



## Diagnosis of Lung Cancer Disease Based on Back-Propagation Artificial Neural Network Algorithm

Hanan A. R. Akkar<sup>a</sup>, Suhad Qasim G. Haddad<sup>b\*</sup>

<sup>a</sup> Prof., Dr., Department of Electrical Engineering, University of Technology, Baghdad, Iraq, hanan.akkarak@uotechnology.edu.iq

<sup>b</sup> Asst. Lecturer, Department of Computer Engineering, University of Technology, Baghdad, Iraq, 120007@uotechnology.edu.iq

\*Corresponding author.

Submitted: 02/04/2020

Accepted: 13/06/2020

Published: 25/12/2020

### KEY WORDS

Back Propagation Artificial Neural Networks (BP-ANN), Computer tomography CT, Gray-Level Co-Occurrence Matrix (GLCM), Histogram equalization, Image pre-processing, Morphological operation.

### ABSTRACT

Early stage detection of lung cancer is important for successful controlling of the diseases, also to offer additional chance to the patients in order to survive. So, algorithms that are related with computer vision and Image processing are extremely important for early medical diagnosis of lung cancer. In current work (CT) computed tomography scan images were collected from several patients Classification was done using Back Propagation Artificial Neural Network (BP – ANN). It is considered as a powerful artificially intelligent technique with training rule for optimization to update the weights of the overall connections in order to determine the abnormal image. Several pre-processing operations and morphologic techniques were introduced to improve the condition of the image and make it suitable for detection cancer. Histogram and (GLCM) Gray Level Co-occurrence Matrix were applied to get best features extraction analysis from lung image. Three types of activation functions (trainlm, trainbr, traingd) were used which gives a significant accuracy for detecting cancer in CT scan lung image related to the suggested algorithm. Best results were obtained with accuracy rate 95.9% in trainlm activation function. Graphic User Interface (GUI) was displaying to show the final diagnosis for lung.

**How to cite this article:** “Diagnosis of Lung Cancer Disease Based on Back-Propagation Artificial Neural Network Algorithm” Engineering and Technology Journal, Vol. 38, Part B, No. 03, pp. 184-196, 2020.

DOI: <https://doi.org/10.30684/etj.v38i3B.1666>

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0>).

## 1. INTRODUCTION

Lung cancer is defined as an uncontrolled cell growing in the tissues of lung, which can be identified as tumor in lung. Lung cancer disease is considered as one of the dangerous diseases in the world which is slowly increasing, it is strongly connected with cigarette smoking and air pollution.

People in countries especially those which have high degree of air pollution suffered from various cancer tumors such as lung cancer, breast Cancer, etc. Various tools were offered for the purpose of obtaining medical images of lung such as Chest Radiography ( X – ray), Computer Tomography (CT), Magnetic Resonance Imaging (MRI scan) or even Sputum Cytology where special doctors can recognize cancer by analyzing these images. Most of these methods are time consuming, physically expensive, located so far from the patient, in addition, most of these systems have the ability to discover cancer in its progressive stage, in that the patient person will have a small chance to live [1]. Various soft computing algorithms were utilized in previous years by researchers to determine and enhance the diagnosis degree of cancer cells in a medical image in order to present a simple design, less budget, time saving in addition to giving hopeful result for patients in rural places [2]. Objective of this work is to support doctors to offer best possible treatment by providing useful visions with the help of predictive analytic technique which can scan medical data in a shorter time and more precisely way to diagnosis lung cancer. Classification of lung cancer using BP – ANN was suggested in this work. BP neural network is a powerful training technique for determining best values of weights that can reduce the error in classify the target condition. BP network can be trained by various activation training function, each of these training function has its own characteristic, such as, speed of learning rate, accuracy of training and testing, iterations for reaching greatest result, and finally updating of the learning rate. Three kind of activation functions were used in this work, which are traingd, trainbr, and trainlm. Best result was obtained from trainlm with accuracy 95.9 %. Several pre-processing technique was utilized to create a change in an image nature for the purpose of improving its representative information for next steps, or even rendering it to be more appropriate for diagnosis technique [3]. In order to give more accuracy and reliability to the work the network was training with real images of CT scanned lung images which were obtained from Baghdad educational hospital / Medical city/ Radiology Institute, then with the assist of (<https://www.radiantviewer.com>) which is specialized site for viewing medical images, we could save the images in order to prepare them for training and testing phases. The work is divided into six sections. Section two some related work will be introducing. Section three is about the methodology of the work which will be discussed and illustrated, contains image pre-processing steps such as: image enhancement, smoothing, segmentation, feature extraction. Section four deals with the classification using BP – ANN and illustrated the performance of activation function that were used. Section five will discuss the result of the work, besides that, comparison was done between the suggested technique and previous works Finally, section six gives the conclusion.

## 2. RELATED WORKS

In the last decade several techniques have been suggested on the detection and diagnostic of cancer in lung images with various classification algorithms. The following gives a brief description of previous works. [4] Lung disease was diagnosed by employing image processing methods to offer good tool for improving the analysis. Enhancement and improvement the quality of images were done by using histogram equalization technique. Morphology method and watershed process for segmentation were presented to improve the accuracy of features. Features such as area, perimeter, roundness, eccentricity were extracted, finally classification was completed.[5] Supposes a technique for combination statistical structure and constant moment features to classify lung cancer in the images. All the features have been separated from X – ray images. First level segmentation of images is done in order to extract the important area outside of the ribs. In the second level, statistical features and constant moment were removed based on the contour properties of the region. Afterward, these properties have been used as an input shape to the Fuzzy Hyper Sphere Neural Network (FHSNN) classifier. [6] Introduced two methods for feature extraction first, gray level co-occurrence matrix where four features were calculated which are, contrast, energy, correlation and homogeneity. Second part of feature extraction was dual tree complex wavelet transform. back propagation neural network was used to classify lung image as normal or abnormal. Fuzzy algorithm used to detect the region that was abnormal and to segment the lesion area in lung. [7] Improvements of CT image were made by using contrast steady adaptable histogram equalization. Furthermore, segmentation with watershed marker procedure have been used. Five stages of wavelet conversion were utilized to extract lung image pattern without noise in order to improve the feature of the matching data in the images and analyzing them as small as possible. Finally, classifications have been done by bee colony optimization technique (BCO).[8] Image enhancement was done with Gabor filter. Adaptive threshold was used for segmentation the image of lung. Area, perimeter and

Eccentricity are considered as the three parameters that were calculate from gray level co-occurrence matrix that used in feature extraction part. Finally, back propagation neural network and (*SVM*) support vector machine were used for classification. [9] Multi scale three-dimension convolution neural network (*CNN*) for diagnostic and classification of lung image were suggested. Three convolution layers with a series of [32,64,128] feature maps were used along with two overall pooling layers of (2x2x2) filters supported by three dimensional convolution filters of scale 3 varying from 3 to 3 range. Padding had been applied to achieve the suitable size of the feature map. Besides, batch normalization was done between layers.

