utilizing surface-based highlights, the created framework could accomplish correctnesses of 96% and 99% with sensitivities of 97% and 100% for the cardiomegaly and hyperaeration cases, separately.

Janudhivya et al [3] proposed a CAD framework technique for identifying pleural emission and pneumothorax, which influence the pleural layers of the lungs. The chest CT cuts are at first preprocessed to evacuate the Gaussian commotion by utilizing a sigma channel. The lungs and the locales influenced by pleural emission are rumored utilizing traditional thresholding procedures like Otsu's and iterative thresholding, trailed by morphological tasks. Mumford shah demonstrate is then connected, to portion the lung parenchyma and additionally to extricate the Region of Interest (ROI). Surface highlights are next separated utilizing Spectral surface extraction technique from the ROIs, which are utilized to process the component vectors. These are utilized to prepare utilizing multi-level cut classifier.

Singh et al [4] in this paper proposed a literature method for lung nodule detection, segmentation and classification/taxonomy using computed tomography (CT) images. One of the most common noise in CT imaging is an impulse noise which is caused by unstable voltage. In this paper, a new decision-based technique called new adaptive median filter is presented which shows better performance than those already being used. The CT slices are initially preprocessed to remove the Gaussian noise by using Gaussian filter. Otsu thresholding is applied to extract the region of Interest (ROI). Malignant (virulent) cell presented in the lungs specified nodules are classified for the therapy processes. Now, the images are classified into nodules and non-nodules and take out the object's feature vectors in chosen /selected boxes. Lastly, the support vector machine (SVM) is applied which will classify the extracted feature vectors. The SVM will classify the images into normal or abnormal based on the second order gray level co-occurance matrix features. The CAD system has been developed for the classification of pleural illness like pleural effusion and pneumothorax. The algorithms or procedures proposed in this work use morphological and arithmetic operations for the extraction of the regions in the lung which are affected by pleural effusion and pneumothorax and the performance measures have been computed. The classification outcomes (results) exhibit an accuracy of 94.25% for pleural effusion and 96.77% for Pneumothorax. The sensitivity of the system for pleural effusion was 85.84% and 92% was for pneumothorax and the specificity was 97.5% and 98.27% for pleural effusion and pneumothorax respectively.

Jianhua Yao et al [5] proposed a**n** automated method to evaluate pleural effusion on CT scans, the measurement of which is prohibitively time consuming when performed manually. The method is based on parietal and visceral pleura extraction, active contour models, region growing, Bezier surface fitting, and deformable surface modelling. Twelve CT scans with three manual segmentations were used to validate the automatic segmentation method. The method was then applied on 91 additional scans for visual assessment. The segmentation method yielded a correlation coefficient of 0.97 and a Dice coefficient of 0.72 ± 0.13 when compared to a professional manual segmentation. The visual assessment estimated 83% cases with negligible or small segmentation errors, 14% with medium errors, and 3% with large errors.

Ruskin et al [6] A prospective analysis of anteroposterior supine radiographs in 34 patients was undertaken to determine the detectability of pleural effusions on supine radiographs. The presence of pleural effusions and quantity of fluid (small, moderate, or large) were evaluated by using the following radiographic signs: increased homogeneous density superimposed over the lung, loss of the hemidiaphragm silhouette, blunted costophrenic angle, apical capping, elevation of the hemidiaphragm, decreased visibility of lower-lobe vasculature, and accentuation of the minor fissure. Decubitus radiographs were performed to identify and to estimate the quantity of pleural fluid. Sixty-two hemithoraces were evaluated by three observers. From a total of 36 pleural effusions shown on decubitus views, 24 were correctly identified on supine radiographs (sensitivity of 67%, specificity of 70%, and accuracy of 67%). The most frequent but least specific criterion for detecting pleural effusions on supine radiographs is blunting of the costophrenic angle. Other helpful signs include loss of the hemidiaphragm and increased density of the hemithorax. A normal supine radiograph does not exclude a pleural effusion. Our results show that supine radiographs are only moderately sensitive and specific for the evaluation of pleural effusions.

