A Hybrid Approach to Detect Lung Cancer using ANN

Surbhi Maan[1], Dr. Shilpa Sethi[2], Dr. C K Nagpal[3]

[1] Research Scholar, Department of Computer Engineering, J. C. Bose University of Science and Technology, YMCA Faridabad

[2] Assistant Professor, Department of Computer Applications, J. C. Bose University of Science and Technology, YMCA Faridabad

[3] Professor, Department of Computer Engineering, J. C. Bose University of Science and Technology, YMCA Faridabad

[1]surbhimaan.77@gmail.com

[2]munjal.shilpa@gmail.com

[3]nagpalckumar@rediffmail.com

Abstract:

In the medical field, the concept of imaging is gaining importance with a rise in the demand for automated, reliable, fast and efficient diagnosis which may provide insight to the image better than human eyes because manual observation and identification of carcinoma is a very difficult task as it consumes lot of hard work, efforts etc. and there are much more chances of occurring false detection, therefore an efficient technique or system is required for this. Carcinoma is the second leading cause for cancer-related deaths. Diagnosis of tumor and its implementation plays a very important part within the diagnosis of benign and malignant tumors. A major reason behind an increase in the number of cancer patients worldwide is the ignorance towards treatment of a tumor in its early stages. This paper covers such a system which will inform the user about details of tumor using basic image processing techniques. These methods include noise removal and sharpening of the image together with basic morphological functions, to get the background. It helps both the medical staff and the patient to understand the seriousness of the tumor.

Keywords:

Lung Cancer Detection, Image Processing, Classification, CT Scan Images, Segmentation

I. INTRODUCTION

Lung Cancer is the most dangerous disease which occurs due to the uncontrolled growth of abnormal cells. These abnormal cells do not carry out the functionality of the body properly and if not detected at early stages can spread into different areas of body such as lung [3]. Normally a doctor goes through the CT images and detects the cancer. However, the manual observation and identification of carcinoma is a very difficult task as it consumes lot of hard work, efforts etc. and there are more chances of occurring false detection, therefore an efficient technique or system is required for this. Although there are various techniques for determining the cancer but image processing is considered to be one of the efficient method.

Image processing is a method used to convert an image into a digitization form by performing a series of operations such as the input is taken as an image / video frame / photograph and the output may be an image or characteristics associated with the same [1]. There are three main steps in image processing. They are listed below:

- **Input:** The image is imported with the help of an optical scanner or by a digital camera.
- **Analysis:** The part of images not visible through naked eyes are analyzed using techniques such as satellite photographs to determine.
- **Output:** The output in the form of altered image is obtained using techniques.

Various image tools such as X-ray, Ultrasound, Computer Aided Tomography (CT) etc[1] can be used for identification of diseases. Some of them are discussed below:

- **Identification for Cardiovascular disease** - The guts size and its shape is required to classify the guts diseases, image analysis techniques are employed to radiographic images.
- **Identification for Lung disease** - In X-rays, the regions that appear dark contains air while the region that appears lighter are solid tissues. Bone are more radio opaque than tissues.
- **Identification for Breast Tumor** - These can be manipulated using Image processing techniques such as segmentation, enhancement, feature extraction, etc.

The various Image Processing techniques listed are[2]: Image processing, Image enhancement, Image segmentation, Feature extraction, Image classification, Diagnostics.

II. LITERATURE REVIEW

Several researchers has proposed and implemented detection of lung cancer using different approaches of image processing and machine learning.

T. Sowmiya et al. [3] Lung Cancer being one of the most dangerous and life taking disease spreads due to the uncontrolled growth of abnormal cells and sometimes these abnormal cells lead to tumor. However, early diagnosis and treatment of these cells can save the life of person. This paper revolves around the modern data mining techniques that are used for prediction of Lung cancer and its optimization. The technique explained is:- Ant Colony Optimization (ACO) and the proposed method is:- Reduced Order Constrained Optimization (ROCO). Here, the ACO technique is used to increase the disease prediction value and vice-versa and its behavior is based on the real ants having the capability how they choose and select the shortest path between their shelter and food products resource and the ROCO method which helps in working in more type of limitless numbers in different solutions. Overall, it provides the basic framework to detect the disease and deals with various data mining techniques namely ACO technique which is used for rule generating and classification purposes.

