IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. III (Sep – Oct. 2014), PP 28-35
www.iosrjournals.org

Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques

Mr.Vijay A.Gajdhane [1], Prof. Deshpande L.M. [2]
[1]Dept. of Electronics and Tele-communication Engineering, TPCT's College of Engineering, Osmanabad, Maharashtra, India
[2]M.E. (Electronics) Dept. of Electronics and Tele-communication Engineering, TPCT's College of Engineering, Osmanabad, Maharashtra, India

Abstract: *Lung cancer seems to be the common cause of death among people throughout the world. Early detection of lung cancer can increase the chance of survival among people. The overall 5-year survival rate for lung cancer patients increases from 14 to 49% if the disease is detected in time. Although Computed Tomography (CT) can be more efficient than X-ray. However, problem seemed to merge due to time constraint in detecting the present of lung cancer regarding on the several diagnosing method used. Hence, a lung cancer detection system using image processing is used to classify the present of lung cancer in a CT- images. In this study, MATLAB have been used through every procedures made. In image processing procedures, process such as image pre-processing, segmentation and feature extraction have been discussed in detail. We are aiming to get the more accurate results by using various enhancement and segmentation techniques.*
Keywords: *LCDS, Watershed Segmentation, ROI, Thresholding, morphologic, Metastasis, CT*

I. Introduction

Lung cancer is a disease of abnormal cells multiplying and growing into a tumor. The mortality rate of lung cancer is the highest among all other types of cancer. Lung cancer is one of the most serious cancers in the world, with the smallest survival rate after the diagnosis, with a gradual increase in the number of deaths every year. Survival from lung cancer is directly related to its growth at its detection time. But people do have a higher chance of survival if the cancer can be detected in the early stages [1]. Cancer cells can be carried away from the lungs in blood, or lymph fluid that surrounds lung tissue. Lymph flows through lymphatic vessels, which drain into lymph nodes located in the lungs and in the centre of the chest. Lung cancer often spreads toward the centre of the chest because the natural flow of lymph out of the lungs is toward the centre of the chest. Lung cancer can be divided into two main groups, non-small cell lung cancer and small cell lung cancer. These assigned of the lung cancer types are depends on their cellular characteristics. As for the stages, in general there are four stages of lung cancer; I through IV. Staging is based on tumor size and tumor and lymph node location. Presently, CT are said to be more effective than plain chest x-ray in detecting and diagnosing the lung cancer. An estimated 85 percent of lung cancer cases in males and 75 percent in females are caused by cigarette smoking.

Objective of this study is to detect lung cancer using image processing techniques. CT scanned lung images of cancer patients are acquired from various hospitals. Using image processing techniques like pre-processing and feature extraction, area of interest is separated. Developing the algorithm, features like area, perimeter and eccentricity are extracted from all the images. The parameter values obtained from these features are compared with the normal values suggested by a physician. From the comparison result, cancer stage is detected. A graphical user interface is developed to scan all the images and display the features and cancer stage. This system can help in early detection of lung cancer more accurately.

II. Literature Survey

2.1 Introduction

The system for detection of lung nodules is proposed here. This chapter presents the basic concepts of the mentioned terms and the work presented by various researchers.

Computed tomography (CT) technique was invented and developed by G.N.Hounsfield at the Central Research Laboratories of EMI Ltd, UK in 1972. Advancement in CT

2.2 Noise Filter

It is the initial step for detecting the lung cancer. In preprocessing step we can do two steps: They are Denoising and Weiner filtering. Image denoising algorithms may be the mostly used in image processing. Many methods, regardless of implementation, share the same basic idea noise reduction through image blurring. Blurring can be done locally, as in the Gaussian smoothing model or in anisotropic filtering by calculating the variations of an image. White noise is one of the most common problems in image processing [3].

2.3 Watershed Segmentation

Ilya Levner [1] has presented the novel approach for creation of topographical function and object markers used within watershed segmentation. According to the author, two key operations in computer vision are segmentation and pixel grouping. While many image segmentation algorithms exist, when objects of the same predefined class are in close proximity to one another, pixel grouping is necessary to cluster the classified pixels into objects. The watershed algorithm is commonly used within the unsupervised setting of segmenting an image into a set of non-overlapping regions.

