ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Detection of Different Stages of Lungs Cancer in CT-Scan Images using Image Processing Techniques

Priyanka Basak[1], Asoke Nath[2]

M.Sc. Student, Dept. of Computer Science, St. Xavier's College (Autonomous), Kolkata, India[1]

Associate Professor, Dept. of Computer Science, St. Xavier's College (Autonomous), Kolkata, India[2]

ABSTRACT: The most trivial cancer seen in both men and women cancer is *Lung cancer* (small cell and non-small cell). The survival rate among people increased by early diagnosis of lung cancer. Due to structure of cancer cell the prediction of lung cancer is the most challenging problem, where most of the cells are overlay each other. Recently, the image processing mechanisms are used extensively in medical areas for earlier detection of stages and its treatment, where the time factor is very crucial. Detection of cancer cell in lung cancer patients in time will cause overall 5-year survival rate increasing from 14 to 49%. In the present paper the authors have proposed Lung cancer detection system using image processing technique to classify the presence of cancer cells in lung and its stages from the CT-scan images using various enhancement and segmentation techniques, aiming for accuracy in result.

KEYWORDS: Lung cancer, extraction, segmentation, accuracy, Watershed, Thresholding, Masking, Classification

I. INTRODUCTION

Lung Cancer, the undisciplined swelling of abnormal cells, which generally initiate in one or both lungs, mostly in the line of air channel. Those abnormal cells never grow to normal or healthy lung tissue; their rapid production cause tumors. Lung cancer has the second most fatality rate among other categories of cancer. Even after diagnosis it has smallest survival rate, thereby continuously increasing the death rate yearly. Growth of lung cancer during diagnosis is related to its survival rate. But if the cancer cells been diagnosed in its early stages ones survival rate increases [1]. Cancer cells can borne from the lungs in blood, or lymph fluid surrounding the lung tissue. Lymph floods through lymphatic vessels, which duct into lymph nodes placed in the lungs and in the centre of the chest. The lung cancer generally spread within the middle of the chest due to the usual outflow of lymph from lungs is toward the centre of the chest. Generally it is grouped into Non-Small Cell Lung Cancer and Small Cell Lung Cancer. Depending on the cellular characteristics these assign the lung cancer types. As the stages, generally there are four stages of lung cancer; I - IV. The Stages are based on size and location of tumor and location of lymph node. Currently, CT are found more efficient than plain chest x-ray in detecting and diagnosing the lung cancer. Discussing various techniques of image processing to detect the stages of lung cancer more accurately [2].

II. STAGES OF LUNG CANCER

A. *Why is it important to know the stage of lung cancer?*

Finding out the stage of lung cancer is important for two reasons. First, staging of lung cancer helps decide which therapy (or therapies) should be used. Second, lung cancer staging tells how much cancer has spread. Knowing the stage of cancer helps the health care team know the risks versus the benefits of different procedures and treatments. Treatments that are good for one stage may not be helpful for another stage, and in fact can be harmful. For example, if cancer has spread outside the lung, surgery to remove part of the lung may not improve the chance of living longer and may cause unnecessary harm [3].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

B. *How does staging differ between small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC)?*

SCLC is divided into "limited" and "extensive" stages. Limited stage SCLC occurs when the lung cancer is limited to one side of the chest. Extensive stage occurs when the lung cancer has spread to the other side of the chest or to other organs such as the liver or brain. NSCLC staging uses the TNM system. The initials TNM stand for the size and location of the **T**umor, the location of cancer in the lymph **N**odes and if and where the cancer has spread (called **M**etastases). The staging system can be complicated but there are general rules that are used.

- The **T** number increases as the tumor gets bigger and how close it is to major structures in the chest like large airways in the lungs. The T number also increases if the cancer is growing into structures like the heart, major blood vessels, or tissues outside the lung.
- The L number says whether your cancer has spread to the lymph nodes. Lymph nodes are part of your immune system and cancer cells can spread into the lymph system. Usually, if the cancer has spread, it spreads to the nodes closest to the main tumor and then goes further away. The lymph node is rated as N1 if the cancer is found in the lymph nodes on the same side as the main tumor in the lung. A rating of N2 means cancer has spread to the middle part of the chest (called the mediastinum). A rating of N3 means the cancer has spread to the opposite lung or outside the chest. [A rating of N0 means the cancer has not been found in the lymph nodes.]
- The **M** says that metastases (spread of cancer) has happened throughout the body and is growing in other tissues or organs. Lung cancer often spreads to the brain, bones, adrenal glands, liver or other areas. The M stage is based on if the cancer has spread and where it has spread [3].

