

Utilizing Deep Learning Techniques to Diagnose Nodules in Lung Computed Tomography (CT) Scan Images

Sugandha Saxena, S.N Prasad, Deepthi Murthy T S

Abstract—There are different methods available for detecting lung cancer including CT (Computed Tomography) scan, MRI(Magnetic Resonance Imaging) scan etc. Among all methods, CT scan images are preferred more because they can detect a very small nodule in the lungs. Early treatment can be given to patients if it is diagnosed at early stages, hence reducing the number of deaths. This paper shows that Median Filter outperformed the Average, Gaussian, Laplacian, and Wiener Filters in the preprocessing stage for the removal of noise from images. Additionally, a study has been conducted on several image segmentation algorithms, such as clustering, watershed, and Thresholding segmentation. This was followed by the extraction and classification of nodules. Different performance parameters have been calculated to validate the results of the model and it is discovered that proposed model has greatest performance.

Indexing Terms—CT scan images, Lung cancer, K-Means Clustering, Otsu Thresholding, Preprocessing Filters, Watershed Segmentation.

I. INTRODUCTION

Even though many different types of cancer have been discovered, lung cancer is considered the most severe and fatal illness. According to the American Cancer Society publication, roughly 1,31,880 deaths from lung cancer were estimated in the United States alone in 2021[1]. It affects approximately 1 in every 13 males and 1 in every 16 females. Early treatment is the only way to cure this disease in time. With the use of an automated cancer detection system, lung cancer can be detected early and successfully. Tumors are formed when a human body cell begins to divide in an uncontrolled manner and the size may exceed abnormally in the lungs. This tumor can be cancerous if not treated timely. Lung cancer can be classified into two main types: SCLC (Small Cell Lung Cancer) and NSCLC (Non-Small Cell Lung Cancer[2].

Manuscript received April 22, 2023; revised April 28, 2023.

Sugandha Saxena is a Research Scholar of School of Electronics and Communication Engineering, REVA University, Bangalore, Karnataka, 560064, India. (phone: +91-9972925838; e-mail: sugandhasxn@gmail.com).

S.N Prasad is a Professor of Department of Electrical and Electronics Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka, 576104, India. (Corresponding author phone: +91 9845177025; e-mail: sn.prasad@manipal.edu).

Deepthi Murthy T S is an Associate Professor of School of Electronics and Communication Engineering, REVA University, Bangalore, Karnataka, 560064,India(phone:+91-9986132129;e-mail: deepthimurthy@reva.edu.in).

Various methods are used by doctors to detect lung cancer. These methods include CT scan, MRI scan, radiography, etc. Among all, a CT scan is chosen because it makes a detailed image of human lungs by using X-rays and can detect a very small lump or nodule in the lung. Unfortunately, noise present in CT scan images lowers the quality of the image, which can make it challenging for radiologists to detect lung cancer in its early stages. Much research has been done to remove different types of noises associated with medical images such as CT scan. A review of the literature reveals that researchers have utilized various pre-processing filters on lung CT scan images with the goal of improving the image quality by increasing PSNR (Peak Signal to Noise Ratio) and decreasing MSE (Mean Square Error).

After the noise has been removed, image segmentation is carried out to detect nodules of different shapes in the lung. Many researchers have used supervised image segmentation approaches such as Hessian-based vascular feature extraction which helped in removing blood vessels and later nodules were detected and classified using a neural network. Although these methods are fast and straightforward to execute, they have resulted in low accuracy as they could not extract nodules of different sizes[3]. Hence, it is crucial to devise a new and dependable method which can diagnose nodules of different shapes with high accuracy so that further it can be classified as malignant or benign.

In this paper, several pre-processing filters have been applied to lung CT scan images. Performance parameter comparison of various pre-processing filters has been done for a few images from public and private databases. Both databases consist of lung CT scan images. Public database images have been taken from LIDC (Lung Image Database Consortium) and a private database has been created by acquiring images of existing patients from A.J Institute of Medical Sciences, Mangalore, India. After this, multiple segmentation methods are applied and compared to identify the optimal technique for detecting lung nodules with high accuracy irrespective of their size and shape. It has been observed that combining the Median Filter in the pre-processing stage with the K-Means Clustering method in the segmentation stage yields better and more accurate results than other techniques. Later, Maximum Sensitivity Neural Network (MSNN) is proposed to detect lung cancer accurately. This paper is organized as follows: Section II provides review of existing methods. Section III describes the methodology adopted for this work. Section IV describes different filtering techniques which are used in pre-processing stage. Section V explains different image

segmentation algorithms used to detect lung nodules. Section VI explains various performance parameters calculated to estimate the image quality. Section VII explains the architecture of MSNN, Section VIII explains the experimental results. Conclusion is explained in section IX.

II. LITERATURE SURVEY

Identification of nodules in lung with high accuracy is very important. This is a challenging task for radiologists and more time is needed for investigations. Patient healthcare is delayed especially in poor nations due to a shortage of radiologists. There are many existing strategies for identifying lung cancer nodules using lung CT scan images. This section summarises recent findings from researchers who created models for detecting lung cancer nodules.

Mohd Firdaus Abdullah et al [4] have done a comparative study between different image segmentation techniques such as Otsu Thresholding, Watershed Segmentation, and K-Means Clustering. The results showed that Watershed Segmentation outperformed other approaches, with an accuracy of 99.8553%. However, K-Means Clustering Algorithm also gave high accuracy of 99.02% but only for uniform nodule shapes.

M. Vas and A Dessai [5] used the morphological operation method for segmenting lung CT scan images. The hospital database gave an accuracy of 92 %. This is due to improper segmentation. In their image, the white portion indicated the aortic region, but it was also considered a nodule part leading to less accuracy.

K. Senthil Kumar et al. [6] showed that the Adaptive Median filter outperformed the Median and Average filters in the preprocessing step for medical CT scan images. Five segmentation algorithms were used, and it was observed that Guaranteed Convergence Particle Swarm Optimization (GCPSO) had 90% accuracy for tumor extraction from lung CT scan images even though this algorithm does not have guaranteed convergence to a local optimum. It means that the tumor may not be detected perfectly and, thus, accuracy can be improved further.

P. Muthamil Selvi and B. Ashadevi [7] used lung CT scan images and applied different filters such as Median, Wiener, and Mean. Based on the analysis, it was found that the Median filter, with its high PSNR and low MSE values, was more efficient than other filters at reducing noise.

Suren Makajua, et al. [8] proposed a system utilizing Watershed segmentation as a means of identifying the cancerous nodule in a lung CT scan image. In this work, nodule was detected with 92% accuracy by using LIDC dataset. Although, the proposed system has still not achieved high accuracy that is near 100%.

Asif Iqbal Khana, et al. [9] employed CoroNet model which can detect COVID-19 infection. ImageNet dataset was used to train the Xception architecture, which serves as the foundation for the model. It was observed that the model achieved 93% precision, 98.2% of recall, and an accuracy of 89.6%.

III. METHODOLOGY

This section outlines the flow of work involved in detecting a nodule. The first part of the work is database acquisition from a reputed research centre. In this work, lung CT scan

images are used from a database that has been studied under the supervision of a radiologist. These are 512X512 pixel colour CT scan images. The pre-processing stage receives the coloured image after it has been transformed to a grayscale image.

The second part of the work is an application of several pre-processing filter like the Average filter, Gaussian filter, Laplacian filter, Median filter and Wiener filter to lung CT scan images. This step aims to enhance the quality of the image by eliminating any noise present in it. Different filters have been evaluated by using various performance metrics such as MSE, PSNR, Speckle Suppression Mean Preservation index and Speckle Suppression Index to compare their performance.

The application of different image segmentation algorithms is carried out in the third part of the work. This is a key stage in medical applications as it helps in the analysis of what is present inside image and identification of region of interest within the image [10]. Since it involves human lives, accuracy is essential and therefore, all the algorithms have been compared to identify the best one among all. Finally, the best segmentation algorithm results are used to extract the tumor. The model's flow is depicted in Figure 1.

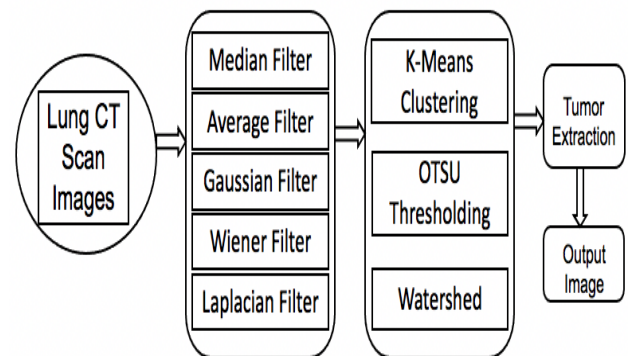


Fig 1 Flow diagram of the Model

IV. PREPROCESSING FILTERS

To remove noise while retaining the image quality, several filtering algorithms are applied.

A. Median Filter

A median filter is a nonlinear approach that may remove various types of noise and restore image clarity to a larger level [11]. A 3x3 mask covers the entire image and substitutes each element with its median value [12]. The mask is applied to the first element of the image to determine the median, and the mask elements are sorted in ascending order [13]. The mask is subsequently moved to the next element until all the image elements have been covered.

B. Average Filter

An additional technique for reducing spatial noise in the input image involves utilizing an average filter. This approach involves a matrix "X" that has V rows and U columns as the input image. The first step is to expand the size of the input image matrix to have V+2 rows and U+2 columns by appending zeros. Next, a 3X3 mask is applied to the first element of the input image matrix X, and the mean of all elements in the mask is calculated. This average value is used to replace the value of the element X (1,1) in the input

image matrix. This process is then repeated by shifting the mask to cover all the elements of the input matrix with the average value of its neighboring pixels, and the steps are repeated accordingly [13].

C. Wiener Filter

It is an efficient method for balancing reverse filtering and noise smoothing. It can be used to eliminate additive noise using a statistical approach. This filter is best for MSE, Local Mean and Variance calculation around each pixel. Equation (1) can be used to express the Wiener filter in the Fourier domain [13].

$$S(f_1, f_2) = \frac{H(f_1, f_2)S_{yy}(f_1, f_2)}{H(f_1, f_2)^2 S_{yy}(f_1, f_2) + S_{nn}(f_1, f_2)} \quad (1)$$

Where,

$S_{yy}(f_1, f_2)$ - original image power spectrum.

