

IMPLEMENTATION OF LUNG CANCER NODULE FEATURE EXTRACTION USING DIGITAL IMAGE PROCESSING

¹B.Kasthuri, ²A.Ahila M.E., (Ph.D)

¹PG Scholar, M.E - Communication Systems,

²HOD, Department of ECE,

^{1,2}Infant Jesus college of Engineering and Technology, Thoothukudi, India.

Abstract: In current days, picture handling strategies are generally utilized as a part of numerous restorative territories for enhancing prior recognition and treatment stages, particularly in different malignancy knobs, for example, the lung growth, bosom disease, cerebrum tumor et cetera. This paper utilizes mean channel and middle channel for picture preprocessing. For picture division, Otsu's thresholding and marker-controlled Watershed division methodology are utilized to portion the lung of CT picture. In highlight extraction step, the physical dimensional measure and the dim level co-event network (GLCM) technique are connected. Contingent upon the lung highlight extraction, choice is made whether the lung has knob or not. Determination is generally in light of CT (registered tomography) pictures. CT examining presents extraordinary open doors for lung growth finding. The primary goal of this paper is to execute lung knob division and highlight extraction utilizing advanced picture handling for the characterization of the malady stages to keep away from genuine stages early and to lessen lung disease rate dissemination.

I. INTRODUCTION

The lungs are a couple of wipe like, cone-formed organs [1]. The right lung has three projections, and is bigger than the left lung, which has two flaps. Life structures of lung is appeared in Fig.1. Lung growth is an illness of strange cells increasing and developing into a knob. Fig.2 portrays the start of the tumor. The sorts of lung disease are isolated into four phases. In stage I, the growth is restricted to the lung. In stages II and III, the malignancy is kept to the mid-section (with bigger and more intrusive tumors named stage III). Stage IV disease has spread from the mid-section to different parts of the body. Of a wide range of growth, lung disease is the most well-known reason for passings, representing 1.3 million passings every year. An expected 159,260 individuals are required to end up from lung malignancy in 2014, representing roughly 27 percent of all tumor. Early recognition of lung disease can build the shot of survival among individuals.

There are numerous systems to analyze the lung malignancy, for example, Chest Radiograph (x-beam), Computed Tomography (CT), Magnetic Resonance Imaging (MRI sweep) and Sputum Cytology. Be that as it may, the greater part of these procedures are costly and tedious. In this way, there is an extraordinary requirement for another innovation to analyze the lung malignancy in its initial stages. Picture handling procedures give a decent quality device to enhancing the manual investigation. The utilization of picture handling strategies can help radiologists and specialists in diagnosing illnesses and to offer a fast access to restorative data picked up significance in a brief timeframe. In this paper, MATLAB has been utilized through each techniques made.

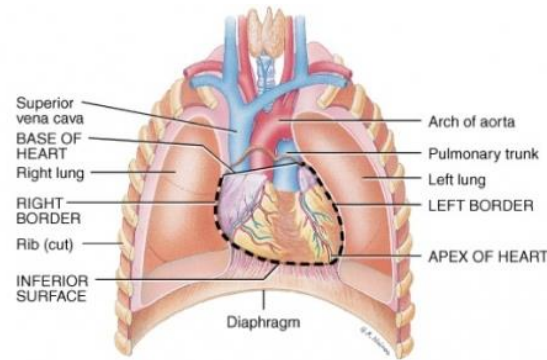


Fig.1 Anatomy of Lung

In specialized writing done by A.Amutha and R.S.D Wahidabanu [3], Level Set-Active Contour Modeling was utilized as a technique as a part of diagnosing lung tumor. Initial step was expelling clamor from picture utilizing part based non-nearby neighborhood denoising work and done component extraction in view of histogram to order amongst ordinary and unusual classes. At the last stride or in tumor location, level set-dynamic form demonstrating with minimized inclination to the picture was presented. In another study [4], Auto improvement, Gabor channel and Fast Fourier change (FFT) were utilized to upgrade the picture and utilized Thresholding and Watershed division to portion the picture. While for highlight extraction, Binarization and Masking methodology were connected. N.A. Memon et. al [5] proposed thresholding technique which select the limit in view of the item and foundation pixel implies. Area developing is utilized then to extricate the definite pit district with exactness

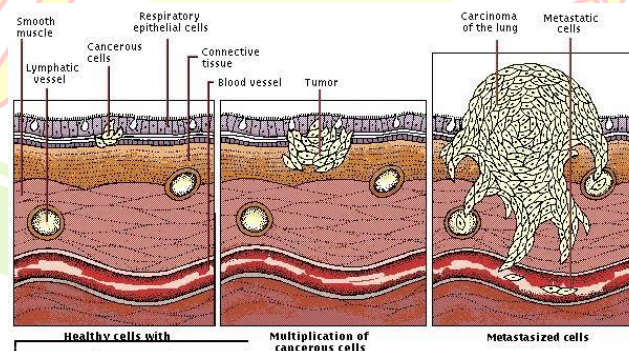


Fig2. The beginning of cancer.