Dasu Vaman et al. [4] Image quality and accuracy are the core factors which are taken in account for obtaining an image which clearly depicts the lung cancer influenced area, furthermore as image quality assessment and its improvement are obsessed with the advance stage which is used within Gaussian rules based on the Gabor filters. To obtain feature extraction, the segmentation principles are taken into account. A normality comparison is created based on the general features. During this research, pixels percentage and mask-labeling are the main detected features.

Disha Sharma et al. [5] developed an automatic CAD system for early detection of lung cancer by analyzing LUNG CT images.

Firstly, the lung regions are extracted from the CT image using several image processing techniques, including bit image slicing, erosion, and Weiner filter. Instead of using the thresholding technique they have used bit plane slicing technique in the extraction process in which they convert the CT images into the binary images. After performing the extraction process, the lung regions which are extracted are segmented using region growing segmentation algorithm which are further analyzed to extract a group of features which are visiting be employed in the rules. Here, they have achieved accuracy of 80%.

Rachid Sammouda et al. [6] developed an automatic CAD system for early detection of lung cancer through the analysis of lung CT images. First, the lung regions are extracted from the CT image using classical image processing techniques including bit plane slicing. Sequentially the procedure applied is: erosion, median filter, dilation, outlining, lung border extraction and flood fill algorithm. The lung regions which are extracted are segmented by an unsupervised modified Hopfield Neural Network Classifier (UMHNNC). Then the clustering of similar segmented lung region is done. After segmentation process, the resultant is used in the nodule detection process. After the detection process is over, three filters are applied which extract only those lung cancer regions which are correct.

Fatma Taher et al. [7] proposed a way which detects carcinoma supported the color images of the sputum. This system follows the method of feature extraction. After extraction, segmentation is finished by the Hopfield Neural Network (HNN) which divides it into 3 clusters namely background, cytoplasm and nuclei clusters. Here, the approach employed in image segmentation is region based approach which provides accuracy higher than the pixel based segmentation. Then the nuclei cells are detected from which some features are extracted and rules are formulated which are employed in the CAD system which discriminate between normal and abnormal cells (cancerous cells). Here, the CAD system is tested on 100 cases and it is observed that the CAD system and the pathologists agreed on 92% of the cases successfully in predicting the carcinoma and 8% disagree. The evaluation criteria like, sensitivity, accuracy and specificity so as to guage the CAD system are defined in terms of True Positives, False Positives, True Negatives and False Negatives as follows:

Sensitivity = True Positives / (True Positives + False Negatives)

Accuracy = (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives)

Specificity = True Negatives / (True Negatives + False Positives)

S. Sivakumar et al. [8] developed an algorithm called weighted Fuzzy C -Means (WFCM) which work upon the weights which are added to the centroid value. The objective is to find the data points which are lightly to be considered less. The proposed WFCM equation is given by eqn(1) and eqn(2):-

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n} W_k u_{ik}^m \;\; ||x_k - v_i||^2 \quad 1 \le m < infinity \; --------(1)$$

$$W_k \quad = \sum_{y=1}^{n} \exp\left(-\frac{||x_k - x_y||}{STD}\right) \; ---------------------(2)$$

where

- W_k is the weight measurement of the data,
- **c** is the number of clusters,
- **n** is the number of data points,
- u_{ik} is the membership of x_k in class i , satisfying $\sum_{ik}^{c} u_{ik}$=**1,**
- **m** is the fuzziness value,
- **v** is the set of cluster centers.