III. METHODOLOGY

A. *Image Acquisition*

At first, from the lung cancer patient the CT scan image is acquired. The main benefit of using computed tomography images is less distortion. CT images have low noise in contrast to X-ray and MRI images; henceforth they are used for developing the technique. 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. Fig.2 shows a typical CT image of a normal chest and a lung cancer patient used for analysis. The acquired images are in raw form and have a lot of noise, thus to improve the contrast, clarity and to separate the background noise, it is required to pre-process the images. Thus, different methods like smoothing, enhancement, segmentation are applied to get image in necessary form [4]. Fig.1 shows the steps of Lung Cancer Detection System.

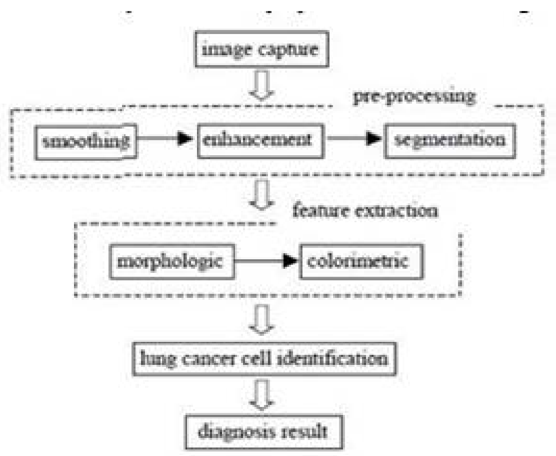

Fig.1 .Lung Cancer Detection System

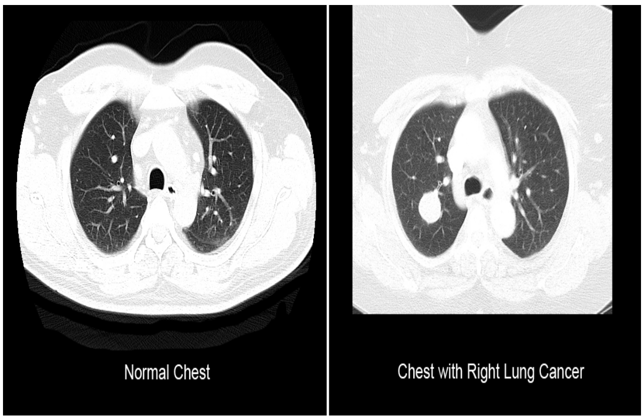

Fig.2 .Acquired CT-Scan images

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

B. *Image Pre-Processing*

This process is sub-divided into 3-steps- '*Smoothing*', '*Enhancement* 'and '*Segmentation*'.

Smoothing: It subdue the noise or further small instabilities in the image; equivalent to the suppressions of high frequencies in the frequency domain. All sharp edges that bear essential data about the image are also blurred during smoothing. Many methods, irrespective 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 dissimilarities of an image. One of the most common problem in image processing is White noise. Even a high resolution Photo is bound to have some noise in it. For a high resolution picture a simple box blur may be sufficient. Mainly the idea of neighborhood filter is to calculate pixel weights, depending on their colors similarity.

We describe two methods: the 'Median Filter' and 'Weiner filters'.

- ***Median Filter*:**

It is a non-linear operation and is often used in image processing to reduce salt and pepper noise. In general, a great deal of high spatial frequency detail allowed to pass by the median filter, 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 comprises of the median value in the m x n neighborhood around the corresponding image pixel [2]. The image is padded with 0's on the edges by Medfilt2, 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.

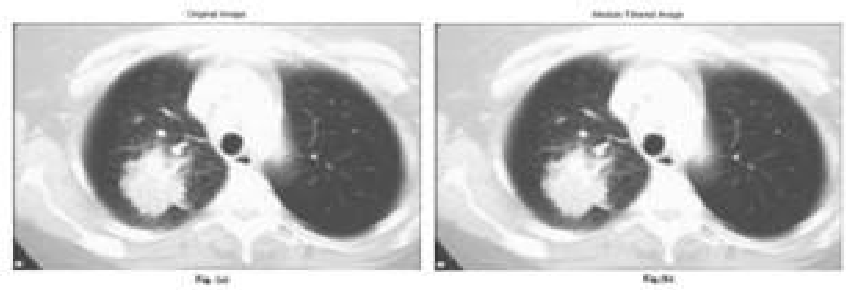

Fig.3. (a) Original Image (b) Median Filtered Image

- ***Weiner Filter*:**

It is the MSE-optimal stationary linear filter, for removing additive noise and blurring from degraded images. The inverse filtering is a restoration procedure for deconvolution, i.e., by inverse filtering or generalized inverse filtering image recovery is possible as image blurring is done by a common low pass filter. However, inverse filtering have sensitivity to additive noise. The approach of reducing one degradation at a time let us to develop a restoration algorithm for each type of degradation and simply combine them. An optimal tradeoff between inverse filtering and noise smoothing is executed by the Weiner filter [2].

