Detection of lung cancer from CT image using image processing and neural network

Conference Paper · May 2015

DOI: 10.1109/ICEEICT.2015.7307530

2 authors:

Md Badrul Alam Miah
Mawlana Bhashani Science and Technology …

Mohammad Abu Yousuf
Jahangirnagar University

Detection of Lung Cancer from CT Image Using Image Processing and Neural Network

Md. Badrul Alam Miah
Department of Information and Communication Technology
Mawlana Bhashani Science and Technology University
Tangail, Bangladesh
badrul.ict@gmail.com

Mohammad Abu Yousuf
Institute of Information Technology (IIT)
Jahangirnagar University
Dhaka, Bangladesh
usuf672@yahoo.com

***Abstract*—Detection of lung cancer is the most interesting research area of researcher's in early stages. The proposed system is designed to detect lung cancer in premature stage in two stages. The proposed system consists of many steps such as image acquisition, preprocessing, binarization, thresholding, segmentation, feature extraction, and neural network detection. At first Input lung CT images to the system and then passed through the image preprocessing stage by using some image processing techniques. In first stage, Binarization technique is used to convert binary image and then compare it with threshold value to detect lung cancer. In second stage, segmentation is performed to segment the lung CT image and a strong feature extraction method has been introduced to extract the some important feature of segmented images. Extracted features are used to train the neural network and finally the system is tested any cancerous and noncancerous images. The performance of proposed system shows satisfactory results and proposed method gives 96.67% accuracy.**

Keywords—Preprocessing; Binarization; Segmentation; Feature Extraction; Neural Network; Lung Cancer Detection

I. INTRODUCTION

Lung cancer is one kind of dangerous diseases of the world. In every year more people die because of lung cancer than any other types of cancer such as: breast, brain, and prostate cancers. Lung cancer is leading cause of death from cancer among people of ages between 45 and 70. Lung cancer is responsible for more than 25% of all cancer-related deaths every year and lung cancer kills more people than breast, colon and prostate cancers combined. There are many existing techniques (most of these are expensive and time consuming) that are used to detect lung cancer in advanced stages, such as Computed Tomography (CT), Chest Radiography (x-ray), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology. So, it is a great needed for a new technology to detect lung cancer in its early stages. The proposed techniques provide a good quality tool to detect lung cancer in early stages. Related works are explained in section II. The Proposed system employed in this paper is in section III. The Result and discussion of this paper explain in section IV. Implementation of this paper explained in section V. Conclusion of this paper is in section VI.

II. RELATED WORKS

The Lung cancer is a disease of abnormal cells multiplying and growing into a tumour. Cancer cells can be carried away from the lungs in blood, or lymph fluid that surrounds lung tissue. Now several systems are proposed and still many of them are conceptual design. Artificial Neural Network based Classification and detection system of lung cancer [1-2], this system is conceptual and provide poor accuracy. Computer-aided diagnosis in chest radiography [6] has classify the lung regions extraction approaches into two different categories; either rule-based or pixel classification based category. Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm [4], it is old model and provides poor detection. CADs can be divided into two groups: density-based and model-based approaches [6]. In some approaches uniformity, connectivity, and position features were extracted [5]. Lung cancer detection by using artificial neural network and fuzzy clustering methods [3], presents two segmentation method and Lung Cancer Detection Using Image Processing Techniques [7-8,11] and Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier [10] have been developed but they provide poor detection and identification. Lung Cancer Detection using Curvelet Transform and Neural Network [12], propose a new technique for LCD identification where curvelet transform can extract the features of lung cancer CT scan images proficiently. In recent year, the latest research's work are done in the field of lung cancer detection such as Lung Cancer detection and Classification by using Machine Learning and Multinomial Bayesian [14], Lung Cancer Detection and Classification by Using Bayesian Classifier [15], Automatic Detection of Lung Cancer in CT Images [16], Lung Cancer Detection Using BPNN and SVM [17], Size Estimation of Lung Cancer Using Image Segmentation and Back Propagation [18], and Gray Coefficient Mass Estimation Based Image Segmentation Technique For Lung Cancer Detection Using Gabor Filters [19]. After surveying different research works, the objective of proposed system is to represent a fast and robust system for detecting Lung Cancer properly in early stage and our proposed system provide more accuracy than many other existing techniques.

