

**PMAS- Arid Agriculture University Rawalpindi**

**Barani Institute of Sciences Sahiwal**

**Department of Computer Sciences**

**Project Title**

**Lung Cancer Detection Using Image Processing in MATLAB**

**BY**

Muhammad Zeeshan Mazhar (15-ARID-4011)

Muhammad Shoaib Sani (15-ARID-4008)

Muhammad Irfan Naseer (15-ARID-4005)

Muhammad Rizwan Ahmed (15-ARID-4007)

**ABSTRACT**

The major cause of cancer death is lung cancer. Detection of cancer in the early phase can provide more treatment options, less invasive surgery and increases the survival rate. For lung cancer, if the disease is detected in time, the survival rate of patient increases from 14 to 49% in recent 5 years. The main aim of this research is to evaluate the various computer-aided techniques, analyzing the current best technique and finding out their limitation and drawbacks and finally proposing the new model with improvements in the current best model. In this study, MATLAB has been used through every procedure 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. In first stage, Binarization technique is used to convert binary image and then compare it with threshold value to detect lung cancer. In second stage, segmentation is performed to segment the lung CT image and a strong feature extraction method (morphological operations) has been introduced to extract some important features of segmented images. The properties of extracted features have been found as well.

**Table of Contents**

Page No.

Chapter 1: Introduction 1

Chapter 2: Background Study 2

Chapter 3: Methodology 3

*3.1 Image Acquisition* 3

* 1. *Gabor Filter* 3

Chapter 4: Results 4

* 1. *Region Growing* 4
  2. *Marker Controlled Watershed using Masking* 4
  3. *Binarization* 5
  4. *Feature Extraction* 5

Chapter 5: Conclusion and Future Work 6

References

1. **Introduction**

Cancer is one of most dangerous disease that causes deaths [1]. Data obtained from Global Burden Cancer shows that in 2012 there are 14.1 million cases of cancer in the world, with lung cancer occupies the first position with a percentage of 13%. While the number of cancer deaths was recorded 8.2 million deaths, with lung cancer cause of death in first place with a percentage of 19% [2]. Lung cancer is a disease of abnormal cells multiplying and growing into a tumor. Cancer cells can spread out from the lungs through the blood stream or lymph fluid that surrounds the lung tissue. Generally, cancer cells often spread toward the center of the chest due to the natural flow of lymph. Metastasis occurs when cancer cells spread to other organs. The process of early detection of cancer plays an important role to prevent cancer cells from multiplying and spreading. Previous researches have been conducted for analyzing lung cancer such as using clustering method in microarray data [3], the detection of lung cancer with general image processing techniques in CT scan data with good results and accuracy [4]. In this study we proposed and evaluated additional image segmentation methods in analyzing lung cancer using image processing techniques.

**Figure 1** shows a general description of lung cancer detection system that contains four basic stages. There are four stages to determine whether there is a lung cancer or not. The first phase is we get CT scan image data. The second phase, we implement image enhancement to improve quality of image. The third phase is image segmentation which is an important step in the detection of cancer. The fourth stage is feature extraction that is give us a conclusion whether there is a lung cancer or not.

(Figure 1)

1

1. **Background Study**

Lusted L.B proposed for the first time by analyzing and recognizing the abnormal and normal photography images [9]. Hence, an analysis of medical images by computer was introduced and it includes some steps such as Image Enhancement, Segmentation and feature extraction. Andreas Savakis and Richard Carbone [10] explained about Discrete the wavelet transform (DWT) and it uses a variation of the lifting scheme techniques and provides an advantage that includes small memory requirements and its core may be used for image enhancement and de-noising. Ada and Rajneet Kaur [11] in their research work Image enhancement implemented by the histogram equalization and feature extraction stage implemented by Binarization approach. Feature extraction stage is an important stage that uses algorithms and techniques to detect and isolate various desired portions of a given image. To predict the probability of lung cancer presence Binarization approach is used. IIya Levner and Hong Zhang [12] explained about watershed segmentation, it extracts seeds indicating the presence of objects or background at specific image location. Nooshin Hadavi and Md. Jan Nordin [13] in their research work Image segmentation implemented by thresholding and it has some benefits such as fast processing. In this research work, the Gabor filter is used for Image enhancement instead of DWT and Histogram equalization. For segmentation watershed segmentation algorithm is used and for feature extraction masking approach used instead of Binarization these methods shows promising results compare to other methods.