It is optimal in terms of the mean square error, i.e. it minimizes the overall mean square error in the process of inverse filtering and noise smoothing. It is a linear estimation of the original image, based on a stochastic framework approach. The orthogonality principle implies that the Wiener filter in Fourier domain can be expressed as follows:

$$W(f_1, f_2) = \frac{H^*(f_1, f_2) S_{xx}(f_1, f_2)}{|H(f_1, f_2)|^2 S_{xx}(f_1, f_2) + S_{\eta\eta}(f_1, f_2)},$$

where $S_{xx}(f_1, f_2)$, $S_{\eta\eta}(f_1, f_2)$ are respectively power spectra of the original image and the additive noise, and $H(f_1, f_2)$ is the blurring filter. It is easy to see that the Wiener filter has two separate part, an inverse filtering part and a noise smoothing part. It not only performs the deconvolution by inverse filtering (highpass filtering) but also removes the noise with a compression operation (lowpass filtering).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Implementation: To implement the Wiener filter in practice we have to estimate the power spectra of the original image and the additive noise. For white additive noise the power spectrum is equal to the variance of the noise. To estimate the power spectrum of the original image many methods can be used. A direct estimate is the periodogram estimate of the power spectrum computed from the observation:

$$S_{yy}^{per} = \frac{1}{N^2}[Y(k,l)Y(k,l)^*]$$

where **Y(k,l)** is the DFT of the observation. The advantage of the estimate is that it can be implemented very easily without worrying about the singularity of the inverse filtering. Another estimate which leads to a cascade implementation of the inverse filtering and the noise smoothing is

$$S_{xx} = \frac{S_{yy} - S_{\eta\eta}}{|H|^2},$$

which is a straightforward result of the fact: $S_{yy} = S_{\eta\eta} + S_{xx}|H|^2$. The power spectrum S_{yy} can be estimated directly from the observation using the periodogram estimate. This estimate results in a cascade implementation of inverse filtering and noise smoothing:

$$W = \frac{1}{H}\frac{S_{yy}^{per} - S_{\eta\eta}}{S_{yy}^{per}}.$$

The disadvantage of this implementation is that when the inverse filter is singular, we have to use the generalized inverse filtering. People also suggest the power spectrum of the original image can be estimated based on a model such as the $1/f^{\alpha}$ model [4].

Enhancement: Enhancement is a technique, 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. Unfortunately, there is no general theory for determining what “good” image enhancement is when it comes to human perception. If it looks good, it is good. However, when image enhancement techniques are used as preprocessing tools for other image processing techniques, the quantitative measures can determine which techniques are most appropriate [4].

In the image enhancement stage we used the following three techniques: 'Gabor filter', 'Auto-enhancement' and 'Fast Fourier transform' techniques.

- ***Gabor filter:***

Image presentation based on Gabor function integrates an outstanding local and multiscale decomposition in terms of logons that are simultaneously (and optimally) localization in space and frequency domains. Gabor filter known to be a linear filter, whose impulse response is specified by a harmonic function multiplied by a Gaussian function. Due to the multiplication-convolution property (Convolution theorem, the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the Harmonic function and the Fourier transform of the Gaussian function [5].

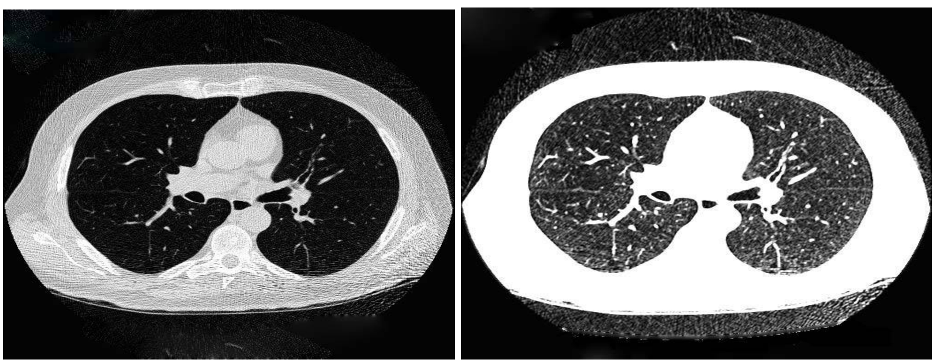

Fig.4 - (a) Original Image (b) Enhanced by Gabor
The result of applying Gabor enhancement technique

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

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**:**
$G(x, y) = \exp(-x'^2 + \gamma^2 y'^2) \cos(2\pi x') 2\sigma 2\lambda$
Where, **$x' = x\cos\theta + y\sin\theta$; $y' = -x\sin\theta + y\cos\theta$** [2]. 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.5.