III. THE PROPOSED SYSTEM

Proposed system focus on detection of Lung cancer from CT image which is automatically detects Lung cancer. The system based on computer vision that is used to detect the Lung cancer. The main contribution of this thesis is to develop a rotation, scaling and translation invariant feature extraction method for detection of Lung cancer. The overall lung cancer detection system architecture is shows in Fig. 1. The whole system of lung cancer detection divided into following steps- Image Acquisition, Image Preprocessing, Binarization, Segmentation, Thresholding Method, Feature Extraction, Neural Network Detection.

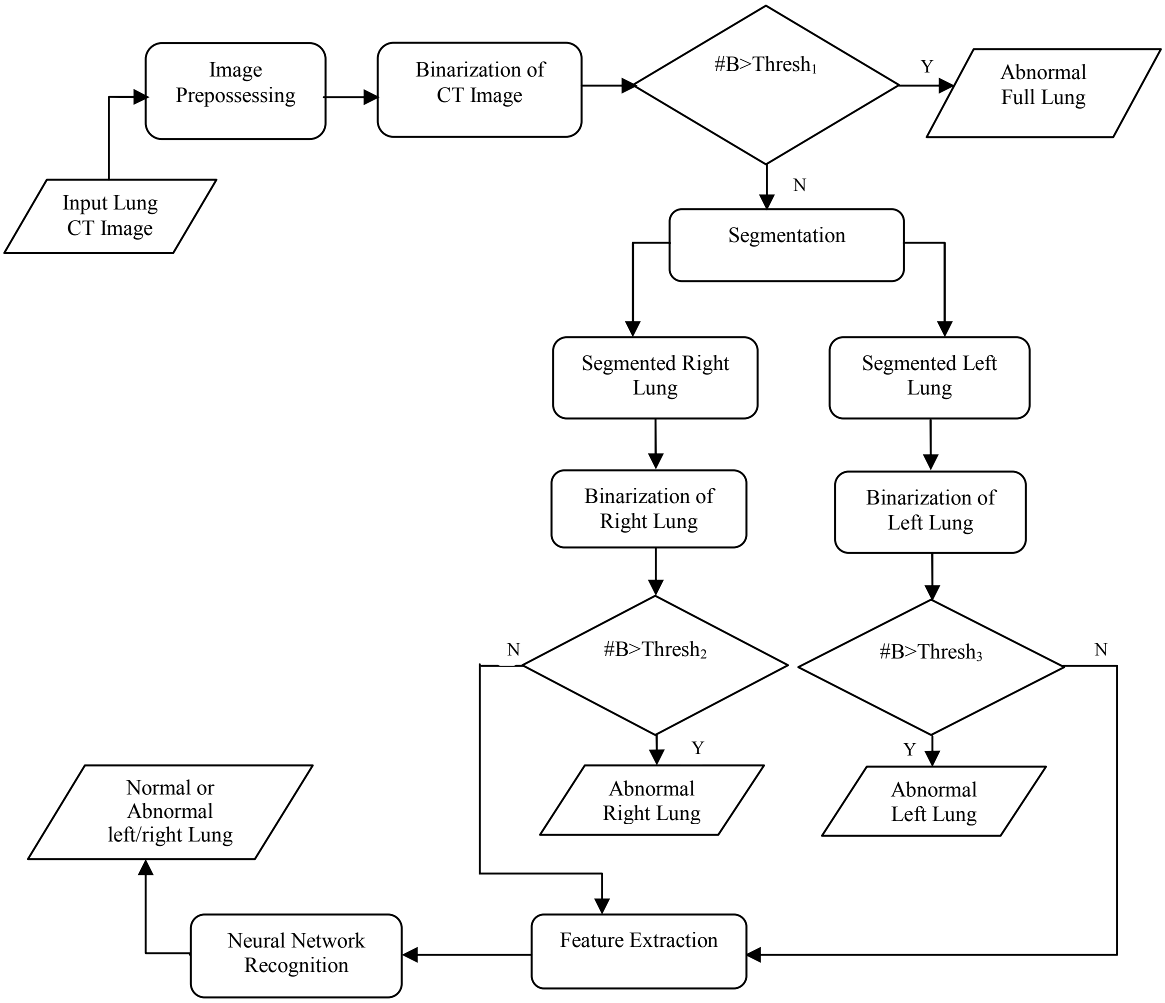


Fig. 1. Lung Cancer Detection System Architecture

A. Image Acquisition

Normally a special type of digital X-Ray machine is used to acquire detailed pictures or scans of areas inside the body called computerized tomography (CT). Computed tomography is an imaging procedure. The system has been collected total 300 Lung CT images that are cancer and normal image of lung from the Internet and Hospital. The system used Lung CT images that are .jpeg file format.