2.6 Thresholding

Feature extraction is an essential stage that represents the final results to determine the normality or abnormality of an image[1]. These features act as the basis for classification process. Ginneken [2] has classified the lung regions extraction approaches into two different categories; either rule-based or pixel classification based category. Most of the proposed approaches belong to rule-based category [4-5], where a sequence of steps, tests and rules are used in the extraction process. Techniques employed are (local) thresholding, region growing, edge detection, and ridge detection, morphological operations, fitting of geometrical models or functions and dynamic programming.

2.7 Stages

Staging involves evaluation of a cancer's size and its penetration into surrounding tissue as well as the presence or absence of metastases in the lymph nodes or other organs. Staging is important for determining how a particular cancer should be treated, since lung-cancer therapies are geared toward specific stages. In stage I, the cancer is confined to the lung. In stages II and III, the cancer is confined to the chest (with larger and more invasive tumors classified as stage III) Stage IV, cancer has spread from the chest to other parts of the body

S.Sivkumar [10] used fuzzy clustering and support vector machine for lung nodule detection. His work presents an image segmentation approach using standard Fuzzy C-Means (FCM), Fuzzy-Possibilistic C-Means and weighted Fuzzy-Possibilistic C-Means

2.8 Area, Parameter, Eccentricity

Hala Al-shamlan and Ali El Zaart [11], explained the categories of features like geometric, texture and gradient. In medical diagnose, geometric features are essential to recognize any objects. Therefore, to configure and distinguish ROI from others we need to know its geometric features. The basic characters of geometric feature are area, perimeter and compactness. These are measured in scalar.

III. Methodology

Overall, there are three main processes used throughout the report; Pre-processing, feature extraction and finally the classification process. MATLAB is used in every process made throughout the project. Process involved in the lung cancer detection system for the project can be view in Figure (3.1).

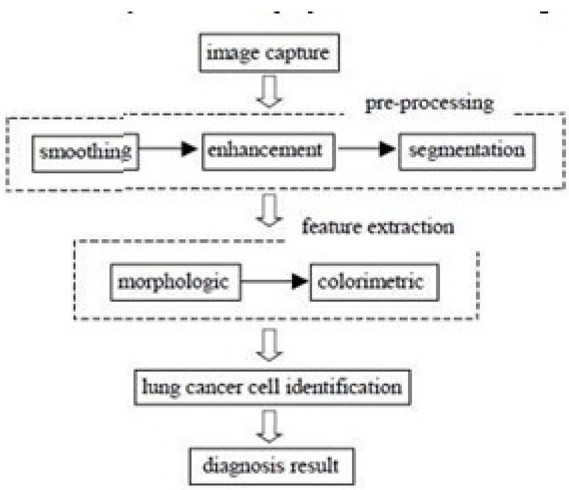

Fig.(3.1) Lung Cancer Detection System

LCDS system uses convolution filters with Gaussian pulse to smooth the cell images. The contrast and color of the images are enhanced. Then the nucleuses in the images are segmented by thresholding. All of those are simple digital image processing techniques. After that, LCDS utilizes morphological and colorimetric to extract feature from image of the nucleuses [3]. The extracted morphologic features include the average intensity, area, perimeter and eccentricity of the nucleuses. On this basis, a lung cancer cell identification nodule is employed to analyze those features to judge whether cancer cells exist in the specimens or not. Moreover, if there are cancer cells, the cancer cell type is identified. The entire diagnosis process of LCDS is shown in Fig. (3.1)

3.1 Image Acquisition

First step is to acquire the CT scan image of lung cancer patient. The lung CT images are having low noise when compared to X-ray and MRI images; hence they are considered for developing the technique. The main advantage of using computed tomography images is that, it gives better clarity and less distortion. For research work, the CT images are acquired from NIH/NCI Lung Image Database Consortium (LIDC) dataset. DICOM (Digital Imaging and Communications in Medicine) has become a standard for medical Imaging. Figure 3.2 shows a typical CT image of lung cancer patient used for analysis.