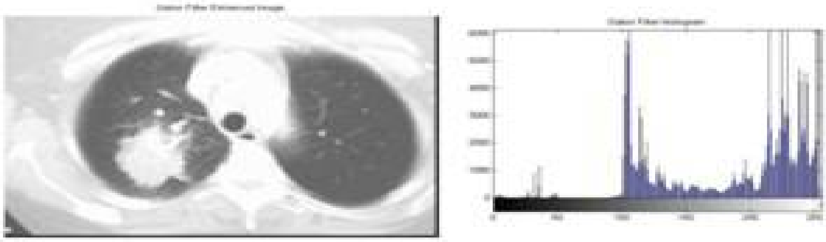

Fig.5. (a) Filtered Image (b) Histogram

- ***Fast Fourier Transform:***

FFT technique operates on Fourier transform of a given image. The frequency domain is a space, where each value of image at image position F specifies the amount that the intensity values in image "I" differ over a specific distance related to F. Here, Fast Fourier Transform is used in image filtering i.e. Enhancement. Fig.6 visualize the effect of applying FFT on original images, where FFT method has an enhancement of 27.51%.

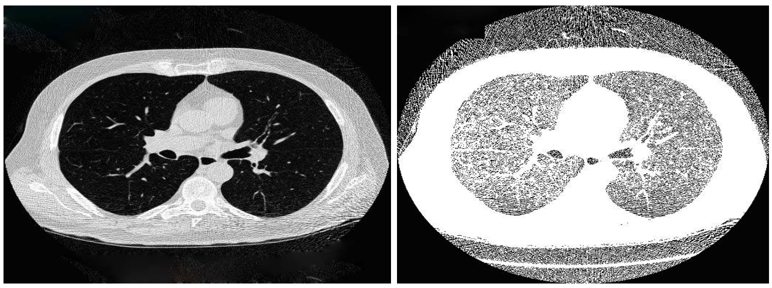

Fig.6. (a) Original Image (b) Enhanced by FFT
Auto enhancement technique using FFT

Subject	Auto Enhancement	Gabor Filter	FFT Filter
Sub1	37.95	80.975	27.075
Sub2	47.725	80	36.825
Sub3	36.825	79.5	25.625
Sub4	34.775	81.8	25.175
Sub5	32.85	81.4	22.85
Final Average	38.025	80.735	27.51

Table.1. A comparison of the techniques used for image enhancement. According to the values it can be concluded that the Gabor Enhancement is the most suitable technique for image enhancement.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Website: www.ijircce.com
Vol. 5, Issue 5, May 2017

Segmentation: Image segmentation is most essential procedure for image analysis subsequent tasks. It divides the image into multiple segments of constituent regions or objects. Segmentation of medical images in 2D, slice by slice has many useful applications for the medical professional such as: visualization and volume estimation of objects of interest, detection of abnormalities (e.g. tumors, polyps, etc.), tissue quantification and classification, and more. [6]. Thegoal of segmentation is to simplify and/or change the representation of the image into something that is 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 visual characteristics [7]. This results in a set of segments that collectively cover the entire image, or a set of contours extracted from the image (edge detection). All pixels in a given region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s).

Segmentation algorithms are based on one of two basic properties of intensity values: discontinuity and similarity. The first category is to partition the image based on abrupt changes in intensity, such as edges in an image. The second category is based on partitioning the image into regions that are similar according to a predefined criterion. 'Histogram thresholding' approach falls under this category.

- ***Thresholding approach:***

Thresholding is one of the most important method to do image segmentation. The segmented image obtained from thresholding has the advantages of smaller storage space, fast processing speed and ease in manipulation, compared with gray level image which usually contains 256 levels. And thus, thresholding techniques is of attention during the past 20 years [8]. Thresholding is a non-linear operation, converts a gray-scale image into a binary image where the two levels are assigned to pixels that are below or above the specified threshold value. Otsu's method uses (gray thresh) function to compute global image threshold. Otsu's method is based on threshold selection by statistical criteria. Otsu suggested minimizing the weighted sum of within-class variances of the object and background pixels to establish an optimum threshold. Recalling that minimization of within-class variances is equivalent to maximization of between-class variance. This method gives satisfactory results for bimodal histogram images. Threshold values based on this method will be between 0 and 1, after achieving the threshold value; image will be segmented based on it. Fig.7 shows the result of applying thresholding technique.