$S_{nn}(f_1, f_2)$ - additive noise power spectrum.

$H(f_1, f_2)$ – filter with the blurring factor.

D. Gaussian Filter

The Gaussian filter uses a 2D pattern to smooth an image while preserving its edges. The image is convolved with 2D Gaussian Distribution Function [14]. When applied to an image, this filter works in two steps. The first step of horizontal filtering in the input image involves placing the filter over each pixel and using its value as the center. Then, the new pixel values are calculated by multiplying the weight of the filter by the values of all the pixels within the filter's range. The final image is produced by vertically filtering each pixel from the horizontally processed image. Equation (2) below can be used to calculate the Gaussian $G(x,y)$ of an image.

$$G(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (2)$$

E. Laplacian Filter

It is used to find edges in the input image and is also known as an edge detector. The filter calculates the second derivative of an image by observing the rate of change in its first derivative. In order to generate new pixel values from an input image, the weights are convolved with each pixel value as well as the neighboring pixels. It is also used to improve features of Image (I) that have a lot of acute discontinuities [14]. Equation (3) can be used for Laplacian filter.

$$L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad (3)$$

V. SEGMENTATION TECHNIQUES

Segmentation is a technique that enables the identification of objects within an image, facilitating the localization of a specific area of interest (RoI). This step is significant as it helps to find the relevant information by partitioning the image into small regions. Small nodules can be separated from the lungs to identify whether it is cancerous or not. This work employs three methods of segmentation: Otsu Thresholding [15], K-Means Clustering, and Watershed.

A. K-Means Clustering Algorithm

1. Started by assuming the number of clusters (k) as $m_1, m_2, m_3, \dots, m_k$.

2. Firstly, Euclidean distance is calculated between every pixel x_i and clusters so that each pixel can be assigned to the nearest cluster. Equation (4) is used to compute the distance.

$$d = \left(\sum_{j=1}^k (x_{ij} - m_j)^2\right)^{1/2} \quad (4)$$

Where i value ranges from 1 to N, j value ranges from 1 to N.

d=computes the distance between every pixel and cluster.

3. Now finding the new cluster center by calculating the mean of pixels in each cluster. This step is performed if the grouping is not done properly. Equation (5) shown below can be used for the same.

$$Mean = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij} \quad (5)$$

Where i ranges from 1 to N

J ranges from 1 to N

N defines Number of samples of current cluster i.

4. Keep repeating step number 2 and 3 until grouping was done properly or a certain condition is achieved.

B. Watershed Algorithm

1. Start the algorithm by reading the image.
2. Convert the image to a grayscale image.
3. Perform morphological operation to subtract background from the image.
4. Adjust the image intensity value.
5. Obtain a binary image by employing the thresholding method.
6. Obtain the inverse image of black and white image.
7. Now calculate the distance transform of the resultant image.
8. Obtain watershed image.

C. Otsu Thresholding Algorithm

1. Start the algorithm by reading the image.
2. Calculate the histogram and probability distribution for each level of intensity.
3. Now step through all possible thresholds where intensity is maximum.
4. Calculate the sum of probabilities for background and foreground of the image.
5. Calculate sigma which is inter-class variance.
6. Calculate desired threshold corresponding to maximum variance
7. Determine the mean value of the image's background.
8. Determine the mean value of the image's foreground.

VI. NODULE DETECTION

The following algorithm steps can be followed to extract nodules (RoI) from lung CT scan image.

1. Start the algorithm by reading the image.
2. Calculate the threshold by using Otsu segmentation technique.
3. Select the seed pixel.
4. Calculate the gray value between the seed and the other pixels.
5. Check the pixel. If the gray value is low compared to the threshold value, then the seed pixel is like the surrounding pixels and the merging of pixels happens to form a region.

6. The algorithm halts if the grayscale value exceeds the threshold value.
7. Repeat from step 3 until all the pixels in the image are checked.

VII. MAXIMUM SENSITIVITY NEURAL NETWORK

Maximum Sensitivity Neural Network (MSNN) is a convolutional neural network which is proposed for the detection of lung cancer using CT scan images. It learns from the pattern of images and categorizes them into cancerous and noncancerous images. The network input layer takes 512X512 grayscale images. The network architecture which consists of five consecutive blocks is shown in Fig 2. Every block is made up of a series of sequential layers including convolution, normalized batch layer, rectified linear unit and max pooling layer.

The following is a description of the different layers used in the network:

The initial layer carries out the convolution operation between the input image(f) and filter size(g) by using equation (6):

$$a(x) * b(x) = \sum_{k=-\infty}^{\infty} a(k).b(x - k) \quad (6)$$

where x and k are spatial variables.

In general, a smaller filter size may lead to an overfitting issue, while a bigger filter size may increase the underfitting issue. Therefore, this layer uses 8 filters with a 6x6 ideal filter size.

The next successive layer is batch normalization (BN) layer which expedites training speed and lessens network sensitivity. Therefore, performing normalization over a batch(v) of m instances for 'i' unit can be done using the following steps.

Firstly, compute batch mean by using equation (7):

$$\mu_i = \sum_{r=1}^m v_i^r / m \quad (7)$$

Where r ranges from 1 to m .

Secondly, compute batch variance by using equation (8):

$$\sigma_i^2 = \sum_{r=1}^m (v_i^r - \mu_i)^2 / m \quad (8)$$

Thirdly, compute normalized batch instances by using equation (9):

$$v_n^r = v_i^r - \mu_i / \sigma_i \quad (9)$$

Lastly, scale with learnable parameters by using equation

$$a_i^r = \gamma_i * v_n^r + \beta_i$$

Rectified Linear Unit (ReLU) layer helps to add nonlinearity to the network by adding a rectifier function which is computing linear operations during convolution. The function works by using equation (10):

$$f(x) = 0 \text{ if } x < 0 \text{ and } f(x) = x \text{ if } x > 0 \quad (10)$$

The purpose of max pooling layers is to reduce the computational expenses by shrinking the size of the convolved feature map. Each input image is given a probability score using a series of FC (Fully Connected), BN, ReLU, FC, and SM (Soft Max) layers. Where FC layer helps in the classification of images into categories. And SM layer converts the output of the last layer of network into a probability distribution.

For this work, a private database has been acquired to fully assess the efficacy of the network. Table I presents the classification accuracy of MSNN while analyzing lung CT scan images.

TABLE I. PERFORMANCE METRICS COMPARISON WITH OTHER METHODS

	Performance Metrics (%)				
	Accuracy	Sensitivity	Precision	Specificity	F-Score
Faster R-CNN[16]	80.1	-	-	-	-
Cascade R-CNN[17]	84	-	-	-	-
SC-Dynamic R-CNN[18]	88.1	-	-	-	-
MSNN (Proposed Method)	96	100	94	93	97

VIII. PERFORMANCE MEASURE

Equations (11-14) shown below helps to calculate various performance parameters of pre-processing filters:

A. Mean Square Error (MSE)

MSE can be calculated by taking the difference between the estimated and true image pixel values [19]. It represents two monochromatic images (X, Y) of size $a \times b$, one of which is a noisy approximation of the other and is determined by equation(11):

$$MSE = \frac{1}{ab} \sum_{i=0}^{a-1} \sum_{j=0}^{b-1} [X(i,j) - Y(i,j)]^2 \quad (11)$$

B. Peak Signal to Noise Ratio (PSNR)

Image maximum pixel value (Max_i) to its MSE value is called PSNR [19]. Equation (12) below can be used to calculate this parameter.

$$PSNR = 10 \log_{10} \left[\frac{Max_i^2}{MSE} \right] \quad (12)$$

C. Speckle Suppression Index (SSI)

This parameter value should be less than 1 which indicates that the filter is capable of successfully removing speckle noise[20]. Equation (13) below can be used to calculate this parameter where I_f and I_o are filtered and the original image, respectively. Var represents the variance of image.

$$SSI = \frac{\sqrt{Var(I_f) * mean(I_o)}}{\sqrt{Var(I_o) * mean(I_f)}} \quad (13)$$

D. Speckle Suppression and Mean Preservation Index (SMPI)

This parameter helps in determining how well the filter works[20]. A low value for this index indicates improved filter performance. Equation (14) below can be used to calculate SMPI where I_o and I_f are the original and filtered images, respectively.

$$SMPI = Q * \frac{\sqrt{Var(I_f)}}{\sqrt{Var(I_o)}} \quad (14)$$

Where $Q = 1 + |mean(I_o) - mean(I_f)|$

Equations (15-18) shown below help to calculate various performance parameters for segmentation and classification.

A. Accuracy

This parameter defines true positive and true negative or correct cases over the total number of cases which includes false cases too. Equation (15) below can be used for accuracy calculation:

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad (15)$$

B. Precision

This parameter determines true positive cases over true positive and true negative cases. Equation (16) below can be used for precision calculation:

$$Precision = \frac{TP}{TP+FP} \quad (16)$$

C. Recall

This parameter measures the true positive rate. A value close to 100% means the test result is positive and the patient has a disease. Equation (17) below can be used for recall calculation:

$$Recall = \frac{TP}{TP+FN} \quad (17)$$

D. F-Score

This parameter determines how many cases are classified correctly. Equation (18) below can be used for calculating this parameter:

$$F\text{-Score} = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (18)$$

IX. RESULTS AND DISCUSSIONS

Table II presents a comparison of the effectiveness of different preprocessing filters, including the Median, Average, Gaussian, Laplacian, and Weiner filters. [21] To select a suitable filtering method, the resulting image should exhibit a low MSE value, a high PSNR value, an SSI value below 1, and the lowest possible SMPI value. These criteria were assessed for several images obtained from public and private datasets. Based on the findings, the Median filter was found to be superior to other filters in both databases.