The acquired images are in raw form. In the acquired images lot of noise is ob-served. To improve the contrast, clarity, separate the background noise, it is required to pre-process the images. Hence, various techniques like smoothing, enhancement are applied to get image in required form

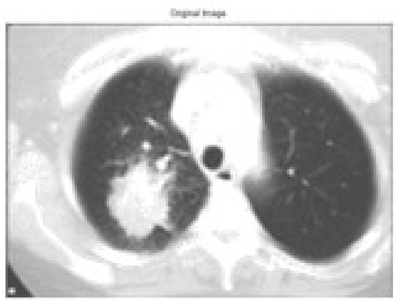

Figure 3.2: Input CT scan Image

3.2 Image Pre-processing

3.2.1 Smoothing

It suppresses the noise or other small fluctuations in the image; equivalent to the suppressions of high frequencies in the frequency domain. Smoothing also blurs all sharp edges that bear important information about the image. To remove the noise from the images, median filtering is used. Median filtering is a non-linear operation often used in image processing to reduce salt and pepper noise. In general, the median filter allows a great deal of high spatial frequency detail to pass while remaining very effective at removing noise on images where less than half of the pixels in a smoothing neighborhood have been affected. **B=medfilt2(A,[m,n])** performs median filtering of the matrix A in two dimensions. Each output pixel contains the median value in the m x n neighborhood around the corresponding pixel in the image. Medfilt2 pads the image with 0's on the edges, so the median values for points within one-half the width of the neighborhood ([m,n]/2) of the edges might appear distorted as shown in fig.3.2.1

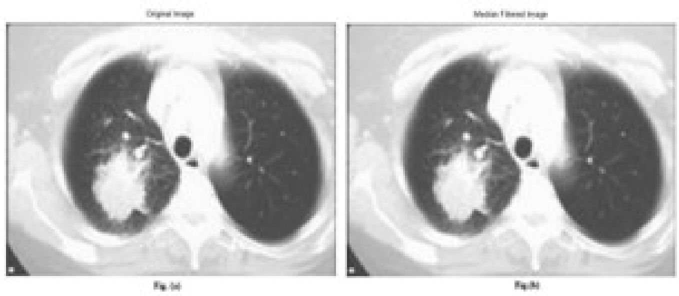

Fig.3.2.1(a)Original Image Fig.(b)Median Filtered Image

3.2.2 Enhancement

Enhancement technique is used to improve the interpretability or perception of information in images for human viewers, or to provide better input for other automated image processing techniques. Image enhancement can be classified in two main categories, spatial domain and frequency domain. Here gabor filter is used for enhancement purpose as it gives better result compared to fast fourier and auto enhancement [2]. A 2D Gabor kernel can be mathematically defined as:

$$G(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}\right)$$

Where, $x' = x\cos\theta + y\sin\theta$; $y' = -x\sin\theta + y\cos\theta$ The parameters involved in the construction of a 2D Gabor filter are: 1.The variance σ of the gaussian function 2.The wavelength λ of the sinusoidal function 3.The orientation θ of the normal to the parallel stripes of the Gabor function 4.The spatial aspect ratio γ specifies the ellipticity of the support of the Gabor function.

For $\gamma = 1$, the support is circular. For $\gamma < 1$, the support is elongated in the orientation of the parallel stripes of the function as shown in fig.3.2.2

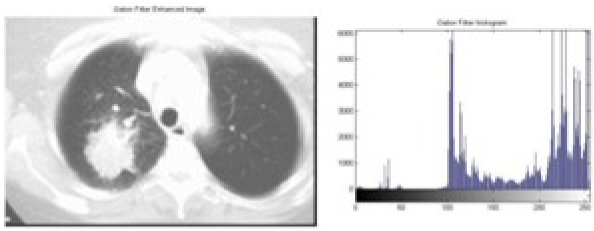

Fig.3.2.2 (a)Filtered Image Fig. (b)Histogram

3.2.3 Image Segmentation

Image segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyze. Segmentation divides the image into its constituent regions or objects. The result of image segmentation is a set of segments that collectively cover the entire image or a set of contours extracted from the image [5]. Marker-controlled watershed segmentation follows this basic procedure and as shown in fig.3.2.3:

1) Compute a segmentation function. This is an image whose dark regions are the objects you are trying to segment.
2) Compute foreground markers. These are connected blobs of pixels within each of the objects.
3) Compute background markers. These are pixels that are not part of any object.
4) Modify the segmentation function so that it only has minima at the foreground and background marker locations.
5) Compute the watershed transform of the modified segmentation function.