Anita et al [7] presents a nerve that emerges from the second, third and fourth lumbar nerves and supplies the muscles and skin of the front area of the thigh is known as femoral nerve. It is the fundamental nerve that reaches

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

out to the knee muscles. The harm of this nerves influences the strolling capacity of a human. A continuous picture are gathered from general society database. As the gathered pictures are more inclined to clamor, they are preprocessed. The preprocessed pictures are next portioned by Active Contour Method. From the divided pictures, shape and surface highlights, for example, territory, unusualness strength, edge, real pivot length, differentiate, connection. Vitality and homogeneity are separated. Highlight extraction chooses the pertinent characteristics for arranging and recognizing the influenced zones. The pictures are named threatening or generous utilizing Random Forest Classifier. Execution lists, for example, affectability, specificity and precision were figured.

Abolmaesumi et al [8] has proposed a novel division system for separating cavity forms from ultrasound pictures. The issue is first discretized by anticipating equispaced radii from a discretionary seed point inside the hole toward its limit. The separation of the depression limit from the seed point is demonstrated by the direction of a moving item. The movement of this moving item is thought to be represented by a limited arrangement of dynamical model's liable to vulnerability. Hopeful edge focuses got along every span incorporate the estimation of the question position and some false returns. The displaying approach empowers us to utilize the interfacing various model estimator alongside a probabilistic information affiliation channel, for shape extraction. The joining rate of the technique is quick since it doesn't utilize any numerical advancement. The power and exactness of the technique are shown by fragmenting shapes from a progression of ultrasound pictures. The outcomes are approved through examination with manual divisions performed by a specialist. An utilization of the strategy in dividing bone forms from figured tomography pictures is additionally introduced.

John et al [9] Advances in imaging technology and computer science have greatly enhanced interpretation of medical images and contributed to early diagnosis. The typical architecture of a Computer Aided Diagnosis (CAD) system includes image pre-processing, definition of region(s) of interest, features extraction and selection, and classification. In this paper, the principles of CAD systems design and development are demonstrated by means of two examples. The first one focuses on the differentiation between symptomatic and asymptomatic carotid atheromatous plaques. For each plaque, a vector of texture and motion features was estimated, which was then reduced to the most robust ones by means of ANalysis of VAriance (ANOVA). Using fuzzy c-means, the features were then clustered into two classes. Clustering performances of 74%, 79%, and 84% were achieved for texture only, motion only, and combinations of texture and motion features, respectively. The second CAD system presented in this paper supports the diagnosis of focal liver lesions and can characterize liver tissue from Computed Tomography (CT) images as normal, hepatic cyst, hemangioma, and hepatocellular carcinoma. Five texture feature sets were extracted for each lesion, while a genetic algorithm-based feature selection method was applied to identify the most robust features. The selected feature set was fed into an ensemble of neural network classifiers. The achieved classification performance was 100%, 93.75% and 90.63% in the training, validation and testing set, respectively. It is concluded that computerized analysis of medical images in combination with artificial intelligence can be used in clinical practice and may contribute to more efficient diagnosis.

Donohue et al [10] A pleural effusion is excess fluid that collects in the pleural cavity, the fluid-filled space that surrounds the lungs. Surplus amounts of such fluid can impair breathing by limiting the expansion of the lungs during inhalation and can cause severe chest pain. Measuring the fluid volume is indicative of the effectiveness of any treatment but, due to the similarity to surrounding regions, fragments of collapsed lung present and topological changes; accurate quantification of the effusion volume is a highly challenging imaging problem. A novel technique is presented that segments the structures of the inner thoracic cage, fits 2D closed curves to the detected pleural cavity features and propagates the curves transversely. Missing structure is interpolated using adjacent slices and orthogonal planes to correctly approximate the pleura. An adaptive slice-by-slice region growth procedure constrained by the bounding curves then extracts the pleural fluid volume. The applicability of the technique is analysed on two concurrent test samples totalling 35 pleural effusion patients. A strong correlation was observed between the pleural fluid volume obtained computationally and manually by a radiologist (R^2 = 0.94). The mean pleural fluid volume for the semi-automated approach was 482.54 mL (±518.32 mL) compared with 485.9 mL (±532.8 mL) for the manual segmentation. The difference was not statistically significant (Student t-test, $p = 0.98$). The mean overlap of the segmented pleural fluid was 82% (±10%) across all datasets, with a false positive rate of 22% (±11%). The mean duration of the semi-automated segmentation with user interaction was 3.3 minutes (±0.9 minutes) compared with 7.1 minutes (±4.1 minutes) for the manual segmentation. Compared with the current literature, this is an acceptable segmentation result with a considerably lower mean execution time and thus takes a significant step towards a clinically usable pleural effusion segmentation system.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