For image1, the public and private databases have low MSE values of 2.5365 and 1.475155, high PSNR values of 44.0885 and 46.4424355, SMPI values of 1.0500 and 1.10719, respectively and SSI value obtained less than 1. For image 2, the public and private databases have low MSE values of 4.9004 and 3.2976, and PSNR value is 42.9488 and 41.2204, respectively. SSI value is less than 1 and the SMPI value obtained is 1.1776 and 1.1966 for public and private databases. Even though the PSNR value was found to be higher for the Laplacian filter when applied to this image, the MSE and SMPI values were substantially higher, and the SSI value was greater than 1, which suggests poor filter performance. Similarly, for image 3, the public and private databases have MSE values of 3.8416 and 1.7622; high PSNR values of 42.2856 and 45.6700, SMPI values of 1.2617 and 1.2527, respectively and SSI value obtained is less than 1. For image 4, the public and private databases have low MSE values of 5.3381 and 2.4859, high PSNR values of 40.8569 and 44.3147, SMPI values of 1.3109 and 1.2783, respectively and SSI value obtained is less than 1. This signifies that using images from any database, Median filter will only outperform compared to other filters.

Furthermore, by using lung CT scan images with nodules from a private database, performance metrics are calculated and compared. Table III, IV, V and VI shows MSE, PSNR, SMPI and SSI values are calculated for different images by using Gaussian, Wiener, Laplacian, Median and Average filters which are plotted for simple analysis. Fig 3-6 clearly demonstrate that the Median filter, which is highlighted, outperforms all other filters. This is evidenced by its lower mean square error values, higher PSNR values, lower SMPI values, and SSI values less than 1 for all images. In contrast, the Laplacian filter had very high SMPI and SSI values, rendering them negligible in the plotted graph.

Tables VII, VIII, IX show the calculation of accuracy, recall, precision, and F-score for different images by using Otsu thresholding, K-means clustering and Watershed segmentation techniques. Fig 7-10 clearly highlights good performance for K-means clustering as it reflects high recall, accuracy and F-score value with a little compromise in precision value.

The average performance of the three segmentation techniques is shown in Table X. Otsu thresholding shows that the average accuracy value is 61%, the precision value is 96 %, the recall value is 56 % and the F-score value is 71%. According to the Watershed method, the average accuracy value is 81 %, the precision value is 88%, the recall value is 89 % and the F-score value is 88%. Whereas K-Means clustering shows an average value for accuracy of 88%, precision value of 88%, recall value of 99% and F-score value of 92 %. After comparison, it is observed that K- Means Clustering outperforms other approaches and could be the best option for image segmentation.

The elbow technique can be used in K-Means Clustering which helps in the selection of several clusters by showing the number of clusters on the x-axis and the distance of each pixel from the centroid on the y-axis. You will receive an elbow-like curve after plotting. Cluster value is determined by the location of the elbow point. Further, nodule extraction is carried out using the algorithm steps discussed in this work. Nodules can have different shapes such as oval, lobulated, polygonal, or ragged. Nodules with round or oval margins are referred to as 'round'; distinct margins with some smooth large convexities are referred to as 'lobulated'. A margin in polygonal shape is referred to as 'polygonal', while irregular margin is referred to as ragged shape[22][23].

MSNN was developed using MATLAB R2021 on a Windows10 computer with an Intel Core i5 2.50 GHz processor. The network was trained on a dataset of 434 lung CT scan images using a batch size of 20, a learning rate of 0.0001, and an epoch value of 5. The dataset was randomly divided into a training set and a validation set. Specifically, split 1 involved using 80% of the images for training and the remaining 20% for testing. Split 2 involved using 85% of the images for training and the remaining 15% for testing. Split 3 involved using 70% of the images for training and the remaining 30% for testing. Finally, Split 4 involved using 75% of the images for training and the remaining 25% for testing. In every case of split, MSNN showed good performance and the result is reflected in Table I with respect to split 1 only. Fig 11 shows accuracy and loss plot during training and testing of dataset.

Hence, MSNN is found to be very effective in detecting lung cancer after training. It shows that model accuracy increased to 94% when compared to other architectures. Additionally, it provided a sensitivity of 94%, precision of 100%, F-Score of 96% and specificity of 100%. Generally, convolution neural network suffers from overfitting. Therefore, by using a global average pooling layer, the network's complexity is controlled.

Fig 12 shows confusion matrix for different splits of dataset for classification. Fig 13 shows (a) original lung CT scan images with ragged and oval-shaped nodules, (b) image after applying Median filter, (c) edges identified for images using Sobel filter, (d) segmented images after applying K-Means clustering algorithm, (e) only lung region extracted, (f) nodule marked manually by radiologist, and (g) nodule detection of different shapes by using algorithm discussed in this work.

MSNN architecture was able to extract detailed features from lung CT scan images which are shown in Fig 14. Convolution layers can extract features from various angles and therefore, the first and second convolution layer extracted fundamental information which contains spots and edges. Results showed that deeper layers like third, fourth, fifth and sixth layers extracted high level and abstracted features by merging earlier features. Hence, features retrieved from the deeper layers are more suited for classification[24].

A sensitivity map has been plotted to find out which part of the image is most important and helps in classifying the image as cancerous or noncancerous. Fig 15 shows a lung CT scan image where the red area depicts more contribution for classification and the remaining area has no or less contribution. This classification has been shown for both abnormal lungs with cancerous nodule and healthy lungs.

X. Conclusion and Future Work

This work employs multiple techniques for the pre-processing stage to remove noise from lung CT scan images to improve their visualization and for the segmentation stage to identify nodules. The Median filter was found to be more effective than other pre-processing filters on both public and private databases, as it generated output images with a low MSE value, a high PSNR value, a low SMPPI value, and SSI values less than one.

Later, by using the lung CT scan images from a private database, image segmentation techniques such as Otsu Thresholding, Watershed and K-Means Clustering are applied. After comparison, K-Means segmentation outperformed the other two methods. A literature survey showed that high accuracy has been obtained for detecting nodules from lung CT scan images but only for uniform shaped nodules. In this work, different shapes of nodules such as oval and ragged have been detected to find the region of interest. The accuracy attained is high (88 %) for K-Means segmentation, however, it can be improved further in future work by experimenting with alternative segmentation algorithms.

To detect lung cancer using CT scan images, different architectures have been developed with several parameters. However, the literature review found that using a lot of parameters can reduce classification accuracy and makes

calculations challenging[25]. This paper presents the development of a deep learning structure that utilizes a convolutional neural network to identify lung cancer. The extracted characteristics were then used to train a K-Nearest Neighbor (KNN) classifier. The results showed that the MSNN architecture had higher accuracy compared to the other structures.

ACKNOWLEDGMENT

The authors express their gratitude to the Interventional Radiology Division, Department of Radio-Diagnosis, A.J. Institute of Medical Sciences in Mangalore, India for providing the patient database and for their guidance in interpreting the medical images. The authors also appreciate the assistance and resources provided by REVA University for conducting this research.

REFERENCES

- [1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer Statistics, 2021", *Cancer Journal for Clinicians*, vol. 71, no. 1, pp 7–33, 2021.
- [2] Prashant Mathur, Krishnan Sathish Kumar, "Cancer Statistics, 2020: Report from National Cancer Registry Programme, India", *National Centre for Disease Informatics and Research, Bengaluru, India*, vol. 6, no. 6, pp 1063–1075, 2020.
- [3] M. Savic, Y. Ma, G. Ramponi, W. Du, and Y. Peng, "Lung Nodule Segmentation with a Region-Based Fast Marching Method", *Journal of sensors*, vol. 21, no. 5, pp 1–32, 2021.
- [4] Mohd Firdaus Abdullah, Muhammad Safwan Mansor and Siti Noraini Sulaiman, "A Comparative Study of Image Segmentation Technique applied for Lung Cancer Detection", *IEEE International Conference on Control System, Computing and Engineering, Penang, Malaysia*, pp 72–77, 2019.
- [5] M. Vas and A. Dessai, "Lung cancer detection system using lung CT image processing", *International Conference on Computing, Communication, Control and Automation, ICCUBEA*, 2018.
- [6] K. Senthil Kumar, K. Venkatalakshmi, and K. Karthikeyan, "Lung Cancer Detection Using Image Segmentation by means of Various Evolutionary Algorithms", *Computational and Mathematical Methods in Medicine*, pp 107-114, 2019.
- [7] P. M. Selvi and B. Ashadevi, "Elimination of Noise in CT Images of Lung Cancer using Image Preprocessing Filtering Techniques", *International Journal of Advanced Science and Technology*, vol. 29, no. 4s, pp 1823–1832, 2020.
- [8] S. Makaju, P. W. C. Prasad, A. Alsadoon, A. K. Singh, and A. Elchouemi, "Lung Cancer Detection using CT Scan Images", *International Conference on Smart Computing and Communications, Kurukshetra, India*, vol. 125, pp 107–114, 2018.
- [9] A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images", *Computer Methods Programs in Biomedicine*, vol. 196, 2020.
- [10] Y. Jalali, M. Fateh, M. Rezvani, V. Abolghasemi, and M. H. Anisi, "ResBCDU-Net: A Deep Learning Framework for Lung CT Image Segmentation", *Journal of Knowledge Based Systems*, vol. 21, no. 1, pp 1–24, 2021.
- [11] X. Huang, Q. Lei, T. Xie, Y. Zhang, Z. Hu, and Q. Zhou, "Deep Transfer Convolutional Neural Network and Extreme Learning Machine for lung nodule diagnosis on CT images", *Journal of Knowledge-Based Systems*, vol. 204, pp 8357-8361, 2020.
- [12] A. Asuntha and A. Srinivasan, "Deep learning for lung Cancer detection and classification", *Multimedia Tools and Applications 2020 79:11*, vol. 79, no. 11, pp 7731–7762, 2020.
- [13] M. S. Rahman, P. C. Shill, and Z. Homayra, "A New Method for Lung Nodule Detection Using Deep Neural Networks for CT Images", *International Conference on Electrical, Computer and Communication Engineering*, 2019.
- [14] L. M. Pehrson, M. B. Nielsen, and C. A. Lauridsen, "Automatic Pulmonary Nodule Detection Applying Deep Learning or Machine Learning Algorithms to the LIDC-IDRI Database: A Systematic Review," *Diagnostics (Basel)*, vol. 9, no. 1, Mar. 2019.
- [15] S. Saraswathi and L. M. I. Sheela, "Detection of Juxtapleural Nodules in Lung Cancer Cases Using an Optimal Critical Point Selection Algorithm," *Asian Pacific Journal of Cancer Prevention: APJCP*, vol. 18, no. 11, pp 3143–3148, 2017.
- [16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,"

[17] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into High Quality Object Detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp 6154–6162, 2018.