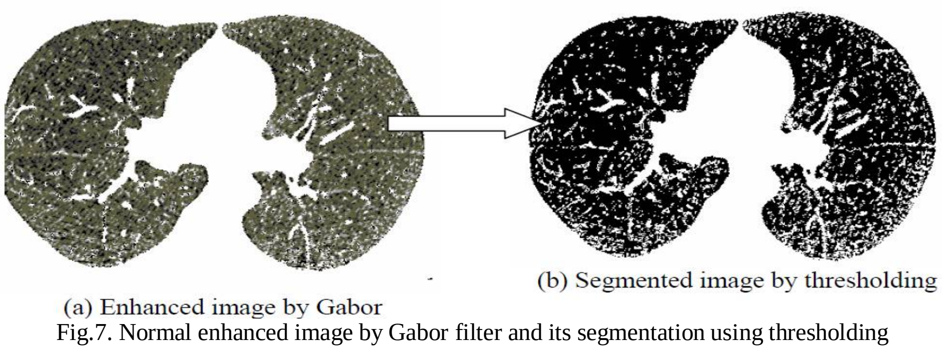

Fig.7. Normal enhanced image by Gabor filter and its segmentation using thresholding

- ***Marker-driven watershed segmentation***

This technique extracts seeds that indicate the presence of objects or background at specific image locations. Marker-controlled watershed approach has two types: External associated with the background and Internal associated with the objects of interest. Image Segmentation using the watershed transforms works well if we can identify or "mark" foreground objects and background locations, to find "catchment basins" and "watershed ridge lines" in an image by treating it as a surface where light pixels are high and dark pixels are low. Marker-controlled watershed segmentation follows this basic procedure –

1) Compute a segmentation function. This is an image whose dark regions are the objects you are trying to segment.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

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.

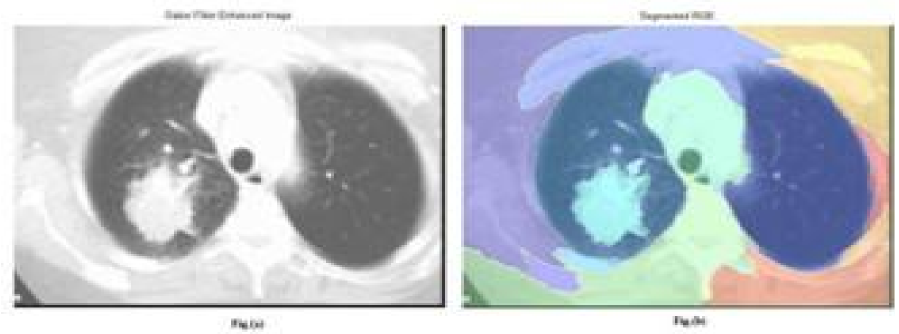

Fig.8. (a) Enhanced Image (b) Segmented Image

Subject	Thresholding	Watershed Filter
Sub1	81.625	85.375
Sub2	82.2	85.25
Sub3	82.125	85.55
Sub4	81.725	84.75
Sub5	81.5	84.9
Final Average	81.835	85.165

Table.2. Image Segmentation Experimental Result

According to the experimental subjective assessment during the segmentation stage as shown in table 2, Marker-Controlled Watershed Segmentation approach has more accuracy (85.165%) and quality than Thresholding approach (81.835%).

C. *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:

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 [9].

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 [9].

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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Where, X edge [ ] and Y edge[ ] are vectors represent the co-ordinate of the i_{th} and j_{th} pixel forming the curve, respectively.

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.

$$Eccentricity = Length\ of\ Major\ Axis \div Length\ of\ Minor\ Axis$$

Features estimated for separated nodule of given sample image above as shown in Fig.8 has been found as follows:

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

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

Eccentricity is used to decide the shape or circularity of the object. Area: 2291 Perimeter: 221 Eccentricity: 0.8289 Stage 3 Detected.

To predict the probability of lung cancer presence, the following two methods are used:

'Binarization' and 'Masking', both methods are based on facts that strongly related to lung anatomy and information of lung CT imaging.

- ***Binarization Approach:***

This approach depends on the fact that the number of black pixels is much greater than white pixels in normal lung images, so we started to count the black pixels for normal and abnormal images to get an average that can be used later as a threshold, if the number of the black pixels of a new image is greater that the threshold, then it indicates that the image is normal, otherwise, if the number of the black pixels is less than the threshold, it indicates that the image in abnormal. The threshold value that is used in this research is 17178.48 and the True acceptance rate (TAR) is (92.86%) and False acceptance rate (FAR) is (7.14%). Figure 9 shows the binarization method procedure and Figure 10 shows binarization check method flowchart [5].

