International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-6 Issue-3, February 2017

A Qualitative Review on Image Processing Algorithms to Detect Early Stage Lung Cancer

Raja Rao.Chella

Abstract: ***Nowa days, the image processing algorithms are being usedwidely in medical systems for detection of lung cancer. It is observed that the life span rate of lung cancer patients increases from 15 to 50% if they were detected at early stages. Detection of cancer cells is the most important issue for medical researchers as it becomes more complex in the treatment process. The detection steps of presence of cancerous cells include image pre-processing, segmentation, feature extraction and classification. In this paper, algorithms for enhancement, segmentation and feature extractionto detect the cancerous tumors which are small and large in size from the lung CT scan images are reviewed. Finally thealgorithms are compared with one another using three parameters called accuracy, sensitivity and specificity.***

Index Terms: CT Images, Image Preprocessing, Segmentation, Enhancement, Feature Extraction and Classification.

I. INTRODUCTION

In the human body the lungs exist as a pair of cone shaped organs on either side which look very smooth like a sponge. The right sided lung has three lobes, and is lightly larger than the left sided lung, which again has two lobes. The lungs take oxygen from the inhaled air and are transported to the bloodstream and to the other parts of the body using lung tissues and leaves the carbon dioxide when the air is exhaled. Lung cancer is a disease of abnormal blood cells which are multiplying and growing into a tumor within the lung. Cancer cells can be carried away from the lungs in blood, or lymph fluid that surrounds lung tissue. Metastasis occurs when a cancer cell leaves the site where it began and moves into a lymph node or to another part of the body through the bloodstream .Primary lung cancer is the cancer that just started in the lung. Lung cancer symptoms consist of shortness of breath, wheezing, chest pain that does not get better, coughing accompanied with blood, difficulty in swallowing, and loss of weight and appetite. The most recent estimate statistics according to the American Cancer Society indicate that 226,160 new cases will be diagnosed (116,470 in men and 109,690 in women) in US, and there will be estimated 160,340 mortalities from lung cancer (87,750 in men and 72,590 among women).Furthermore, based on statistics from the World Health Organization (WHO), deaths caused by cancer will reach about 12 million people in 2030.

Manuscript published on 28 February 2017.
***** Correspondence Author (s)
Raja Rao. Cheela* Research Scholar, Department of Electronics & Communication Engineering, Sri Satya Sai University of Technology and Medical Sciences, Sehore, Bhopal (M.P.) India.

There are two main groups as small cell lung cancer and Non-small cell lung cancer

a) *Small cell lung cancer:* Recent study says that 20% of cancer cases come under this type. The primary cause for this type of cancer is Smoking. For this type doctors suggest the chemotherapy procedure instead of surgery.

b) *Non-small cell lung cancer:* This type of cancer is formed in the tissues of the lung. There are several types of non-small cell lung cancer. Signs of this type of cancer include a cough that arises very often and problem of effective breathing. The CT images are shown below for Lung with tumor in Fig.1

Fig.1. CT Scan Image of Lungwith a Nodule

II. SURVEY OF RESEARCH WORK

The detection of cancerous tumor in the CT images includes four steps which is shown below in Fig.2.

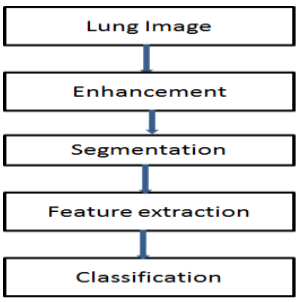

Fig.2. A Basic Lung Detection Process

There are two image databases which are freely available for the researchers on medical image processing and they are Lung Image Database Consortium (LIDC) and ImageDatabase Resource Initiative (IDRI). At present, for the early detection of malignancy or cancer of lung and treatment in the early stage, the digital image processing approaches are most widely used and applied in medical fields in which the time factor is very important to discover the disease in the patient as soon as possible, especially in tumors. The Tumor stages is depicted in Table. I.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