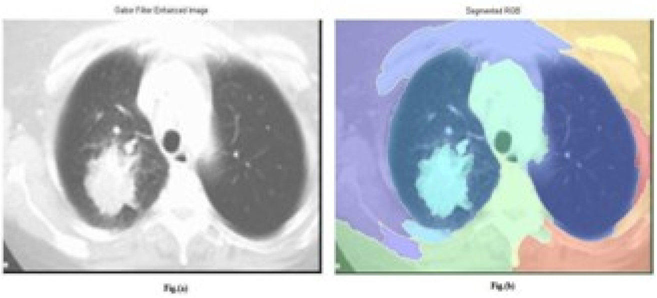

Fig.3.2.3 (a) Enhanced Image Fig.(b)Segmented Image

3.2.4 Feature Extraction

This stage is an important stage that uses algorithms and techniques to detect and isolate various desired portions or shapes of a given image. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant, then the input data will be transformed into a reduced representation set of features. The basic characters of feature are area, perimeter and eccentricity. These are measured in scalar. These features are defined as follows:

A) Area: It is the scalar value that gives actual number of overall nodule pixel in the extracted ROI. Transformation function creates an array of ROI that contains pixels with 255 values.

Area = A = ($A_{i,j}$, X ROI[Area] = I , Y ROI[Area] = j)

Where, i, j are the pixels within the shape. ROI is region of interest. X ROI[ ] is vector contain ROI x position, Y ROI[ ] is vector contain ROI y position [6].

B) Perimeter: It is a scalar value that gives actual number of the nodule pixel. It is the length of extracted ROI boundary. Transformation function create array of edge that contain pixel with 255 values that have at least one pixel which contain 0 values [6].

$$\textbf{Perimeter} = \bar{P} = (P_{i,j}\ ,\ \bar{X}\ \text{edge}[\bar{P}\ ] = i,\ \bar{Y}\ \text{edge}[\bar{P}\ ] = j)$$

Where, $\bar{X}$ edge [ ] and $\bar{Y}$ edge[ ] are vectors represent the co-ordinate of the i^{th} and j^{th} pixel forming the curve, respectively.

C) Eccentricity: This metric value is also called as roundness or circularity or irregularity complex (I) equal to 1 only for circular and it is less than 1 for any other shape.

$$\textbf{Eccentricity} = \frac{\textbf{Length of Major Axis}}{\textbf{Length of Minor Axis}}$$

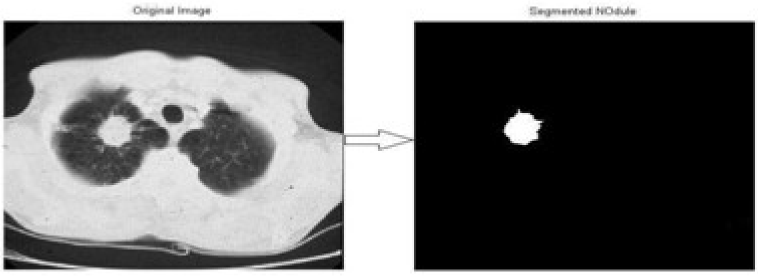

Fig.3.2.4 Segmented Nodule for Feature Extraction.

Features estimated for separated nodule of given sample image above as shown in fig.3.2.4 has been found as follows:

1) **Area:** It is the simplest property and by its given size. Therefore, it is the total number of white pixels in the extracted area.

2) **Perimeter:** It is another simple property defined by the perimeter of the region. It is the length of extracted ROI boundary.

3) **Eccentricity:** It is used to decide the shape or circularity of the object.

Area: 2291 Perimeter: 221 Eccentricity: 0.8289 Stage 3 Detected

3.2.5 Classification

Support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification. The basic SVM takes a set of input data and for each given input, predicts which of two classes forms the input, making it a non-probabilistic binary linear classifier. SVM uses a kernel function which maps the given data into a different space; the separations can be made even with very complex boundaries. The different types of kernel function include polynomial, RBF, quadratic, Multi-Layer Perceptron (MLP). Each kernel is formulated by its own parameters like γ, σ, etc. Figure 3.2.5 shows maximum margin hyper planes. The original hyper plane algorithm is a way to create non linear classifier by applying the kernel trick to maximum margin hyper planes.