If a data point is not nearer to input data then W_k will be larger. The modified equation for WFCM is given by eqn(3).

$$v_i = \frac{\sum_{k=1}^{n} W_k \ (u_{ik})^2 \cdot x_k}{\sum_{k=1}^{n} W_k \ (u_{ik})^m} \qquad ----------------------(3)$$

The analysis on CT scan images shows that clustering of FCM is less sensitive to outliers. Moreover, the iteration counts through WFCM which converges to fewer iterations and thus produces somewhat good clustering results. Finally, the output datasets shows that WFCM deals with huge amount of data and produces more accuracy in clustering.

Jue Jiang et al. [9] proposed two neural networks namely Multiple Resolution Residual Network (MRRN) and dense MRRN which are used to part the lung tumor obtained from the CT images. Experiments showed that performance observed with high accuracy. The approach used can also be applied to those CT images in which the size of tumour and its appearance can be altered.

Moffy Vas et al. [10] developed an automated system. The various steps that are carried out are image acquisition (Unprocessed CT image), pre-processing, segmentation, feature extraction and classification to detect the carcinoma. The noise in the images are removed with the help of median filter. The classifier used is the Artificial neural networks which provided accuracy of 92% with the aim of detecting the cancer at early stages.

The main findings from the literature review are listed in Table 1.

Table 1: Summary of Literature Review

S. No.	Name of Author	Methodology	Working	Pros	Cons
1.	T. Sowmiya et al. [3]	Ant Colony Optimization Algorithm	Basic frame work to detect the disease & deals with data mining techniques for rule generation & classification purposes.	It helps to work in more type of limitless numbers in different solutions.	The planner time savings of ROCO can be improved.
2.	Dasu Vaman et al. [4]	Gabor filter, Auto enhancement algorithm, FFT Fast Fourier Transform, Thresholding approach, Marker-Controlled Watershed Segmentation approach, Binarization, Masking Approach.	Initially, image is enhanced then these images are divided and segmented to obtain general features.	The proposed technique is efficient for segmentation principles.	Often lack the sensitivity and specificity needed for accurate classification.
3.	Disha Sharma et al. [5]	Image processing techniques such as bit image slicing, erosion, Weiner filter.	An automatic CAD system is developed for early detection in which lung regions are extracted to convert CT image into binary then these are segmented.	The accuracy achieved is 80% for cancerous nodules in 2.5-7.0 mm.	If the nodules size is greater than 2.5-7.0 mm then the proposed system fails.
4.	Rachid Sammouda et al. [6]	Hopfield Artificial Neural Network Classifier	An automatic CAD system is developed for early detection by analyzing lung CT images	The detection process is very promising.	This system does not distinguish between benign and malignant cancer.
5.	Fatma Taher et al.	Bayesian	Technique is proposed	Over diagnosing	Bayesian

	[7]	classification, Hopfield Neural Network	for detecting the cancer based on its size and shape. It is followed by extracting the nuclei and cytoplasm then the connected regions are determined and from them features are analyzed to formulate diagnostic rules.	100 cases, the CAD system and the pathologist agreed on 92% of the cases and 8% disagree.	decision theory can be used for detecting the lung cancer cells.
6.	S. Sivakumar et al. [8]	Weighted Fuzzy C-means clustering	The performance of the standard FCM and the proposed weighted FCM is examined on LIDC[8] model.	WFCM produces more result in less time with more accuracy of clustering.	WFCM deals with lot of noise data.
7.	Jue Jiang et al. [9]	U-net (Convolutional networks) , Full Residual Neural Network	Developed a multi-scale CNN approach which identifies accurate tumor volume if present in the lungs.	The approach used can also be applied to those CT images in which the size of tumour and its appearance can be altered.	There is no difference in the volumetric tumor estimations.
8.	Moffy Vas et al. [10]	Image processing techniques	First images were collected then the cropping is done to eliminate noise, various filters are used for this then segmentation is done to convert it into binary and at last rules are generated. mages	The achieved accuracy is 92%. It helps in detecting the cancer at earlier stages.	Further accuracy can be achieved by using p-tile thresholding and watershed segmentation.