### 3. PROPOSED METHODOLOGY

The methodology of diagnosing lung cancer was done in six phases as shown in Figure 1. All the operation was done using MATLAB 2019b (9.7) which is derived from Matrix Laboratory. MatLab programming language is assumed to be extremely straightforward since almost every data object is considered as a matrix. It's a cooperative matrix-based system for scientific and engineering numeric computation and conception. Powerful operations can be performed using just one or two commands [10] . Steps of methodology will be discussed briefly in the next section.

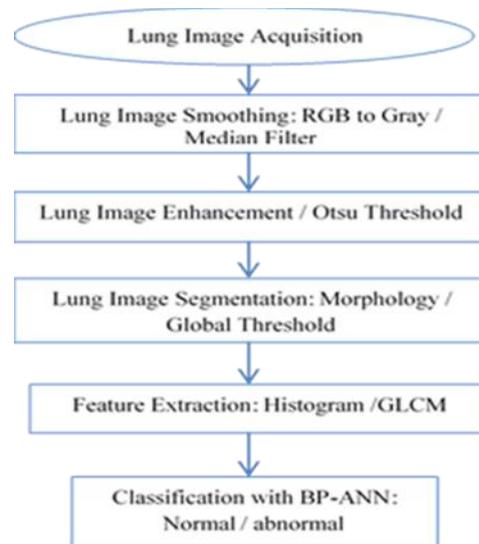


Figure 1: Methodology of Lung Cancer Diagnosis for the Suggested Technique

#### I. Image Capture

First phase is to obtain CT scan image of lung cancer. Generally, an image can be defined as a picture that makes a representation of something valuable. It can be a picture of any person, an outdoor scene, an electronic element microphotograph, or any medical image. The main advantage of using CTcomputed tomography images in this work is because it's simple and provides less distortion and low noise when compared with X-ray and MRI images [11]. Hence, they are considered in this work. In order to give more accuracy, durability and reliability to the proposed work, real dataset were used and collected from Baghdad educational hospital /medical city / Radiology department. Then with the use of (<https://www.radiantviewer.com>) which is a specialized site for viewing medical images, we can simply open the image file and save it .One hundred samples of images for several patients were collected. 60 were mentioned as abnormal images with cancer diseases and 40 as normal images. Each image is in size 512x512 pixels, with final size 255x255 pixels.

#### II. Image Pre-Processing

Second phase is the image pre-processing. The goal of this phase is to smooth, improve and enhance the observation of data in images to offer suitable input for other phases in image processing. Generally, the quality and features of the image is impacted by different parameters such as, non-uniform inputs, change in signals, different noises, besides that all kinds of electronic apparatuses might have an effect on data passing through them which will influence final output

image. Image preprocessing describes any type of processing performed to the image to prepare or convert it into an appropriate format that will be more easily and effectively recognized. The preprocessing of image aims to selectively eliminate the redundancy existing in scanned images without disturbing the details which play an important role in the diagnosis procedure [12]. Two methods of processing were used in this step which are smoothing and enhancement.

### 1) *Smoothing of Image*

Two kinds of methods were utilized in this phase: first method is to convert the color image RGB to gray scale image because gray image reduces the processing time and can produce fast algorithms. Red green and blue are the three colors that can define the situation of any colored image, in MatLab Syntax [rgb2gray] function can be used to achieve the conversion from RGB to gray image. The function transforms color images to gray scale by the means of removing capacity information while keeping the luminance by creating a weighted amount of the R, G, and B categories, Eq. (1), was used to get the transformation process where the first quantity will multiply by (R), the second multiply by (G), the third by (B), and (I) is the final image that we got, as shown below:

$$I = 0.2989R + 0.5870G + 0.114B \quad (1)$$

Second method was done using median filter. Noise is a random error in images so removing the noise from images will be an important factor. Image filtering can be classified as linear or nonlinear. Median filter is a nonlinear process used to eliminate different noise such as Gaussian white noise from images, it is very effective method for eliminating noise while protecting edges. It is chiefly effective at removing impulse noise which is called salt and pepper, salt = 255, pepper = 0 gray levels. Median filters searching through the image pixel by pixel, substituting each value with the median value of adjacent pixels [13].

### 2) *Image Enhancement*

To eliminate noise from CT lung image, Otsu's process was utilized. It enables us to decrease the grey level into a binary image. The procedure supposes that there are two categories of pixels present in image includes foreground pixels and background pixels then it calculates optimum threshold value which divides the two categories. suppose that threshold is (  $t$  ), and each section are ( $W_f$ ) and ( $W_b$ ) for foreground weight and background weight respectively, means are ( $\mu_f$ ) and ( $\mu_b$ ). Then the variance between the two segments is defined in, Eq. (2), and probability will define in, Eq. (3) as follows:

$$\sigma_t = W_f W_b (\mu_f - \mu_b)^2 \quad (2)$$

$$P_i = \frac{n_i}{N} \quad (3)$$

Where ( $n_i$ ) is considered as the pixels of an image with gray value ( $i$ ), and ( $N$ ) is considered as the overall number of pixels in an image, and ( $P_i$ ) will be the probability of a pixel which have ( $i$ ) of gray value. So the background weights will be computed by collective amount of all the  $P_i$  values [10,14]. Finally, the image will be in binary value between 0 and 1 only.