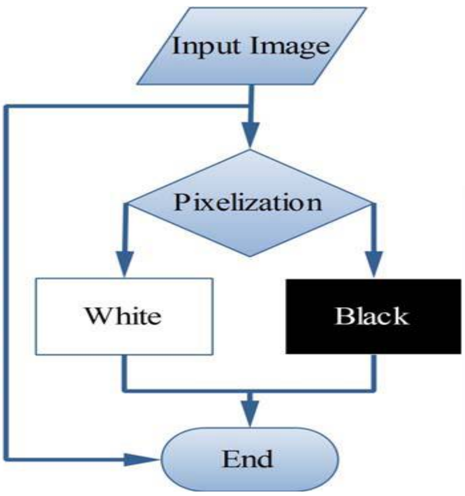

Fig.9. Binarization method procedure

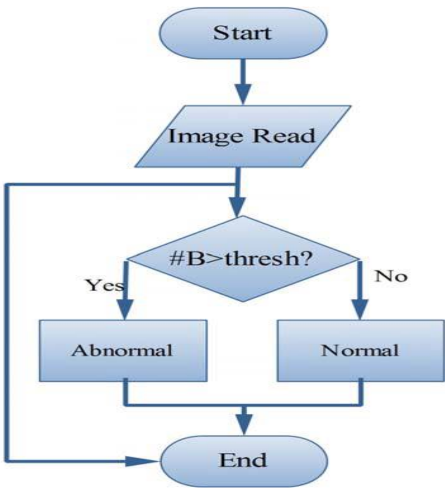

Fig.10. Binarization check method flowchart

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

- ***Masking Approach:***

This approach depends on the fact that the masses are appeared as white connected areas inside ROI (lungs), as they increase the percent of cancer presence increase. The appearance of solid blue colour indicates normal case while appearance of RGB masses indicates the presence of cancer, the TAR of this method is (85.7%) and FAR has (14.3%).

D. *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. Fig.11 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 [2].

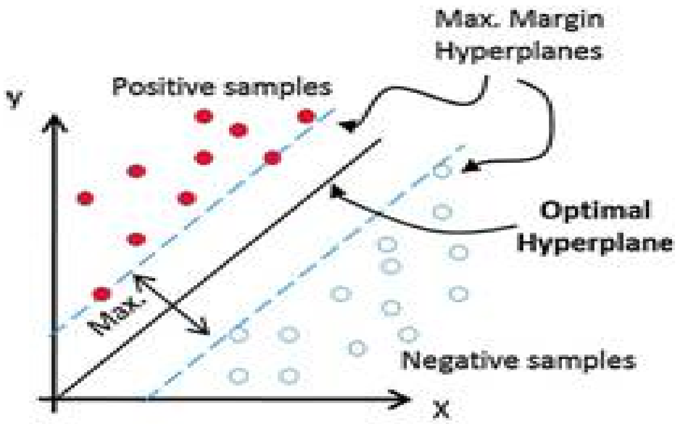

Fig.11. Maximum Margin Classifier

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,
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 [10]. 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. [2]

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 3 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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

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.3. Cancer Cell Criterion

V. RESULTS

In this paper Lung CT images are used, obtained from a specialist medical imaging centre. The image smoothing is done using Median filter, enhancement is done by Gabor filter. After enhancement, segmentation is done using Watershed lung region or (ROI) is extracted. The steps applied on CT images are shown –

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

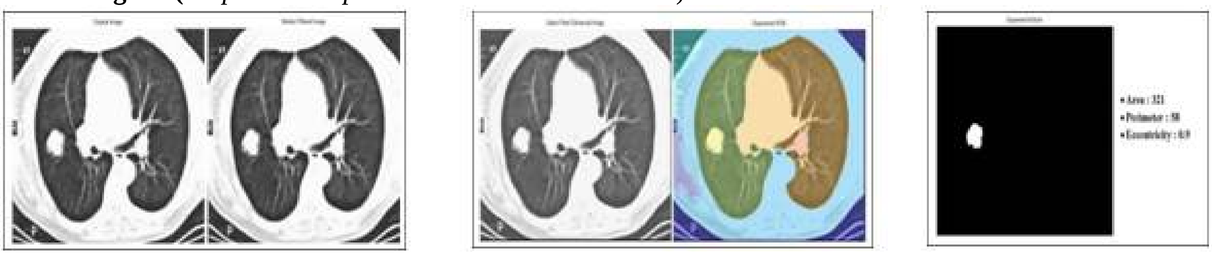

Fig.12.a) Original and median filtered Image b) Enhanced and segmented Image c) Binary Image

Cancer is in initial stage, small tumor detected. Ex- for Fig.12, the Area is 321; Perimeter is 58; Eccentricity is 0.9. So, it indicates Stage 1 is detected.