III. MATERIALS AND METHODS

The proposed system was implemented by MATLAB using GUI interface. First collects a CT image from database. The database consists of 9 test images. Next, preprocessing is an important step in image analysis process to eliminate the noise from an image. Then, thresholding is applied to segment the lung region and the difference between the height of the left lung and right lung is calculated. The features are extracted for the segmented image which is overlapped with the original CT image. Finally, KNN classifier classifies the image either as normal lung or disease affected lung such as pneumothorax or pleural effusion. The block diagram of the proposed system is shown in the figure 4.

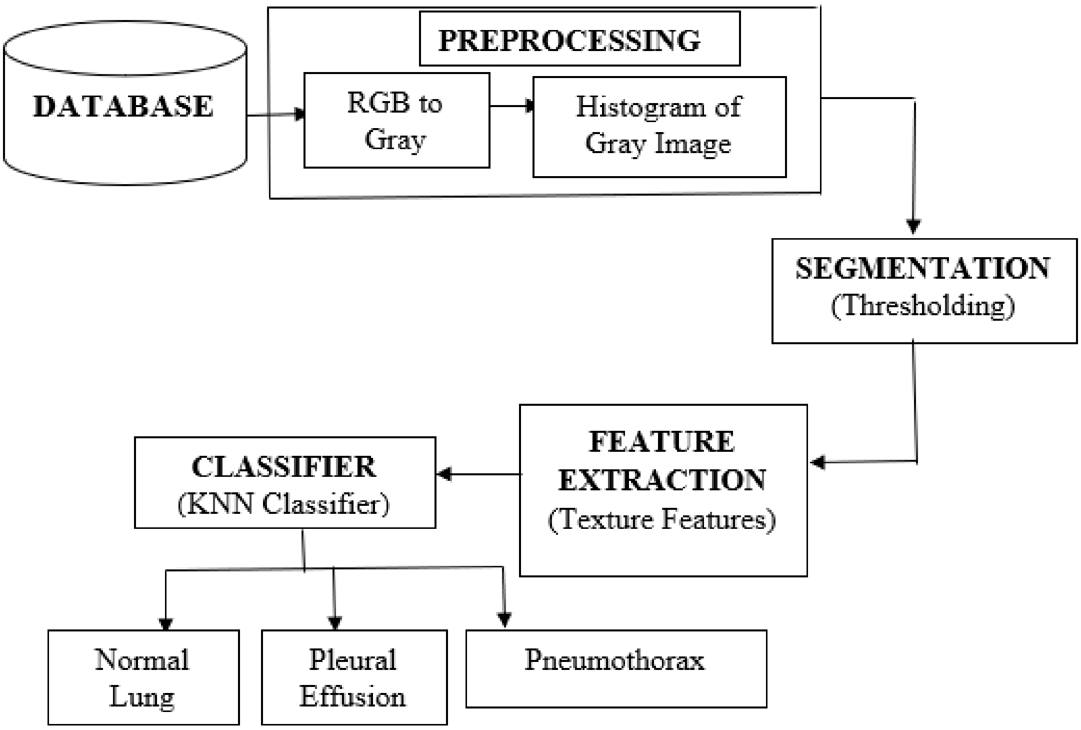

Figure 4. Block diagram of the proposed system

A. Preprocessing

Preprocessing is an important step in image processing to eliminate the noise from an image.