TABLE I. TUMOR STAGES

TUMOR STAGE	SIZE OF TUMOR
T1	< 30MM
T2	30 TO 70 MM
T3	> 70MM

Cancerous tumors are found in lung with the more people including men and women is quite common because of several reasons which include smoking, food habits, pollution and many more. There are certain methods available to detect cancerous cells at different stages of life or time. Diagnosis is mostly based on Computed Tomography (CT), MRI and PET images. The qualitative review presented in this paper focuses on detection of nodules andearly symptoms of the diseases appearing in patient's lungs.The first stage in detection is the preprocessing. The preprocessing of a CT image is done with help of the following filters.

i. *Mean (or) average filter*
ii. *Median filter*
iii. *Adaptive median filter*
iv. *Weiner filter*

These filters are used to calculate the parameters like Mean Square Error(MSE),Peak Signal to Noise ratio (PSNR), etc. For segmentation process there are some articles are reviewed forevaluation tumor levels in the lung. For enhancement, and feature extraction there are certain approaches and algorithms are also reviewed in this paper.The requiredqualitative values are tabulated and analyzed and parameters are compared. The following three parameters are used in the analysis; they are accuracy, specificity and sensitivity which are tabulated from papers that are reviewed. The parameters are calculated with the following expressions.

Accuracy= [TP+TN] / [TP+FP+TN+FN]
Sensitivity= TP / [TP+FN] and
Specificity= TN / [TN+FP].

Where TP is True positive, FP is False Positive, TN is True Negative and FN is False Negative.

III. RECENT ALGORITHMS PROPOSED

Image enhancement, segmentation, feature extraction and finally the classification methods that are reviewed is described below in brief for the tumors which are in big size.

Ada and Rajneet kaur [1] proposed a hybrid technique based on feature extraction and Principal Component Analysis (PCA) for lung detection in CT scan images. TheImage enhancement is implemented with histogram equalization and feature extraction and implemented by Binarization approachthrough which the features are extractedwith number of white pixels having intensity more or less than the threshold value to check the normal or abnormal lungs. To predict the probability of lung cancer is done with binarization and GLCM approach.

Avinash et al.[2]proposed a method for detection of cancerous tumor in lung CT image using Gabor filter for Image enhancement. Instead of DWT and Histogram equalization, watershed segmentation algorithm was used and for feature extraction masking approach used instead of binarization. With this approaches they were able to improve the detection time of cancer.

Sanghamitra et al.[3] came with new method of detection of tumor with four stages of processing. First CT images are collected with different features, preprocessing is done with median filter, wiener filter or Gabor filter. Next, for segmentation, EK-means clustering method was used and the features extraction was done using gray-level co-occurrence matrix (GLCM) method and Back Propagation Network (BPN) is appliedfor classification of disease stages. They classified the given CT image as normal lung or cancerous lung.EK-Means clustering algorithm is one of the unsupervised clustering algorithm. In this method first clusters number 'k' is defined with an integer and randomly the cluster center is chosen. Then calculation of distance between each pixel and cluster center is done. The distance is specified as Euclidean function. Every single pixel is compared to all cluster centers using the distance formula. The pixel is moved to cluster which has shortest distance among all. Then the centroid is reevaluated. Again each pixel is compared to all centroids.The process continuous until the center converges.The performance rate of this algorithm depends onchoice of Structural Element. Segmented lung image and its gray scale are shown in in fig.3. K-Means clustering image is show in fig.4.