B. Image Preprocessing

After Image Acquisition, images are passed through the image preprocessing steps. Fig. 2 shows the block diagram of image preprocessing steps.

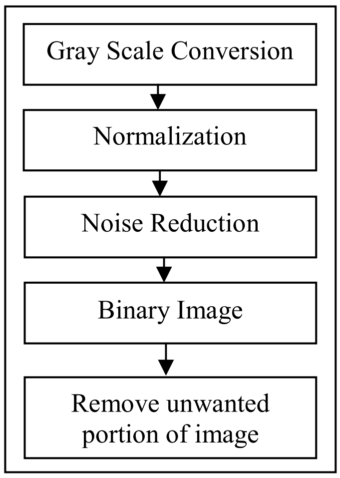


Fig. 2. Block Diagram of image preprocessing

1) Gray Scale Conversion

RGB image converted into gray scale image by using the Matlab function *rgb2gray.* It converts RGB image or color image to grayscale by eliminating the hue and saturation information while retaining the luminance.

2) Normalization

Normalize the acquired image by using the Matlab function *imresize.* The system uses *imresize* function with the value of 150 × 140 pixels and 200 × 250 pixels. This size gives enough information of the image when the processing time is low.

3) Noise Reduction

To remove the noise the system used median filter i.e. medfilt2. Medfilt2 is 2-D median filter. Median filtering is a nonlinear operation often used in image processing to reduce "salt and pepper" noise. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges.

4) Binary Image

Noise free gray scale image is converted to binary image, that is an image with pixels 0's (white) and 1's (black). To convert gray scale image into binary image, the system use the Matlab function *im2bw.*

5) Remove unwanted portion of the image

Converting into binary image, we have to remove the unnecessary pixels (0) from original image. This is done because we need to develop size independent algorithm.

Algorithm

a) Start from top-left corner; repeat for each column and row.

If sum of all black pixels in row/column>0

Then save column and row

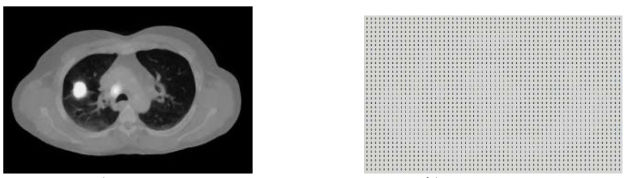
b) Else don't save column/row.

C. Binarization

Image binarization is a technique that converts an image gray level to a black and white image. Binarization is used as a pre-processor frequently. Typically the two colors used for a binary image are black and white. Binary images are also called *bi-level* or *two-level*. This means that each pixel is stored as a single bit i.e. a 1 or 0, the names *black-and-white shown in Fig. 3.*

- Covert into Binary image. Let *f(x, y)* is an input image. T is the threshold value and *g(x, y)* is the output image of thresholding process then the mathematical equation of this conversion is:-

$$g(x, y) = 1, \text{ if } f(x, y) \geq T \text{ otherwise } 0 \quad (1)$$



a) b)

Fig. 3. (a) Gray scale image, (b) Binarized image

D. Segmentation

Image Segmentation in computer vision system, is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. In the proposed system, segmentation processes consists of different steps. At first convert the original grayscale image into edge only image shown in Fig. 4.Then convert the edge only image into dilated image and then filled image shown in Fig. 5. And finally Right Lung and Left Lung are segmented from the image shown in Fig.6.

a) b)

Fig. 4. a) Original grayscale image, b) Edge Only

a) b)

Fig. 5. a) dilated image, b) Filled image

a) b)

Fig. 6. a) Segmented Right Lung, b) Segmented Left Lung

E. Thresholding Method

Thresholding method is based on a threshold value to turn a gray-scale image into a binary image. The key idea of this method is to select the threshold value (or values when multiple levels are selected). Recently, methods have been developed for thresholding computed tomography (CT) images. The simplest method of image segmentation is called the thresholding method. The proposed system used three types of threshold value i.e. $Thresh_1$, $Thresh_2$, and $Thresh_3$. In binary CT image, if the percentage of white pixels is greater than the $Thresh_1$, then full lung is affected. In segmented binary image, if the percentage of white pixels is greater than the $Thresh_2$ and $Thresh_3$, then the right lung and left lung respectively is affected.