[18] X. Wang, L. Wang, and P. Zheng, "SC-Dynamic R-CNN: A Self-Calibrated Dynamic R-CNN Model for Lung Cancer Lesion Detection," *Computer Mathematical Methods Medicine*, 2022.

[19] N. S. Nadkarni and S. Borkar, "Detection of lung cancer in CT images using image processing," *Proceedings of the International Conference on Trends in Electronics and Informatics*, pp 863–866, 2019.

[20] J. Dehmeshki, H. Amin, M. Valdivieso, and X. Ye, "Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach," *IEEE Trans Med Imaging*, vol. 27, no. 4, pp. 467–480, 2008.

[21] K. Munir, H. Elahi, A. Ayub, F. Frezza, and A. Rizzi, "Cancer Diagnosis Using Deep Learning: A Bibliographic Review," *Cancers (Basel)*, vol. 11, no. 9, 2019.

[22] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognition*, vol. 85, pp 109–119, 2019.

[23] G. A. P. Singh and P. K. Gupta, "Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans," *Neural Computing and Applications* 2018 31:10, vol. 31, no. 10, pp 6863–6877, 2018.

[24] J. Donahue et al., "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition," *31st International Conference on Machine Learning, ICML 2014*, vol. 2, pp 988–996, 2013.

[25] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp 6232–6251, 2016.

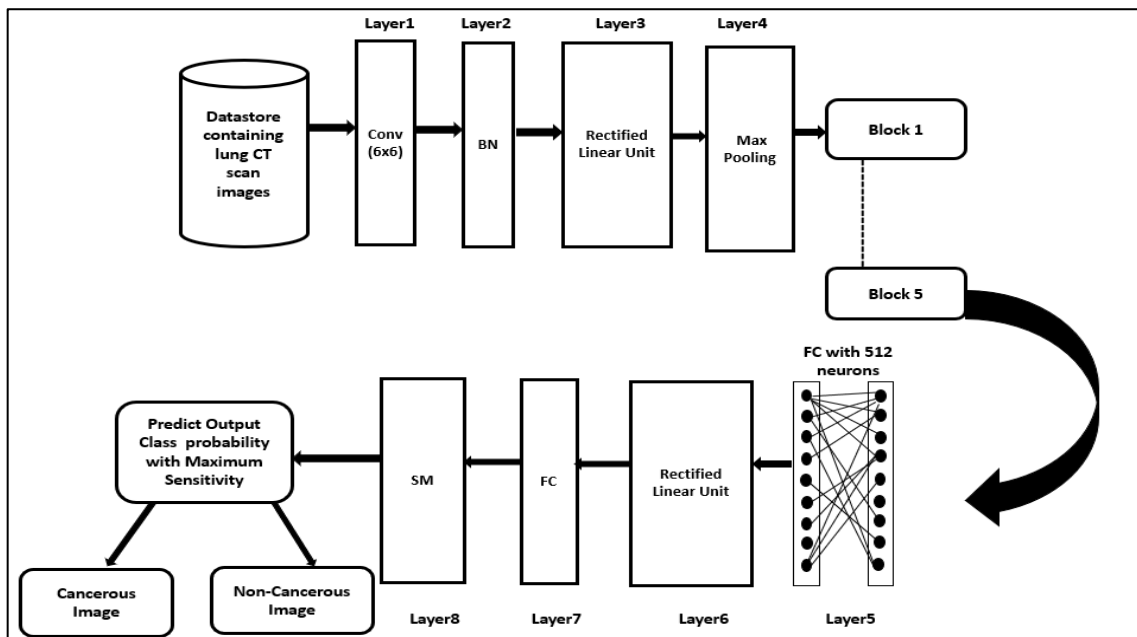


Fig 2 Maximum Sensitivity Neural Network Architecture.

TABLE II. COMPARISON BETWEEN PUBLIC AND PRIVATE DATASET FOR VARIOUS PREPROCESSING FILTERS.

SAMPLE IMAGES		MSE		PSNR		SMPI		SSI	
		Public	Private	Public	Private	Public	Private	Public	Private
IMAGE1	Database	Public	Private	Public	Private	Public	Private	Public	Private
	Median filter	2.2365	1.4751	44.0885	46.4424	1.0500	1.1071	0.9026	0.918389
	Average filter	6.3060	6.1334	40.1333	40.2537	1.1029	1.1487	0.9170	0.911269
	Gaussian filter	4.2167	3.9287	41.8811	42.1882	1.0739	1.1488	0.9211	0.912985
	Laplacian filter	4.6770	1.6379	44.0963	45.9877	12.5476	14.8878	3.6859	4.441867
IMAGE2	Wiener filter	4.8635	4.7959	41.2613	41.3220	1.1091	1.1500	0.9184	0.913179
	Median filter	4.9094	3.2976	42.9488	41.2204	1.1776	1.1966	0.9348	0.4194
	Average filter	8.4472	8.4202	38.8636	38.8775	1.1427	1.4307	0.9259	0.4314
	Gaussian filter	5.5205	6.4503	40.7109	40.0349	1.1028	1.3210	0.9278	0.4276
IMAGE3	Laplacian filter	8.8814	7.8254	45.3857	43.6200	14.1200	25.5903	3.3318	1.3177
	Wiener filter	6.09635	6.4541	40.2801	40.0324	1.1317	1.1903	0.9277	0.4180
	Median filter	3.8416	1.7622	42.2856	45.6700	1.2617	1.2527	0.9413	0.2224
	Average filter	7.3260	6.5975	39.4821	39.9369	1.1357	1.3919	0.9310	0.2400
	Gaussian filter	4.7440	5.0832	41.3693	41.0693	1.0988	1.2897	0.9329	0.2348
IMAGE4	Laplacian filter	1.8473	2.6990	45.4652	43.8186	14.1181	31.7776	3.5172	1.9137
	Wiener filter	4.9811	5.3954	41.1574	40.8105	1.1488	1.1699	0.9328	0.2220
	Median filter	5.3381	2.4859	40.8569	44.3147	1.3109	1.2783	0.9220	0.4357
	Average filter	11.4491	11.0555	37.5430	37.6950	1.5727	1.4666	0.9040	0.4442
	Gaussian filter	8.5870	8.7065	38.7923	38.7323	1.4102	1.3457	0.9069	0.4412
	Laplacian filter	6.1582	5.9833	40.2362	40.3613	25.2660	32.7460	4.7816	1.7324
	Wiener filter	7.4873	10.4674	39.3875	37.9323	1.1658	1.2860	0.9120	0.4330

TABLE III. MSE VALUES CALCULATED FOR 20 LUNG CT SCAN IMAGES

Images	Gaussian filter	Laplacian Filter	Wiener Filter	Average Filter	Median filter
Image 1	10.4107	8.0476	16.4134	14.3986	5.629514
Image 2	4.2594	1.6310	16.4134	6.7403	1.76783
Image 3	3.9366	1.6448	4.9831	6.1898	1.503216
Image 4	3.9287	1.6379	4.7959	6.1334	1.475155
Image 5	3.9466	1.6394	5.0215	6.2053	1.57069
Image 6	4.1693	1.7372	5.2580	6.61253	1.767757
Image 7	3.9898	1.6275	4.8373	6.3127	1.533886
Image 8	3.9846	1.8267	4.7717	6.2992	1.795907
Image 9	5.0760	2.1471	5.5763	7.7779	2.189621
Image 10	5.3464	2.6278	5.2915	8.1579	2.583865
Image 11	5.3401	2.9329	5.3445	8.1589	2.864231
Image 12	5.1214	1.6372	4.9772	7.7951	2.32745
Image 13	5.5205	1.8818	6.0963	8.4472	3.297607
Image 14	5.4622	3.9279	5.6064	8.3463	2.866573
Image 15	5.3648	2.6274	5.4720	8.1413	2.533028
Image 16	5.5526	3.6321	5.7118	8.4839	3.0891
Image 17	19.8258	16.4474	23.7791	25.8741	17.81281
Image 18	11.7531	8.8564	15.4171	16.4925	8.735854
Image 19	8.9496	8.9759	10.1341	11.7138	7.969679
Image 20	5.8520	1.5991	6.0918	9.0387	1.150322

TABLE IV. PSNR VALUES CALCULATED FOR 20 LUNG CT SCAN IMAGES

Images	Gaussian filter	Laplacian Filter	Wiener Filter	Average Filter	Median filter
Image 1	37.956	39.07411	35.9788	36.54757	40.62609
Image 2	41.83726	45.00623	35.9788	39.84398	45.6564
Image 3	42.17952	45.06944	41.15574	40.21404	46.36059
Image 4	42.18823	45.98775	41.32203	40.25379	46.44243
Image 5	42.16854	45.98379	41.12242	40.20313	46.1699
Image 6	41.93015	45.58972	40.92255	39.92713	45.65658
Image 7	42.12124	46.01551	41.28476	40.12861	46.27287
Image 8	42.12686	46.01758	41.34404	40.13793	46.38176
Image 9	41.07553	44.46576	40.66731	39.22213	44.72711
Image 10	40.85013	43.01457	40.895	39.01499	43.53105
Image 11	40.85525	42.00098	40.85167	39.01448	43.56072
Image 12	41.03691	43.9896	41.16091	39.21258	44.462
Image 13	40.71095	41.38505	40.2801	38.86363	42.94881
Image 14	40.75709	42.01447	40.64389	38.91586	43.55717
Image 15	40.83526	44.01561	40.74928	39.02386	44.0944
Image 16	40.68578	43.0033	40.56304	38.84483	43.23248
Image 17	35.15849	35.56983	34.36883	34.00214	35.62348
Image 18	37.42925	38.65822	36.25074	35.95794	38.71775
Image 19	38.61275	42.13636	38.07292	37.4438	41.1164
Image 20	40.4577	43.09201	40.28333	38.56971	43.14725