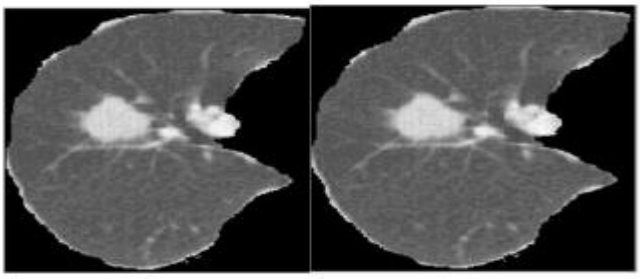

Fig.3. Original and Gray Scale Images

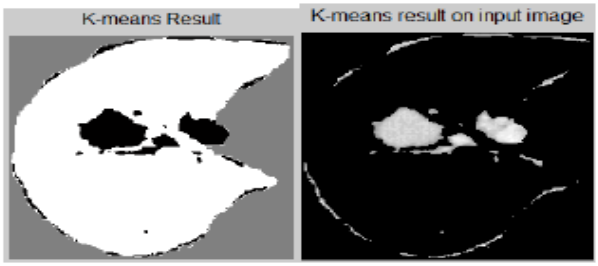

Fig. 4. K-Means Clustering Image

In supervised learning algorithms, Back Propagation Network (BPN) is one method in which theinputs and outputs will be given to the neural network for classification. This method or algorithm consists of two phases: the forward phase and the backward phase to calculate and modify the weights and bias values respectively.In this method,the features likeentropy, correlation, homogeneity, PSNR and SSIM are extracted from segmented images and Back Propagation Network is used for the classificationof Imagewhether it is normal Image or the tumorimageand this gives an accuracy of about 90.87%.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Md .Badrul et al. [4] gave another processing method in which the median filter was used to eliminate noise and to preserve the edges at the same time. They have proposed threetypes of threshold values.Threshold value '1' is used to detect the whether the lung is fully effected or not and '2' and '3' values are used to detect whether the left or right lungs are effected or not respectively.To calculate how accurate this technique compare to the others, they used Neural networks with some parameters as weight factors, and finally they concluded that the average accuracy was 96.67% with several tests based on the images collected from internet and hospitals as well.

Taruna etal [5]presented a methodfor detectionand classification of lung nodules from CT scan images. In this approach, a median filter is used for enhancing the image to get negligible noise and optimal thresholding with morphological closingoperations.The gray level characteristics are used for finding the Region of Interest (ROI) ie lung tumor. Then from the segmented image geometrical features are analyzed by Linear Discriminative Analysis (LDA) classifier along with the statistical features. Finally the results were given as accuracy of 84%, sensitivity of 97.4% and specificity of 53.3% which show that the prior classification of lung nodule can be detected effectively.

Elmar Rendon-Gonzalez and Volodymr Ponormaryov[6]defined the SVM (Support Vector Machine) Classifier based lung nodule detection system. The preprocessing involves selection of the threshold which is -500 HU followed by dilation and further erosion to get only lung image from the all other details. Using this approach, the False Positives can be reduced so that the sensitivity improves which further improves the accuracy of the tumor detection system. For the feature extraction GLCM matrices were used to give the gray level variations. The method estimates a function to classify the data in two classes. The decision function used was

$$D(x) = W^T + b \quad (1)$$

where x_i is the training inputs, W^T is an *m*-dimensional vector, *b* is a bias term, and for $i = 1, \ldots M$

$$w^T x_i + b \begin{cases} \geq 1 & for \quad y_i = 1, \\ \leq -1 & for \quad y_i = -1. \end{cases}$$

Support vectors are called the values which follow the above equations.this method provides the sensibility of 84.93%, specificity of 80.92% and an accuracy performance of 78.08%. The performance of various methods for large tumors are depicted in table.II shown below.

TABLE II. PERFORMANCE COMPARISION OF DETECTION ALGORITHMS (FOR THE TUMORS LARGE IN SIZE)

Method/Technique	Parameter		
	Accuracy	Sensitivity	Specificity
Wavelet algorithm	79.45	87.67	83.56
Linear Discriminative Analysis(LDA)	84	97.14	53.3
Wavelet SVM	82.2	90.09	73.9
Voxel based CAD	----	78.9	-----
Deep continuous neural network	89.3	-----	-----
Support Vector Machine (SVM)	78.4%	-----	------

For Image enhancement, segmentation, feature extraction and finally the classification methods described below in brief for the tumors small in size. Messay et al.[7] proposed an intensity thresholding with morphological processing for image segmentation and detection and for classification, they used Fisher Linear discriminator (FLD) classifier for classification. With an average of 517.5 nodule candidates the method was able to detect 92.8% of all the nodules in the LIDC/testing dataset. Overall, with a specificity of 3 false positives (FPs) per case/patient on average, the CAD system was able to correctly identify 80.4% of the nodules (115/143) using 40 selected features. A 7-fold cross-validation performance analysis using the LIDC database only shows CAD sensitivity of 82.66% with an average of 3 FPs per CT scan/case.