II. BACKGROUND METHODOLOGY

Fig3 delineates a general depiction of lung tumor knobs division and highlight extraction framework that contains five essential strides. The initial step begins with taking an accumulation of CT pictures (typical and irregular) from the accessible database. The second step applies picture preprocessing, to get best level of value and clearness. The third step is picture division which assumes a powerful part in picture handling steps, and the fourth step

contains highlight extraction. The last stride gives picture determination consequence of the pointers of typicality or variation from the norm of lung pictures.

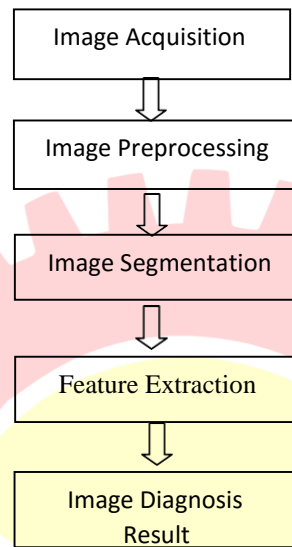


Fig3. Block diagram of Lung Cancer Nodule Segmentation and Feature Extraction System.

A. Image Acquisition

The premier stride in restorative picture preparing is picture procurement. The pictures are gathered from a database of Lung Image Database Consortium (LIDC)[6] furthermore from different doctor's facilities. The therapeutic information is more often than not in DICOM design, which is the standard for capacity and exchange of medicinal pictures [7]. The lung CT pictures having low commotion when contrasted with X-beam filter picture and MRI picture. In this way the CT pictures are taken for distinguishing of the lungs. The principle focal points of figured tomography picture have better clarity, low clamor and contortion. Lung CT pictures are given as info. Measurements of pictures are 512x512 pixels in size. The information CT picture contains commotions, for example, background noise, what's more, pepper commotions and so on. Along these lines, picture preprocessing stage is expected to dispense with clamors.

B. Image Preprocessing

Picture preprocessing is one of the classes of picture taking care of, try to make definite more plainly obvious. Picture preprocessing is a way to deal with upgrade the way of picture, so that the vital picture is better than the first. In this paper, mean channel and center channel are displayed for preprocessing of selecting the CT pictures. The center channel is a non-straight instrument, while the ordinary channel is an immediate one. Mean isolating is a clear, instinctual and easy to realize of smoothing pictures, i.e. reducing the measure of force assortment between one pixel and the accompanying. The center channel is ordinarily used to diminish salt-and-pepper upheaval in a photo. It as often as possible makes a better appearing with respects than the mean channel of defending accommodating unpretentious component in the photo. The center is figured by first sorting all the pixel values from the enveloping neighborhood into numerical solicitation and a short time later supplanting the pixel being

considered with the inside pixel regard [8]. If the zone under thought contains an essentially number of pixels, the ordinary of the two focus pixel qualities is used. Both mean and center channels are used to clear uproar. This preprocessing picture is used as the commitment for picture division.

C. Image Segmentation

Picture division is an essential method for most picture examination resulting errands. Division disengages a photo into its constituent locale or inquiries [8]. The goal of division is to take off less mind boggling or change the representation of a photo into something that is more essential and less requesting to dismember. The division estimations rely on upon two fundamental properties of power qualities irregularity and likeness [9]. At first class is to allocate picture in perspective of sudden changes in force, for instance, edges in a photo. Useless relies on upon distributing picture into zones that are similar according to a game plan of predefined criteria. In this paper, marker-controlled watershed division and Otsu's thresholding methodology are used. Confining touching things in a photo is one of the more troublesome picture taking care of operations. The watershed change views the power as a limit describing "slants" and "valleys" and tries to perceive the valleys.

In fundamental watershed, the issue of over division is gone up against [9]. In marker-based watershed division, markers are used. Marker territories are then set to be regional minima inside the topological surface (commonly, the inclination iginal input image), and the watershed algorithm is applied [10]. The placement of internal and external markers into regions of interest can easily cope with the over-segmentation problem. The flowchart of Marker-Controlled Watershed segmentation algorithm is shown in Fig.4.)