TABLE V. SMPI VALUES CALCULATED FOR 20 LUNG CT SCAN IMAGES

Images	Gaussian filter	Laplacian Filter	Wiener Filter	Average Filter	Median filter
Image 1	1.294603	13.92687	1.37016	1.298645	1.064346
Image 2	1.146684	13.11891	1.37016	1.11942	1.107688
Image 3	1.104776	14.79063	1.151386	1.146634	1.103744
Image 4	1.14883	14.8878	1.150077	1.148765	1.10719
Image 5	1.154547	14.83277	1.149414	1.145893	1.101582
Image 6	1.101705	14.91183	1.15418	1.141685	1.092064
Image 7	1.200961	14.76328	1.140694	1.140216	1.109842
Image 8	1.139585	14.85694	1.160133	1.152783	1.115066
Image 9	1.398108	13.4531	1.275212	1.136603	1.236237
Image 10	1.178796	14.46772	1.10578	1.109548	1.129562
Image 11	1.174096	14.43168	1.101911	1.105273	1.135625
Image 12	1.178117	14.30164	1.105351	1.109196	1.143828
Image 13	1.198777	45.38505	1.13172	1.142732	1.177622
Image 14	1.175833	14.64185	1.100584	1.107133	1.11512
Image 15	1.186019	14.7206	1.119324	1.119423	1.123658
Image 16	1.18307	14.83477	1.108753	1.115681	1.119091
Image 17	1.254297	14.92802	1.332866	1.356091	1.206737
Image 18	1.267971	14.35027	1.218229	1.233681	1.257511
Image 19	1.324968	33.53021	1.22989	1.435646	1.299309
Image 20	1.105202	15.26368	1.16768	1.146892	1.061226

TABLE VI. SSI VALUES CALCULATED FOR 20 LUNG CT SCAN IMAGES

Images	Gaussian filter	Laplacian Filter	Wiener Filter	Average Filter	Median filter
Image 1	0.878084	2.444955	0.880571	0.871408	0.895633
Image 2	0.910793	3.805185	0.880571	0.909212	0.916187
Image 3	0.913835	4.232766	0.913942	0.912092	0.919436
Image 4	0.912985	4.441867	0.913179	0.911269	0.918389
Image 5	0.9123	4.271089	0.912245	0.910604	0.917799
Image 6	0.912521	3.987196	0.912615	0.910786	0.918478
Image 7	0.913384	4.198172	0.913548	0.91166	0.918844
Image 8	0.913741	4.346538	0.913692	0.911982	0.919156
Image 9	0.925771	3.353325	0.926028	0.923839	0.934616
Image 10	0.924119	3.608169	0.924991	0.922358	0.92989
Image 11	0.922509	3.500805	0.923293	0.92071	0.928428
Image 12	0.921608	3.598895	0.922454	0.919775	0.927878
Image 13	0.927868	3.33207	0.927761	0.925973	0.934839
Image 14	0.922191	3.605417	0.922666	0.920354	0.928008
Image 15	0.922303	3.782351	0.922942	0.920437	0.928227
Image 16	0.921976	3.674299	0.922549	0.92019	0.927793
Image 17	0.926053	1.793156	0.927901	0.918894	0.941524
Image 18	0.959076	2.288282	0.962668	0.953761	0.97417
Image 19	0.471231	1.746785	0.462526	0.474583	0.46482
Image 20	0.903099	3.234979	0.903707	0.901092	0.911693

TABLE VII. PERFORMANCE METRICS (RECALL, ACCURACY, PRECISION, F-SCORE) OBTAINED THROUGH OTSU THRESHOLDING SEGMENTATION METHOD FOR 20 CT SCAN IMAGES OF THE LUNGS.

Images	Recall	Accuracy	Precision	F-Score
Image 1	0.6175	0.8146	0.99	0.7635
Image 2	0.4852	0.5461	0.98	0.6534
Image 3	0.5258	0.5699	0.99	0.6892
Image 4	0.5541	0.6046	0.94	0.7131
Image 5	0.5331	0.5666	0.89	0.6955
Image 6	0.5407	0.5742	0.99	0.7019
Image 7	0.5407	0.5742	0.97	0.7019
Image 8	0.5687	0.5687	0.99	0.6769
Image 9	0.4909	0.5360	0.99	0.6585
Image 10	0.5580	0.5938	0.99	0.7163
Image 11	0.6430	0.6430	0.99	0.7613
Image 12	0.6588	0.6785	0.99	0.7943
Image 13	0.4993	0.5424	0.95	0.6660
Image 14	0.6592	0.6795	0.99	0.7946
Image 15	0.5981	0.6295	0.99	0.7485
Image 16	0.6452	0.6671	0.92	0.7843
Image 17	0.4972	0.6720	0.99	0.6641
Image 18	0.7742	0.8628	0.91	0.8727
Image 19	0.4664	0.4664	0.99	0.6361
Image 20	0.4638	0.5310	0.96	0.6337

TABLE VIII. PERFORMANCE METRICS (RECALL, ACCURACY, PRECISION, F-SCORE) OBTAINED THROUGH K-MEANS CLUSTERING SEGMENTATION METHOD FOR 20 CT SCAN IMAGES OF THE LUNGS.

Images	Recall	Accuracy	Precision	F-Score
Image 1	0.99	0.8462	0.8462	0.6528
Image 2	0.99	0.8815	0.8815	0.9370
Image 3	0.99	0.9071	0.9071	0.9513
Image 4	0.99	0.8867	0.8867	0.9399
Image 5	0.99	0.9281	0.9281	0.9627
Image 6	0.99	0.9281	0.9281	0.9627
Image 7	0.99	0.9270	0.9270	0.9621
Image 8	0.99	0.8829	0.8829	0.9378
Image 9	0.99	0.9113	0.9113	0.9536
Image 10	0.99	0.9189	0.9189	0.9577
Image 11	0.99	0.9264	0.9264	0.9618
Image 12	0.99	0.9422	0.9422	0.9702
Image 13	0.99	0.9138	0.9138	0.9549
Image 14	0.99	0.9403	0.9403	0.9692
Image 15	0.99	0.9219	0.9219	0.9593
Image 16	0.99	0.9382	0.9382	0.9681
Image 17	0.99	0.6522	0.6522	0.7895
Image 18	0.99	0.6077	0.6077	0.7560
Image 19	0.99	0.9912	0.9912	0.991
Image 20	0.99	0.8746	0.8746	0.9331

TABLE IX. PERFORMANCE METRICS (RECALL, ACCURACY, PRECISION, F-SCORE) OBTAINED THROUGH SEGMENTATION METHOD FOR 20 CT SCAN IMAGES OF THE LUNGS.

Images	Recall	Accuracy	Precision	F-Score
Image 1	0.4732	0.4680	0.8634	0.6114
Image 2	0.9473	0.8378	0.8782	0.9114
Image 3	0.8825	0.8126	0.9083	0.8952
Image 4	0.9627	0.8546	0.8837	0.9215
Image 5	0.9548	0.8875	0.9262	0.9403
Image 6	0.9549	0.9548	0.9020	0.9277
Image 7	0.8985	0.8397	0.9264	0.9122
Image 8	0.8858	0.8012	0.8886	0.8872
Image 9	0.9252	0.8499	0.9114	0.9183
Image 10	0.9227	0.8555	0.9202	0.9215
Image 11	0.9239	0.8594	0.9243	0.9241
Image 12	0.9355	0.8849	0.9419	0.9387
Image 13	0.9248	0.8481	0.9104	0.9175
Image 14	0.9227	0.8719	0.9399	0.9312
Image 15	0.9224	0.8583	0.9237	0.9231
Image 16	0.9171	0.8670	0.9397	0.9283
Image 17	0.8625	0.6013	0.6454	0.7383
Image 18	0.8747	0.5678	0.5989	0.7110
Image 19	0.9326	0.9326	0.9911	0.9651
Image 20	0.9095	0.8032	0.8711	0.8899

TABLE X. AVERAGE PERFORMANCE OF IMAGE SEGMENTATION TECHNIQUES

Performance Parameters	Otsu thresholding Method	Watershed Method	k-means Method
Accuracy	0.61	0.81	0.88
Precision	0.96	0.88	0.88
Recall	0.56	0.89	0.99
F-Score	0.71	0.88	0.92