F. Feature Extraction

The system has been used a rotation and size independent feature extraction method to extract the feature of the lung cancer and finally obtain 33 features for each type of lung cancer CT images.

Center of the image

Center of the image can obtain by using following equation:

$$Center\_x = width/2 \tag{2}$$

$$Center\_y = height/2 \tag{3}$$

Feature1

The first feature is the ratio of height and width of the segmented lung.

$$feature\ 1 = height/width \tag{4}$$

Feature 2 - 25

These features check how the black pixels are distributed in the image. First the number of pixels inside the image is calculated that is *total_pixels* of segmented lung.

$$Total\_pixel = height \times weight \tag{5}$$

The feature 2 and 3 are the percentage of black pixels located in the upper and lower areas of segmented lung.

$$feature2 = up\_pixels/\ total\_pixels \tag{6}$$

$$feature3 = down\_pixels/\ total\_pixels \tag{7}$$

The feature 4 and 5 are the percentage of black pixels located in the left and right areas of segmented lung.

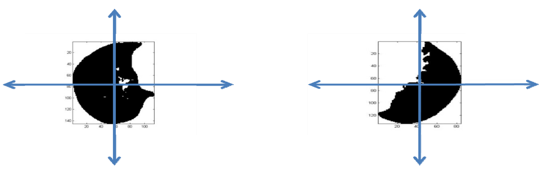
$$feature4 = left\_pixels/\ total\_pixels \tag{8}$$

$$feature5 = right\_pixels/\ total\_pixels \tag{9}$$

Similarly, now split the image into 4 sub regions and calculate the percentage of black pixels located in every region. Again sub divides every region into four and calculates the percentage of black pixels of those regions shown in Fig. 7. The features 6 to 25 are the percentage of black pixels located in the every sub areas of lung cancers.

$$feature_n = sub\_area\_pixels_n/\ total\_pixels \tag{10}$$

Where n = 6 to 25



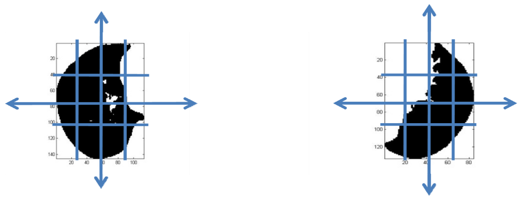


Fig. 7. Split the image region

Feature 26

The feature 26 is the average of the distance between all the black pixels and the central point.

$$feature26 = \frac{1}{Total\_Pixels} \times \sum_y \sum_x \sqrt{(x-i)^2 \times (y-j)^2} \tag{11}$$

Where (i, j) are the coordinates of a point and (x, y) are the coordinates of central point.

Feature 27-33

Calculate the central moments of segmented lung. For *f(x, y)* 2-dimntional function of $M \times N$ binary image, the moment of order $(p + q)$ is defined by:

$$m_{pq} = \sum_{x=1}^{M} \sum_{y=1}^{N} (x)^p (y)^q f(x,y) \tag{12}$$

Where p, q = 0, 1, 2, 3 …. …. ……..

Central moment obtains by following equation:

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{13}$$

Where $\bar{x} = \frac{m_{10}}{m_{00}}$ and $\bar{y} = \frac{m_{01}}{m_{00}}$

For scaling normalization the central moment changes as following equation:

$$\eta_{pq} = \mu_{pq} \big/ \mu_{00}^{\ \gamma} \tag{14}$$

Where $\gamma = \left[(p+q) \big/ 2 \right] + 1$

seven values, computed by normalizing central moments through order three, that are invariant to object scale, position, and orientation. In terms of the central moments, the seven moments are given as,

$$M_1 = \eta_{20} + \eta_{02} \tag{15}$$

$$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^{\ 2} \tag{16}$$

$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (17)$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (18)$$

$$\begin{aligned} M_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30}+\eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \end{aligned} \quad (19)$$

$$\begin{aligned} M_6 = {} & (\eta_{20} - \eta_{20})\left[(\eta_{30}+\eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \end{aligned} \quad (20)$$

$$\begin{aligned} M_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left[(\eta_{30}+\eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] \\ & - (\eta_{30} + 3\eta_{12})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \end{aligned} \quad (21)$$

G. Neural Network Detection

After the Thresholding method, rest of the Lung Cancer Detection System uses neural network which is very efficient and reliable. After the feature extraction process, these features are passed through the neural network to train up the system for classification purpose or detection purpose. The whole proposed training system of lung cancer detection consist of the following steps- Image Acquisition, Image Preprocessing, Segmentation, Feature Extraction, Neural Network Classification.