Input:

CT images collected from public database is the input.

Process:

Step 1: Convert the computed tomography RGB image to a gray scale image.

Step 2: Represent the image in bar graph using histogram.

Output:

The output of the preprocessing module is the histogram of gray image.

The steps involved in the preprocessed image are shown in the figure 5.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

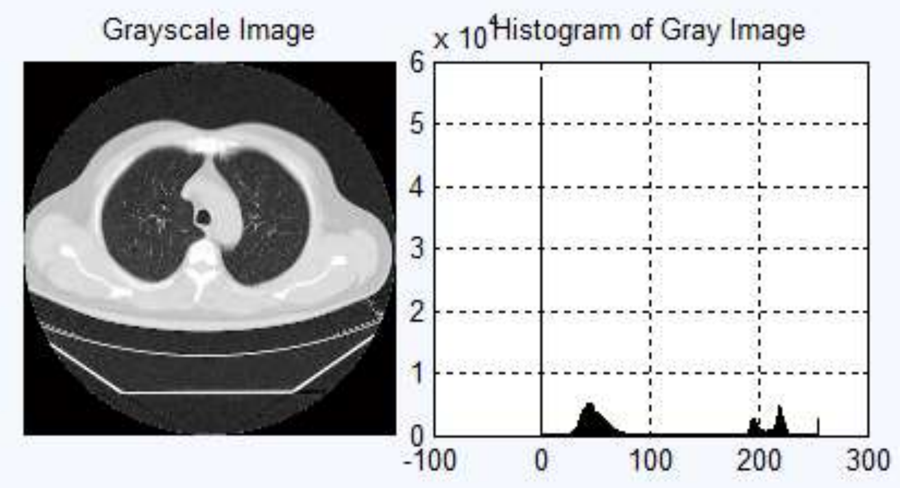

Figure 5. Steps involved in preprocessed image

B. Segmentation

After preprocessing stage, thresholding method is employed to segment the lung region in the CT images.

Input:

Histogram of Gray image.

Process:

Step 1: Apply thresholding on the histogram of gray image.

Step 2: Get rid off stuff touching the border and objects smaller than 1000 pixels.

Step 3: Extract only the two largest blobs and find the centroid of each lung.

Step 4: Calculate the height of each lung and find the difference between them.

Output:

Segmented image overlapped with original gray image.

The steps involved in segmentation is shown in the figure 6.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