III. PROPOSED MODEL

The probable proposed framework for designing the intelligent CADs system consists of following 5 components shown in Fig.1:

- Image Acquisition (Unprocessed CT image)
- Image pre-processing
- Image Segmentation
- Feature Extraction from CT images
- Classification using ANN

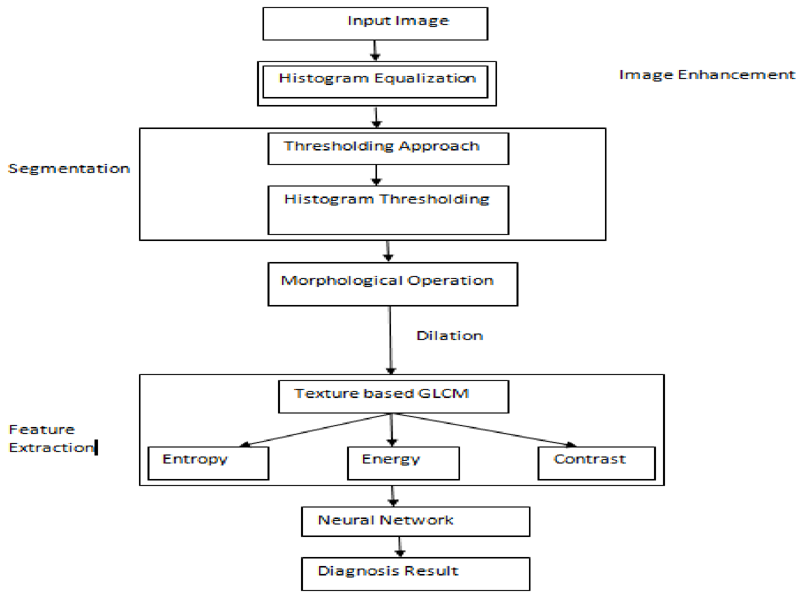

Fig.1 Proposed framework for detecting Carcinoma using Image Processing Techniques

Proposed framework Description:

1. Input Acquisition

It comprises of acquiring the CT scan images from the LIDC-IRDI dataset.

2. Image Pre-processing

It consists of 2 steps:

- Removal of noise from input CT lung image.
- Conversion of RGB image into grey scale image.

For further enhancement of the image, here the filter used is Histogram equalization.

➢ **Histogram Equalization**

The given CT scan image is equalized using histogram. The Histogram of an image represent the relative frequency of occurrences of pixel in a given image. The non-uniform varying image due to external conditions is equalized to a uniform variation.

3. Segmentation

In this, a digital image is divided into multiple segments from which some meaningful information is extracted by which we can locate objects and limits. In this, a label is assigned to each pixel of an image which provide some useful characteristics such as color, texture or intensity. As a result, a set of segments is obtained which can be used to create 3D images.

For segmentation of the image, the thresholding technique is used:

- **Thresholding approach**

This method is predicted on a clip-level to obtain a binary image. The key is to pick the edges value (or values when multiple-levels are selected). The popular thresholding methods which is used in the industry are the maximum entropy method, Otsu's method and et al. k-means clustering can also be used.

4. Morphological Operation

This is used as a image processing tools for sharpening the regions and filling the gaps for binarized image. It is performed into two operations that are :- Dilation and Erosion.

- **Dilation Operation**

Dilation operation is used for filling the broken gaps at the edges and to have continuities at the boundaries. A structuring element of 3x3 square matrix is used to perform dilation operation.

5. Feature Extraction

To the extracted region the feature extraction process is applied for the calculation of 3 invariant features.

- Area
- Contrast
- Entropy

Grey Level Co-occurrence Matrix (GLCM) features are extracted from the matrices.

6. Classification using Neural network

For the automated recognition of tumor cell in given CT scan image a neuro classifier is realized. The classifier module implements a hybrid algorithm neural network based on the value of Entropy, Energy and Contrast which results into number of person suffering from cancer.