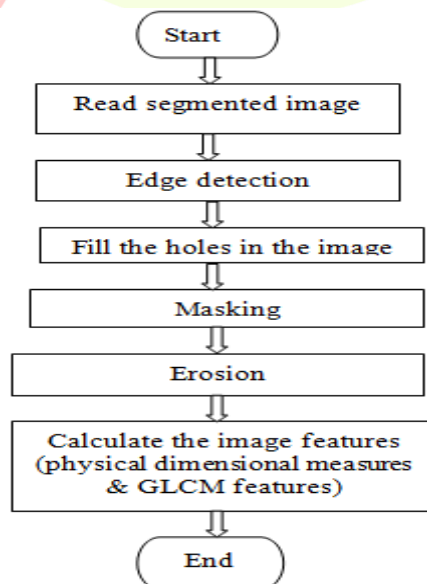


Fig4. Flowchart of Feature Extraction Program.

Unconventionality:

The erraticism is the proportion of the separation between the foci of the oval and its real pivot length. The worth is somewhere around 0 and 1. In the wake of ascertaining of physical dimensional measure, then surface element extraction is performed on the quantized picture by utilizing Gray level co-event framework (GLCM) strategy, a standout amongst the most known composition examination technique. A dim level co-event network is a second request factual measure presented by Haralick [16]. GLCM is the dark level co-event lattice (GLCM), otherwise called the dim level spatial reliance framework. The Gray-Level Co-event Matrix (GLCM) depends on the extraction of a dark scale picture. The GLCM capacities describe the composition of a picture by figuring how regularly combines of pixel with particular qualities and in a predefined spatial relationship happen in a picture, making a GLCM, and afterward extricating factual measures from this matrix[17]. Measurable parameters computed from GLCM qualities are as per the following:

The first picture of CT strange lung with knob is appeared in Fig7. This figure is taken from Mandalay General



Fig5. Original image of CT lung with nodule.

The yield picture is obscured than the first picture. Along these lines, other channel is further utilized. Fig.8(b) is the yield of middle separated picture. To figure out which the best appropriate channel to be utilized, mean square blunder (MSE) and crest sign to clamor proportion (PSNR) are computed for channels. Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) are the two blunder measurements used to look at picture quality. MSE and PSNR conditions are appeared in eq(17) and eq(18). The correlation of normal channel and middle channel are portrayed in Table .1.

TABLE1: Comparison of Average Filter & Median Filter

| | MSE | PSNR |
|----------------|--------|---------|
| Average filter | 0.3608 | 52.5577 |
| Median filter | 0.0809 | 59.0493 |

Based on the results shown in Table 1, median filter has the smallest MSE and the highest PSNR. Meanwhile, error is differences between value of estimator of original image and filtered image. Small value of MSE means small error of function used. PSNR is the ratio of maximum power of signal and power of noise. High PSNR means that the reconstruction of image is better since the signal has higher power than the noise. Thus, the median filter, which gives the smallest MSE and the highest PSNR, is selected. Fig.9 are described Marker-controlled watershed segmentation result is the watershed transform of gradient magnitude. The sobel masking operator is used on the CT gray image in both horizontal and vertical directions to create the gradient picture. Markers and article limits superimposed on unique picture is appeared. Along these lines, object limits are more obvious to show the mark framework as a shading picture by utilizing straightforwardness to superimpose this mark lattice on top of the first power picture. Fig.6 demonstrates a divided picture by Otsu's thresholding. Otsu's thresholding changes over a dark scale picture into a paired picture.

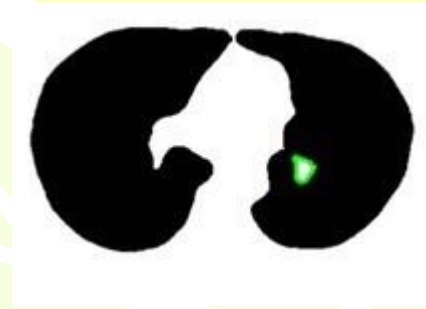


Fig 6. Output of Otsu's thresholding image.

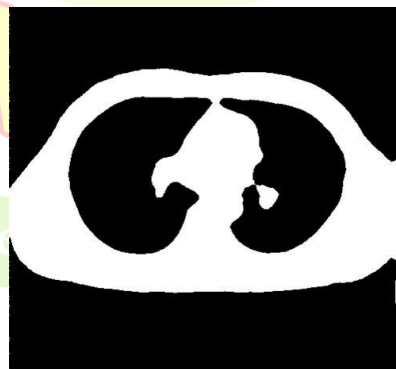


Fig 7. Output of Morphological operation for abnormal lung.