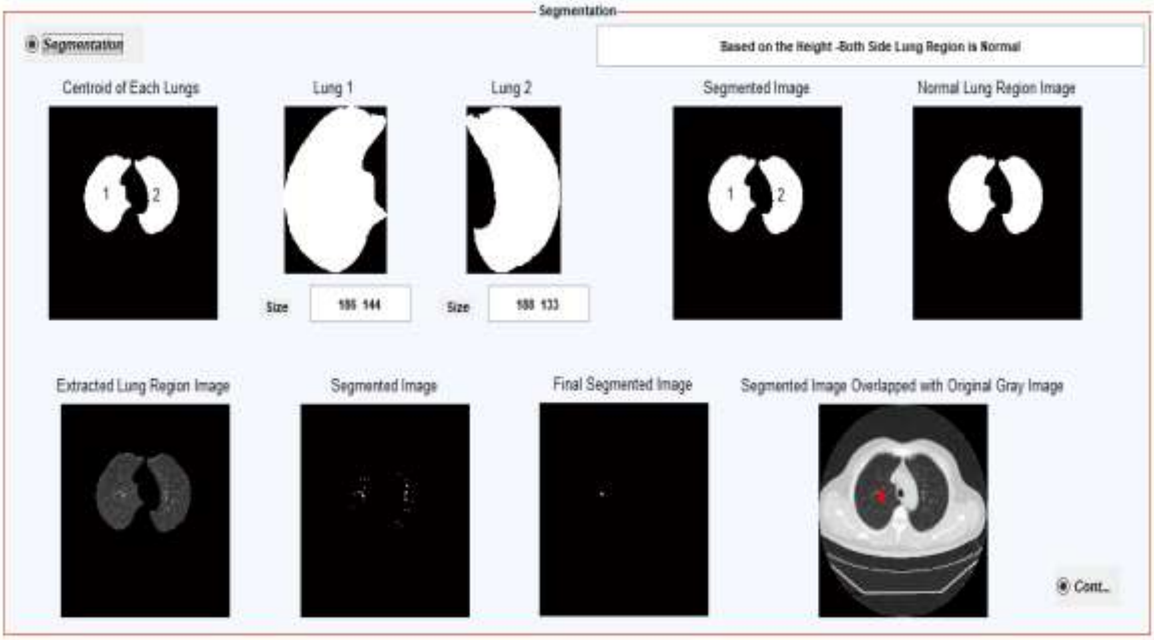

Figure 6. Steps involved in segmentation

C. Feature Extraction

From the segmented image, features like energy, contrast, correlation, homogeneity and area are extracted.

Input:

Segmented image overlapped with original gray image.

Process:

Step 1: Compute the 'Energy' feature which is the highest value when all values in the co-occurrence matrix are all equal using eqn (1),

$$\boldsymbol{Energy} = \sum_{i,j=0}^{N-1} (\boldsymbol{Pij})^2 \quad (1)$$

Step 2: Compute the 'Contrast' feature is the element difference moment of order II, which has a relatively low value when the high values of P are near the main diagonal using eqn (2),

$$\boldsymbol{Contrast} = \sum_{i,j=0}^{N-1} \boldsymbol{Pij}(i-j)^2 \quad (2)$$

Step 3: Compute the 'Correlation' feature that measures the linear dependency of gray levels of neighboring pixels using eqn (3),

$$\boldsymbol{Correlation} = \sum_{i,j=0}^{N-1} \boldsymbol{Pij} \frac{(i-\mu)(j-\mu)}{\sigma^2} \quad (3)$$

Step 4: Compute the 'Homogeneity' feature that measures the closeness of the distribution of elements in the gray level matrix and helps in better segmentation using eqn (4),

$$\boldsymbol{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{\boldsymbol{Pij}}{1+(i-j)^2} \quad (4)$$

Step 5: Compute the 'area' feature which is the actual number of pixels in an image.

Output:

Five different features are computed.

D. Classification

The extracted features are next used for classification. The database consists of 9 test CT images. Among those nine images 4 are lung affected with pneumothorax, 3 are lung affected with pleural effusion and remaining are normal lung. KNN classifier classifies the image either as normal lung or disease affected lung such as pneumothorax or pleural effusion.

IV. RESULTS AND DISCUSSION

The system was implemented using MATLAB GUI interface. Initially, CT image is collected from database which consists of nine CT images. The preprocessing is the important step in image processing to eliminate the noise from an image. After preprocessing stage, thresholding method is employed to segment the lung region in the CT images. The inner white region and outer white region of the CT image is removed using image clear border. The left and right blob of the lung are segmented and if there are any holes in lung region will also be filled. The centroid and their heights of each lung are found. The height of left lung is denoted by ns1 and the heights of right lung is denoted by ms1. The difference between the height of left lung and right lung is less than the threshold value 11 then it is denoted as normal lung if it is greater than the threshold value 11 then is denoted as the disease affected lung. The segmented image overlapped with original gray image. Then the features like area, contrast, correlation, energy and homogeneity are extracted. The last process is classification. The extracted features are next used for classification. The database consists of 9 test CT images. Among those nine images 4 are lung affected with pneumothorax, 3 are lung affected with pleural effusion and remaining are normal lung. KNN classifier classifies the image either as normal lung or disease affected lung such as pneumothorax or pleural effusion. The feature extracted values of five different features are shown in the figure 7.

The overall performance analysis is shown in the figure 8. Here the KNN classifier is compared with other classifiers like SVM (Support Vector Machine) and RF (Random Forest) classifier. From the performance analysis, KNN Classifier is highly accurate than SVM and RF classifier and it is simplest algorithm easy to explain and understand. The accuracy and precision using KNN Classifier is 90% & 100%.