## III. *Segmentation*

Segmentation is used primarily to separate a digital image into several sets of pixels and to locate the boundaries of features. The key purpose of segmentation is to turn the representation of an image into a clear and functional segment that is easy to manage and interpret the image data, resulting in a series of regions that entirely cover the main image [11]. Two steps of segmentation were utilized as follows:

### 1) *Morphological image operation*

It can be characterized as collecting non-linear procedures that will relate the shape or morphology of image features. Morphological techniques can examine an image with a specific area called the basic unit that can cover all possible positions of the image and equate it to the corresponding pixel neighborhoods [15]. Some procedures evaluate whether the element suits inside the neighborhood, while others examine if it reaches or crosses the neighborhood. Erosion removes minor information from a binary image, but shrinks the scale of regions of interest immediately, so the differences between various characteristics become greater. The erosion of a binary image ( $Z$ ) by a structuring element ( $S$ ) can be represented as new binary image ( $G$ ), as shown in, Eq. (4):

$$G = Z \ominus S \quad (4)$$

Dilation results in the reverse reaction from erosion. It adds a level of pixels for internal and external margins in areas, that will grow or thickens the object. The dilation of an image ( $Z$ ) by a structuring element ( $S$ ) can be represented as a different binary image ( $G$ ), as shown in, Eq. (5):

$$G = Z \oplus S \quad (5)$$

So, the dilation will expand the components of an image while the erosion shrinks them. Opening which is a combination of erosion and dilation will smooth the contour of an object. Opening is so called because it can open up a gap between objects connected by a thin bridge of pixels. Therefore, the opening ( $f$ ) by ( $S$ ) is the erosion of ( $f$ ) by ( $S$ ), followed by a dilation of the result by ( $S$ ) [15], as shown in, Eq. (6):

$$(f \circ S) = (f \ominus S) \oplus S \quad (6)$$

## 2) Global threshold

Most important uses of thresholding are to isolate features from their background to measure the size of the features. Therefore, this operation depends on choosing a suitable threshold level. In case of low threshold, this will decrease the size of some important features in the image. Otherwise if we choose a higher value, this will result in including inessential background issues. Global threshold was used in order to segment the tumor of cancer from the whole image of lung. Many experiments were done to choose the suitable quantity for threshold, the best results are based on trial and error, starting from (800) until (1500) which give the best and optimum result for isolation of the cancerous tumor. Thresholding was shown in, Eq. (7), where the intensity of the tumor is higher than the background of the existing part of lung [16].

$$I_{(x,y)} = \begin{cases} 1 & \text{if } I(x,y) \geq \text{Th} \\ 0 & \text{if } I(x,y) < \text{Th} \end{cases} \quad (7)$$

## IV. Feature Extraction

Feature extraction procedure is a crucial step for classification, it can extract the important fundamental features from the segmented image. Different methods and algorithms were used for feature extraction in this work. The extracted region of interested can be separate as either normal or not depending on their essential properties [17]. Two kind of feature extractions were used in this work as explained below:

### 1) Histogram Equalization

Mean histogram equalization was used, it is the average score on a given pixels of image, which will be calculated as follows:

$$(\mu_m = \sum X_i / N) \quad (8)$$

Where the symbol  $\mu_m$  represents the mean,  $i$  represent the quantity of pixels in image, and  $N$  is overall number of pixels in image [18].

### 2) Gray-Level Co-occurrence Matrix (GLCM)

Another important arithmetic method used for feature extraction which takes into consideration the spatial relation of pixels is the gray scale co-occurrence matrix. The gray comatrix function was

used in MatLab, which forms a gray measure co-occurrence matrix by the way of calculating the concurrence of a pixel with gray level value (  $i$  ) that it takes place in an identified spatial relation to a pixel with the value (  $j$  ). Every single component (  $i, j$  ) in the subsequent GLCM can be considered as the sum of how often the pixel with value (  $i$  ) may occur in a specific spatial relation to a pixel with value (  $j$  ) in an input image. In this work, GLCM function will make a calculation of the image homogeneity with various guidelines which cover (00 ,450,900 ,1350) [18]. Finally, the total numbers of features from each image were collected and will be the input neurons for the proposed network.

#### 4. CLASSIFICATION

Back Propagation algorithm is the most common method of supervised learning in ANN. It is commonly used for solving many real world problems by using the principle of multilayer perceptron (MLP). In general, BP is an algorithm for gradient descent where the network is pushed along the negative gradient of the function output. There are several factors to be considered, such as the number of neurons inputs, hidden layers and neurons in output, learning rate, momentum rate and feature activation. These elements can influence how BP – ANN learning converges. BP learning consists of two loops, a forward and a backward pass, through the various layers of the network. On the first pass the network's synaptic weights are all set. The synaptic weights are all modified according to an error-correction law during the second pass. The learning rate and momentum are changed during the training process to carry the network out of its local minima and to promote network convergence [8, 19] . The following steps shown the learning part:

- 1) The input layer contains an input vector.
  - 2) A set of chosen output is offered at the output layer.
  - 3) After a forward part is done, the errors between the chosen and real output are matched.
  - 4) The comparison results are used to regulate weight changes according to the learning rules.
- In this work tan-sigmoid transfer function will be used. Besides that, several activation functions will have already utilized in order to get the best results and ideal training performance. The structure of the proposed BP – ANN will have five neurons, two hidden layers, and two outputs targets. Algorithm of BP – ANN are shown below:

Step 1: Initialize network weight values. The Learning weight = 0.02 and the Momentum term = 0.95.

Step 2: Summarize the weighted input and use activation function to measure hidden layer output

$$h_i = f \sum_i X_i W_{ij} \quad (9)$$

Where:  $h_i$  : the real output of hidden neuron (  $j$  ) for input node (  $i$  ).  $X_i$  input node of input neuron (  $i$  ).  $W_{ij}$  :synaptic weight between input neuron (  $i$  ) and hidden neuron (  $j$  ).  $f$ : the activation function.

Step 3: Sum weighted output of hidden layer and apply activation function to compute output of output layer.

$$O_k = f \left[ \sum_j h_j W_{jk} \right] \quad (10)$$

Where:  $O_k$  : the real output of output neuron  $k$  .  $W_{jk}$  Synaptic weight between hidden neuron ( ) and output neuron  $k$ .

Step 4: Calculate the back propagation error (  $\delta_k$  ), as follows:

$$\delta_k = (d_k - O_k) \dot{f} \left[ \sum_j h_j W_{jk} \right] \delta_k \quad (11)$$

Where:  $\dot{f}$  = the derivative of the activation function.  $d_k$ = the desired of output neuron  $k$ .

Then, the error will have propagated backward to update the weights and also the biases of the network.