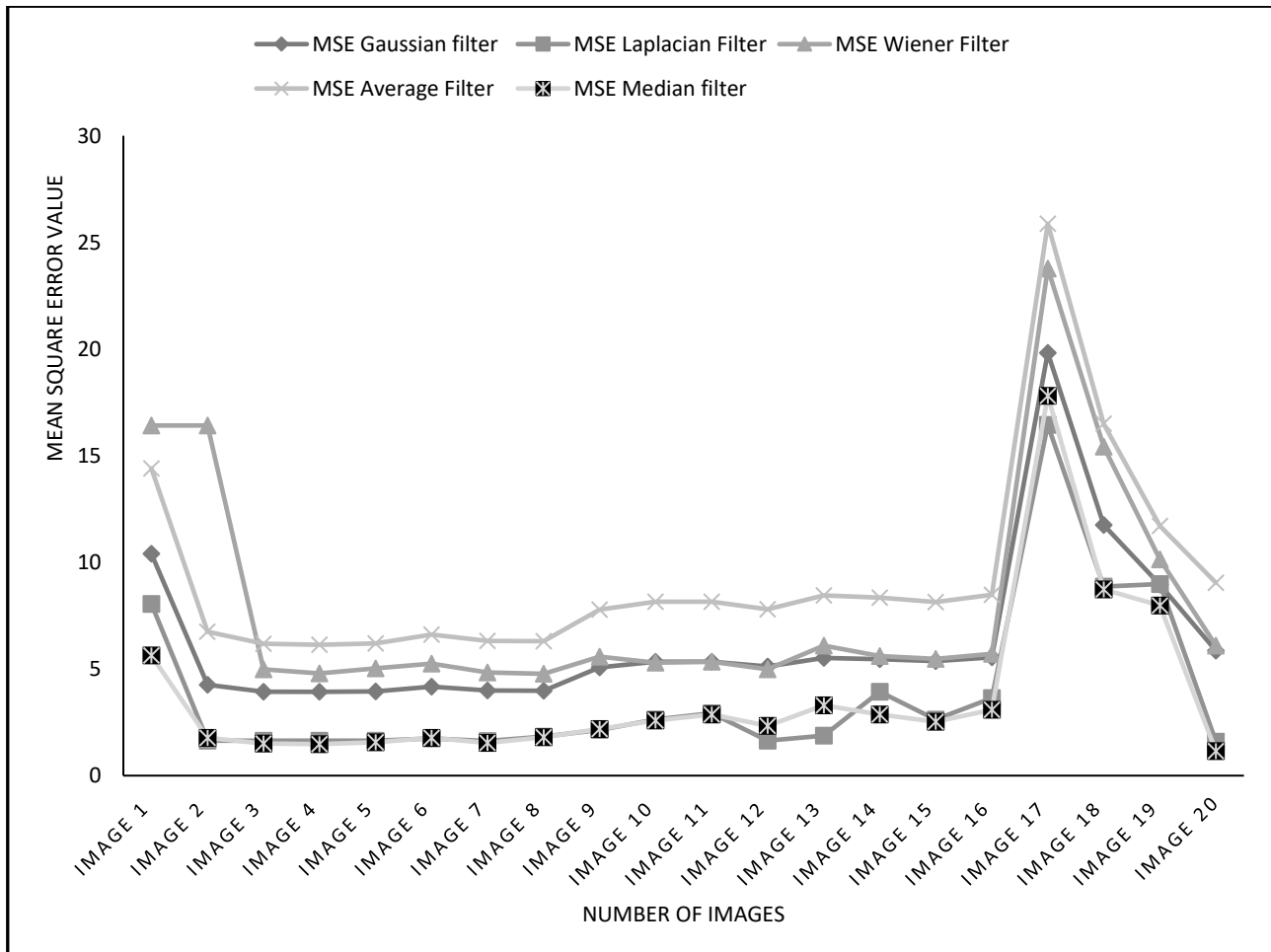


Fig 3 Comparison results of Mean Square Error (MSE) value for different filters

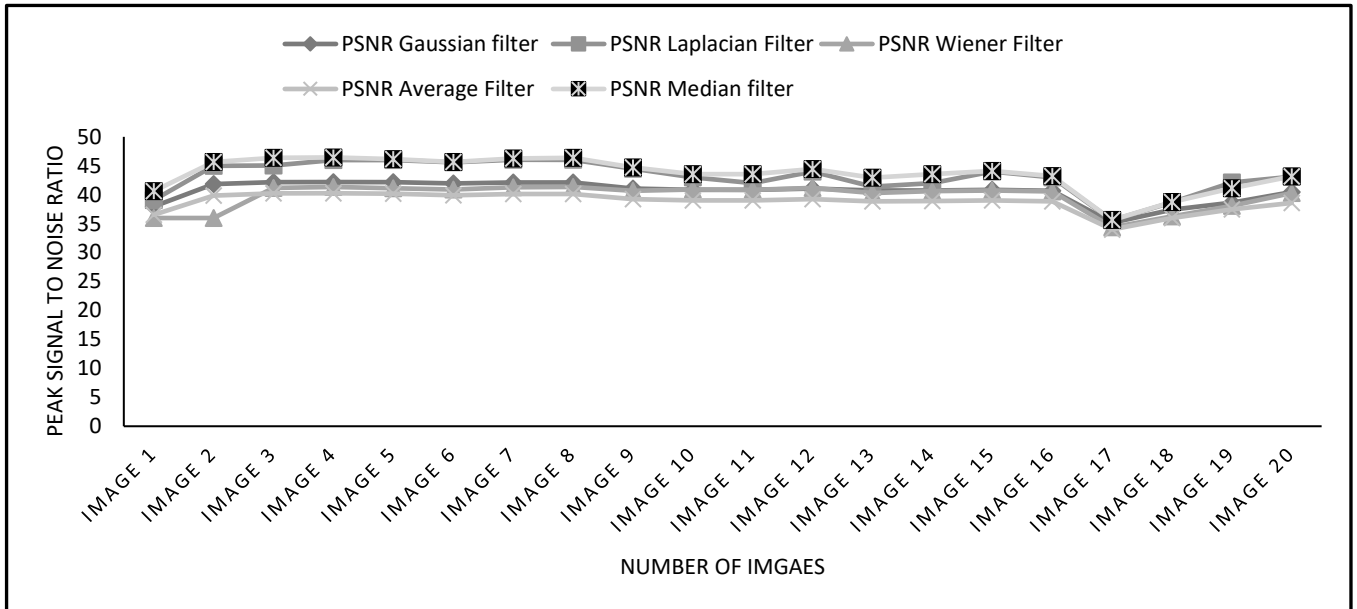


Fig 4 Comparison results of Peak Signal to Noise Ratio (PSNR) value for different filters.

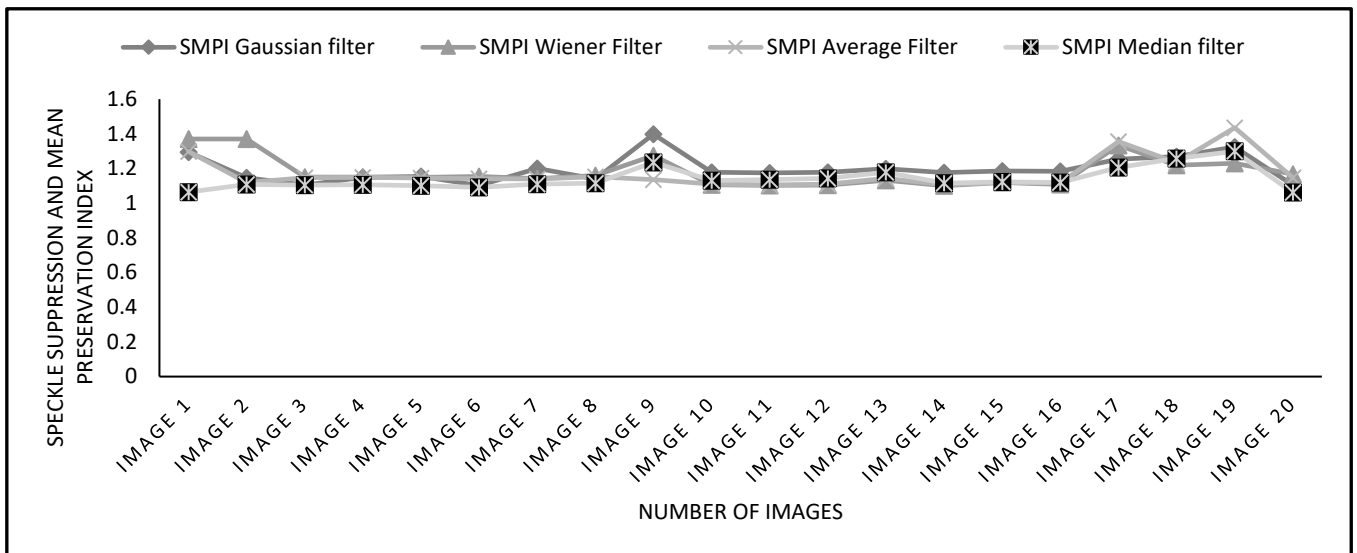


Fig 5 Comparison results of Speckle Suppression and Mean Preservation Index (SMPI) value for different filters.

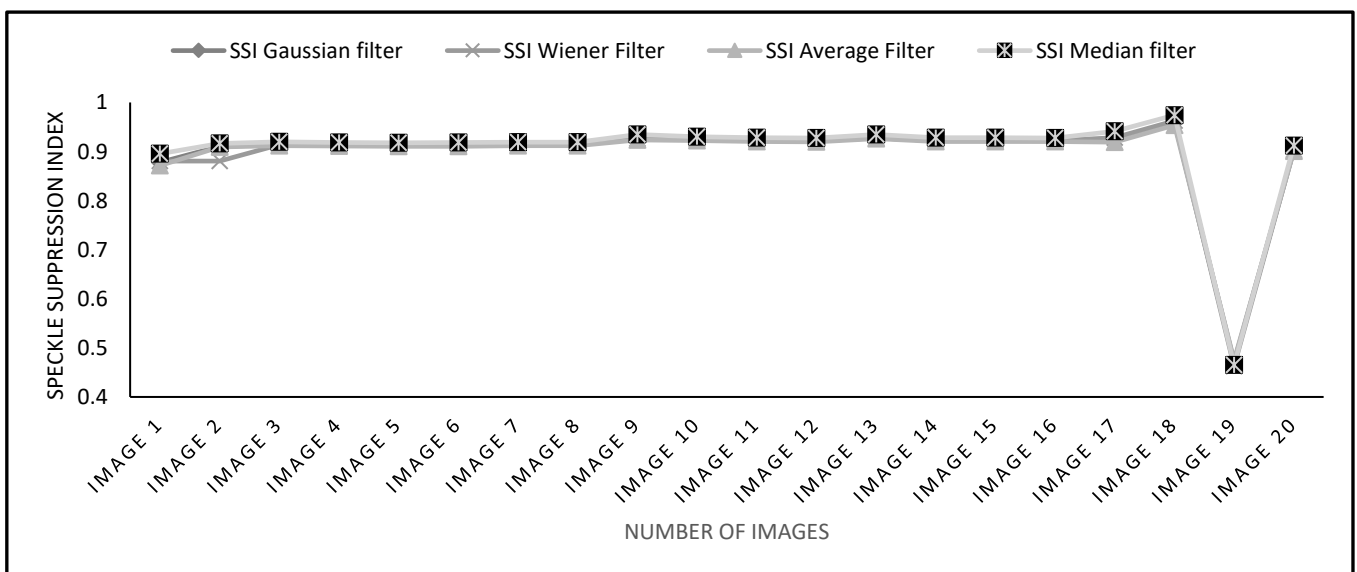


Fig 6 Comparison results of Speckle Suppression Index (SSI) value for different filters.