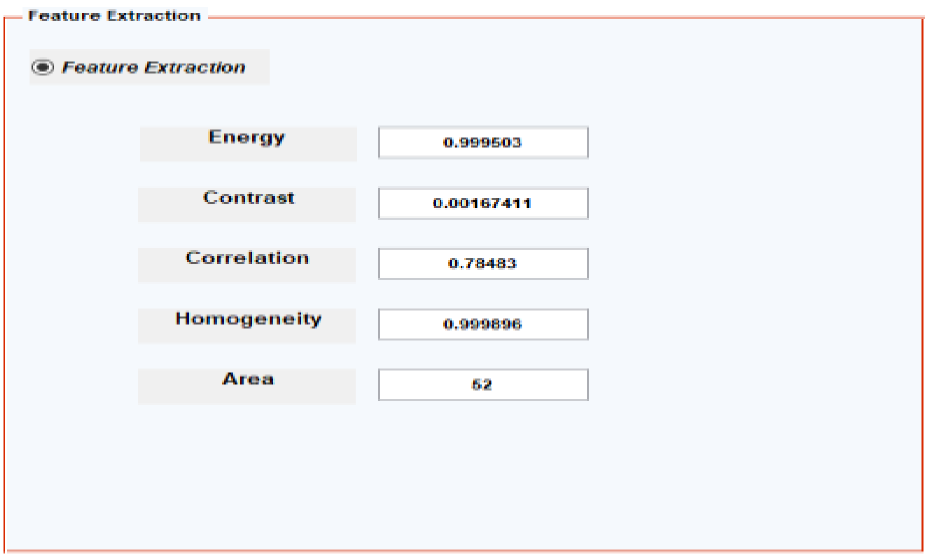

Figure 7. Feature extracted values

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

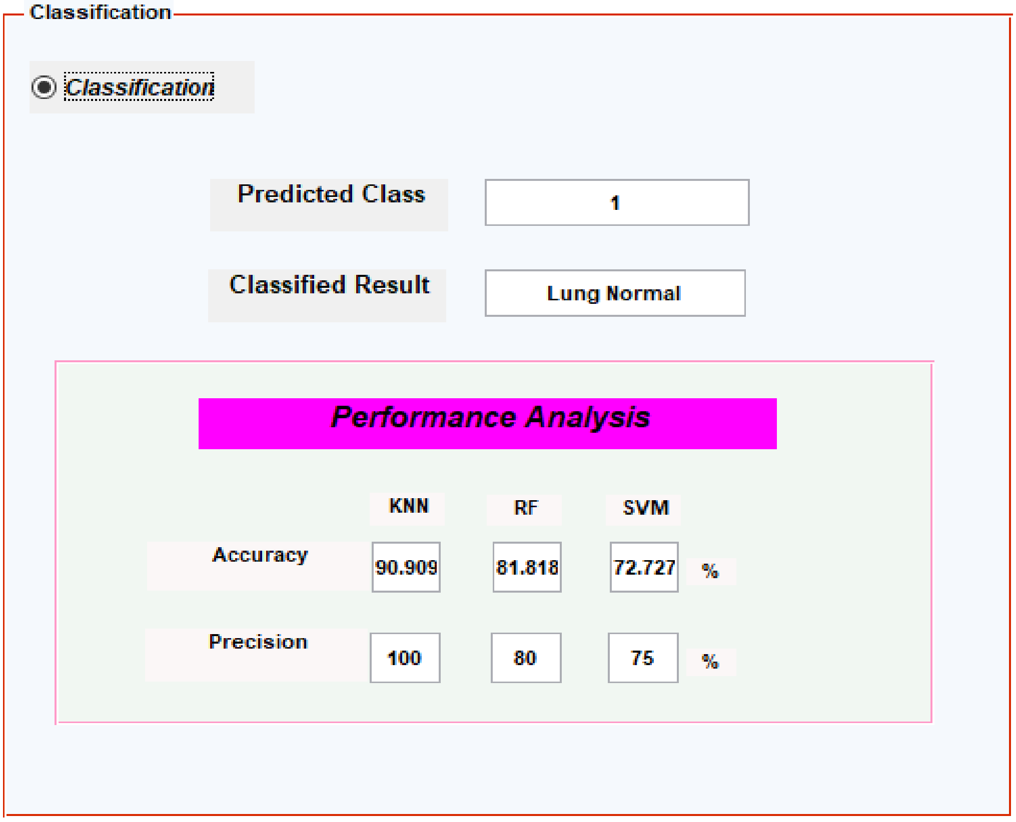

Figure 8. Overall system performance analysis

V. CONCLUSION

KNN Classifier has numerous experts, for example, high precision, adaptable and cons like high memory prerequisite, forecast organize is moderate, delicate to superfluous highlights and size of the information and computationally high costly. KNN classifier is utilized to characterize the picture either as typical lung or infection influenced lung, for example, pneumothorax or pleural emission. The thresholding is utilized to fragment the lung and it is covered with the first dark picture to extricate the highlights. Here the KNN classifier is contrasted and different classifiers like SVM (Support Vector Machine) and RF (Random Forest) classifier. KNN stores the whole preparing dataset which it utilizes as its portrayal. From the execution examination, KNN Classifier is very precise than SVM and RF classifier and it is least complex calculation simple to clarify and get it. The exactness and accuracy utilizing KNN Classifier is 90% and 100%. Future work incorporates to enhance the exactness and the nature of the pictures.

VI. REFERENCES

1. Anita Titus, H. Khanna Nehemiah, A Kannan and J. Jabez Christopher "A Morphological Approach for the Extraction of the Disorders of the Pleura from Chest CT Slices", Emerging Research in Computing, Information, Communication and Applications (ERCICA) 2014.
2. Gil Paulo Melendez and Macario Cordel, "Texture-based detection of lung pathology in chest radiographs using local binary patterns," IEEE Transaction, 978-1-4673-9158-0115, Nov2016.