Step 5: Calculate weight correction term and finally update the weights.

$$W_{ik}(n+1) = W_{ik}(n) + \Delta W_{ik}(n) \quad (12)$$

$$W_{ij}(n+1) = W_{ij}(n) + \Delta W_{ij}(n) \quad (13)$$

Where:  $W_{ij}(n + 1)$  is the weight of the next terms for input neuron ( $i$ ) with hidden neuron ( $j$ ),  $W_{ik}(n + 1)$  is the weight of the next terms for hidden neuron ( $j$ ) with output neuron ( $k$ ) in sequence.  $W_{ik}(n)$  and  $W_{ij}(n)$  are the weights for the current terms,  $\Delta W_{ij}(n)$  and  $\Delta W_{ik}(n)$  are the weights for the correction terms. Three types of activation functions were used which are traingd, trainbr, and trainlm. Descriptions for each one are shown in Table 1. Best result was obtained from trainlm with accuracy rate 95.9 % with 9 iterations and one second for training. Figure 2 shows details simulation of training in BP – ANN with trainlm function. Figures 3 - 5 shows the best training performance, mean square error training, validation checks and epoch for trainlm activation function.

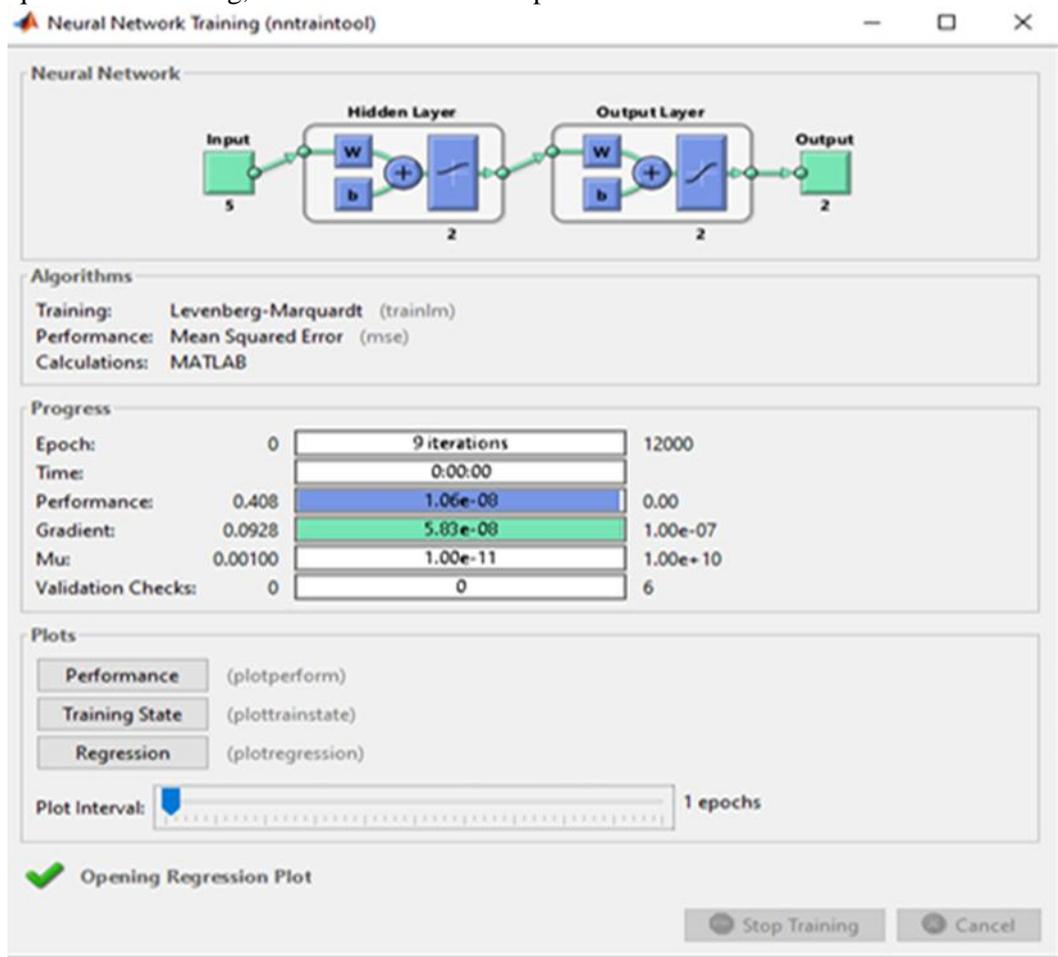


Figure 2: Neural Network Simulation for Training Phase Algorithm with Trainlm Activation Function

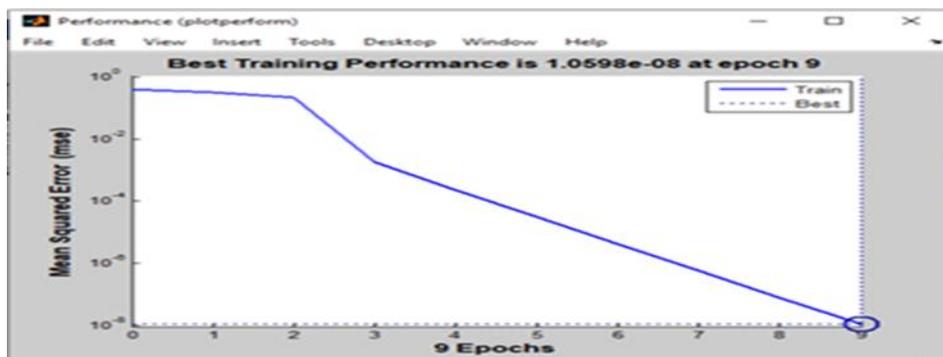


Figure 3: Mean Square Error Training for trainlm Activation Function



Figure 4: Neural Network Regression for trainlm Activation Function

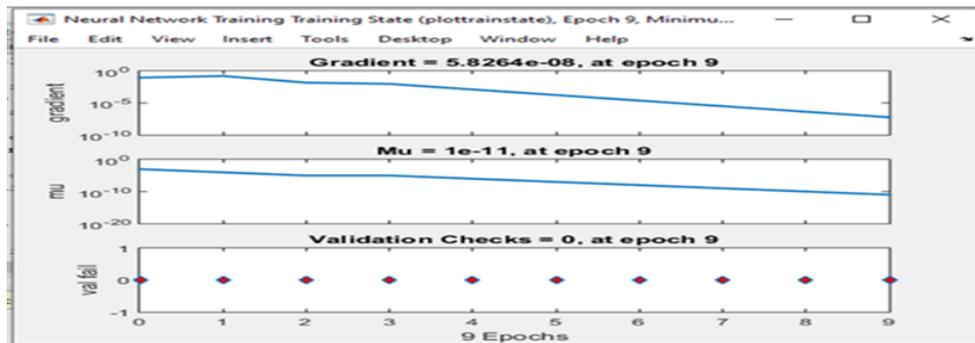


Figure 5: Validation Checks and Epoch for trainlm Activation Function

Table1: Definition and Performance of Activation Functions that used in the proposed technique