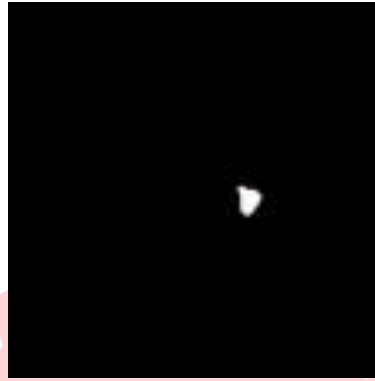


Fig 8. Extracted Feature Values from the Lung Benign Nodule Image

After portioned lung knob, components, for example, geometric properties, textural properties and scientific properties are calculated from the segmented image by their respective formula. In this step, physical dimensional measures Gray Level Co- occurrence Matrix (GLCM) features are used to separate out lung nodules. It is used for calculating image feature values. The extracted features for the nodule images are tabulated in Fig.13 and Fig.14 in MATLAB. In this paper it is tried to extract as many information as possible in the form of features using digital image processing on CT image. These feature extraction results are used as a base for the lung cancer classification system, which improves the chances of survival rate for the patient. For implementing the algorithm GUI toolbox of MATLAB software are used in Fig.15 and Fig.16. In GUI, nodule location, nodule size and program run time are calculated. By using this method, lung cancer disease can detect at early stages and can increase the survival rate.

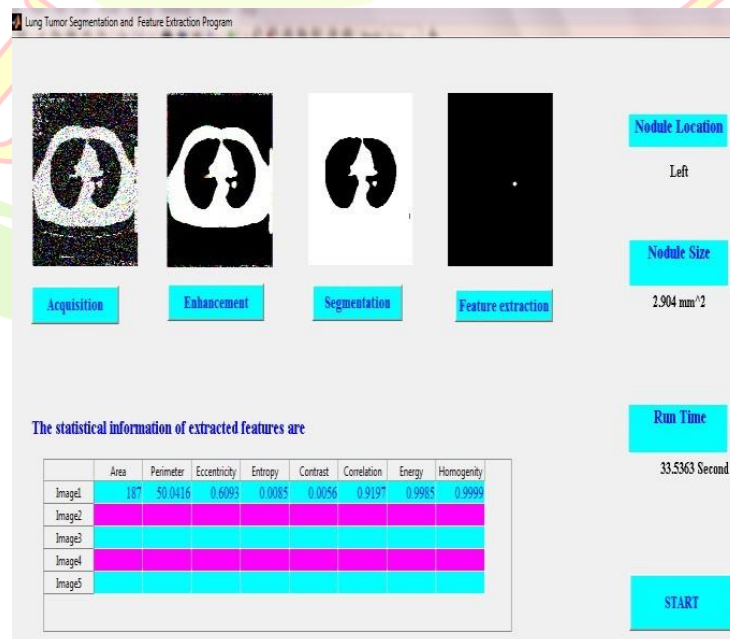


Fig 9. Lung Tumor Segmentation and Feature Extraction Program for left lung using GUI.

The diagnostic system can lead to one of the four possible categories. Table.2 shows the diagnosis result

TABLE2: Diagnosis Result

| Diagnosis | Abnormal condition of patient | Normal condition of patient |
|--------------------|-------------------------------|-----------------------------|
| Positive(Abnormal) | TP | FP |
| Negative(Normal) | FN | TN |

1. Accuracy

Accuracy is a statistical measure of how well a classifier correctly identifies or excludes a condition. The accuracy is the proportion of true results (both true positive and true negative) in the population. The accuracy obtained in this system is 90.909%.

True positive (TP): the diagnostic system yields positive test result for the sample and the sample actually has the disease, **False positive (FP):** the diagnostic system yields positive test result for the sample but the sample does not actually have the disease,

True negative (TN): the diagnostic system yields negative test result for the sample and the sample does not actually have the disease, and

False negative (FN): the diagnostic system yields negative test result for the sample but the sample actually has the disease.

2. Sensitivity

It measures the proportion of actual positives which are correctly identified. That is the percentage of cancerous region correctly identified as cancerous region. The sensitivity obtained is 100%.