IV. IMPLEMENTATION

The implementation for the proposed **Detection of Lung Cancer using Image Processing techniques** will be done using MATLAB R2018a. The dataset for the same has been collected from LIDC-IDRI[11]. It is freely available. The dataset is of 2.5 GB. The dataset consists of CT scan images, DICOM Metadata Digest (CSV), Radiologist Annotations / Segmentations (XML format), Nodule Size List (web), Nodule Counts by Patient (XLS), Patient Diagnoses (XLS).

It consists of 1000 CT scans of large and small tumors. The lung CT image is much faster than X-ray and MRI image. The CT images have better clearance, low noise. The CT images are of 512×512 pixels in size. The 110 nodules of size less than 3mm are used.

V. RESULT

By developing the proposed framework, the demarcation of the tumour in the CT images is obtained. The database for the same are shown in Figure 2, 3 & 4.

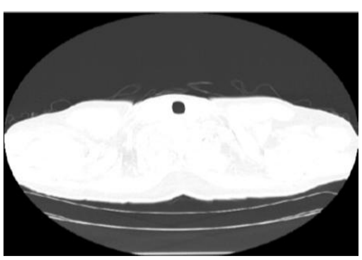

Fig.2 Test Image1

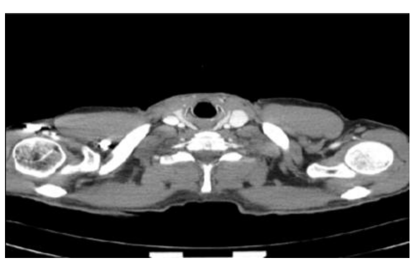

Fig.3 Test image2

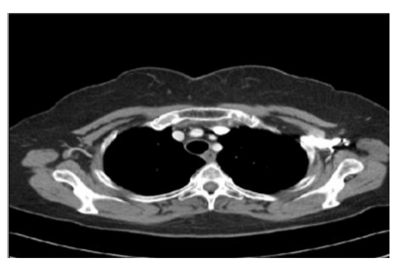

Fig.4 Test image3

The following results shows the outputs received after each step in the algorithm.

After the original image undergoes pre-processing transformation, we get fig.6 from fig.5. These basic pre-processing transformation include:

1. Changing the image to greyscale, as we need to find contour of the final image which works on grey scale images.

2. Applying low pass filter, to remove any noise, if present, in the image.

3. Applying high pass filter, to obtain sharpened image with clear-defined boundaries.

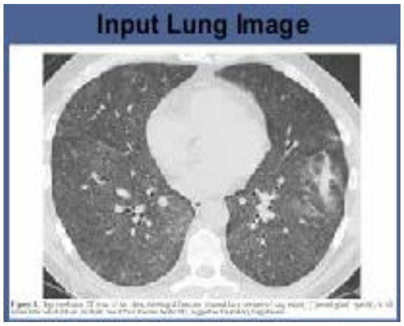

Fig.5 Original image

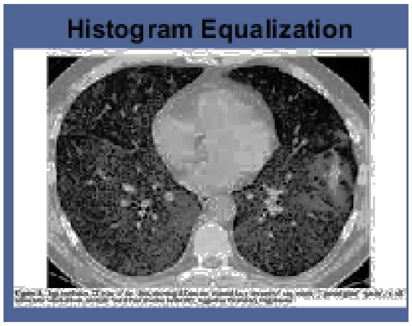

Fig.6 Pre-processed image

Series of steps that lead to Fig.7 from Fig.6 are:

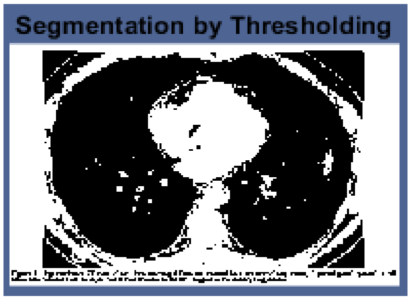

Fig.7 Thesholding image

Transitions from fig.7 to fig.8, fig.9 and fig.10 happens by following steps mentioned below:

1. Dilating with structural element of small radius, say 3, to dilate.

2. Use morphological open to get same amount of dilation, to estimate background.

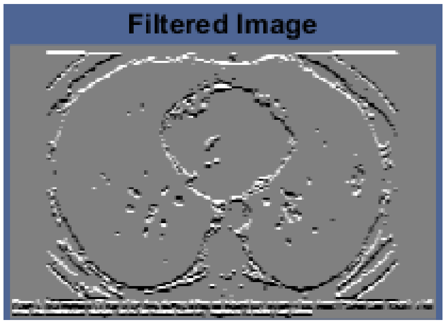

Fig.8 Filtered image

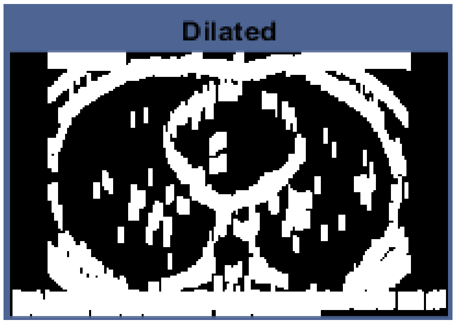

Fig.9 Dilated image

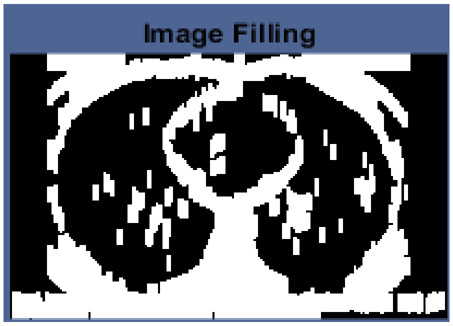

Fig.10 image filing

Hence, the entropy value achieved is shown below:

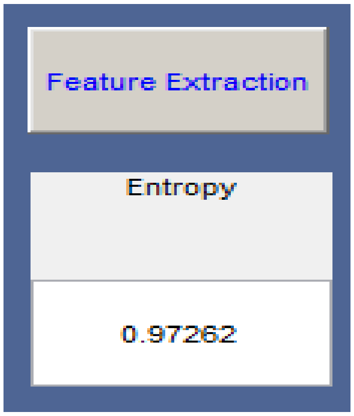

Fig.11 Entropy image

TABLE 1. ENTROPY VALUE OF THE TEST IMAGES

Images	Entropy value extracted
2	0.9726
3	0.956
4	0.9643

The neural network training model for the corresponding CT image is depicted below .i.e., fig.12

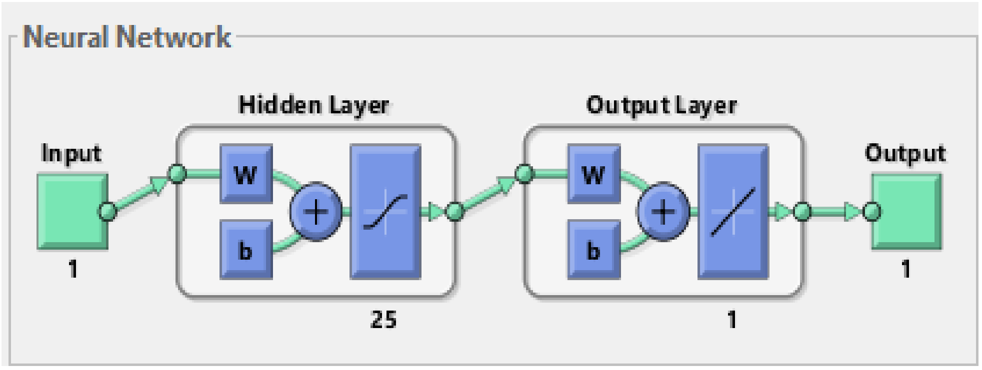

Fig.12 Neural Network Training model

The neural network training state for the corresponding CT image is depicted below .i.e., fig.13