Activation Function	Description	Epochs	Performance	Time(Sec)
<i>trainlm</i>	Algorithm Levenberg-Marquardt. Fastest training algorithm for moderately scaled networks. Has memory reduction mechanism for use with large training collection.	9 Iteration	081.0598e-08	1 Sec
<i>trainbr</i>	Regularization in Bayesian culture. Modifying the Levenberg-Marquardt training algorithm to generate well distributed networks decreases the complexity of deciding the optimal layout of the network.	16 Iteration	5.1821e-08	5 Sec
<i>traingd</i>	In incremental mode training and slow response the basic gradient descent can be used.	2325 Iteration	0.019993	8 Sec

### 5.RESULT AND DISCUSSION

The results are achieved with the aid of using suggested BP – ANN algorithm for detection and diagnosis of lung cancer diseases which was useful to optimize the synaptic weights connections of the network. Five input neurons for different features extraction, two hidden layers, two output layers and two output neurons were utilized for normal and abnormal lung with cancer. Five types of features were extracted from histogram equalization and Gray-Level Co-occurrence Matrix (GLCM) were used as input neurons. Three types of activation functions were tested and utilized that are *traingd*, *trainbr*, and *trainlm*, each one has different algorithm and performance. Table 1 shows the specification of activation functions, it clarifies the description, definition of each one, with epochs, performance and time of implementation of each kind of activation function that was used in this work. The result shows that *trainlm* was the best with 9 iterations, it takes only one second to implement with accuracy rate of 95.9%. Besides that, *trainbr* take 5 second with 16 iterations for implementation and finally *traingd* take 8 seconds for implementation and 2325 iterations. Figure 2 shows neural network training algorithm with *trainlm* activation function. Figures 3 - 5 shows best training performance for *trainlm* activation function, neural network regression, validation checks

and epoch in sequence. Three kinds of important parameters were obtained from the confusion matrix for each type of activation functions used in the proposed technique of *BP – ANN* classification. The classification focuses on accuracy, sensitivity and specificity, which were calculated by using the term of, true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*) that shown below [6].

$$Accuracy = \left( \frac{TP+TN}{TP+TN+FP+FN} \right) \times 100\% \quad (14)$$

$$Sensitivity = \left( \frac{TP}{TP+FN} \right) \times 100\% \quad (15)$$

$$Specificity = \left( \frac{TN}{TN+FP} \right) \times 100\% \quad (16)$$

Figure 6 shows result of these parameters for trainlm activation function in confusion matrix performance. Table 2 illustrates comparison between these parameters for the proposed technique with some previous works such as (*BP – ANN*), (*SVM*) support vector machine, and *ANN* as in [4, 6, 8]. Significant progress in the results was achieved with the proposed technique as shown in Table 2. Figure 7 describe the chart that illustrated accuracy, sensitivity and specification of proposed and previous techniques. Different steps of image processing for lung cancer, with graphical user interface (*GUI*) for diagnosis and classification of lung image shown in Figures 8 -9 -10 and 11.

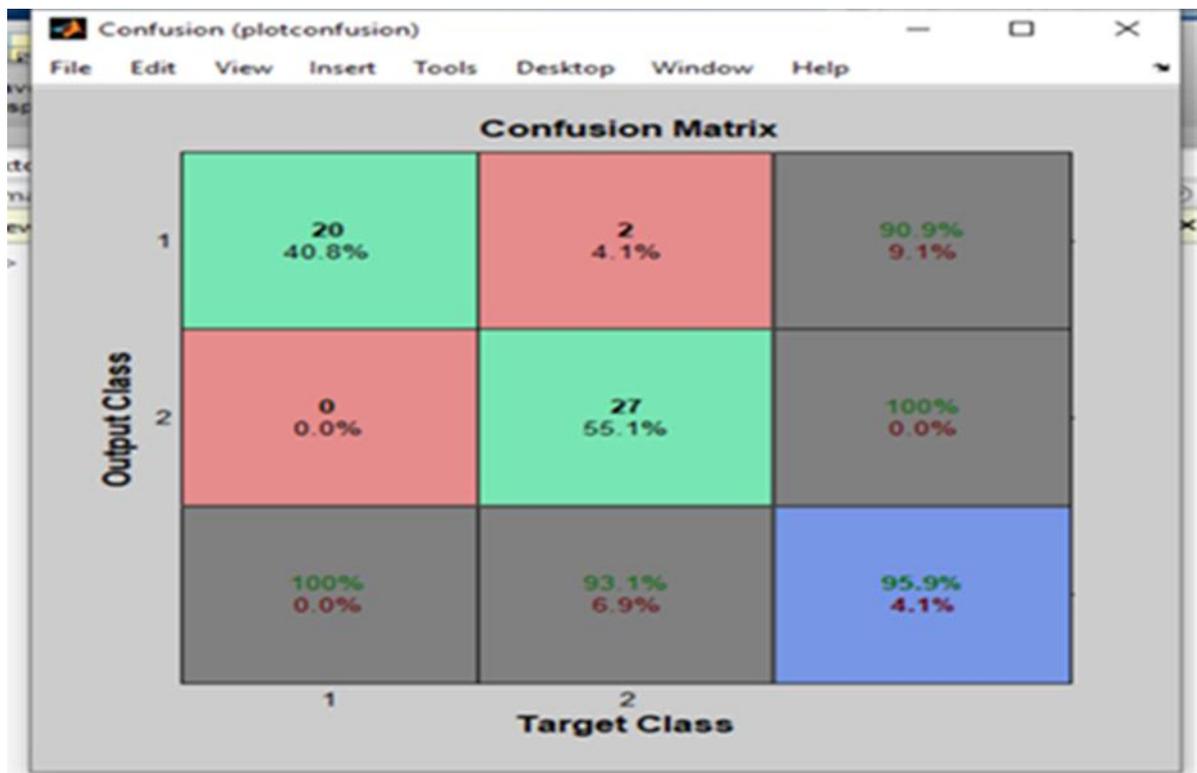


Figure 6: Confusion Matrix for Testing Phase for trainlm Activation Function

Table 2: Evaluation Parameters of Proposed Method and Previous Methods with Different Algorithm