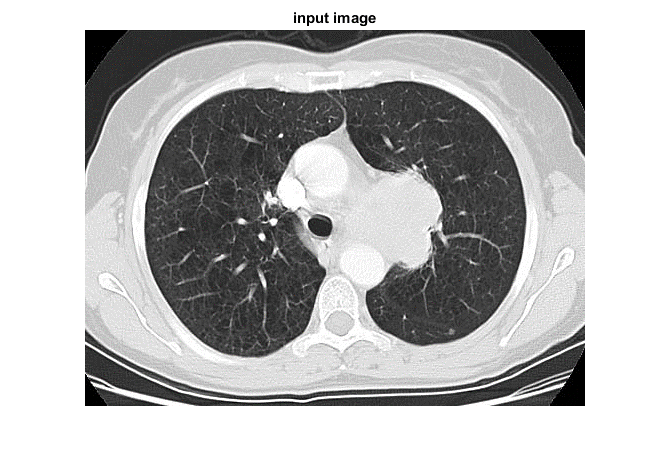
2

1. **Methodology**

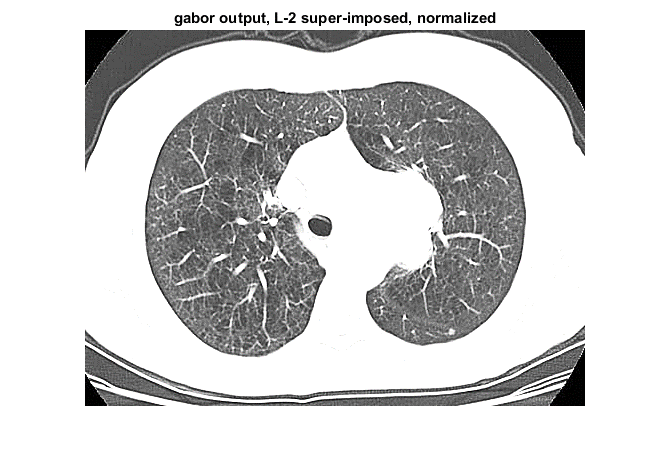
Following methods were used to do the complete detection procedure.

**3.1 Image Acquisition**

The imagery used is from CT scan database. Our data is from the VIA and ELCAP Public Access. Database [5]. Image is in digital imaging and communications in medicine (dicom) format. There are 50 patients and each patient have about 250 pieces of images.



**3.2 Gabor Filter**

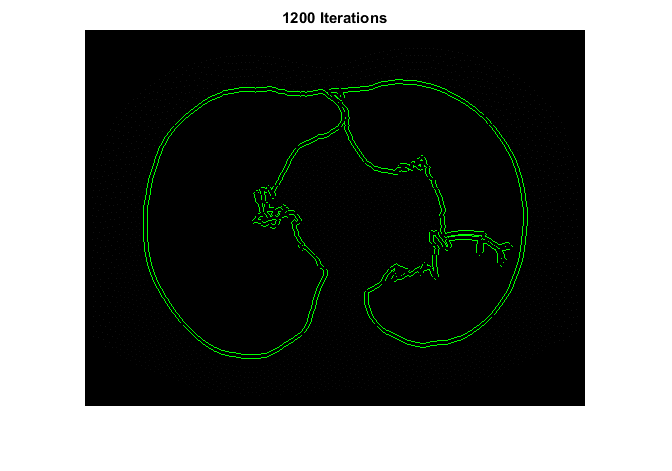
Gabor filter named by Dennis Gabor, is a linear filter is used for edge detection. Representation of Gabor filter is similar to the human visual system. In the spatial domain, 2D Gabor filter is a Gaussian filter function modulated by a sinusoidal function [4]. In the process of this cancer detection imagery used is a 2D image, so using 2D Gabor filter.

3

1. **Results**

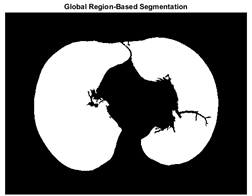
**4.1 Region Growing**

Region growing is a procedure to extend the covered area by collecting pixels or sub-regions based on predetermined criteria. Basically, the approach is to initiating a set point (seed) and then expanding the area surrounding the seed which has the same properties as the seed, for example, the area that has the same range of color or gray level [6]. Figure 4 shows three stages of segmentation by region growing method.



**4.2 Marker Controlled Watershed**

To get the segmentation result of watershed method, the steps are as follows; firstly, calculate the distance gradient for edge detection, and then mark the target object by using a morphological technique called opening by reconstruction and closing by reconstruction. After marking object is revealed, other areas can be discarded. Modules and example of marker controlled watershed segmentation method is existing in MATLAB software that can be directly used. Figure below shows the result of segmentation by using Marker- Controlled Watershed.

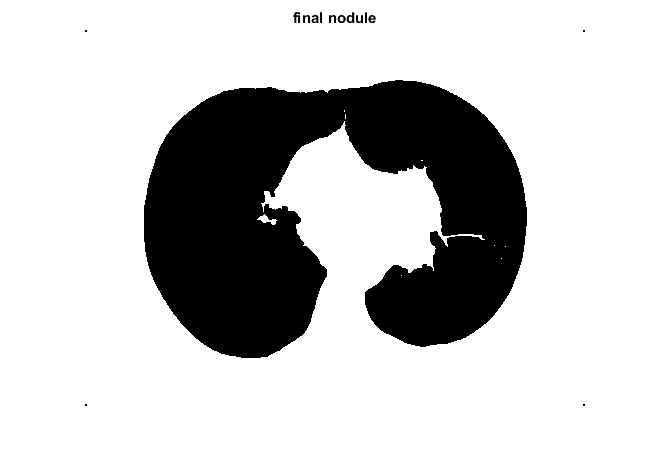
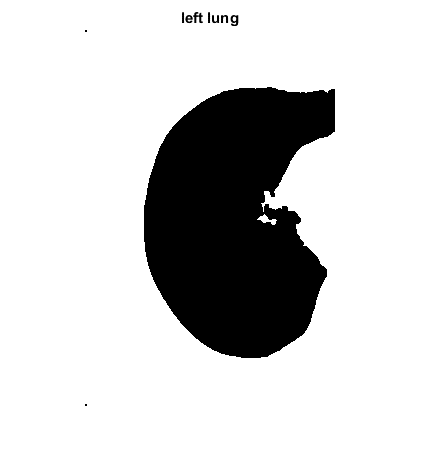
(Final Segmentation including Thresholding)