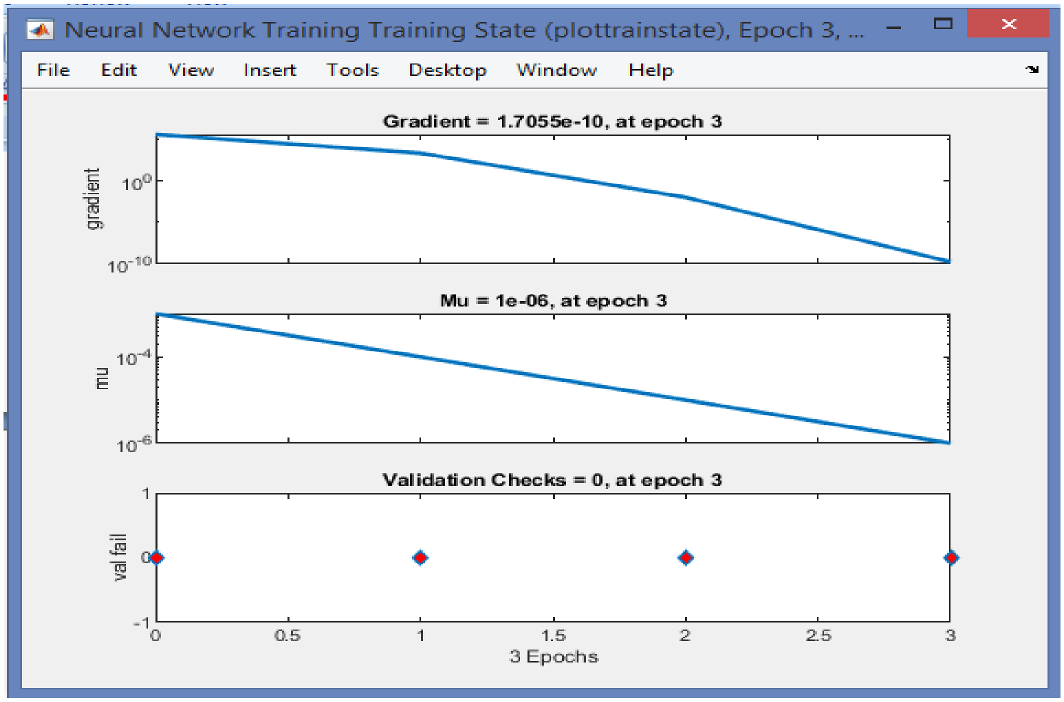

Fig.13 Neural Networking Training State

The neural network training Regression for the corresponding CT image is depicted below .i.e., fig.14