3. R. Janudhivya, S. Gomathi, P. Madhu Mathi, J. Seetha, "A New Approach For Lung Cancer Cell Detection Using Mumford-Shah Algorithm" IEEE Transaction, Nov 2016.
4. Sonali Singh and I. Mary Sajin Sanju , "The Detection of Pulmonary Nodules in CT Images Using Heuristic Approach Segmentation And Classification", ARPN Journal of Engineering and Applied Sciences, VOL. 11, NO. 9, MAY 201
5. Yao J, Bliton J, Summers RM., "Automatic segmentation and measurement of pleural effusions on CT," IEEE Transaction BioMed Eng, 60(7): 1834–1840.Jan 2013.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

6. J. A. Ruskin, J. W. Gurney, M. K. Thorsen and L. R. Goodman, Detection of Pleural Effusion on Supine Chest Radiographs, American Journal of Roentgenology, vol. 148, pp. 681–683, (1987).
7. Anita Titus, Nirmaladevi R, “Reflexive spotting of deep vein thrombosis in the femoral nerve using Rf Classifier Model,” IOSR Journal, vol.13, Issue 2, May 2018.
8. P. Abolmaesumi and M. R. Sirouspour, “An interacting multiple model probabilistic data association filter for cavity boundary extraction from ultrasound images,” IEEE Trans. Med. Imag., vol. 23, no.6, pp. 772–784, Jun. 2004.
9. John Stoitsis,, Ioannis Valavanis , Stavroul G. Mougiakakou , Spyretta Golemati , Alexandra Nikita , Konstantina S. Nikita “Computer aided diagnosis based on medical image processing and artificial intelligence methods” Nuclear Instruments and Methods in Physics Research A 569 (2006) 591–595,Sept 2006.
10. Donohue, Rory James, “Semi-automated segmentation of pleural effusions in 3D thoracic X-ray computed tomography”, NCBI, 2014.

CITE AN ARTICLE

Titus, A., & Nirmaladevi, R. (2018). COMPUTER AIDED DIAGNOSIS OF LUNG DISEASES USING KNN CLASSIFIER. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*,*7*(6), 199-208.