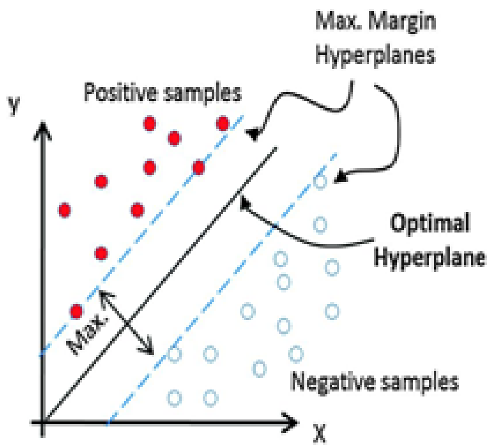

Figure 3.2.5: Maximum Margin Classifier

IV. Staging of Lung Cancer

Lung nodules are the smallest growths in the lung that measure between 5 mm to 25 mm in size. In abnormal images nodule size is greater than 25 mm,

Primary Tumor(T)	Criteria
T1	< 3cm in diam; T 1a <= 2cm; T 1b > 2cm <= 3cm
T2	> 3cm <= 7cm; T 2a > 3cm <= 5cm; T 2b > 5cm <= 7cm
T3	> 7cm
T4	Any Size greater than above

Table 4.1: Cancer Stage Criterion

Staging involves evaluation of a cancer size and its penetration into surrounding tissues as well as presence or absence of metastasis in the lymph nodes or other organs [7]. Stages from I to IV in order of severity:

- Stage I : cancer is confined to the lung
- Stage II and III : cancer is confined to the chest
- Stage IV, cancer has spread from the chest to other parts of the body.

According to the medical field, non-small cell lung cancer is staged using TNM system (T for extent of primary tumor, N for regional lymph node involvement and M for metastasis). Table 4.1 shows the criteria decided by doctors for the classification of lung cancer stages. Different types of stages are as shown in following figures according to their parameters

1. For stage: 1 (the pictorial representation is shown as below):

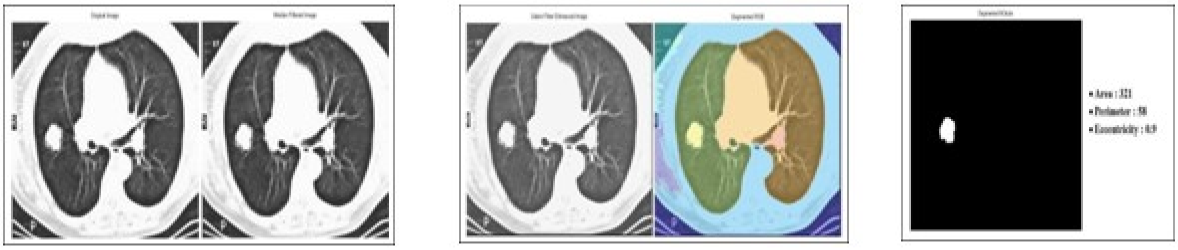

Fig4.1.1 a) Original and median filtered Image b) Enhanced and segmented Image c) Binary Image

2. For Stage: 2 (the pictorial representation is shown as below):

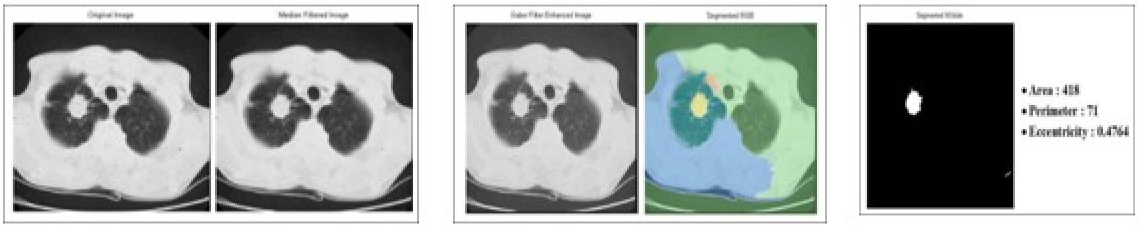

Fig4.1.2 a) Original and median filtered Image b) Enhanced and segmented Image c) Binary Image