1) Network Design

A neural network is employed for lung cancer detection. A multilayer feed forward neural network with supervised learning method is more reliable and efficient for this purpose. Neural Network design of the proposed system is shown in Fig. 8.

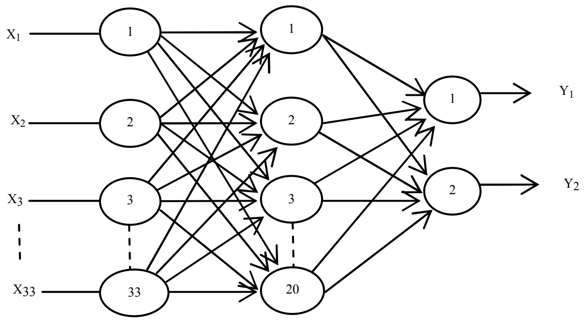


Fig. 8. Neural Network design for the system

IV. RESULTS AND DISCUSSION

A. Training and Testing

To train the neural network extracted features are used. The proposed system is designed such that it can detect which lunch is affected left lung or right lung specifically. In positive samples (cancerous Left lung) are trained by only 0, 1 and negative samples (Non cancerous Left lung) by otherwise. In positive samples (cancerous Right lung) are trained by only 1, 0 and negative samples (Non cancerous Right lung) by otherwise. The system used 20 types of lung CT images to train up the network so that the system detects the lung cancer accurately. The system classifies the cancerous and non cancerous CT scan images after training stage and specified that which lung is affected (Right lung or Left lung). And finally the system is tested any positive and negative samples and it gives proficient results.

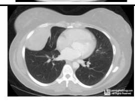
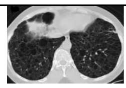
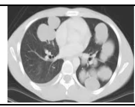
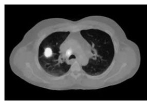
B. Result

The result of the proposed system is performing by two ways: Binarization Technique and Neural Network

1) Binarization Technique: It provide 99% (approx.) correct result for this system.

2) Neural Network: The experimental results of the system are given Table I.

TABLE I. OVERALL DETECTION RATE OF THE SYSTEM

Type of Images	**No. of frames**	**Correct detection rate (%)**	**Error rate (%)**
	30	96.67	3.33
	30	93.33	6.67
	30	100.00	0.00
	30	93.33	6.67
	30	100.00	0.00
Total	**150**	**96.67**	**3.33**

C. Compare to the Existing System

The proposed system introduce a binary thresholding technique, strong feature extraction method and compare to other existing system and achieve better performance for Lung Cancer Detection (LCD) system. The proposed system provides more accurate result compare than other existing system shown in the following Table II.

TABLE II. COMPARE TO THE EXISTING SYSTEM

Lung Cancer Detection System	**Accuracy (%)**
Lung Cancer Detection using Curvelet Transform and Neural Network	90%
Automatic Detection of Lung Cancer in CT Images	96%
Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier	96.04%
Gray Coefficient Mass Estimation Based Image Segmentation Technique For Lung Cancer Detection Using Gabor Filters	83%
Identifying Lung Cancer Using Image Processing Techniques	80%
Proposed System	**96.67%**

V. IMPLEMENTATION

To develop a strong lung cancer detection system, the system use MATLAB, which is a high performance language for computing. It is the standard instructional tool for high-productivity research, development, and analysis. Toolboxes allow learning and applying specialized technology. Toolboxes are comprehensive collections of MATLAB functions that extend the MATLAB environment to solve particular classes of problems. It includes among others image processing and neural networks toolboxes. For these toolboxes, the proposed system is developed using MATLAB.

VI. CONCLUSION

Lung cancer is one kind of dangerous diseases, so it is necessary to detect early stages. But the detection of lung cancer is most difficult task. From the literature review many techniques are used for the detection of lung cancer but they have some limitations. In our proposed method pursue approaches in which first step is binary thresholding, and then feature extraction, and then these features are used to train up the neural network and test the neural network. The proposed system successfully detects the lung cancer from CT scan images. At the end of the system can say that the system achieve its desired expectation. The proposed system test 150 types of lung CT images and obtains the result where overall success rate of the system is 96.67% which meet the expectation of system. In future this technique can be used in the detection of brain tumor, breast cancer etc.