Sequences	Parameters	Proposed Method ( <i>BP – ANN</i> )	Previous Method ( <i>BP – ANN</i> )	Previous Method ( <i>SVM</i> )	Previous Method ( <i>Feed F. ANN</i> )
1	Accuracy	<b>95.918 %</b>	90%	83.07%	94%
2	Sensitivity	<b>100%</b>	100%	65.64%	88.7%
3	Specificity	<b>93.1 %</b>	79%	92.36%	97.1%

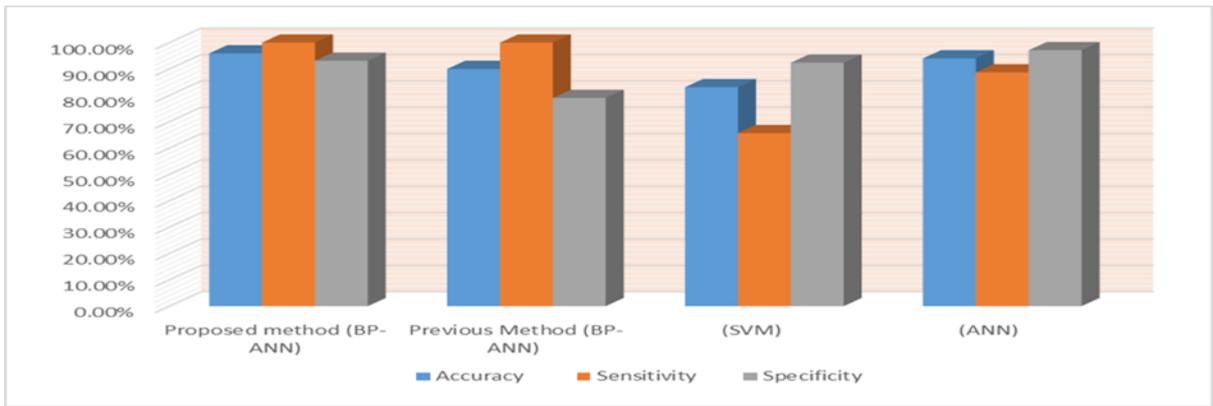
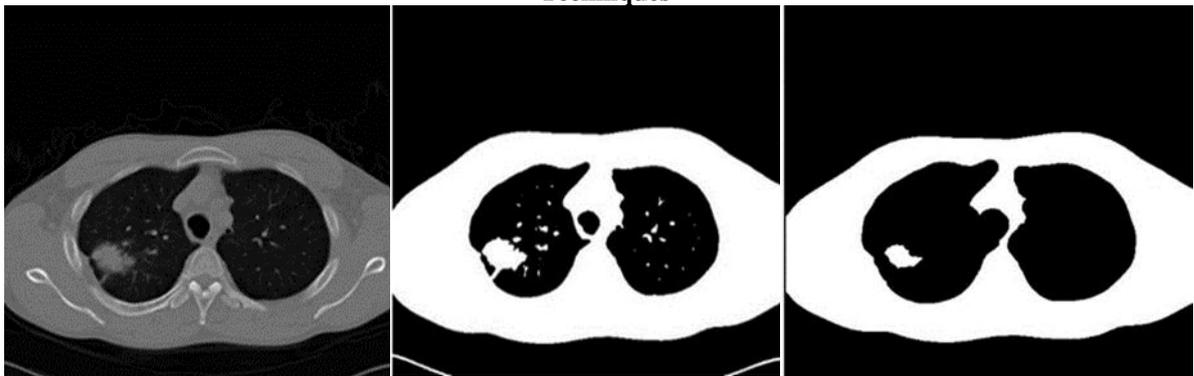


Figure 7: Chart Illustrated Accuracy, Sensitivity and Specification of Proposed Technique and Previous Techniques



a-Image of Lung with Cancer      b- Image after Pre-Processing      c-Final Feature Extraction