3. For Stage: 3 (the pictorial representation is shown as below):

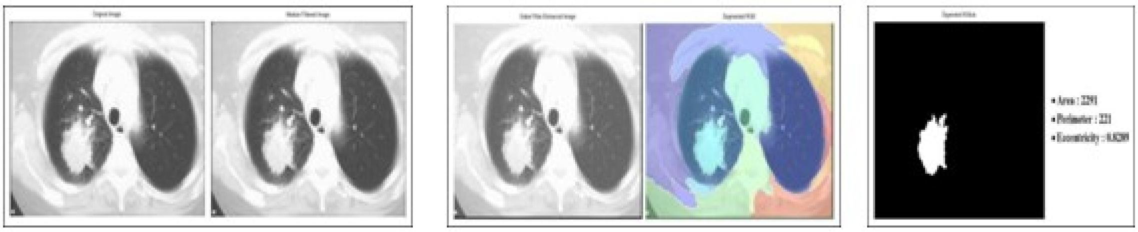

Fig4.1.3 a) Original and median filtered Image b) Enhanced and segmented Image c) Binary Image

4. For Stage: 4 (the pictorial representation is shown as below):

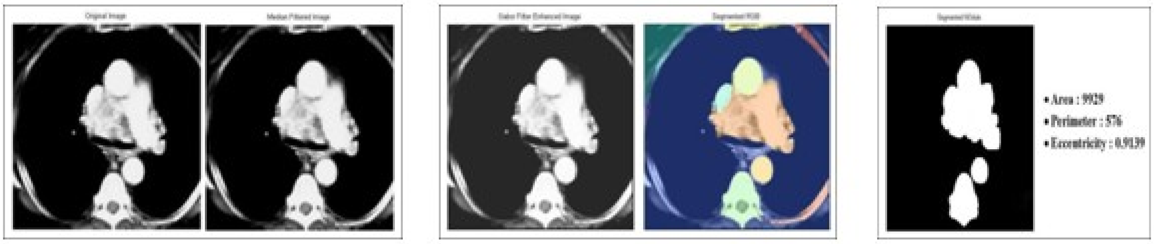

Fig4.1.4:a) Original and median filtered Image b)Enhanced and segmented Image c)Binary Image

3.2.6 Gaussian Radial Basis Function

Radial basis functions have received significant attention, most commonly with a Gaussian of the form,

$$\mathbf{k}(\mathbf{x},\mathbf{x}') = \mathbf{exp}\left(-\frac{\mathbf{kx}'-\mathbf{k}^2\mathbf{x}'}{2\sigma^2}\right)$$

Typically a method of clustering is first employed to select a subset of centre. An attractive feature of the SVM is that this selection is implicit, with each support vectors contributing one local Gaussian function, centred at that data point. Basic equation for support vector machine is :

$$\mathbf{Y = W^{T} \times X + b}$$

Where, Y= Output Class; W= Hyperplane; X= Input Feature; b= Bias

V. Conclusion and Future Scope

Lung cancer is one of the most dangerous diseases in the world. Correct Diagnosis and early detection of lung cancer can increase the survival rate. The present techniques include study of X-ray, CT scan, MRI, PET images. The expert physicians diagnose the disease and identify the stage of cancer by experience. The treatment includes surgery, chemotherapy, radiation therapy and targeted therapy. These treatments are lengthy, costly and painful. Hence, an attempt is made to atomize this procedure to detect the lung cancer using image processing techniques. CT scan images are acquired from various hospitals. These images include less noise as compared to X-ray and MRI images. An image improvement technique is developing for earlier disease detection and treatment stages; the time factor is taken in account to discover the abnormality issues in target images. The CT captured images are processed. The region of interest i.e., tumor is identified accurately from the original image. Gabor filter and watershed segmentation gives best results for pre-processing stage. From the extracted region of interest, three features are extracted i.e., area, perimeter and eccentricity. These three features help to identify the stage of lung cancer. The results indicate that the tumors are of different dimensions. By measuring the dimensions of the tumor the lung cancer stage can be detected accurately using the proposed method. The results show good potential for lung cancer detection at early stage.