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

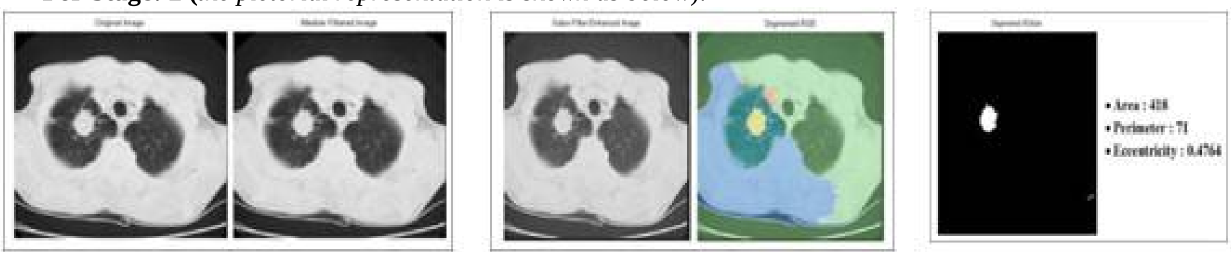

Fig.13.a) Original and median filtered Image b) Enhanced and segmented Image c) Binary Image

Cancer is confined to the Lung, Stage 2 is detected. Ex- for Fig.13, the Area is 428; Perimeter is 71; Eccentricity is 0.4764. So, it indicates Stage2 is detected.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

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

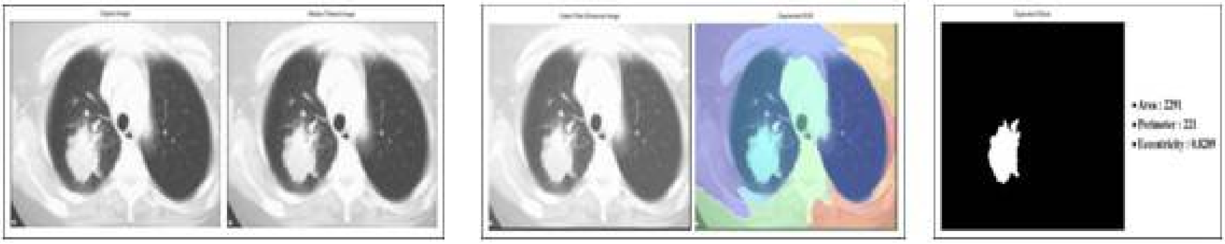

Fig.14.a) Original and median filtered Image b) Enhanced and segmented Image c) Binary Image

Cancer has been confined to chest, Stage 3 is detected. Ex- for Fig14, Area is 2291; Perimeter is 221; Eccentricity is 0.8289. So, it indicates Stage 3 Detected.

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

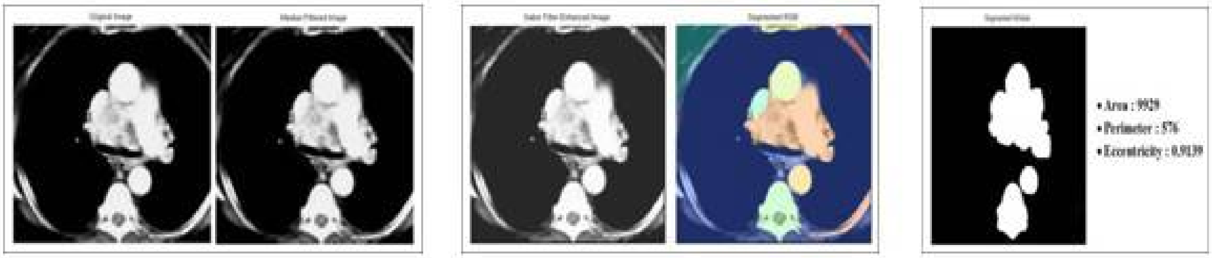

Fig.15.a) Original and median filtered Image b) Enhanced and segmented Image c) Binary Image

Cancer has been spread from chest to other parts of the body, Stage 4 is detected. Ex- for Fig.15, Area is 9929; Eccentricity is 0.9199. So, it indicates Stage 4 Detected.

VI. LIMITATIONS

The analysis process of the three-dimensional images obtained by this technique is still suffering from some limitations caused by the scanning process. The limitation of the PET scan is that it can be time consuming. Using CT scans in detecting lung cancer has a number of limitations with high false-positive rate, because it detects a lot of non-cancerous nodules and it misses many small cancer nodules.

The techniques which were used to analyse these images have a number of limitations namely: the high number of false negatives representing the missed cancer cells, and the high number of false positives representing cells classified as cancerous, resulting in putting a patient through unnecessary radiation and surgical operations. In addition, most techniques fail to consider the outer pixels which may sometimes represent a class in themselves. Moreover, the pre-processing techniques need further enhancement to discard the debris cells in the background of the images, and to remove all noise from the images.