Figure 8: Different Steps of Image Processing for Lung with Cancer Diseases

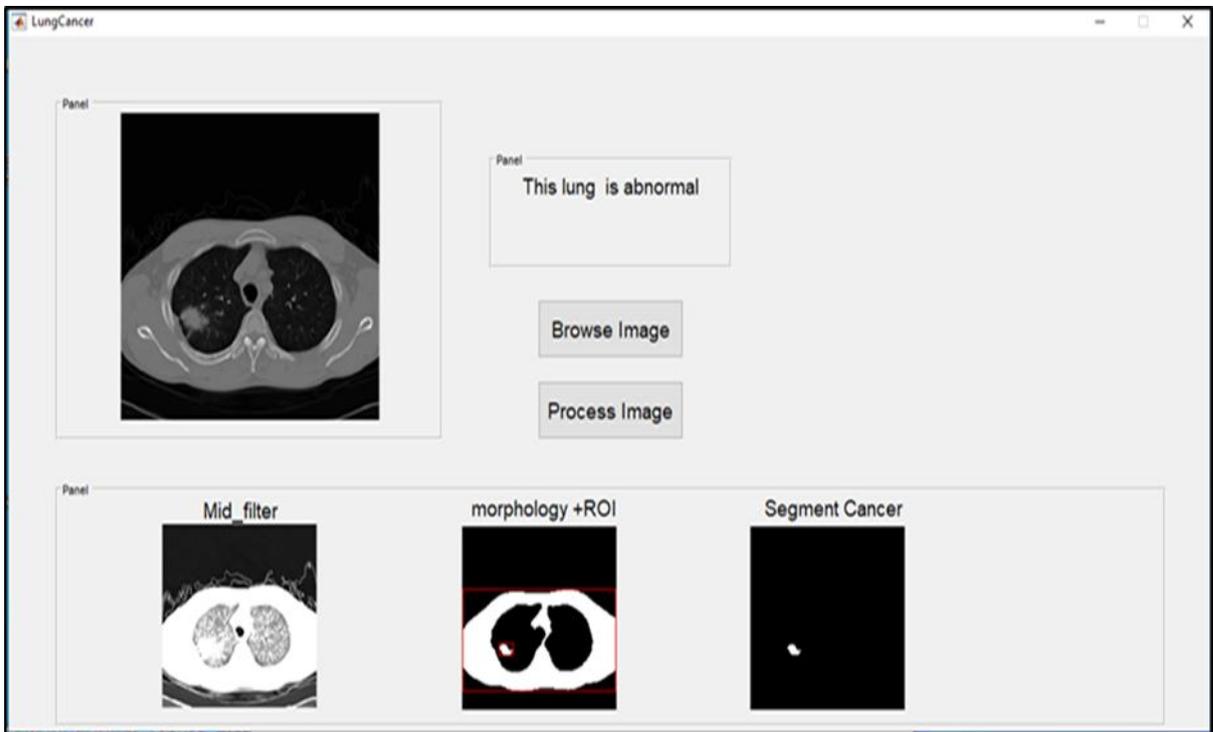
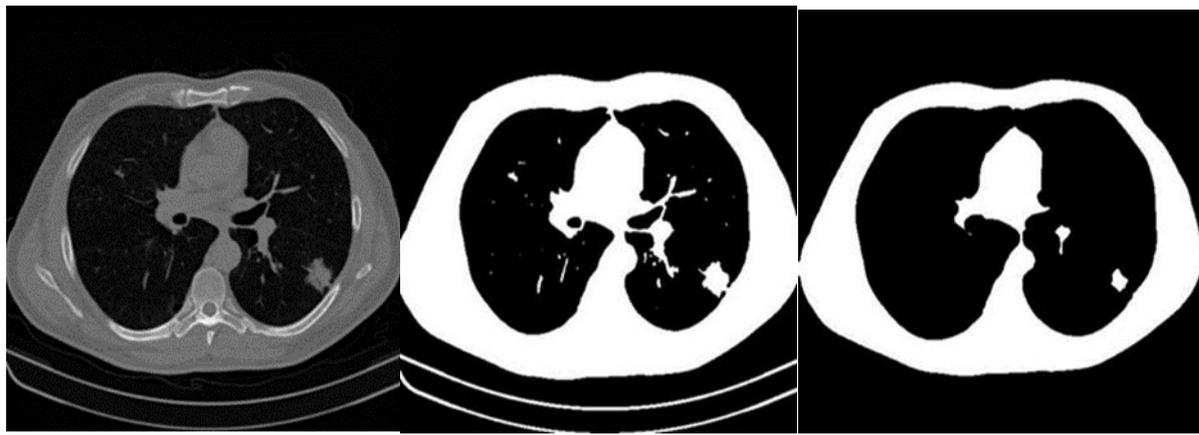


Figure9: Graphical User Interface (GUI) for Diagnosis of lung Image with Cancer Diseases that Shown in Figure 8 using BP-ANN



a-Image of Lung with Cancer

b- Image after Pre-Processing

c- Final Feature Extraction

Figure 10: Different Steps of Image Processing for Lung with Cancer Diseases

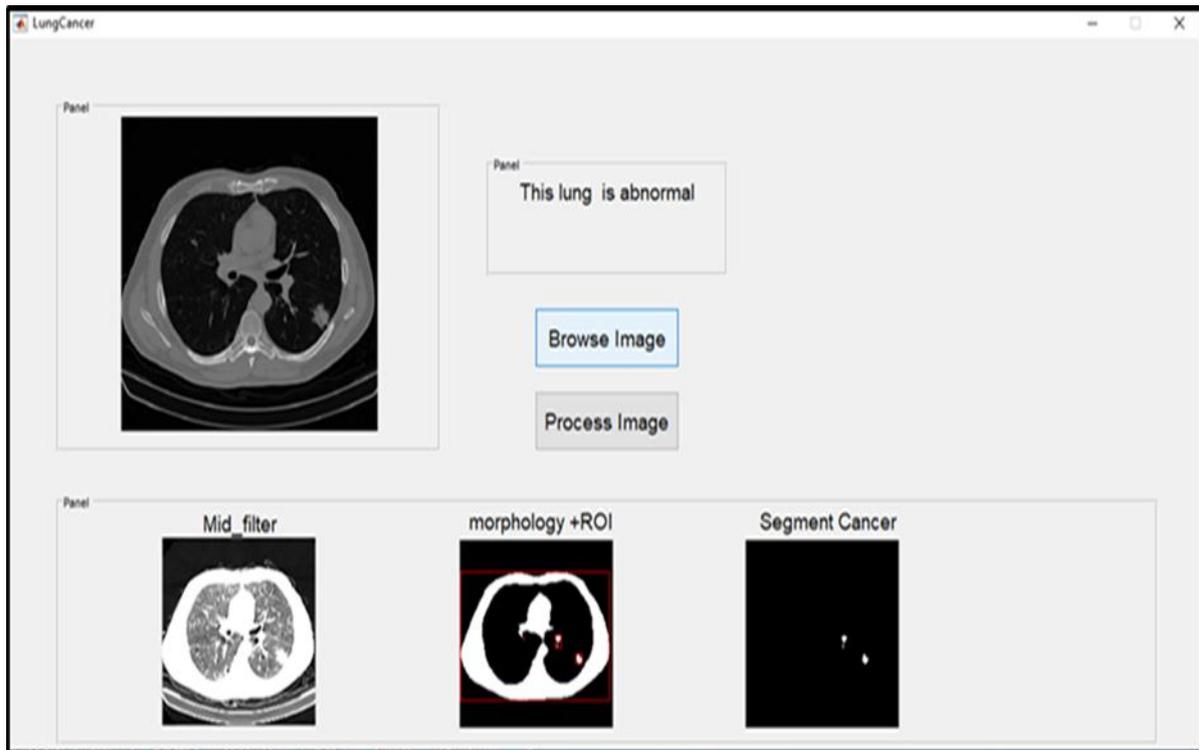


Figure11: Graphical User Interface (GUI) for Diagnosis of lung Image with Cancer Diseases that Shown in Figure 10 using BP-ANN

## 6. CONCLUSION

In this paper an effective and fast classification algorithm for diagnosing lung cancer is proposed. Some of the important points are shown below:

- 1) In order to give the proposed method reliability and accuracy real dataset were used. It was collected from several patients in Baghdad Educational Hospital /medical city / Radiology department. One hundred samples of images have been collected.
- 2) Using of ([https:// www.radiantviewer.com](https://www.radiantviewer.com)) which is specialized site for viewing medical images we can save the real dataset.
- 3) Several image processing was done, such as smoothing, enhancement, and segmentation.
- 4) Using Otsu's process for enhancement and eliminate noise led to an acceleration in the proposed algorithm.
- 5) Five features were selected and extracted by using two feature extractions techniques which are: mean histogram equalization and *GLCM*, both techniques gave best characteristics for the tumors.
- 6) Three types of activation functions were used trainlm, trainbr, and traingd. They gave different mean square error, epoch, iteration, and performance.