Also for classification purpose, Support Vector Machines are an attractive approach to data modeling. They combine generalization control with a technique to address the curse of dimensionality. The kernel mapping provides a unifying framework for most of the commonly employed model architectures, enabling comparisons to be performed. In classification problems generalization control is obtained by maximizing the margin, which corresponds to minimization of the weight vector in a canonical framework.

For future work, we can implement this technique on some more images. Increasing the number of images used for the process, can improve the accuracy. Also MRI, X-ray, PET images can be considered for this technique. Comparison can be done for all these images. So one can justify which types of images gives better result for lung cancer detection

Bibliography

[1]. Ilya Levner, Hong Zhangm ,"Classification driven Watershed segmentation ", IEEE TRANSACTIONS ON IMAGE PROCESSING VOL. 16, NO. 5, MAY 2007
[2]. Anita chaudhary, Sonit Sukhraj Singh,"Lung Cancer Detection on CT Images Using Image Processing ", International transaction on Computing Sciences, VOL 4, 2012
[3]. 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,DEC-2001
[4]. Disha Sharma, Gagandeep Jindal,"Identifying Lung Cancer Using Image Processing Techniques ", International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2011), vol: 17, pp: 872-880,2011
[5]. Nguyen, H. T.,et al,"Watersnakes: Energy-Driven Watershed Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 25, Number 3, pp.330-342, March 2003
[6]. Suzuki K., "False-positive Reduction in Computer-aided Diagnostic Scheme for Detecting Nodules in Chest Radiographs",Academic Radiology, Volume 13, Number 10, pp.10-15, February 2005
[7]. D. Lin and C. Yan, "Lung nodules identification rules extraction with neural fuzzy network", IEEE, Neural Information Processing, vol. 4, (2002)
[8]. 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).
[9]. R. A. Blechschmidt, R. Werthschtzky, and U. Lrcher, "Automated CT image evaluation of the lung: A morphology-based concept", IEEE Transaction on Medical Imaging., vol. 20, no. 5, pp. 434442, May 2001
[10]. S.Sivkumar,"Lung Nodule Detection Using Fuzzy Clustering and Support Vector Machine ", International Journal of Engineering and Technology Volume 5, No. 1, ISSN: 0975-4024, March 2013.
[11]. Hala Al-shamlan, Ali El Zaart, "Feature Extraction Values for Breast cancer Images",International Conference on Bioinformatics and Biomedical Technology,May 25, 2005
[12]. Mokhled S. AL-TARAWNEH, "Lung Cancer Detection Using Image Process-ing Techniques",Leonardo Electronic Journal of Practices and Technologies,(ISSN 1583-1078), Issue 20, p. 147-158, January-June 2012.
[13]. The DICOM Standards Committee. DICOM homepage. http://medical.nema.org/, September 2004.
[14]. Shiying Hu, Eric A. Hoffman, Joseph M. Reinhardt,"Automatic Lung Segmentation for Accurate Quantization of Volumetric X-Ray CT Images", IEEE TRANS-ACTIONS ON MEDICAL IMAGING, VOL. 20, NO. 6, JUNE 2001
[15]. Hye Suk Kim,Hyo-sun Yoon, "Automatic Lung Segmentation in CT Images using Anisotropic Diffusion and Morphology Operation", Seventh International IEEE Conference on Computer and Information Technology, 0-7695-2983-6/07, july 2007.
[16]. Rushin Shojaii, Javad Alirezaie,"Automatic Lung Segmentation in CT Images using Watershed Transform",International Conference on Biomedical Technology, IEEE , 0-7803-9134-9, May 2005.

AUTHORS

Mr. Vijay A. Gajdhane, Dept. of Electronics and Tele-communication Engineering, TPCT's College Of Engineering, Osmanabad- 413501, Maharashtra, India

Prof. Deshpande L.M, M.E. (Electronics) **,** Dept. of Electronics and Tele-communication Engineering, TPCT's College Of Engineering, Osmanabad- 413501, Maharashtra, India