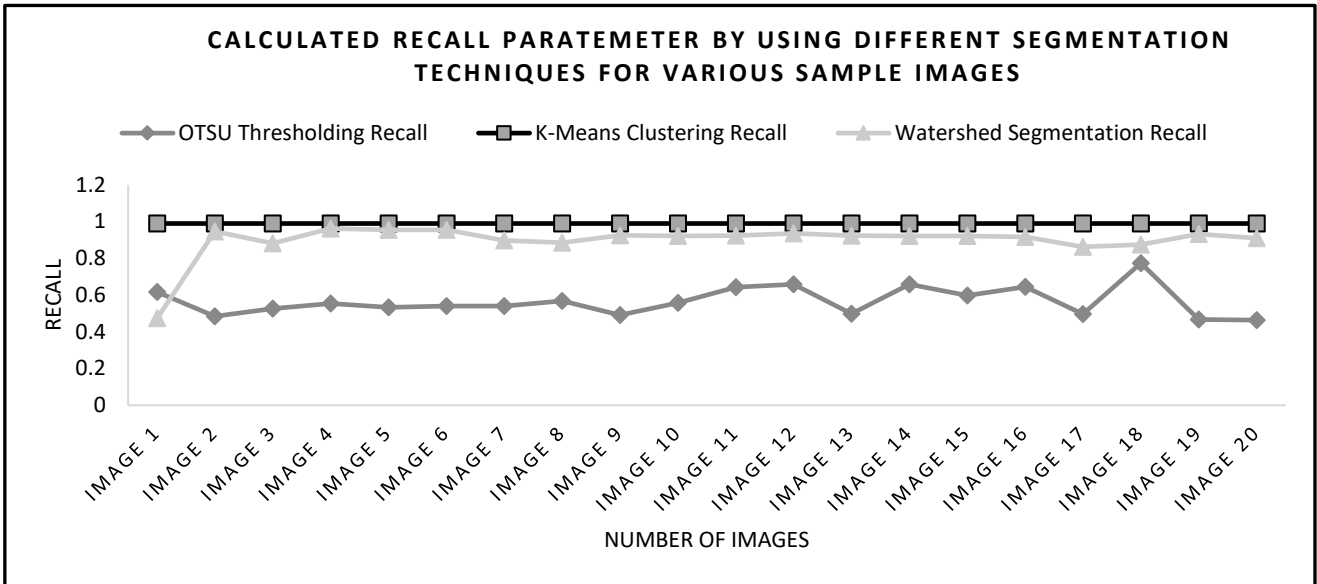


Fig 7 Recall parameter calculated by using different segmentation techniques.

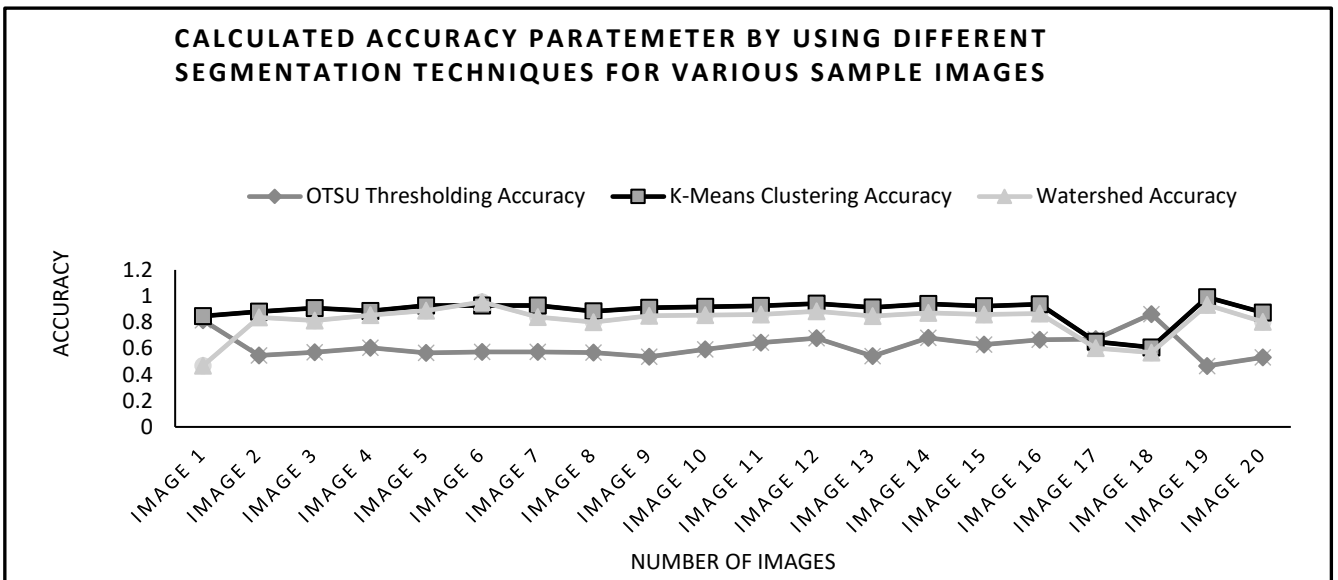


Fig 8 Accuracy parameter calculated by using different segmentation techniques.

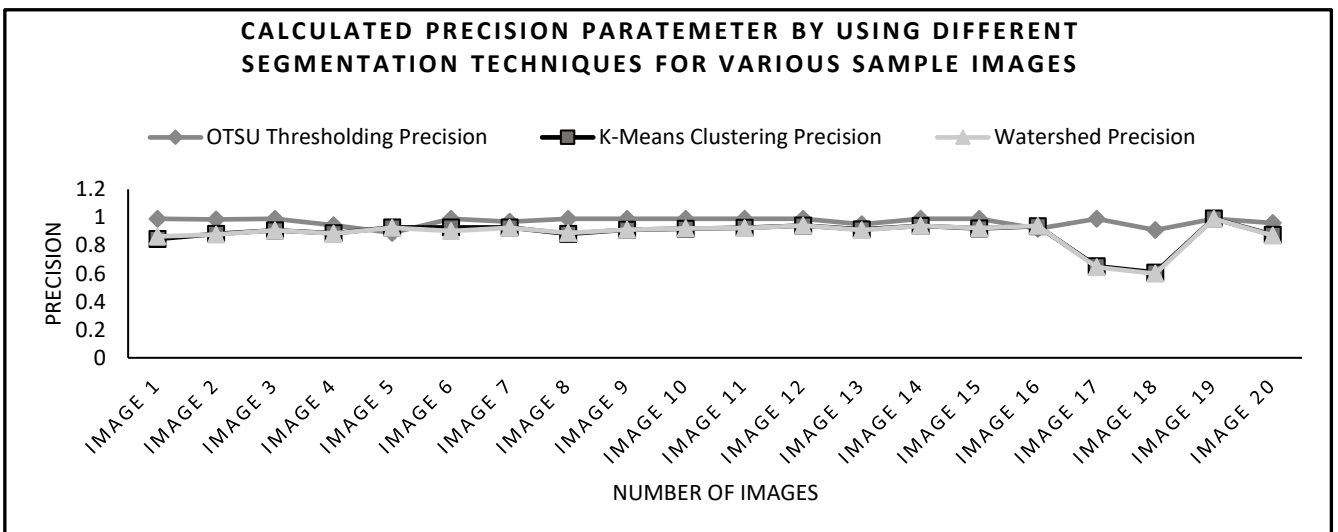


Fig 9 Precision parameter calculated by using different segmentation techniques.

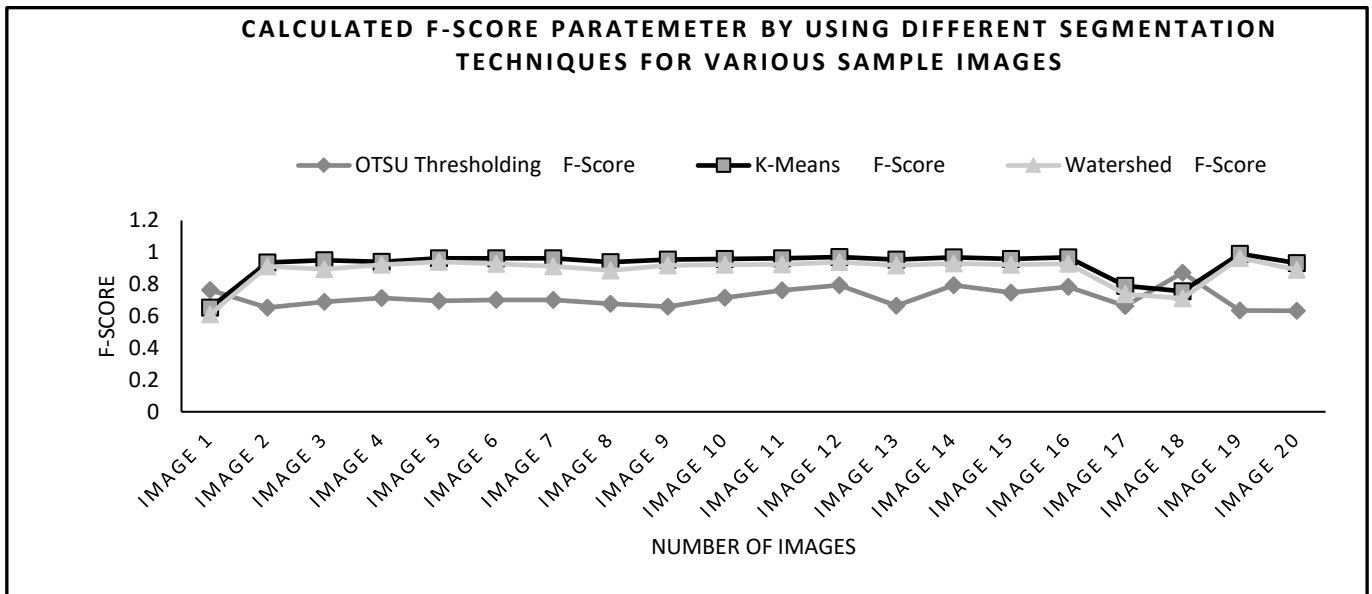


Fig 10 F-Score parameter calculated by using different segmentation techniques.