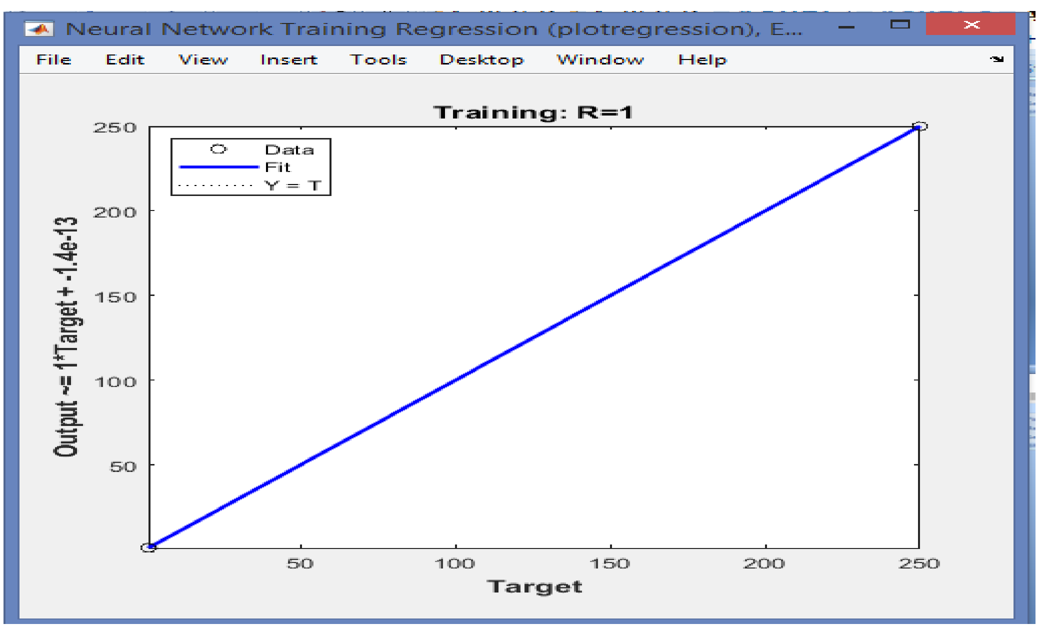

Fig.14 Neural Networking Training Regression

VI. CONCLUSION:

In this paper, a brief glimpse of various image processing techniques is presented to detect lung cancer at earlier stages so as to overcome various health issues further related to it. By applying image processing techniques on the CT scan images of the patients, the nodules are detected which possibly identify whether there is possibility of cancer or not. After applying the prediction model, the dataset is obtained from which feature extraction process goes through to know how many people are suffering from cancer or not. It provides accurate information which not only helps the radiologists but also the doctors to take correct precautions on time. Not only this process is less costly but also less time consuming. An automated system is developed. The various steps that are carried out in image processing are image acquisition (Unprocessed CT image), image pre-processing, image segmentation, feature extraction (area, perimeter, entropy) and classification to detect the lung cancer. The noise in the images are removed with the help of median filter. The classifier used is the Artificial neural networks which provided accuracy of 90% with the aim of detecting the cancer at early stages. For future work, the proposed system can be trained with large number of datasets to diagnose the type of cancer with its size and shape. It also includes the stages of the lung cancer.

REFERENCES:

[1] Available from http://www.engineersgarage.com/articles/image-processing-tutorial-applications, accessed August 2019.

[2] Abhigna B. S, Nishant Sai, Prof. Tapas Kumar, 2016, "Lung Cancer Detection through Image processing", International Journal of Scientific & Engineering Research, vol. 7, no. 12, ISSN 2229-5518.

[3] T. Sowmiya, M. Gopi, M. New Begin, L. Thomas Robinson, 2014, "Optimization of Lung Cancer using Modern data mining techniques", International Journal of Engineering Research Vol. No. 3, Issue No. 5, pp. 309-314.

[4] Dasu Vaman Ravi Prasad, 2013, "Lung Cancer detection using image processing techniques", International Journal of Latest Trends in Engineering and Technology.

[5] Disha Sharma, Gagandeep Jindal, 2011, "Identifying Lung Cancer using Image Processing Techniques", International Conference on Computational Techniques and Artificial Intelligence (ICCCT AI).

[6] Rachid Sammouda, Computer Sciences Department, 2016, "Segmentation and Analysis of CT Chest Images for Early Lung Cancer Detection", Global Summit on Computer & Information Technology (GSCIT), Sousse, pp. 120-126.

[7] Fatma Taher, Rachid Said Sammouda, 2007, "Identification of Lung Cancer based on Shape and Color", International Conference on Innovation in Information Technology, pp. 481-485.

[8] S. Sivakumar, C. Chandrasekar, 2012, "Lung Image Segmentation through Weighted FCM", International Conference on Recent Advances in Computing and Software Systems, pp. 109-113.

[9] Jue Jiang, Yu-chi Hu, Chia-Ju Liu, Darragh Halpenny, Matthew D. Hellmann, Joseph O. Deasy, Gig Mageras and Harini Veeraraghavan, 2019, "Multiple Resolution Residually Connected Feature Streams For Automatic Lung Tumor Segmentation From CT Images", IEEE Transactions on Medical Imaging, vol. 38, no. 1, pp. 134-144.

[10] Moffy Vas, Amita Dessai, 2017, "Lung Cancer detection system using lung CT image processing", International Conference on Computing, Communication, Control and Automation (ICCUBEA), pp.1-5.