Cascio et al [8] proposed a region growing algorithm followed by opening process for segmentation of selected image. The segmentation and the extraction of the suspected nodular lesions from CT images constitutes a new 3D segmentation, using deformable MSMs For distinguishing the real nodules among nodule candidates, an additional classification step is applied; furthermore, a neural network is applied to reduce the false positives (FPs). The system performance was tested on a set of 84 images and detection rate with an accuracy of 97%. A reduction to 2.5 FPs/CT is achieved at 88% sensitivity.

Saleem Iqbal et al [9] came with new idea in 4014 i.e. automatic detection and segmentation of lung small sized nodules, nodules in close vicinity of the diaphragm and lung wall and subsequent features extraction and classification for low and high contrast nodules on CT images. The methods include multistep threshold for the nodule detection and shape index threshold for false positive reduction. They used 60 CT scans of LIDC dataset and correctly detected 92% of total nodules in the case of 3-30mm size nodules.

Mehdi Alilou et al [10] used multiple thresholds followed by morphological opening and 3D region growing algorithm for segmentation of lung regions. Finally, a combination of a rule-based procedure and support vector machine classifier (SVM) were utilized to classify the nodules. The proposed method was applied on CT images of 60 patients, containing the total of 211 nodules, selected from the publicly available Lung Image Database Consortium (LIDC) image datasetthe proposed framework demonstrated acceptable detection performance of 80% (Sensitivity: 0.80; Fp/Scan: 3.9). After analyzing the recently proposed methods of detection of small size nodules a table is formed with the needed details.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

TABLE III. PERFORMANCE COMPARISION OF DETECTION ALGORITHMS (FOR THE TUMORS SMALL IN SIZE)

Study (Year)	CT scan cases	Nodules	Nodules detected	Nodule size	Sensitivity (%)
Cascio et al (2012)	84	148	130	>3 mm	88
Saleem et al (2014)	60	222	211	3- 30 mm	92
Alilou et al (2014)	60	211	169	> 4 mm	80
Gupta et al (2016)	60	315	301	1 – 30 mm	95.5
	1010	3615	3058	1-30 mm	84.5

Premalatha et al. [11] proposed a system consists of pre-processing, segmentation, feature extraction and final classification. The proposed marker controlled watershed segmentation technique separates the touching objects in the image. It provides best identification of the main edge of the image and also avoids over segmentation. It gives high percentage of accuracy compared to the thresholding algorithm. So it is efficient for segmentation. The proposed technique gives very promising results comparing with other used techniques. The CT scan images which are used for processing are collected from the hospital. This image datasetcontainslung CT scan images with tumor and without tumor. The Fig 5 shows some of the lung CT scan imageswithtumor and withouttumor.

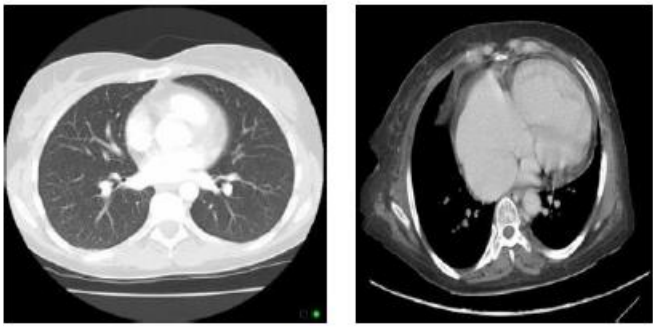

Fig 5. (a) Lung Without Tumor (b) Lung With Tumor

In the image pre-processing stage, image smoothing is the first step. For smoothing, Gaussian filter is apply don the input image. Gaussian smoothing is very effective for removing noise. Gaussian removes high frequency components from the image. So it is a low pass filter. Smoothing reduces the noise and giving us a more accurate intensity surface. The mathematical equation fortheGaussian filter is as given in equation2.