REFERENCES

[1] G. Bhat, V. G. Biradar , H. S. Nalini, "Artificial Neural Network based Cancer Cell Classification (ANN – C3)," *Computer Engineering and Intelligent Systems*, vol. 3, no. 2, 2012.

[2] A. Pathan, B. K. Saptalkar, "Detection and Classification of Lung Cancer Using Artificial Neural Network," *International Journal on Advanced Computer Engineering and Communication Technology,* vol. 1 Issue. 1, ISSN 2278–5140.

[3] D. Lin and C. Yan, "Lung nodules identification rules extraction with neural fuzzy network", *IEEE, Neural Information Processing*, vol. 4, 2002.

[4] B. Zhao, G. Gamsu, M. S. Ginsberg, L. Jiang and L. H. Schwartz, "Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm," *journal of applied clinical medical physics*, vol. 4, 2003.

[5] A. El-Baz, A. Farag, R. Falk, and R. L. Rocco,,"detection, visualization, and identification of lung abnormalities in chest spiral CT scans: phase I," *Information Conference on Biomedical Engineering*, Egypt, 2002.

[6] B.V. Ginneken, B. M. Romeny and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey," *IEEE, transactions on medical imaging,* vol. 20, no. 12, 2001.

[7] D. Vaman, R. Prasad, "Lung Cancer Detection Using Image Processing Techniques", *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, Vol. 3 Issue 1 September 2013.

[8] D. Sharma and G. Jindal, "Identifying Lung Cancer Using Image Processing Techniques," *International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2011)*, 2011, pp. 115-120.

[9] A. Chaudhary and S. S. Singh, "Lung Cancer Detection Using Digital Image Processing," *International Journal of Research in Engineering and Applied Sciences (IJREAS)*, vol. 2 , no. 2, Feb. 2012 , pp. 1351-1359.

[10] Ada and R. Kaur, "Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier," *International Journal of Application or Innovation in Engineering and Management (IJAIEM)*, vol. 2, no. 6, June, 2013, pp. 385-383.

[11] D. Vaman, R. Prasad, "Lung Cancer Detection Using Image Processing Techniques," *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, vol. 3, no. 1, September, 2013 , pp. 372-378.

[12] B. Gupta and S. Tiwari, " Lung Cancer Detection using Curvelet Transform and Neural Network," *International Journal of Computer Applications,* vol. 86 , no.1, Jan. 2014, pp. 15-17.

[13] C. M. Niranjana, J. Dhananandhini, K. Rajeswari, A. Dhivya, "A New Approaches of Lung Segmentation Using Neuro-Fuzzy Network," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, no. 6, Jan. 2014, pp. 52-54.

[14] Mr. S. A. Dwivedi, Mr. R. P. Borse and Mr. A. M. Yametkar, "Lung Cancer detection and Classification by using Machine Learning and Multinomial Bayesian," *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE),* vol. 9, no. 1, Jan. 2014, pp. 69-75.

[15] A. M. Yametkar and R. D. Patane, "Lung Cancer Detection and Classification by Using Bayesian Classifier," *Proceedings of IRF International Conference*, Feb. 2014, pp.7-13.

[16] G. Vijaya, A. Suhasini and R. Priya, "Automatic Detection of Lung Cancer in CT Images," *IJRET: International Journal of Research in Engineering and Technology*, vol. 3, no.7, May, 2014, pp. 182-186.

[17] G. Kaur and H. Singh, "Lung Cancer Detection Using BPNN and SVM," *International Journal of Latest Scientific Research and Technology,* vol. 1, no. 2, Jul. 2014, pp.95-98.

[18] N. Sharma and Sheenam, "Size Estimation of Lung Cancer Using Image Segmentation and Back Propagation," *International Journal for Research in Technological Studies,* vol. 1, no. 9, Aug. 2014, pp. 14-17.

[19] K. Sankar and DR. M. Prabakaran, "Gray Coefficient Mass Estimation Based Image Segmentation Technique For Lung Cancer Detection Using Gabor Filters," Journal of Theoretical and Applied Information Technology, vol. 66 , no. 2, Aug. 2014, pp. 638-644.