4

**4.3 Binarization**

Binarization is the process of changing the color of the pixel values into two classes [8], such as black and white. After getting the quantity of black and white pixels on segmentation results, then we compared it with a threshold value to determine the condition of lung (normal or cancer). The threshold value is obtained from observations on normal lung. The threshold value that is used in this research is 17178.48. If the number of black pixels more than the threshold, then we conclude the lung is normal, otherwise there is a chance of a cancerous nodule in the lungs.

**4.4 Feature Extraction**

Feature extraction is the process of extracting some specific portions of our interest from an image. To make the coding simpler, we have used simple morphological operations to extract the possible nodule, rather than other algorithms like Gray Level Co-occurrence Matrix (GLCM). Then, to calculate properties of the possible nodule, we have used built in MATLAB function regionprops(). The main focus in extraction of features was simplicity, while there was not much focus put on efficiency.

**Table 1.**  Details of Patient Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Patient Name\*** | **Nodules** | **Area(px^2)** | **Eccentricity** | **Perimeter** |
| Patient 1 | Yes | 92 | 0.9399 | 47.86 |
| Patient 2 | No | - | - | - |
| Patient 3 | No | - | - | - |

5

1. **Conclusion and Future Work**

In this study, we evaluated the image taken from the image database, we enhance it using Gabor filter. In addition, we also use Gaussian filter but only for a limited number of images based on their intensity and pixel characteristics. Then, we applied a combination of region based and marker controlled watershed segmentation using open source segmentation code from the MATLAB File Exchange. We also did the thresholding using Otsu’s Thresholding. After that, we did the Binarization process to see if the number of black pixels are greater than the number of white pixels or not. If black pixels are greater than white, then the Lung is normal. If not, then there is a chance we will find nodules in the lungs. Then, we did more processing on the image by decreasing brightness and using Median Filter. Once we’ve separated the nodule from the lung, we separate it and find its regional properties using the regionprops().

In future, we are going to take the data of the nodule and other characteristics of the image to train the Support Vector Machine (SVM) Classifier and classify a nodule as benign or malignant.

6

**References**

[1] Permata T S and Bustamam A 2015 Clustering protein-protein interaction network of TP53 tumor suppressor protein using Markov clustering algorithm International Conference on Advanced Computer Science and Information Systems (ICACSIS) pp 221–226

[2] GLOBOCAN 2012, International Agency for Research on Cancer, World Health Organization. http://globocan.iarc.fr/

[3] Khalid R 2014 Clustering analysis of cancerous microarray data Journal of Chemical and Pharmaceutical Research 6(9) pp 488-493, ISSN: 0975-7384

[4] Kajal N et al 2015 Early Detection of Lung Cancer Using Image Processing Technique: Review International Journal of Advent Research in Computer and Electronics (IJARCE) 2(2), E-ISSN: 2348-5523

[5] Lung Cancer Database, Available at: https://eddie.via.cornell.edu/cgibin/datac/signon.cgi

[6] Gonzalez R C and Woods R E 2008 Digital Image Processing Upper Saddle River (New Jersey: Prentice Hall)

[7] Allaoui A E and Nasri M 2012 Medical Image Segmentation by Marker Controlled Watershed and Mathematical Morphology 1LABO MATSI, ESTO, B.P 473, University Mohammed I OUJDA, Maroko

[8] Sarwinda D and Bustamam A 2016 Detection of Alzheimers disease using advanced local binary pattern from hippocampus and whole brain of MR images International Joint Conference on Neural Networks (IJCNN) IEEE pp 5051–5056

[9] L. B. Lusted, "Logical analysis in roentgen diagnosis," Radiology, vol. 74, pp. 178-93, 1960.

[10] Andreas Savakis and Richard Carbone, “Discrete Wavelet Transform Core for Image Processing Applications”, Department of Computer Engineering, Rochester Institute of Technology, Rochester, NY 14623.

[11] Ada and Rajneet Kaur ,”Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013.

[12] IlyaLevner and Hong Zhang,”Classification- Driven Watershed Segmentation”, IEEE Transactions on Image Processing, Vol. 16, No. 5, May 2007.

[13] Nooshin Hadavi and Md.jan Nordin ,”Lung Cancer diagnosis using CTscan Images Based on Cellular Learning Automata”. University Kebangsaan Malaysia.