$$G(x,y)= [e^{-(x^2+y^2)/(2\sigma^2)}]1/(2\pi\sigma^2) \text{----------(2)}$$

where x is the distance from the center on the horizontal axis, y is the distance from the center on the verticalaxis,and σ is denoted as the standard deviation of thegaussiandistribution. It determines the amount of smoothing. Theoutput of Gaussian filter for figure 5(b) is shown in figure 6.

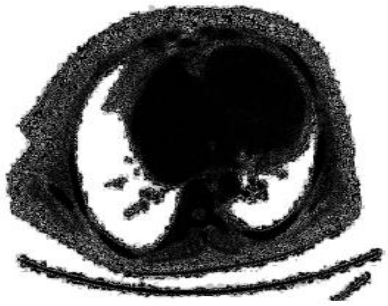

Fig.6. Output Image of Gabor Filter.

Segmentation is used to divide an image into different small regions or objects. It has many applications in the medical field for the segmentation of the 2D medical images. After segmentation process, binarization process is done. In this approach, the total number of black pixels & white pixels are counted. If the total number of black pixels of input image is more than threshold, then the tumor is normal tumor. Otherwise, if the total number of the black pixels is less than the threshold then the tumor is cancerous tumor.

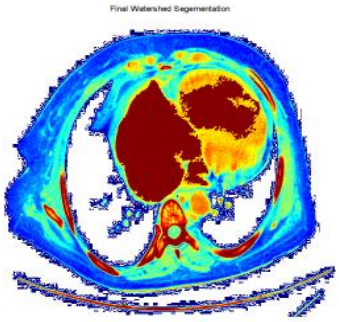

Fig 7. Marker-Controlled Watershed based Segmented Image

To improve the lung tumor detection quality certain methods were followed in the preprocessing. One such method is to use filters to make the image better and enhance it from noising,corruption or interference.

Vicky Ambule etal discussed the about Median filterswhich can be used to reduce impulse noise level from corrupted images. Median filters are used to remove the salt-and-pepper noise. The median filter is a simpler nonlinear smoothing operation that takes a median value of the data inside a moving window of finite length. Median filter can be used to evaluate the averaging value of filter. A methodology based on median filter for the removal of salt and pepper noise by its detection followed by filtering in binary images has been proposed in this paper. An input image has been taken and tested with the approach of median filter. The CT image of tumor affected lung is shown in Fig.8.which is further given to preprocessing using median filter.The image output of median filter is shown in the following Fig.9

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Fig 8. CT Scan Image of Tumor Affected Lung

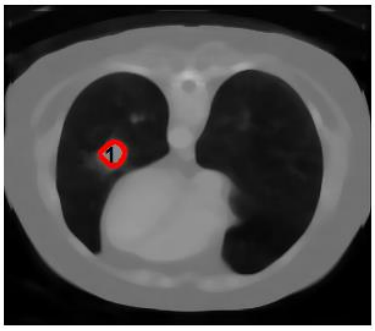

Fig 9. Segmented Image of Tumor Affected Lung using Median Filter

Raajan.P etal developed a method derived from the Wiener filter techniques. Appling the wiener filters in an image adaptively, tailoring itself to the local image variance. It smoothen the image at low variance. Similarly, it also smoothen the image more when the variance high. This filter provides better results compared to the median filter.The image of weiner filtered is shown in Fig.10.

Fig 10. Segmented Image of Tumor affected Lung using Wiener Filter

Finally, the marker controlled watershed algorithm is applied for edge detection and control over segmentation in gradient image. Using wiener filter provides the accuracy of segmentation high.