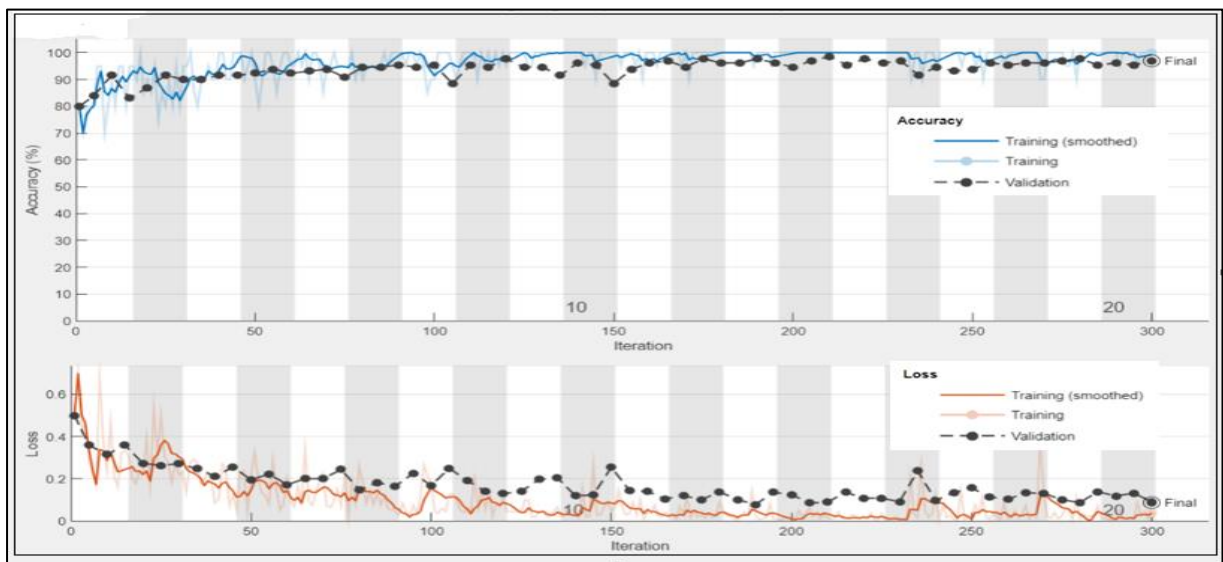


Fig 11 Accuracy and loss plot during training and validation of the network

		Split 1		Split 2	
Predicted Labels	Cancerous	71	0	71	0
	Non-Cancerous	4	55	4	55
		Cancerous	Non-Cancerous	Cancerous	Non-Cancerous
		True Labels		True Labels	

		Split 3		Split 4	
Predicted Labels	Cancerous	70	0	70	0
	Non-Cancerous	4	56	4	56
		Cancerous	Non-Cancerous	Cancerous	Non-Cancerous
		True Labels		True Labels	

Fig 12 Confusion Matrix for different splits of dataset for classification.

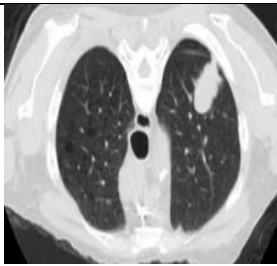
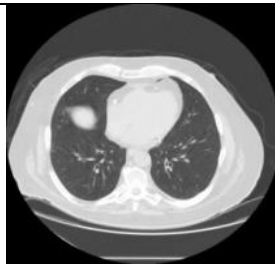
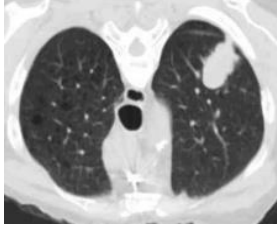

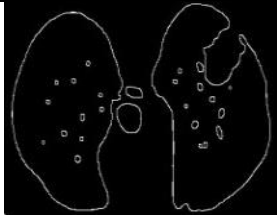



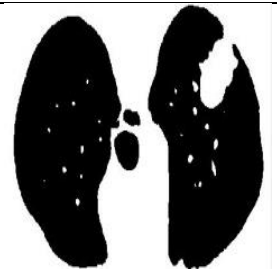

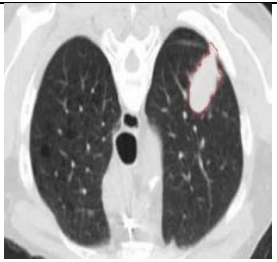
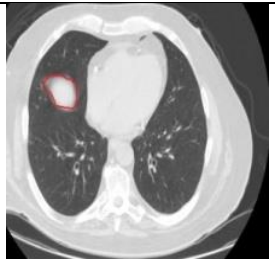


a)	Original Image		
b)	Median filtered Image		
c)	Edge detection Image by Sobel Filter		
d)	Segmented Image after applying K-Means Clustering algorithm.		
e)	Lung field extracted		
f)	Manually Marked Nodule		
g)	Nodule extraction		
		(i) Image with ragged shaped nodule	(ii) Image with oval shaped nodule

Fig 13 (a)Original lung CT scan image, (b)Resultant image after applying Median filter on (a),(c)Edge detection by applying Sobel filter on (b),(d)Segmented image after applying K-Means Clustering segmentation on (b), (e)Right and left lung extraction, (f) Lung nodule marked by radiologist manually and (g) Nodule extracted automatically with the help of algorithm discussed in this paper.

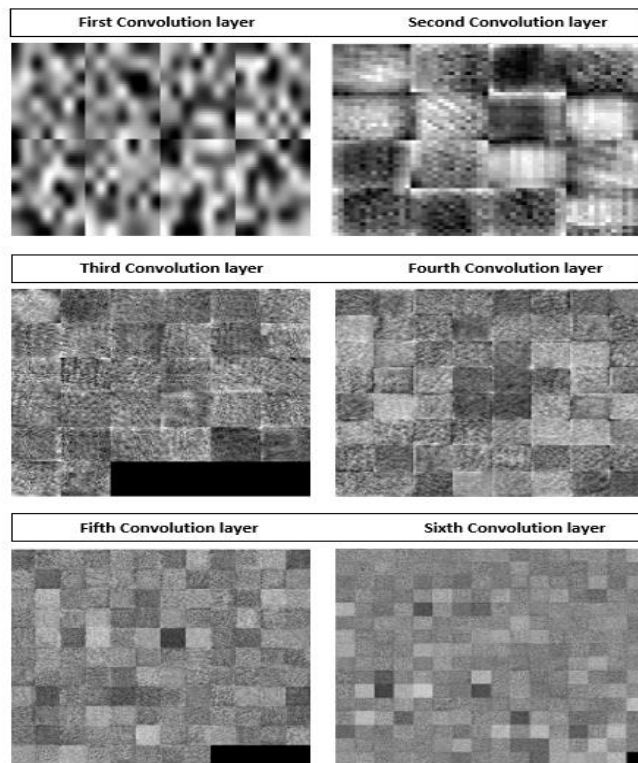
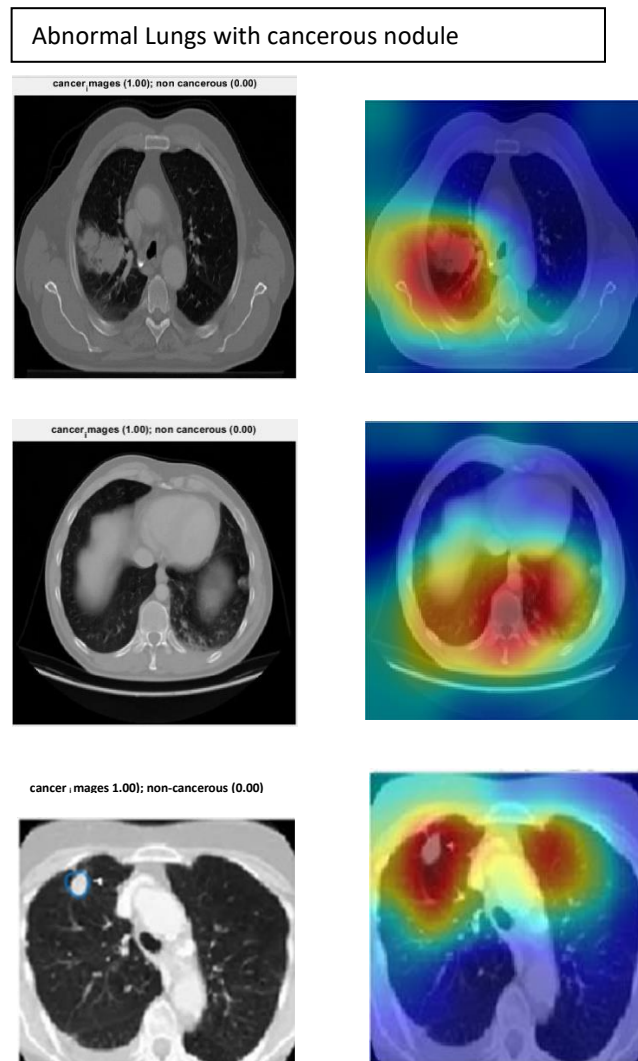


Fig 14 Visual representations of features extracted at different convolutional layers of MSNN.



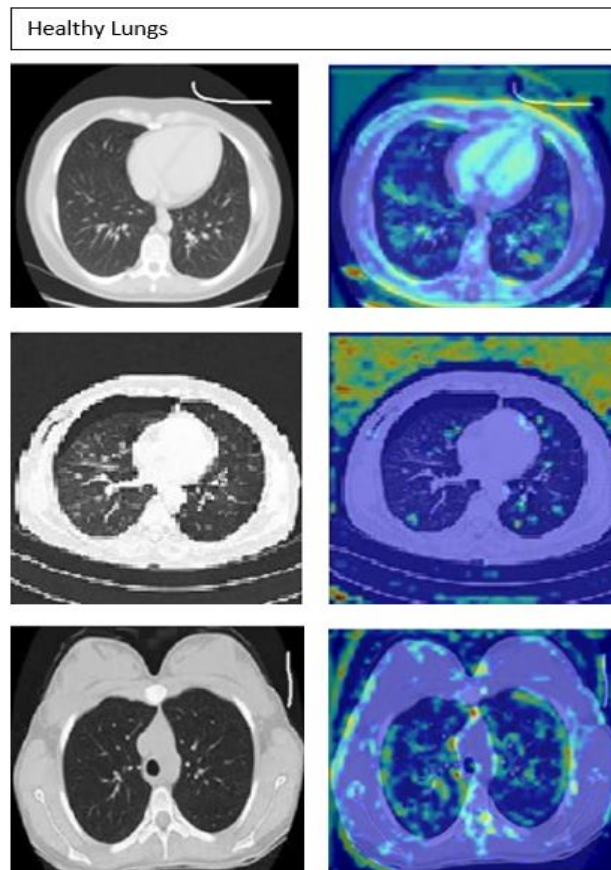


Fig 15 MSNN model classification by using maximum sensitivity maps for (a)Abnormal lungs with cancerous nodules and (b)Healthy lungs.