IV. DISCUSSIONS

In the preprocessing step, both the mean channel and the middle channel can be utilized to expel commotion from a picture. As indicated by the yields of picture preprocessing pictures in Fig.8, middle channel is more reasonable than the normal channel for this examination on the grounds that the primary point of preference of middle separating is that even after pixel power qualities are changed the edges of the pictures are protected. The expanding veil size is more compelling in minimizing the effect of commotion. The cover size of these figure results is 15x15. Additionally, the middle channel has the littlest MSE and the most noteworthy PSNR. So, the middle channel is decided for this exploration. As per the yields of the division step, marker-controlled watershed division approach has higher precision and preferable quality over thresholding approach. After division of picture, morphological operation is utilized to get singular lung and to take out pointless parts.

By doing morphological operations, it gets the individual lung as well as clear the lung knob. For highlight extraction step, highlight separated from the picture is utilized for deciding and

grouping the suspicious territory of the lung (knob). Components are figured from their recipes. Computed components and knob size and knob area will help specialists to take the right choice. At long last, the execution of the framework is assessed. Three execution assessments are precision, affectability and specificity for the picture handling procedures. This framework offers affectability of 100%, specificity of 83.33% and precision of 90.909%. The consequence of this framework and the examination of specialist together build the precision of recognizing threatening and considerate lung knobs. In the event that the harmful knob is identified in its before stage, then the possibility of surviving the patient increments.

V. CONCLUSION

The death rate of lung growth is the most elevated among all different sorts of disease; it can be recognized ahead of schedule by distinguishing the lung knobs. In this paper, picture preparing and picture division are executed to get the finding result. By utilizing these strides, the knobs are distinguished and some components are separated. The separated components can be utilized for arrangement of ailment stages. Deciding the knob elements can give to know more data of the state of lung malignancy at the early stages. This system helps the radiologists and the specialists by giving more data and taking right choice for lung disease tolerant in brief time with precision. Hence, this technique is less exorbitant, less tedious and simple to actualize.

VIII. REFERENCES

- [1] Non-Small Cell Lung Cancer, [online available], <http://www.katemacintyrefoundation.org>, Adapted from National Cancer Institute (NCI) and Patients Living with Cancer (PLWC), 2007, accessed on 13 July 2013.
- [2] Anatomy of lung picture and beginning of cancer, [online available], www.allreferhealth.com, accessed on 4 June 2013.
- [3] A.Amutha and R.S.D Wahidabanu, "A Novel Method for Lung Tumor Diagnosis and Segmentation using Level Set- Active Contour Modelling", *European Journal of Scientific Research*, Vol.90, No.2, November 2012, pp.175-187
- [4] Mokhled S. Al-Tarawneh, "Lung Cancer Detection Using Image Processing Techniques", *Leonardo Electronic Journal of Practices and Technologies*, Issue 20, January-June 2012, p.147-158. ISSN 1583-1078
- [5] N.A. Memon et. al, "Segmentation of Lungs from CT Scan Images for Early Diagnosis of Lung Cancer," *World Academy of Science, Engineering and Technology*. 2006.
- [6] Lung Cancer Database Consortium, [online available], www.eddie.via.cornell.edu.com, accessed on 22 July 2013. [7] *Digital Imaging and Communications in Medicine (DICOM) Part 1: Introduction and Overview*, National Electrical Manufacturers Association, 2006, pp. 11.

- [8] S Jayaraman, S Esakkirajan and T Veerakumar, "Digital Image Processing", 3rd edition, Tata McGraw Hill, 2010, ISBN(13):978-0-07-014479-8, ISBN (10): 0-07014479-6.
- [9] Gonzalez R.C., Woods R.E., "Digital Image Processing using MATLAB", Upper Saddle River, NJ Prentice Hall, 2008.
- [10] Levner I., Zhang H., "Classification-Driven Watershed Segmentation", IEEE Transactions on Image Processing, 2007, 16(5), 1437-45.
- [11] 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. [12] Huang Q., Gao W., Cai W., "Thresholding technique with adaptive window selection for uneven lighting image", Pattern Recognition Letters, Elsevier, p. 801-808.
- [13] Sudha.V, Jayashree.P., " Lung Nodule Detection in CT Images Using Thresholding and Morphological Operations", International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319–6378, Volume-1, Issue-2, December 2012.
- [14] Pratt, William K., "Digital Image Processing", New York, John Wiley & Sons, Inc., 1991, p.634.
- [15] Cigdem Demir and Bulent Yener., "Automated cancer diagnosis based on histopathological images: a systematic survey", Technical report, Rensselaer Polytechnic Institute, Department of Computer Science, TR-05-09.
- [16] Haralick(1979), "R.M. Statistical and structural approaches to texture", Proceedings of the IEEE, vol.67, pp.786-804.
- [17] Haralick, R.M., and L.G. Shapiro., "Computer and Robot Vision", Addison-Wesley, vol.1, p-459, 1992.