TABLE.IV. COMPARISON OF MEDIAN AND WEINER FILTERS WITH SAME METHOD OF DETECTION

Filter Type	Region number	Area	Perimeter	Centroid	Diameter
Median	1	307	63.8	90.3 122.8	19.8
Weiner	1	314	71.6	90.9 122.4	20.0

The method is experimented with Berkley Segmentation Dataset and provides better results. The results show that the proposed method effectively reduce the over segmentation effect and achieve more accurate segmentation results than the existing method.The table.IV shows the performance measure of watershed segmentation using with median filter and wiener filter.

IV. CONCLUSIION

This paper focused on the recently algorithms developed by several authors along with the results which are tabulated above. Finally it is observed that the accuracy is more in the case of Deep Neutral Network classification Algorithm and Sensitivity is high with the wavelet based Support Vector Machine and specificity is LDA classifier method in the case of large tumors. In the case of small size tumors a new approach having three masks in the segmentation followed by flood fill algorithm has given the average sensitivity is nearly equals to 90%. Hence there is a need of highly accurate algorithms at the segmentation, feature extraction and even in the last stage the classification stage.

REFERENCES

1. Ada, Rajneet Kaur “Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier”IJAIAM. Volume 2, Issue 6, June 2013
2. Avinash. S, Dr. K. Manjunth, Dr. S. Senthil Kumar,” An Improved Image Processing Analysis for the Detection of Lung Cancer using Gabor Filters and Watershed Segmentation Technique”,IEEE,2016.
3. P.B. Sangamithraa, S. Govindaraju.,” Lung Tumour Detection and Classification using EK-Mean Clustering”, IEEE –WiSPNET conference,2016.
4. Md. Badrul Alam Miah, Mohammad Abu Yousuf,” Detection of Lung Cancer from CT Image Using Image Processing and Neural Network”, Electrical Engineering and Information &Communication Technology (ICEEICT) 2015.
5. Taruna Aggarwal, Asna Furqan, Kunal Kalra,” Feature Extraction and LDA based Classification of Lung Nodules in Chest CT scan Images”,IEEE,2015.
6. Elmar Rendon-Gonzalez and Volodymyr Ponomaryov,”Automatic Lung
7. Nodule Segmentation and Classification in CT Images Based on SVM”,IEEE-2016.
8. T. Messay, R. Hardie and S. Rogers, “A new computationally efficient CAD system for pulmonary nodule detection in CT imagery,”Med Image Anal, vol. 14, pp. 390–406, 2010.
9. D. Cascio, R. Magro, F. Fauci, M. Iacomi, and G. Raso, “Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models,” Computers in Biology and Medicine, vol. 42, no. 11, pp. 1098–1109, 2012
10. Saleem Iqbal et al,”Potential Lung Nodules Identification for Characterization by Variable Multistep Threshold and Shape Indices from CT Images”, Computational and Mathematical Methods in Medicine Volume 2014 .
11. M. Alilou, V. Kovalev, E. Snezhko, and V. Taimouri, “A comprehensive framework for automatic detection of pulmonary nodules in lung CTimages,” Image Anal Stereol, vol. 33, pp. 13-27, 2014.
12. Dasari Hemalatha, Raja Rao.Ch, S.J.Sugumar,” Detection of Lung Cancer Using Marker-Controlled Watershed Transform”, International Journal & Magazine Engineering, technology, management Research,Volume 3,Isuue no.10,2016.
13. Vicky Ambule, Minal Ghute, Kanchan Kamble, Shilpa Katre,” Adaptive Median Filter for Image Enhancement”, International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 1, January 2013.
14. Raajan.P, Muthuselvi.S, Agnes Saleema. A,” An Adaptive Image Enhancement using Wiener Filtering with Compression and Segmentation”, International Journal of Computer Applications, 2015

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

About the Author:

Raja Rao. Cheela, is a Research Scholar (SSSEC1537) inthe Department of Electronics & Communication Engineering at Sri Satya Sai University of Technology and Medical Sciences, Sehore, Bhopal, Madhya Pradesh. India.