VII. 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. Image improvement techniques are developing for earlier disease detection and treatment stages; the time factor was taken in account to discover the abnormality issues in target images. Image quality and accuracy are the important factors. Techniques presently used include study of X-ray, CT scan, PET, MRI images. Treatments include surgery, chemotherapy, radiation therapy and targeted therapy. These treatments are lengthy, costly and painful, thus, an attempt is made to atomize this procedure to detect the lung cancer using image processing techniques. CT scan images are acquired from different hospitals. These images have less noise as compared to X-ray and MRI images. . The CT captured images are preprocessed. The region of interest i.e., tumor is identified with

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

accuracy 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. Feature extraction techniques can be done by binarization or masking approach. By measuring the dimensions of the tumor the lung cancer stage can be detected accurately. For classification purpose, Support Vector Machines is one of the best approach to data modeling. They combine generalization control with a technique to address the curse of dimensionality. A unifying framework for most of the commonly employed model architectures, enabling comparisons to be performed by using the kernel mapping. In classification process the problems in generalization control is obtained by maximizing the margin, which corresponds to minimization of the weight vector in a canonical framework. For Future work, increasing the number of images used for the process, can improve the accuracy.

REFERENCES

1. Ilya Levner, Hong Zhangm ,"Classification driven Watershed segmentation ", IEEE TRANSACTIONS ON IMAGE PROCESSING VOL. 16, NO. 5, MAY 2007
2. Mr.Vijay A.Gajdhane, Deshpande, "Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques"IOSR Journal of Computer Engineering (IOSR-JCE),e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. III , Sep – Oct. 2014
3. Am J Respir Crit Care Med Vol. 182, P1-P2, 2010. Online Version Reviewed September 2013, ATS Patient Education Series © 2010 American Thoracic Society
4. http://www.owlnet.rice.edu/~elec539/Projects99/BACH/proj2/wiener.html (lastaccess:4/30/2017||4:52am)
5. Mokhled S. AL-TARAWNEH, "Lung Cancer Detection Using Image Processing Techniques", Leonardo Electronic Journal of Practices and Technologies,(ISSN 1583-1078), Issue 20, p. 147-158, January-June 2012.
6. Venkateshwarlu K., Image Enhancement using Fuzzy Inference System, in Computer Science & Engineering, Master thesis, 2010.
7. Shapiro L.G., Stockman G.C., Computer Vision: Theory and Applications, Prentice Hall,2001.
8. Huang Q., Gao W., Cai W., Thresholding technique with adaptive window selection for uneven lighting image, Pattern Recognition Letters, Elsevier, 2004, 26, p. 801-808.
9. 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.
10. Levner I., Zhang H., Classification-Driven Watershed Segmentation, IEEE Transactions on Image Processing, 2007, 16(5), 1437-45.
11. Non-Small Cell Lung Cancer, Available at: http://www.katemacintyrefoundation.org/ pdf/non-small-cell.pdf, Adapted from National Cancer Institute (NCI) and Patients Living with Cancer (PLWC), 2007, (accessed July 2011).
12. Tarawneh M., Nimri O., Arqoub K., Zaghal M., Cancer Incidence in Jordan 2008, Available at: http://www.moh.gov.jo/MOH/Files/Publication/Jordan%20Cancer%2 0Registry_2008%20Report_1.pdf, 2008, (accessed July 2011).
13. Lung Cancer Database, Available at: https://eddie.via.cornell.edu/cgibin/ datac/signon.cgi, (accessed July 2011).
14. Gonzalez R.C., Woods R.E., Digital Image Processing, Upper Saddle River, NJ Prentice Hall, 2008.
15. Cristobal G., Navarro. R., Space and frequency varient image enhancment based in Gabor representation, Pattern Recognition Letters, Elsevier, 1994, 15, p. 273-277.
16. Krishan A., Evaluation of Gabor filter parameters for image enhancement and segmentation, in Electronic Instrumentation and Control Engineering, Master. Punjab: Thapar University, 2009, p. 126.
17. Nunes É.D.O., Pérez M.G., Medical Image Segmentation by Multilevel Thresholding Based on Histogram Difference, presented at 17th International Conference on Systems, Signals and Image Processing, 2010.
18. Venkateshwarlu K., Image Enhancement using Fuzzy Inference System, in Computer Science & Engineering, Master thesis, 2010.
19. Anita chaudhary, Sonit Sukhraj Singh,"Lung Cancer Detection on CT Images Using Image Processing ", International transaction on Computing Sciences, VOL 4, 2012
20. 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
21. 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.
22. The DICOM Standards Committee. DICOM homepage. http://medical.nema.org/, September 2004.

BIOGRAPHY

Priyanka Basak is a M.Sc. Computer Science student from Department of Computer Science, St. Xavier's College (Autonomous), Kolkata, is currently in 4th Semester; interested in Web Technologies, Cryptography, Steganography, and Image Processing with successful experience in Tech Entrepreneurship.

Dr. Asoke Nath is Associate Professor in the Department of Computer Science, St. Xavier's College (Autonomous), Kolkata. Dr. Nath is involved in research work in Cryptography and Network Security, Steganography, Green Computing, Mathematical modelling of social networks, Big data analytics, Cognitive Radio, Data Science, e-learning, MOOCs etc. He has published more than **215** papers in Journals and conference proceedings.