- 7) Trainlm function has the fastest and stable training algorithm for suggested network with reasonable size, and memory lowering feature for use when the training set is huge.
- 8) The result obtained that *BP – ANN* gave the best and fast diagnosis through activation function trainlm with accuracy rate 95.9%.
- 9) Using confusion matrix to show the parameters of accuracy, sensitivity and specification. which achieved a better results when compared with previous methods, such as *SVM, ANN, BCO, CNN*.
- 10) *GUI* was presented to show multi-steps of diagnosis and final decision whether the lung is normal or abnormal with cancerous cells.
- 11) Using too many features in the training set leads to confusion in the network, decreases the overall performance and Increase processing time.
- 12) Increasing number of training set for lung images will improve the classification rate but will cause an increase in the processing time.
- 13) Since medical images dataset diagnosis must be very accurate so it can be collected from different sources, besides that dataset can be increased for future work to give better accuracy.

### Acknowledgments

This paper was supported by University of Technology /Iraq /Baghdad/ Electrical Department where the most experiments were done in intelligent systems neural laboratory. Also, the data were collected from Baghdad educational hospital /Medical city / Radiology Institute. So, we would like to offer our warmest regards for all of the support given from all sections for extending their support and providing us with the resources necessary for completing this paper.

### Reference:

- [1] Dr. Hanan A. R. Akkar, "Multiple Step of Image Processing and ANN for Diagnosing Lung Diseases: Review Study," *Int. J. Eng. Res.*, Vol. 8, No. 12, pp. 16–21, 2019.
- [2] T. a Keziah and P. Haseena, "Lung Cancer Detection Using SVM Classifier and MFPCM Segmentation," *Int. Res. J. Eng. Technol.*, Vol.5 ,No.4, pp. 3114–3118, 2018.
- [3] S. Sindhu Priya and B. Ramamurthy, "Lung cancer detection using image processing techniques," *Research Journal of Pharm.and Technol.*, Vol. 11, No. 5, pp. 2045–2049, 2018.
- [4] S. Makaju, P. W. C. Prasad, A. Alsadoon, A. K. Singh, and A. Elchouemi, "Lung Cancer Detection using CT Scan Images," *Procedia Comput. Sci.*, Vol. 125, pp. 107–114, 2018.
- [5] D. N. Sonar and U. V Kulkarni, "Lung cancer classification | Kaggle," *International J.of Com. Sciences and Eng.*, Vol.4, No. 12, pp. 51–55, 2016.
- [6] K. Sathiesh, P. Santhoshini, "Lung Pattern Classification for Interstitial Lung Disease using ANN-BPN and Fuzzy Clustering" , *Journal- Global Research and Development J. for Eng.*, Vol. 2, No. 5, pp. 298–304, 2017.
- [7] S. Perumal and T. Velmurugan, "Lung cancer detection and classification on CT scan images using enhanced artificial bee colony optimization," *Int. J. of Eng. and Technol.*, Vol. 7, No. 2.26, p. 74, 2018.
- [8] L. Li, Y. Wu, Y. Yang, L. Li, and B. Wu, "A New Strategy to Detect Lung Cancer on CT Images" , 3rd IEEE Int. Conf. Image, Vis. and Comput. ICIVC, Vol.5 ,No.12 ,pp. 716–722, 2018.
- [9] T. Jin, H. Cui, S. Zeng, and X. Wang, "Learning Deep Spatial Lung Features by 3D Convolutional Neural Network for Early Cancer Detection" , *Int. Conf.on Digit. Image Comput. Tech. and Appl.*, vol. 2017-Decem, pp. 1– 6, 2017.
- [10] G. Blanchet, M. Charbit , "Digital Signal and Image Processing using MATLAB " , European Environment Agency (EEA). Antony Rowe Ltd, Chippenham, Wiltshire, Great Britain ,2006.
- [11] R. Huidrom, Y. J. Chanu, and K. M. Singh, "A Study on Lung Segmentation for the Detection and Diagnosis of Lung Cancer " , *International Journal of Advanced Research in Computer Science and Softwar Engineering* , Vol. 6, No. 11, pp. 55–66, 2016.
- [12] M. V. A.Gajdhane and P. D. L.M, "Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques " , *IOSR J. of Comput. Eng.*, Vol. 16, No. 5, pp. 28–35, 2014.
- [13] N. Kumar, H. S. Shukla, and R. P. Tripathi, "Image Restoration in Noisy free images using fuzzy based median filtering and adaptive Particle Swarm Optimization - Richardson-Lucy algorithm " , *Int. J. of*

*Intell. Eng. and Syst.*, Vol. 10, No. 4, pp. 50–59, 2017.

- [14] R. Mastouri, N. Khelifa, H. Neji, and S. Hantous-Zannad, “Deep learning-based CAD schemes for the detection and classification of lung nodules from CT images: A survey ” , *J. of Xray. Sci.and Technol.*, Vol. 28, No. 4, pp. 591–617, 2020.
- [15] G. V Tcheslavski, “Morphological Image Processing : Basic Concepts ”, pp. 1–13, 2009.
- [16] C. D. S. S. Baboo and E. Iyyapparaj, “Analysis of Classification Methods For Diagnosis of Pulmonary Nodules In Ct Images ” , *IOSR J. of Electr. and Electron. Eng.*, Vol. 12, No. 02, pp. 42–50, 2017.
- [17] M. Computing, N. Panpaliya, N. Tadas, S. Bobade, R. Aglawe, and A. Gudadhe, “a Survey on Early Detection and Prediction of Lung Cancer ” ,*Int. J. Comput. and Sci. Mob. Comput.*, Vol. 4, No. 1, pp. 175–184, 2015.
- [18] M. Vas and A. Dessai, “Lung cancer detection system using lung CT image processing ” ,*Int. Conf. Comput. Commun. Control and Autom. ICCUBEA 2017*, pp. 1–5, 2018.
- [19] S. Mangrulkar, “Artificial neural systems,” West Publishing Company ISA Trans., Vol. 29, No. 1, 1990.